



ISSN: 2301-5373  
E-ISSN: 2654-5101  
Volume 12 • Number 4 • May 2024

# JELIKU

**Jurnal Elektronik Ilmu Komputer Udayana**

**Informatics Study Program**

Faculty of Mathematics and Natural Sciences

Udayana University

## Table of Contents

<i>Klasterisasi Customer Lifetime Value Pelanggan Menggunakan K-Means pada Online Travel Agency</i>	
I Putu Ananta Wijaya, I Ketut Gede Suhartana.....	741-748
<i>Forecasting Saham Perbankan Dengan Penerapan Multilayer Backpropagation Neural Network</i>	
I Putu Ryan Paramaditya, Cokorda Rai Adi Pramatha, I Gede Arta Wibawa, I Gede Santi Astawa, Ida Bagus Gede Dwidasmara, I Dewa Made Bayu Atmaja Darmawan .....	749-760
<i>Klasifikasi Biji Jagung Berdasarkan Tekstur Dan Warna Menggunakan Metode Backpropagation Berbasis Web</i>	
M . Syafiih; Nadiyah Nadiyah .....	761-774
<i>Penerapan Metode MFCC dan LSTM untuk Pengenalan Emosi melalui Suara</i>	
I Dewa Agung Prawangsa, Dr. Anak Agung Istri Ngurah Eka Karyawati.....	775-782
<i>Perancangan Ulang Website Itinerary untuk Daerah Wisata Kabupaten Badung Menggunakan Metode User Centered Design</i>	
Putu Audy Cipta Pratiwi, Luh Arida Ayu Rahning Putri, Luh Arida Ayu Rahning Putri, I Wayan Santiyasa, Gst Ayu Vida Mastrika Giri .....	783-794
<i>Implementasi Sistem Informasi E-Learning Berbasis OdoO Studi Kasus MI YPPI 1945 Babat</i>	
Ahmad Assrorul Abidin, Nufan Balafif, Eddy Kurniawan .....	795-808
<i>Perbandingan Metode Ensemble Learning Random Forest Dan Adaboost Pada Pengenalan Chord Instrumen Piano Dan Gitar</i>	
I Dewa Agung Adwitya Prawangsa, Dr. Anak Agung Istri Ngurah Eka Karyawati.....	809-816
<i>Analisis Keranjang Belanja Untuk Meningkatkan Penjualan Fashion Di E-Commerce Australia</i>	
Azzahra Putri Maharani, Apriade Voutama .....	817-832

*Sistem Pendukung Keputusan Gerai Di Pusat Perbelanjaan Modern Provinsi Bali: Metode ANP*

Andien Rachma Fadillah, I Komang Ari Mogi, Cokorda Rai Adi Pramatha, I Gusti Ngurah Anom Cahyadi Putra..... 833-844

*Ekstraksi Fitur Dengan Convolutional Neural Network Dan Rekomendasi Fashion Menggunakan Algoritma K-Nearest Neighbours*

I Gede Teguh Permana, Ida Bagus Gede Dwidasmara, Made Agung Raharja, I Wayan Santiyasa ..... 845-856

*Penerapan SVM dengan Seleksi Fitur Mutual Information untuk Memprediksi Sentimen PEMILU 2024*

I Gusti Bgs Darmika Putra, Cokorda Pramatha, Anak Agung Istri Ngurah Eka Karyawati, Made Agung Raharja..... 857-866

*Perancangan Sistem Informasi Pemasok Barang Rongsokan Dengan Pendekatan User Centered Design*

Tri Adi Ningsih, I Made Widiartha, I Ketut Gede Suhartana, I Gusti Ngurah Anom Cahyadi Putra..... 867-878

*Analisis Dan Visualisasi Data Penjualan Pada NYC Property Menggunakan EDA*

Meliana Endang Nyimas Lisna, Apriade Voutama..... 879-866

*Klusterisasi Fitur Tanya Dokter Pada Website Alodokter Menggunakan Metode Latent Dirichlet Allocation*

I Putu Fajar Tapa Mahendra, I Gusti Ngurah Anom Cahyadi Putra ..... 887-898

*Analisis Dan Visualisasi Data Untuk Meningkatkan Penjualan Menggunakan Exploratory Data Analysis Dan Looker Studio (Studi Kasus : Nies Collection)*

Melanie Putri, Apriade Vaoutama..... 899-914

*Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Klasifikasi Penerima Beasiswa Studi Kasus di SMKS 2 Adi Luhur*

Lia Susanti, Khoirunnisa ..... 915-922

<i>Pengembangan Sistem Informasi Digital Zoonosis Melalui Pendekatan Web Semantik</i> Komang Kartika Noviyanti, Luh Gede Astuti, I Gede Surya Rahayuda, Agus Muliantara .....	923-934
<i>Pemanfaatan UML dalam Perancangan Sistem Objek Wisata Kabupaten Bandung Barat Berbasis Website</i> Tazkia Damayanti, Apriade Voutama .....	935-946
<i>The Single Sign On Model Using SAML and OAuth for Online Application of UPNYK</i> Ahmad Taufiq Akbar, Hari Prapcoyo, Rifki Indra Perwira .....	947-954
<i>Perancangan Ontologi Semantik: Representasi Digital Film Bioskop Indonesia</i> Andien Rachma Fadillah, Komang Kartika Noviyanti, I Putu Agus Arya Wiguna, Cokorda Pramartha .....	955-966
<i>Citra Digital Voice Recognition Menggunakan SVD</i> Dominggus Mangngi, Putra Prawira Yohanes Puka, Yampi R. Kaesmetan .....	967-972
<i>Perancangan UI/UX Aplikasi Perpustakaan Online Berbasis Mobile Menggunakan Software Figma</i> Muhammad Putra Tristan, Apriade Voutama .....	973-980

This page is intentionally left blank.

**SUSUNAN DEWAN REDAKSI  
JURNAL ELEKTRONIK ILMU KOMPUTER  
UDAYANA (JELIKU)**

**Penanggung Jawab :**

Dra. Ni Luh Watiniasih M.Sc., Ph.D.

**Redaktur :**

Gst. Ayu Vida Mastrika Giri, S.Kom., M.Cs.

**Penyunting/Editor :**

Agus Muliantara, S.Kom., M.Kom

I Komang Ari Mogi, S.Kom., M.Kom

I Gusti Agung Gede Arya Kadyanan, S.Kom., M.Kom

Dr. Anak Agung Istri Ngurah Eka Karyawati, S.Si., M.Eng

**Disain Grafis :**

I Gede Yogananda Adi Baskara

I Gusti Agung Ayu Gita Pradnyaswari Mantara

**Fotografer :**

I Kadek Agus Candra

Widnyana I Komang

Dwipayoga

**Sekretariat :**

Ni Ketut Alit Widiastuti, S.Kom.

Anak Agung Raka Darmawan, S.Kom.

I Putu Herryawan, S.Kom.

*This page is intentionally left blank.*

# Klasterisasi Customer Lifetime Value Pelanggan Menggunakan K-Means pada Online Travel Agency

I Putu Ananta Wijaya<sup>a1</sup>, I Ketut Gede Suhartana<sup>a2</sup>

<sup>a</sup>Informatika, Universitas Udayana  
Kampus Bukit Jimbaran, Indonesia  
<sup>1</sup>anantaw81@gmail.com  
<sup>2</sup>suhartana@gmail.com

## Abstraksi

Penjualan produk didapati pada perusahaan yang bergerak di bidang e-commerce, salah satunya adalah online travel agency (OTA). Setelah covid-19, perusahaan sedang bangkit dengan persaingan yang semakin banyak muncul dan pelanggan yang beralih ke e-commerce pesaing. Menjaga hubungan dari satu pelanggan ke pelanggan lainnya diperlukan. Data historis transaksi pelanggan dapat digunakan untuk melakukan klasterisasi. Sebelum itu, dilakukan pembentukan model RFM yaitu recency, frequency, dan monetary supaya bisa dilakukan perhitungan customer lifetime value. Klasterisasi yang digunakan adalah K-Means karena sederhana dan efisien. Penentuan jumlah klaster optimal menggunakan metode Gap Statistic. Jumlah klaster akan divalidasi separasinya oleh metode davies bouldin index. Setiap data klaster kemudian dicari rata-rata untuk nilai setiap fitur model RFM. Hasilnya akan dikali dengan bobot yang didapat dari AHP. Klaster dengan loyalitas tertinggi didapatkan pada klaster 2 yaitu 0.0297. Nilai fitur frequency pada klaster 2 juga merupakan nilai tertinggi. Hal ini menandakan klaster 2 merupakan pelanggan dengan nilai loyalitas yang tertinggi. Perusahaan dapat merancang strategi untuk pelanggan dengan program seperti pemberian diskon atau poin. Sedangkan, pada klaster dengan nilai loyalitas terendah bisa dibuatkan program personalized product.

**Kata kunci:** Image Processing, Klasterisasi, Customer Lifetime Value, RFM, K-Means, Gap Statistic

## 1. Pendahuluan.

Pelanggan merupakan kunci dari penjualan produk. Penjualan produk didapati pada perusahaan yang bergerak di bidang e-commerce, salah satunya adalah online travel agency (OTA). Pada saat covid-19, perusahaan OTA mengalami penurunan pelanggan dikarenakan lockdown [10]. Pada saat ini, perusahaan sedang bangkit dengan persaingan yang semakin banyak muncul dan pelanggan yang beralih ke e-commerce pesaing. Oleh karena itu, salah satu tujuan dari perusahaan e-commerce adalah menjaga hubungan dari satu pelanggan ke pelanggan lainnya. Hal ini menunjukkan perusahaan harus mampu mengidentifikasi pelanggan karena setiap pelanggan itu unik yaitu punya preferensi dan perilaku berbeda.

Dalam data mining, penggunaan data historis transaksi pelanggan dapat digunakan untuk melakukan klasterisasi. Metode klasterisasi yang digunakan bertipe non hierarki karena tidak mempunyai tingkatan. Metode yang dipilih adalah k-means. Menurut Bao Chong [3], k-means mempunyai keuntungan yaitu sederhana, sangat efisien, dan gampang diimplementasikan sehingga dipakai di banyak bidang. Algoritma ini juga lebih cepat dibandingkan algoritma klasterisasi yang lain dari segi waktu eksekusi. Algoritma k-means sendiri tidak bisa menentukan jumlah klaster yang akan dibuat. Oleh karena itu, digunakan gap statistic yang mengambil nilai gap terbesar dari nilai perhitungan data asli dan data referensi untuk digunakan sebagai jumlah klaster optimal [7]. Setelah itu, jumlah klaster akan divalidasi dengan davies bouldin index (DBI) untuk menentukan apakah data sudah terseparasi dengan baik. Nilai DBI akan optimal jika semakin kecil [9].

Untuk mendapatkan klaster dengan nilai loyalitas terbaik, maka digunakan perhitungan customer lifetime value (CLV). Sejumlah nilai potensial yang dimiliki oleh setiap pelanggan dalam bentuk mata uang, yang dapat diperoleh oleh perusahaan dari pembelian produk atau layanan, berdasarkan tingkat kesetiaan pelanggan tersebut di masa depan. Nilai inilah yang diamati karena manfaat yang diberikan

saat ini hingga di masa depan bagi perusahaan [5]. Perhitungan CLV sendiri memerlukan pembuatan model sebelum klasterisasi, yaitu model *recency*, *frequency*, dan *monetary* (RFM). Setelah klasterisasi, setiap klaster akan memiliki rata-rata setiap fitur model. Rata-rata ini akan dikalikan dengan bobot yang didapatkan dari proses *analytical hierarchy process* (AHP). AHP adalah pendukung keputusan yang akan menguraikan masalah multi faktor atau multi kriteria yang kompleks menjadi suatu hirarki. Nilai yang terbesar didapatkan oleh CLV akan menjadi klaster dengan loyalitas terbaik. Klaster ini dapat diberikan perlakuan khusus seperti memberikan diskon dan poin.

## 2. Metode Penelitian

Penelitian menggunakan data historis transaksi pelanggan dari sebuah perusahaan OTA dengan rentang waktu Juni 2022 hingga Juni 2023. Kemudian data akan dilakukan *preprocessing* yaitu transformasi ke model RFM dan normalisasi dengan *min-max*. Setelah itu, dilakukan klasterisasi dengan *k-means* dan bantuan *gap statistic* untuk menentukan jumlah klaster optimal. Hasil klaster kemudian akan divalidasi dengan DBI untuk menentukan data per klaster sudah terseparasi dengan baik. Setelah itu dilakukan pencarian nilai CLV dengan bantuan bobot dari AHP.

### a. K-Means

Metode ini merupakan salah satu yang tertua dan sangat populer di kalangan praktisi karena kemudahan implementasi dan kecepatan prosesnya [6]. *K-means* berbasis titik pusat (*centroid*) dengan metode partisi (*partitioning method*). *clustering* data menggunakan dengan metode *k-means* bertujuan untuk mempartisi atau membagi  $n$  objek data ke dalam  $k$  kelompok data (*klaster*). Kemiripan data pada suatu *klaster* diukur sesuai dengan kedekatan obyek terhadap nilai rata-rata pada *klaster* atau disebut sebagai *centroid*. Berikut ini merupakan langkah-langkah dari metode *clustering* menggunakan algoritma *k-means*:

1. Algoritma *k-means* menerima inputan dari pengguna berupa nilai jumlah klaster sebanyak  $k$  klaster, dan suatu dataset yang ingin dikelompokkan.
2. Sebanyak  $k$  data dipilih secara *random* dari suatu dataset yang kemudian ditentukan sebagai *initial centroid*.
3. Hitung jarak masing-masing titik objek data ke masing-masing *initial centroid*. Perhitungan jarak dapat dilakukan menggunakan persamaan ukuran jarak *euclidian distance* ditampilkan pada (1).

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

4. Mengalokasikan masing-masing titik data ke dalam *klaster* dengan jarak paling minimum.
5. Menghitung rata-rata dari semua data yang terdapat di dalam *klaster* tersebut sebagai pusat klaster baru.
6. Ulangi langkah 4, 5, dan 6 hingga tidak ada lagi objek yang berubah di dalam suatu klaster.

### b. Gap Statistic

Dalam menentukan  $k$  klaster optimal untuk klasterisasi, digunakan metode *gap statistic*. *Gap statistic* dipublikasikan oleh Tibshirani [7] *Gap statistic* adalah metode yang digunakan untuk mengukur "gap" antara kualitas klaster dari data asli dan kualitas dari data referensi. *Gap statistic* memerlukan dua data matriks yaitu data asli dan data referensi. Berikut langkah-langkah algoritma *gap statistic*

1. Buat data referensi (*references*) menggunakan *sampling* secara *uniformly* (distribusi secara rata) dari data asli. Batas data referensi yang dibuat mengikuti jarak terjauh titik data pada data asli.
2. Lakukan klasterisasi pada data referensi dengan batas sampai  $n\_references$  (nilai dapat disesuaikan). Hitung *euclidean distance* untuk setiap subset hasil klaster dan rata-ratakan. Ulangi sampai  $n\_references$  dan jumlahkan kemudian dirata-ratakan. Nilai ini dimasukkan ke *array* baris ke 1. Lakukan sampai batas  $k\_max$  (batas jumlah klaster yang ingin dicari nilai *gap statistic*-nya) sehingga hasilnya berupa *array* dengan panjang  $k\_max$ .
3. Lakukan klasterisasi pada data asli. Hitung *euclidean distance* untuk setiap subset hasil klaster. Setelah itu hasilnya langsung masuk ke *array* baris ke 1, dilakukan sampai batas  $k\_max$  sehingga tercipta juga *array* dengan panjang  $k\_max$ .
4. Hitung nilai logaritma setiap *item* dari dua *array* (data referensi dan data asli).
5. Kurangi setiap *item* pada *array* hasil data referensi dengan *array* hasil data asli. Didapatkan *array* baru berisi nilai *gap* dan ambil yang paling besar.

### c. Davies Bouldin Index

*Davies bouldin index* (DBI) adalah metode evaluasi untuk mengukur validitas kluster. Pengukuran DBI dilakukan dengan mengukur nilai maksimal jarak *inter-cluster* dan nilai minimal jarak antar kluster. Apabila jarak *inter-cluster* maksimal, maka karakteristik masing-masing kluster akan terlihat jelas. Menurut Bates dan Kalita [4], semakin kecil nilai DBI yang diperoleh atau semakin mendekati nol, maka semakin baik kluster tersebut dalam kepadatan dan separasi. Rumus DBI dapat dilihat pada persamaan (2).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2)$$

#### d. Model RFM

Model RFM digunakan untuk analisis perilaku pelanggan [8]. Penjelasan setiap fitur dari model tersebut sebagai berikut.

1. *Recency*  
Tanggal pembelian terakhir pelanggan dan tanggal referensi pada periode penelitian. Pelanggan yang baik, yang berulang kali mengunjungi perusahaan akan memiliki nilai yang kecil.
2. *Frequency*  
Mewakili berapa kali pembelian yang dilakukan oleh pelanggan. Pelanggan dianggap setia asalkan nilai frekuensinya besar.
3. *Monetary*  
Mewakili jumlah total yang dibelanjakan pelanggan selama periode penelitian. Semakin besar, semakin banyak uang yang dihasilkan perusahaan.

#### e. Analytical Hierarchy Process dan Customer Lifetime Value

*Analytical hierarchy process* (AHP) adalah sistem pendukung keputusan dengan sebuah hierarki fungsional dengan input utamanya persepsi manusia [2]. Dengan hierarki, suatu masalah kompleks dan tidak terstruktur dipecahkan ke dalam kelompok-kelompok tersebut diatur menjadi suatu bentuk hierarki. Persepsi manusia yang di input adalah setujunya suatu fitur RFM dalam menentukan loyalitas suatu pelanggan. Hasilnya adalah pembobotan setiap fitur yang ada. *Customer Lifetime Value* (CLV) digunakan untuk mengukur profitabilitas dari setiap pelanggan. Penggunaan CLV dapat membantu perusahaan untuk menentukan strategi terbaik untuk mencetak laba (Kurniawan, dkk 2021). Perhitungan CLV dilakukan setelah menyelesaikan proses klusterisasi. Setelah itu untuk setiap kluster akan dirangking. Rumus CLV untuk suatu kluster  $k$  ditampilkan pada persamaan (3).

$$CLV_k = W_r C_r^k + W_f C_f^k + W_m C_m^k \quad (3)$$

Keterangan

$CLV_k$  : Nilai CLV untuk kluster  $k$

$C_r^k, C_f^k, C_m^k$  : Nilai rata-rata R, F, M untuk kluster  $k$

$W_r, W_f, W_m$  : Bobot yang dihasilkan dari AHP

### 3. Hasil dan Pembahasan

#### 3.1. Data dan Model RFM

Data yang didapatkan adalah data primer historis transaksi pelanggan dari salah satu perusahaan OTA berlokasi di Bali dari Juni 2022 hingga Juni 2023. Data tersebut saat digabungkan berjumlah 10946 baris. Cuplikan singkat datanya berupa nama kolom dan isi data dapat dilihat pada tabel 1.

**Tabel 1.** Cuplikan Data Historis Transaksi

<i>id</i>	<i>user_id</i>	<i>price</i>	<i>payment_date</i>	<i>currency</i>	<i>type</i>
50662	16923	169	01/06/2022 03:19	AUD	flight
50686	26945	3018	01/06/2022 07:27	AUD	flight
50724	42968	1711	01/06/2022 10:40	AUD	flight

50821	37038	141	02/06/2022 10:22	AUD	flight
50844	43410	4196	02/06/2022 13:15	AUD	flight
50858	43424	170	02/06/2022 21:06	AUD	flight
...	...	...	...	...	...

Atribut yang diproitasakan untuk transformasi data adalah kolom *price*, *payment\_date*, dan *user\_id*. Model RFM didapatkan dari selisih tanggal sekarang dengan *payment\_date*, jumlah transaksi yang pernah dilakukan *user\_id*, total *price* setiap *user\_id*. Kode *sql* yang digunakan untuk transformasi adalah pada tabel 2. Kode ini melakukan pembuatan tabel dengan pengelompokkan berdasarkan *user\_id* setelah itu diambil dari tabel *combined\_transactions* berupa *user\_id*, *recency* dari perbedaan tanggal dari tanggal sekarang dengan tanggal maksimal yang ada pada *user\_id* tersebut, *frequency* dari jumlah id yang ada pada *user\_id* tersebut, *monetary* dari jumlah *price* yang ada pada *user\_id* tersebut.

**Tabel 2.** Kode Transformasi RFM

```
CREATE TABLE rfm_transactions AS SELECT user_id, DATEDIFF(CURDATE(),
MAX(payment_date)) AS recency, COUNT(id) AS frequency, SUM(price) AS monetary
FROM combined_transactions GROUP BY user_id
```

Dari hasil tabel baru tersebut, didapatkan statistik datanya untuk setiap fitur RFM (*recency*, *frequency*, *monetary*) berupa jumlah, rata-rata, standar deviasi, nilai terkecil, kuartil, nilai terbesar. Statistik tersebut ditunjukkan oleh tabel 3. Cuplikan data pada tabel baru bernama *rfm\_transactions* ditunjukkan oleh tabel 4 yang berisi tipe data numerik untuk fitur RFM-nya.

**Tabel 3.** Statistik Deskriptif

	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
jumlah	7808	7808	7808
rata-rata	402.626537	2.292008	1430.975282
standar deviasi	114.431750	2.261620	2155.253033
nilai terkecil	162.000000	1.000000	5.000000
kuartil 1 (25%)	304.000000	1.000000	474.000000
kuartil 2 (50%)	395.000000	2.000000	840.500000
kuartil 3 (75%)	501.000000	2.000000	1612.500000
nilai terbesar	612.000000	54.000000	69994.000000

**Tabel 4.** Cuplikan Data RFM

<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
97	272	20	23135
443	326	3	1749
487	265	2	5540
619	274	4	3224
633	410	1	996
639	566	1	222
...	...	...	...

Setelah melalui proses transformasi, selanjutnya adalah normalisasi data. Normalisasi data dapat membantu proses klasterisasi agar lebih cepat. Metode *min-max* digunakan untuk proses nilai minimum

dan maksimum dari setiap variabel sehingga hasilnya berada di rentang 0 sampai 1. Ilustrasi proses diberikan untuk *user\_id* 97 (1) dan *user\_id* 443 (2) untuk kolom *recency*.

$$x'_1 = \frac{X_1 - X_{min}}{X_{max} - X_{min}} = \frac{612 - 162}{272 - 162} = 0,24444$$

$$x'_2 = \frac{X_1 - X_{min}}{X_{max} - X_{min}} = \frac{326 - 162}{612 - 162} = 0,36444$$

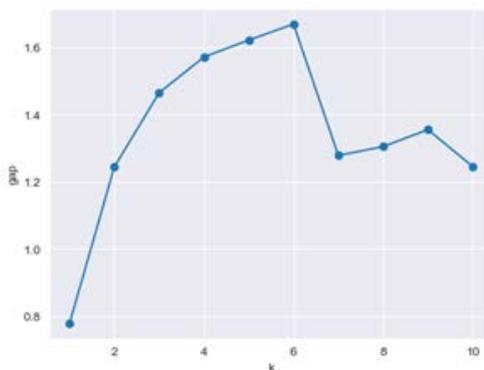
Didapatkan hasil normalisasi masing-masing 0,24444 dan 0,36444 untuk *user\_id* 97 dan 443 untuk kolom *recency*. Hasil cuplikan beberapa data setelah normalisasi *min-max* yang datanya terlihat memiliki rentang di antara 0 hingga 1 ditunjukkan oleh tabel 5.

**Tabel 5.** Cuplikan Tabel Normalisasi RFM

	<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
0	97	0.244444	0.358491	0.330481
1	443	0.364444	0.037736	0.024918
2	487	0.228889	0.018868	0.079084
3	619	0.248889	0.056604	0.045993
4	633	0.551111	0.000000	0.014159
5	639	0.897778	0.000000	0.003100
6	650	0.175556	0.018868	0.043435
...	...	...	...	...

### 3.2. Klasterisasi dengan *K-Means* dan *Gap Statistic*

Klasterisasi pelanggan menggunakan salah satu metode klasterisasi non hirarki *k-means* yang akan membagi data menjadi beberapa klaster tertentu. Penentuan jumlah klaster yang optimal akan dibantu dengan metode *gap statistic* yang mengambil selisih terbesar dari nilai total jarak *euclidean* seluruh klaster untuk data referensi distribusi dengan data asli. Hasilnya jumlah klaster (*k*) optimal yang didapatkan adalah 6 seperti pada gambar 1. Hasil optimal dapat terlihat dari nilai *gap* tertinggi yang didapatkan.



**Gambar 1.** Grafik Hasil Penentuan *k* Optimal *Gap Statistic*

Klasterisasi akan dilakukan menggunakan *method KMeans* pada *library* *sklearn* di *python*. Kode yang digunakan ditunjukkan oleh tabel 6. Kode ini melakukan pembuatan objek *kmeans* yang selanjutnya inisialisasi parameter jumlah klaster sebanyak 6 di *n\_clusters*. Setelah itu, data yang telah dinormalisasi diubah terlebih dahulu ke tipe *numpy* yang sebelumnya bertipe *dataframe*. Setelah itu, dipanggil metode *fit\_predict* untuk melakukan klasterisasi *k-means* tersebut.

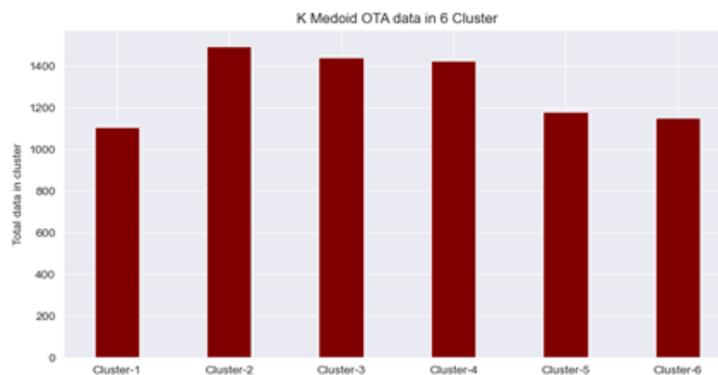
**Tabel 6.** Kode Klasterisasi dengan *Sklearn*

```

from sklearn.cluster import KMeans
kmeans = KMeans()
kmeans.n_clusters = 6
data_target_array = df_normalized[['recency', 'frequency',
'monetary']].to_numpy()
cluster_no_every_data = kmeans.fit_predict(data_target_array)

```

Hasil klasterisasi menciptakan 6 klaster dengan pembagian jumlah tiap klaster yang merata ditampilkan pada gambar 2. Hasil cuplikan data setelah melakukan klasterisasi ditampilkan pada gambar 3. Terlihat data berupa kolom *user\_id*, kolom RFM, serta nomor klaster yang didapatkan.

**Gambar 2.** Jumlah Data Setiap Klaster**Tabel 7.** Cuplikan Hasil Klasterisasi

	<i>user_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>	<i>cluster</i>
0	97	0.244444	0.358491	0.330481	6
1	443	0.364444	0.037736	0.024918	2
2	487	0.228889	0.018868	0.079084	6
3	619	0.248889	0.056604	0.045993	6
4	633	0.551111	0.000000	0.014159	3
5	639	0.897778	0.000000	0.003100	4
6	650	0.175556	0.018868	0.043435	6
...	...	...	...	...	...

Setelah mendapatkan nomor klaster untuk setiap data, selanjutnya dilakukan validasi jumlah klaster dengan *davies bouldin index* untuk menentukan hasil klaster memiliki kepadatan yang baik dan terseparasi. Semakin mendekati nol, maka semakin baik hasil klasterisasi. Penentuan nilai menggunakan metode *davies\_bouldin\_index* dari *library perimetrics* pada bahasa pemrograman *python*. Hasilnya, jumlah klaster 6 mendapatkan nilai 0,45419 yang mendekati nol sehingga masih bisa dibilang klaster memiliki kepadatan baik dan terseparasi dengan baik.

### 3.3. Pembobotan AHP dan Perangkingan CLV

AHP mengambil keputusan secara hirarki yang dipilih dari berbagai kriteria dan alternatif. Dalam permasalahan klasterisasi, hierarki yang dievaluasi adalah model RFM. Dalam hal ini, peneliti menggunakan bobot yang sama dengan penelitian [11] dengan nilai  $W_r$ ,  $W_f$ ,  $W_m$  yaitu 0.058, 0.546, dan 0.395. Penentuan nilai CLV tertinggi akan dilakukan dengan penjumlahan hasil perkalian setiap

rata-rata fitur model dengan setiap bobot. Rata-rata setiap fitur kluster ditampilkan pada tabel 8. Hasil perkalian rata-rata fitur dengan bobot fitur ditampilkal pada tabel 9.

**Tabel 8.** Rata-rata Nilai Fitur per Kluster

<i>cluster</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
1	0.929360	0.018136	0.015502
2	0.189759	0.035291	0.023710
3	0.492069	0.021958	0.018664
4	0.343677	0.028830	0.024190
5	0.784051	0.019539	0.017239
6	0.637803	0.018672	0.021350
Rata-rata	0.56278638	0.02373775	0.02010918

**Tabel 9.** Perangkingan Customer Lifetime Value (CLV)

No cluster	$W_r C_r$	$W_f C_f$	$W_m C_m$	CLV	Peringkat
1	0.0053820800	0.009901456	0.00612439	0.0214	6
2	0.0011005820	0.019283686	0.00934895	0.0297	1
3	0.0028510022	0.011978188	0.00736328	0.0222	4
4	0.0019940986	0.015743580	0.00952755	0.0273	2
5	0.0045419458	0.010659294	0.00680005	0.0220	5
6	0.0036960474	0.010178112	0.00841675	0.0223	3
Rata-rata	0.00292	0.0139	0.00712	0.0249	

Didapatkan hasil kluster dengan nilai CLV dan peringkatnya. Pemeringkatan dilakukan untuk mengetahui kluster yang memiliki nilai hidup pelanggan tertinggi. Hasilnya, kluster 2 memiliki nilai CLV tertinggi sebesar 0,0297 yang berarti kluster 2 merupakan pelanggan dengan nilai loyalitas tertinggi.

#### 4. Kesimpulan

Klasterisasi data pelanggan OTA untuk Juni 2022 hingga Juni 2023 menggunakan *k-means* dan *gap statistic* diperoleh jumlah kluster optimal sebanyak 6, dari kluster kecil masing-masing jumlahnya 1109, 1497, 1441, 1428, 1180, 1153. Hasil validasi dari *davies bouldin index* juga nilai 0,45419 yang artinya mendekati 0 sehingga dapat dikatakan kluster terseparasi dengan baik dan memiliki kepadatan yang baik.

Nilai *customer lifetime value* (CLV) yang tertinggi didapatkan pada kluster 2 yaitu 0.0297. Nilai fitur *frequency* pada kluster 2 juga merupakan nilai tertinggi. Hal ini menandakan kluster 2 merupakan pelanggan dengan nilai loyalitas yang tertinggi. Perusahaan dapat merancang strategi untuk mempertahankan pelanggan dengan program-program seperti pemberian diskon atau poin sehingga pelanggan sehingga keuntungan meningkat. Sedangkan, untuk nilai CLV terendah ada pada kluster 1, yaitu 0.0214. Perusahaan dapat membuat *personalized product* yaitu menampilkan produk-produk terkait untuk diberikan kepada pelanggan kluster 1 sehingga mereka akan lebih menjelajahi produk tersebut.

#### Referensi

- [1] A. R. Kurniawan, R. Mayasari and J. Febriantoko, "CUSTOMER LIFETIME VALUE METRICS AS CUSTOMER RELATIONSHIP MANAGEMENT INDICATORS" *International Journal of Economics, Business and Accounting Research (IJEBAR)*, vol 5, no. 3, p. 956-966, 2021.

- [2] A. Natalis, Y. Natalian, "Pemanfaatan k-Means Clustering dan Analytic Hierarchy Process terhadap Penilaian Prestasi Kerja Pegawai" *Jurnal Teknik Informatika dan Sistem Informasi*, vol 8, no. -, p. 88 – 99, 2022.
- [3] B. Chong, "K-means clustering algorithm: a brief review" *Academic Journal of Computing & Information Science*, vol. 4, no. 5, p. 37-40, 2021.
- [4] Bates, Andrew and K. Jugal, "Counting clusters in twitter posts" *Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies*, vol. -, no. -, p. 85, 2016
- [5] E. A. Pratomo, M. Najib and H. Mulyati, "CUSTOMER SEGMENTATION ANALYSIS BASED ON THE CUSTOMER LIFETIME VALUE METHOD" *Journal of Applied Management*. vol. 17, no. 3, p. 408 – 415, 2019.
- [6] Suyanto, MACHINE LEARNING TINGKAT DASAR DAN LANJUT, 1 ed., Bandung: Informatika, 2018, pp. –
- [7] R. Tibshirani, G. Walther and Hastie, "Estimating the number of clusters in a data set via the gap statistic" *J. R. Statist. Soc. B*, vol. 63, no. 2, p. 411–423, 2001.
- [8] W. A. Taqwim, N. Y. Setiawan and F. A. Bachtiar, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, p. 1986 -1993, 2019.
- [9] Y. Asriningtias, J. Aryanto, "K-Means Algorithm with Davies Bouldin Criteria for Clustering Provinces in Indonesia Based on Number of Events and Impacts of Natural Disasters" *International Journal of Engineering, Technology and Natural Sciences*, vol. 4, no. 1, p. 75 – 80, 2022.
- [10] Z. Liu, J. Zeng and J. Zhou, "Impact of COVID-19 on OTA Platforms and Development: A Case Study of Ctrip" *International Conference on Economic Management and Green Development*, vol. -, no. -, p. 55-60, 2023.
- [11] S. Monalisa, P. Nadya and R. Novita, "Analysis for Customer Lifetime Value Categorization with RFM Model" *The Fifth Information Systems International Conference*, vol. 161, no. -, p. 834 – 840, 2019.

# Forecasting Saham Perbankan Dengan Penerapan Multilayer Backpropagation Neural Network

I Putu Ryan Paramaditya<sup>a1</sup>, Cokorda Rai Adi Prammartha<sup>a2</sup>, I Gede Arta Wibawa<sup>a3</sup>,  
I Gede Santi Astawa<sup>a4</sup>, Ida Bagus Gede Dwidasmara<sup>a5</sup>, I Dewa Made Bayu Atmaja Darmawan<sup>a6</sup>

<sup>a</sup>Informatics Department, Faculty of Mathematics and Natural Sciences,  
Udayana University  
Badung, Bali, Indonesia

<sup>1</sup>ryanparamaditya@gmail.com

<sup>2</sup>cokorda@unud.ac.id

<sup>3</sup>gede.arta@unud.ac.id

<sup>4</sup>santi.astawa@unud.ac.id

<sup>5</sup>dwidasmara@unud.ac.id

<sup>6</sup>dewabayu@unud.ac.id

## Abstract

*The use of the Neural Network algorithm with Backpropagation is used to predict stock price data based on the closing price of the following day, as a reference for buying shares in the future. The time-series dataset that was utilized includes stock data from Yahoo Finance for state-owned banks including Bank BNI (BBNI). Based on the findings from the model training, the lowest loss was 0.0011 at epoch 29, 33, 41, 43, 46, 47, and 49 and the highest was 0.0243 at epoch 0. The lowest Val Loss was 0.0011 at epoch 5, 10, and 46 and the highest was 9.555 at epoch 44. The model test score results showed a Median Absolute Error (MAE) of 85.57 and a Mean Absolute Error Percent (MAE%) of 1.97%. Root Mean Squared Error (RMSE) is 103.85 and Root Mean Squared Error Percent (RMSE%) is 2.39%. This score is considered good because it is below 50%. Prediction results reach an average of above 90%. To get the best prediction results, the percent change must be above -4.35% and the percentage above 95.65%.*

**Keywords:** Stocks, Forecasting, Time Series, Close Price, Neural Network, Backpropagation.

## 1. Introduction

Saham merupakan jenis investasi yang salah satunya populer di seluruh dunia, termasuk Indonesia yang berkontribusi pada pertumbuhan ekonomi dan pendapatan negara, dengan berbagai jenis saham seperti indeks, reksadana, bahkan ETF [1]. Saham memiliki tingkat risiko yang cukup rendah, dan volatilitas yang masih dapat dipertahankan [2], asalkan kita memahami cara berinvestasinya. Indonesia mempunyai aturan dan jaminan untuk investasi saham. Dengan membeli saham, kita dapat menjaga perputaran uang yang kita miliki dan dapat melipatgandakan aset kita [3]. Berdasarkan pada data Kustodian Sentral Efek Indonesia (KSEI) bahwa saat ini terdapat 4 juta investor saham di pasar modal Tanah Air. Pada akhir Juni 2022, terdapat 4.002.289 investor saham, naik 15,96% dari 3.451.513 pada akhir tahun 2021. Pada akhir semester I 2022, terdapat 81,64% investor saham berusia di bawah 40 tahun yang mendominasi, terutama kalangan milenial dan generasi Z, dengan nilai aset sebesar Rp 144,07 triliun [4].

Pada penelitian sebelumnya, telah dilakukan oleh Ramadhan dan Pamuji tahun 2022 dalam membandingkan penggunaan *Neural Network* dan Regresi Linier dalam analisis algoritma *Forecasting* untuk memprediksi harga saham BMRI. Berdasarkan data yang dikumpulkan tersebut, bahwa *Neural Networks* merupakan prediktor yang lebih akurat dibandingkan Regresi Linier [5]. Kemudian, penelitian yang dilakukan Saifuddin dan Hermawan tahun 2019 menggunakan *Artificial Neural Network* dalam melakukan prediksi harga saham BBRI dengan memanfaatkan empat atribut: nilai *open*, *high*, *low* sebagai prediktor dan *close* sebagai kelas. Maka algoritma *Artificial Neural Network* mampu menghasilkan prediksi yang terbaik dengan mengoptimalkan nilai parameter yang diberikan untuk mengantisipasi harga saham Bank BRI dengan jumlah 16 neuron, dan nilai RMSE program adalah 0,266 +/- 0,000 [6]. Penelitian yang

pernah dilakukan oleh Santoso dan Hansun tahun 2019 bertujuan dalam peramalan IHSG dengan *Backpropagation Neural Network*. Bahwa pendekatan peramalan *Backpropagation* menggunakan nilai pengujian MSE minimum, sehingga memberikan hasil peramalan yang cukup baik. Dimana hasil pengujian MSE sebesar 320.49865083640924, dan learning rate yang digunakan sebesar 0.3 [7].

Dari berbagai penelitian tersebut, permasalahan yang dihadapi yaitu, kompleksitas data yang dimana data harga saham memerlukan metode yang sesuai untuk melakukan suatu prediksi. Pada akurasi prediksi yang masih bisa ditingkatkan sehingga hasil prediksi mendekati dengan nilai aktualnya yang mempengaruhi hasil *forecasting*. serta memperhatikan faktor eksternal yang mempengaruhi pergerakan harga saham. Maka dari itu, penelitian ini dilakukan untuk mengoptimalkan kompleksitas data yang bermula proses pengambilan data secara manual menjadi otomatisasi melalui API sehingga data harga saham dapat yang terbaru secara terus-menerus. Kemudian penggunaan metode *Backpropagation Neural Network* diharapkan dapat memperbaiki tingkat akurasi dalam melakukan prediksi yang mendekati nilai sebenarnya dengan tingkat eror yang rendah.

*Forecasting* adalah salah satu aplikasi umum dari analisis deret waktu. digunakan banyak skenario aplikasi dalam penjualan ritel, indikator ekonomi, perkiraan cuaca, pasar saham, dan lainnya [8]. *Neural Network* digunakan untuk mengenali pola dan hubungan pada data dalam jumlah besar, dan algoritma tersebut dapat diterapkan untuk memprediksi kenaikan dan penurunan harga saham [9]. *Backpropagation* merupakan salah satu metode yang digunakan pada *neural network* untuk proses pelatihan yang unggul dalam kemampuan pembelajarannya [10]. Dalam melakukan regresi pada *neural network* menggunakan pendekatan secara *non-linear regression* yang disebut sebagai *General Regression Neural Network* (GRNN). Dimana estimasi pada nilai output ditentukan oleh nilai input [11]. LSTM (*Long Short Term Memory*) digunakan dalam mempelajari dependensi jangka panjang yang sangat tepat untuk mengklasifikasikan, pemrosesan, dan pembuatan prediksi dengan menggunakan data deret waktu [12].

## 2. Reseach Methods

### 2.1. Akuisisi Data

Tabel 2.1 Dataset Saham Bank BNI (BBNI)

Date	Open	High	Low	Close	Adj Close	Volume
01/01/2018	4950	4950	4950	4950	3651,7073	0
02/01/2018	4950	5000	4837,5	4887,5	3605,6001	24868000
03/01/2018	4887,5	4900	4687,5	4712,5	3476,4995	53786400
04/01/2018	4675	4712,5	4575	4700	3467,2783	50780200
05/01/2018	4700	4725	4625	4650	3430,3921	46092600
...	...	...	...	...	...	...
21/02/2023	4575	4600	4525	4525	4150,8252	72588200
22/02/2023	4500	4525	4400	4425	4059,0945	92300400
23/02/2023	4425	4500	4400	4475	4104,96	52724200
24/02/2023	4500	4562,5	4450	4450	4082,0271	53277000
27/02/2023	4450	4500	4425	4475	4104,96	48120000

Dilihat pada Tabel 2.1 dimana merupakan data harga saham Bank BNI (BBNI) kategori perbankan BUMN berasal dari *Yahoo Finance*, pada 01 Januari 2018 hingga tanggal 28 Februari 2023 dengan jumlah data 1289 data. Dalam penelitian ini, menerapkan metode Kuantitatif pada dataset yang digunakan yaitu bertipe *time-series* karena berupa data harga perdagangan saham yang terdiri atas waktu transaksi, harga buka (*Open*), harga tertinggi (*High*), harga terendah (*Low*), harga tutup (*Close*), harga tutup yang disesuaikan (*Adj Close*), dan volume transaksi.

## 2.2. Preprocessing Data



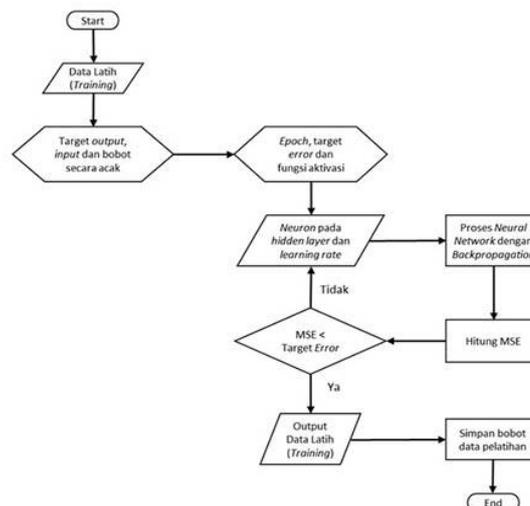
Gambar 2.1 Alur proses dari *pre-processing* data dengan *MinMaxScaler*.

Pada *flowchart* yang dilihat pada Gambar 2.1 dalam melakukan *pre-processing* data yang diawali dengan pemilihan fitur yang dibutuhkan pada dataset. Proses normalisasi data dengan *MinMaxScaler* pada data harga yang diset pada kisaran antara 0 dan 1. Kemudian melakukan pembagian data menjadi data latih (*training*) sebesar 80% dan 20% data lainnya sebagai data uji (*testing*) yang merujuk pada penelitian dari R. Meidyta [13].

## 2.3. Pelatihan Model

Dalam melakukan pelatihan model pada *Neural Network* dengan metode *Backpropagation* yang terlihat pada Gambar 2.2, dimana alur proses pelatihan model dijelaskan sebagai berikut ini [13]:

- Inisialisasikan target *output*, *input*, dan bobot secara acak.
- Menetapkan *epoch* maksimum, target *error* dan fungsi aktivasi pada *hidden* dan *output layer*.
- Masukan jumlah *neuron* pada *hidden layer* dan *learning rate*.
- Menghitung MSE untuk pengujian *neuron* pada *hidden layer*
- Jika nilai MSE < target *error*, maka prosedur pelatihan akan berakhir.
- Simpan bobot data pelatihan untuk proses pengujian.



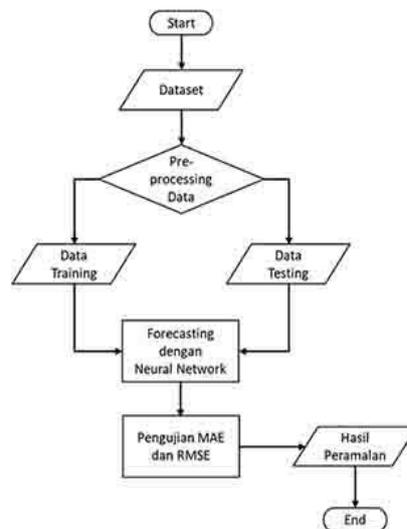
Gambar 2.2 Alur proses dari model NN dengan metode *Backpropagation*.

## 2.4. Pengujian Model

Melakukan proses pengujian pada model NN dengan metode *Backpropagation* dengan cara Inialisasikan input dari data pengujian, lalu menetapkan bobot hasil pelatihan, kemudian lakukan pengujian data. Setelah adanya hasil perkiraan data pengujian, dilakukan perhitungan nilai keakuratan suatu model prediksi yang diterapkan dalam memprediksi saham tersebut dengan MAE dan RMSE. Semakin kecil nilai MAE dan RMSE, maka semakin baik dalam melakukan akurasi pada model yang digunakan [14] [15].

## 2.5. Gambaran Alur Penelitian

Pada *flowchart* yang terdapat pada Gambar 2.3, proses metode dalam membangun program tersebut dalam memprediksi saham tersebut, Pertama mengumpulkan data-data yang digunakan pada penelitian, kemudian melakukan *preprocessing data* kepada data-data tersebut dan memisahkan datanya sebagai data latih (*training*) dan data uji (*testing*), kemudian membangun program dengan menerapkan algoritma *Neural Network* dalam melakukan proses prediksi (*forecasting*), dengan *backpropagation*. Melakukan pengujian error dengan MAE dan RMSE pada masing-masing data saham yang sudah dilakukan prediksi yang ditampilkan dengan hasil angka dan grafik [16].



Gambar 2.3 Alur penelitian prediksi data saham.

## 2.6. Implementasi Sistem

Pada implementasi *Forecasting* yang digunakan bahasa pemrograman *Python* dengan framework *Jupyter Notebook* dan *Streamlit* sehingga python bisa berjalan sebagai back-end pada sistem berbasis web dengan *localhost*

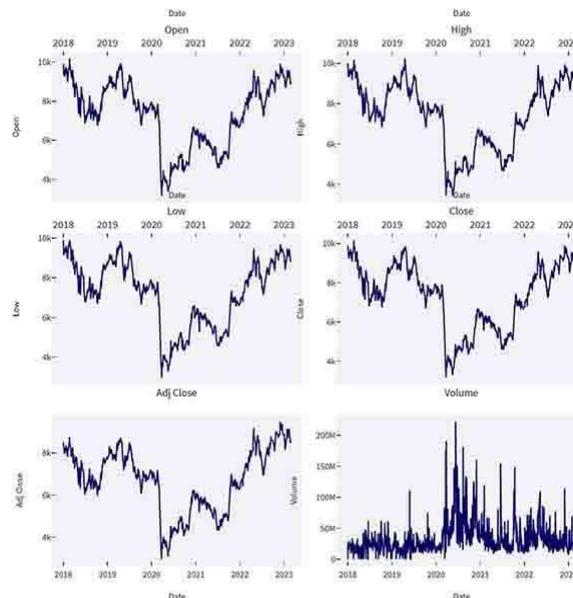
## 3. Result and Discussion

### 3.1. Hasil Akuisisi Data

Pada hasil mengakuisisi data tersebut, dapat melihat dataset secara keseluruhan yang telah dieksplorasi sebelum melakukan pengolahan data. Data tersebut yang digunakan dari data harga buka, tertinggi, terendah dan harga tutup pada saham Bank BNI (BBNI.JK) dengan periode tahun 2018 sampai awal 2023 dengan jumlah data 1289 data. Melakukan eksplorasi data pada dataset BBNI.JK untuk mengetahui perbandingan keseluruhan harga saham dari tahun 2018 hingga tahun saat ini, baik itu harga buka (*open*), tinggi (*high*), rendah (*low*), tutup (*close*), *adj close* dan juga volume transaksi, dalam bentuk plot grafik.

Dilihat pada Gambar 3.1 menunjukkan data harga saham, bahwa harga buka (*open*) tertinggi adalah 10.175 pada periode Februari 2018, dan terendah adalah 3.160 pada periode Maret 2020. Harga tinggi (*high*) tertinggi adalah 10.250 pada periode April 2019, dan terendah adalah 3.410 pada periode Mei 2020. Harga rendah (*low*) tertinggi adalah 9.900 pada periode Desember 2017 dan Februari 2018, dan terendah adalah 2.970 pada periode Maret 2020. Harga tutup (*close*) tertinggi adalah 10.175 pada periode Mei 2018, dan terendah adalah 3.160

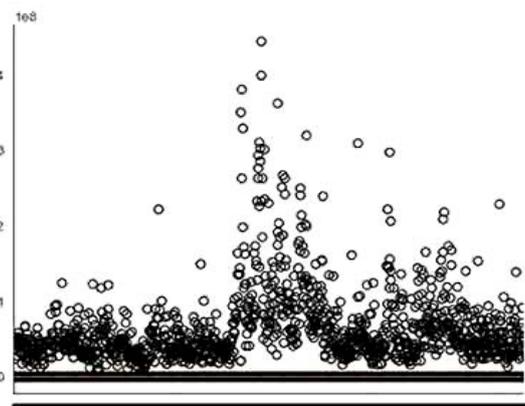
pada periode Maret 2020. *Adj Close* tertinggi adalah 9.900 pada periode November 2022 dan terendah adalah 2.951 pada periode Maret 2020. Pada *volume* transaksi tertinggi adalah 222 juta pada periode Juni 2020.



Gambar 3.1 Grafik harga buka (*open*), rendah (*low*), tutup (*close*), *Adj Close* dan *Volume*.

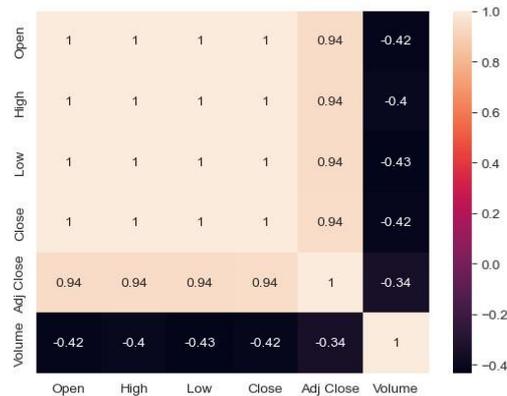
### 3.2. Preprocessing Data

Pada tahapan *preprocessing* data hal pertama yang dilakukan yaitu, Melihat persebaran data yang dihasilkan berdasarkan pada hasil akuisisi data, dilihat pada Gambar 3.2 masih banyaknya data yang bernilai nol pada tabel data ketika data awalan diambil dengan API dari *library yahoo finance* yang digambarkan pada pola grafik. Pada titik data yang mewakili nilai individu dalam set data masih terkonsentrasi di bagian bawah grafik pada garis referensi dengan posisi antara nilai 0 dan 1, sementara pada titik data yang lain tersebar lebih tinggi di sumbu y hingga posisi nilai 4 yang menunjukkan juga terdapat variasi dalam data.



Gambar 3.2 Plot persebaran data pada harga saham sebelum dinormalisasi.

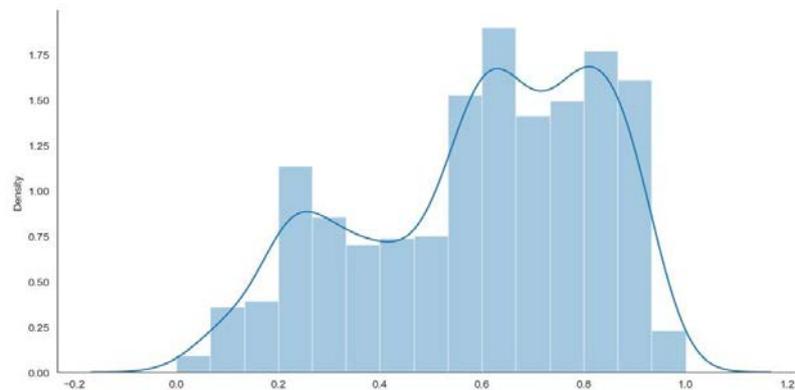
Melakukan korelasi data pada data saham Bank BNI (BBNI.JK) dilihat pada Gambar 3.3, dengan data dan fitur data yang digunakan yaitu *Open*, *High*, *Low*, *Close* memiliki nilai korelasi 1, pada *Adj Close* memiliki nilai korelasi keseluruhan 0.94 yaitu sangat baik, sedangkan untuk *Volume* memiliki nilai korelasi keseluruhan antara -0.44 hingga -0.39. Sehingga disimpulkan bahwa dalam melakukan prediksi harga saham tersebut menggunakan fitur yaitu *Open*, *High*, *Low*, *Close* dan *Adj Close*. Maka pada Gambar 3.3, pemilihan fitur yang berdasarkan pada pengujian korelasi data dalam menentukan fitur *Open*, *High*, *Low*, *Close* dan *Adj Close* digunakan dalam hal ini tidak akan merubah isi dari dataset tersebut dengan tetap terlihat sederhana tanpa ada penambahan apapun.



Gambar 3.3 Matriks korelasi data.

Kemudian, melakukan normalisasi data dengan *MinMaxScaler* untuk normalisasi dataset saham berdasarkan harga tutup (*Close*) dengan kisaran 0 dan 1. Dalam melakukan skalalisasi pada hasil prediksi setelah pelatihan, dengan membuat dua *scaller* yang berbeda, pertama satu data latih yang membutuhkan lima kolom dan satu untuk data *output* yang menskalakan satu kolom pada harga tutup (*Close*).

Berikut hasil dari normalisasi data dengan *MinMaxScaler* pada Gambar 3.4 tersebut. Pada sumbu X ini merupakan nilai data yang dinormalisasi berkisar dari -0.2 hingga 1.2, dimana plot densitas merujuk pada distribusi nilai data yang sudah dinormalisasi berada antara 0.0 hingga 1.0. Pada sumbu Y berkisar dari 0 hingga 1.75 menunjukkan jumlah atau frekuensi nilai data tertentu muncul, dimana dengan jumlah nilai yang dinormalisasi sebesar diatas 1.75 pada nilai data sekitar 0.6 dari kisaran 0 dan 1 dengan *MinMaxScaler*.



Gambar 3.4 Plot grafik data pada harga saham setelah dinormalisasi.

Selanjutnya, melakukan pembagian dataset menjadi data latih (*training*) sebesar 80% dan 20% sebagai data uji (*testing*). Kemudian, melakukan prediksi yang ditentukan dari berdasarkan harga pasar pada 60 hari terakhir. Kemudian membagi beberapa data latih ( $x_{train}$ ) menjadi beberapa batch kecil yang akan diproses dengan jaringan saraf tiruan satu persatu selama proses training. Untuk mengurangi kesalahan prediksi pada algoritma dengan menyesuaikan koneksi dengan neuron yang diproses. Serta menambahkan data latih kedua ( $y_{train}$ ) untuk mengevaluasi hasil prediksi dengan data latih pertama ( $x_{train}$ ).

### 3.3. Penggunaan *Neural Network Backpropagation*

Menentukan lapisan pada neural network dengan menggunakan *input layer* LSTM yang merupakan bagian dari *Neural Network Regression* yang dapat memproses data deret waktu diawal lapisan memiliki 60 *neuron*. Pada *hidden layer* menggunakan *Dense* dan *Dropout* sebanyak tiga layer dengan *Dense* yang diset memiliki 90 *neuron* dan *Dropout* dua *neuron* yang digunakan sebagai *backpropagation* yang terhubung pada *Dense* masing-masing *neuron*. Untuk *output layer* menggunakan *Dense* memiliki satu *neuron* yang diaktivasi dengan *sigmoid*. *Sequential*, *LSTM*, *Dense*, *Dropout* merupakan *library* pada *Keras* bagian dari *Tensorflow*.

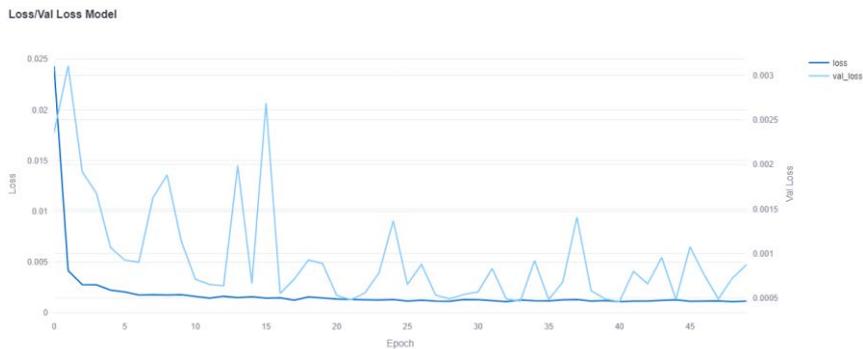
### 3.4. Pelatihan Model

Pada pelatihan model yang dilakukan setelah *compile* pada *neural network* dengan optimasi *adam* dengan *Loss/Val Loss* yang dihitung dengan *Mean Squared Error*. *Epochs* yang digunakan adalah *epoch* 50 dengan *batch size* sebesar 4, penggunaan *epoch* dan *batch size* tersebut berdasarkan penelitian dari W. W. K. Wardani [17].

Pada pengujian *Loss/Val Loss Model* untuk hasil keseluruhan dilihat pada Tabel 3.1, dimana skor yang didapatkan pada setiap *epoch* semakin rendah skornya maka semakin baik hasil prediksinya. Dari hasil pelatihan model yang dilakukan, bahwa untuk *Loss* yang terendah menghasilkan skor berada di bawah 0,005 pada *epoch* 29, 33, 41, 43, 46, 47 dan 49 sebesar 0,0011, sedangkan skor yang tertinggi di atas 0,02 pada *epoch* 0 sebesar 0,0243. Untuk *Val Loss* yang menghasilkan skor terendah pada *epoch* 5, 10, dan 46 sebesar 0,0011, sedangkan pada skor tertinggi pada *epoch* 44 sebesar 9,555. Namun, pada grafik yang ditunjukkan pada Gambar 3.5, garis yang mewakili *Loss* menghasilkan skor cenderung konsisten dari tinggi menuju rendah pada setiap *epoch*, sedangkan pada garis yang mewakili *Val Loss* menghasilkan skor cenderung tidak stabil pada setiap *epoch*.

Tabel 3.1 Hasil *Loss* dan *Val Loss Score*.

Epoch	Loss	Val_Loss
01/50	0,0243	0,0024
02/50	0,0042	0,0031
03/50	0,0028	0,0019
04/50	0,0028	0,0017
05/50	0,0022	0,0011
...	...	...
10/50	0,0018	0,0011
...	...	...
29/50	0,0011	4,91E+00
...	...	...
33/50	0,0011	4,89E+00
...	...	...
41/50	0,0011	4,59E+00
...	...	...
43/50	0,0011	6,60E+00
44/50	0,0012	9,56E+00
...	...	...
46/50	0,0011	0,0011
47/50	0,0011	7,61E+00
48/50	0,0012	4,91E+00
49/50	0,0011	7,24E+00
50/50	0,0012	8,77E+00



Gambar 3.5 Grafik *loss/Val Loss* pada model NN.

### 3.5. Pengujian Model

Hasil skor pengujian model yang dilakukan dengan menghitung *Median Absolute Error* (MAE) mendapatkan skor 85,57 dan *Mean Absolute Error Percent* (MAE%) menjadi 1,97%. Sedangkan, pada *Root Mean Squared Error* (RMSE) mendapatkan skor 103,85 dan *Root Mean Squared Error Percent* (RMSE%) menjadi 2.39%. Bahwa dari hasil skor pengujian tersebut merupakan skor yang sangat baik karena di bawah 50% menurut hasil pengujian MAE dan RMSE yang dikonversi hitung kembali dalam bentuk persentase.

**Mean Absolute Error (MAE): 85.57**

**Mean Absolute Error Percent (MAE%): 1.97 %**

**Root Mean Squared Error (RMSE): 103.85**

**Root Mean Squared Error Percent (RMSE%): 2.39 %**

### 3.6. Hasil Prediksi

Pada Gambar 3.6, grafik tersebut menampilkan perbandingan dengan data pertumbuhan saham yang sebenarnya dengan yang prediksinya. Dimana pada keterangan di bawah warna biru tua keunguan berarti *y\_train* (data latih/*training*), warna biru muda berarti *y\_pred* (data hasil prediksi), warna merah berarti *y\_test* (data uji/*testing*), warna hijau atau merah pada chart grafik batang di bawah berarti *residuals* dimana adanya nilai prediksi mendekati nilai aktual namun masih adanya penyimpangan.



Gambar 3.6 Grafik perbandingan prediksi dan data sebenarnya.

Untuk mendapatkan nilai hasil prediksi yaitu dengan memanggil harga tutup (*close*) pada variabel *y\_test*, dan ditampilkan dengan waktunya. Memanggil *prediction* pada variabel *y\_pred* untuk menampilkan data hasil prediksi. Hasil prediksi (*prediction*) dibandingkan dengan harga tutup untuk menghasilkan *residuals*, yaitu selisih antara prediksi dan harga tutup. Selanjutnya, menghitung *Percent Change*, yaitu persentase perbedaan antara harga tutup dan hasil prediksi. Jika *Percent Change* positif, berarti prediksi lebih tinggi dari harga tutup. Sebaliknya, jika negatif, berarti prediksi lebih rendah. Kemudian, menghitung persentase (*Percentage*) dengan membandingkan akurasi antara prediksi dan harga tutup. Semua hasil ini ditampilkan pada Tabel 3.2.

Tabel 3.2 Hasil prediksi pada harga tutup (*Close*).

Date	Close	Prediction	Residuals	Percent Change %	Percentage %
11/02/2022	3850	3896,0227	46,0227	1,2	101,2
14/02/2022	3850	3911,2485	61,2485	1,59	101,59
15/02/2022	3962,5	3911,0046	-51,4954	-1,3	98,7
16/02/2022	3975	4017,5144	42,5144	1,07	101,07
17/02/2022	3950	4043,8105	93,8105	2,37	102,37
...	...	...	...	...	...
12/05/2022	4100	4501,5327	401,533	9,79	109,79
...	...	...	...	...	...

30/11/2022	4950	4734,4453	-215,555	-4,35	95,65
...	...	...	...	...	...
21/02/2023	4525	4660,5981	135,598	3	103
22/02/2023	4425	4617,5483	192,548	4,35	104,35
23/02/2023	4475	4531,9019	56,9019	1,27	101,27
24/02/2023	4450	4563,1943	113,194	2,54	102,54
27/02/2023	4475	4538,7036	63,7036	1,42	101,42

Dari hasil prediksi pada Tabel 3.2, bahwa menghasilkan capaian persentase rata-rata di atas 90%. Persentase terendah berada di 95,65% dengan hasil prediksi terendah berdasarkan pada Percent change berada di -4,35% pada harga saham tanggal 30 November 2022. Persentase tertinggi berada di 109,79% dengan hasil prediksi tertinggi berdasarkan pada Percent change berada di 9,79% pada harga saham tanggal 12 Mei 2022. Sehingga untuk mendapatkan hasil prediksi yang terbaik, maka *percent change* harus berada di atas -4,35% dan persentase di atas 95,65%.

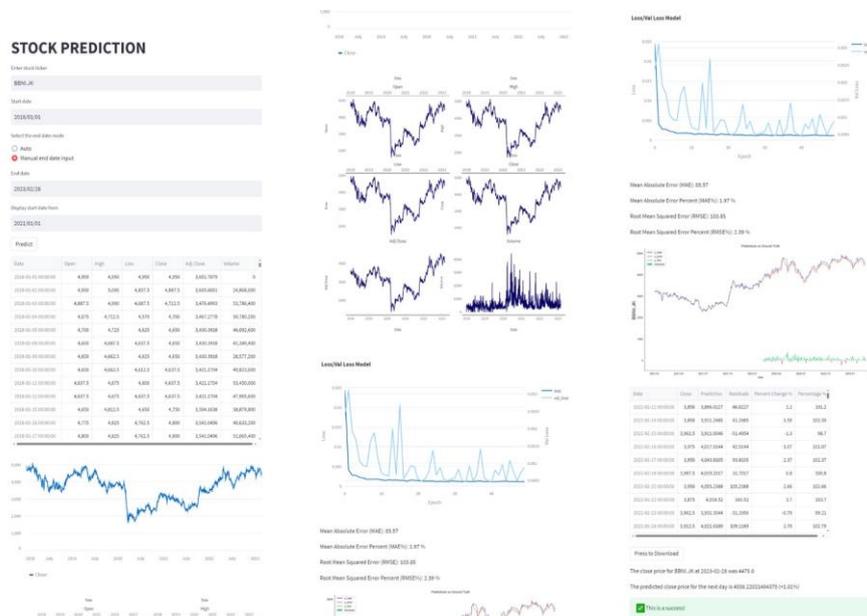
Dalam menampilkan prediksi harga tutup untuk hari berikutnya. Pertama, menampilkan harga tutup hari ini dengan mengambil harga tutup (*close*) terakhir dari *DataFrame df\_new* dan dibulatkan hingga dua angka decimal yang disimpan dalam variabel *price\_today*. Kemudian, harga prediksi hari berikutnya didapatkan dari prediksi harga pertama dari array *pred\_price\_unscaled*, meratakan array tersebut menggunakan metode *ravel*, dan membulatkannya hingga dua angka desimal. Prediksi harga ini disimpan dalam variabel *predicted\_price*. Menghitung perubahan persentase antara harga penutupan hari ini dan prediksi harga penutupan untuk hari berikutnya, dan dibulatkan hingga dua angka desimal. Perubahan persentase ini disimpan dalam variabel *percent\_change*. Pada *prefix* digunakan untuk indikator apakah harga prediksi diperkirakan naik atau tetap/turun dari harga hari ini.

**The close price for BBNI.JK at 2023-02-28 was 4475.0**  
**The predicted close price for the next day is 4556.22021484375 (+1.81%)**

Serta menampilkan harga prediksi untuk hari sekarang dan juga peramalan harga tutup pada hari berikutnya beserta dengan presentase residualsnya. Misalkan pada tanggal 28 Februari 2023 perkiraan harga tutup sebesar 4475.0 dan juga perkiraan harga tutup pada hari besoknya pada 1 Maret 2023 diperkirakan sebesar 4556.22021484375 dengan presentase +1.81%.

### 3.5. Tampilan Program

Berikut ini pada gambar 3.7 dimana bentuk tampilan dari program yang dibangun dengan berbasis web menggunakan bahasa pemrograman Python dan framework Streamlit.



Gambar 3.7 Tampilan Program.

#### 4. Kesimpulan

Berdasarkan pada penelitian yang dilakukan mengenai *Forecasting Saham Perbankan* dengan Penerapan *Multilayer Backpropagation Neural Network* dengan menggunakan dataset saham Bank BNI (BBNI.JK), bahwa didapatkan kesimpulan sebagai berikut:

1. Model yang dipergunakan adalah *Neural Network* dengan *Backpropagation* dan *input layer LSTM*. Model ini memiliki 60 *neuron* di awal lapisan dan menggunakan *Dense 90 neuron* dan *Dropout* dua *neuron* sebagai *backpropagation* di *hidden layer* sebanyak tiga *layer*. *Output layer* menggunakan *Dense* dengan satu *neuron* yang diaktivasi dengan *sigmoid*.
2. Hasil pelatihan model menunjukkan bahwa *Loss* terendah adalah 0,0011 (pada *epoch* 29, 33, 41, 43, 46, 47, dan 49) dan tertinggi adalah 0,0243 (pada *epoch* 0). *Val Loss* terendah adalah 0,0011 (pada *epoch* 5, 10, dan 46) dan tertinggi adalah 9,555 (pada *epoch* 44). Garis *Loss* cenderung konsisten dari tinggi ke rendah, sedangkan garis *Val Loss* cenderung tidak stabil.
3. Skor pengujian model menunjukkan *Median Absolute Error (MAE)* sebesar 85,57 dan *Mean Absolute Error Percent (MAE%)* sebesar 1,97%. *Root Mean Squared Error (RMSE)* adalah 103,85 dan *Root Mean Squared Error Percent (RMSE%)* adalah 2,39%. Skor ini dianggap baik karena di bawah 50%.
4. Hasil prediksi mencapai rata-rata di atas 90%. Persentase terendah adalah 95,65% (dengan prediksi terendah berdasarkan *Percent change* -4,35% pada tanggal 30 November 2022) dan persentase tertinggi adalah 109,79% (dengan prediksi tertinggi berdasarkan *Percent change* 9,79% pada tanggal 12 Mei 2022). Untuk mendapatkan hasil prediksi terbaik, *percent change* harus di atas -4,35% dan persentase di atas 95,65%.

#### References

- [1] adminlp2m, "Exchange Traded Funds (ETF) - Definisi dan Perbedaannya," Lembaga Penelitian dan Pengabdian Masyarakat. Accessed: Oct. 24, 2022. [Online]. Available: <https://lp2m.uma.ac.id/2022/10/14/exchange-traded-funds-etf-definisi-dan-perbedaannya/>
- [2] Tsm, "Investasi di Dunia Digital: Blockchain, Crypto, dan NFT," TSM. Accessed: Oct. 24, 2022. [Online]. Available: <https://www.tsm.ac.id/investasi-di-dunia-digital-blockchain-crypto-dan-nft/>
- [3] A. Bimantara, U. A. Nadhiroh, and E. Komaruljannah, "Strategi Peningkatan Daya Saing Saham dan Obligasi Syariah dalam Menghadapi Masa Pandemi dengan Metode SWOT," *J. Ilm. Ekon. Islam*, vol. 7, no. 3, Art. no. 3, Nov. 2021, doi: 10.29040/jiei.v7i3.3614.
- [4] R. M. Ramyakim and A. Widyasari, "Didominasi Milenial dan Gen Z, Jumlah Investor Saham Tembus 4 Juta," *Divisi Sekretaris Perusahaan, Komunikasi dan Edukasi*, PT Kustodian Sentral Efek Indonesia (KSEI), Jul. 09, 2022. [Online]. Available: [https://www.ksei.co.id/files/uploads/press\\_releases/press\\_file/id-id/208\\_berita\\_pers\\_didominasi\\_milenial\\_dan\\_gen\\_z\\_jumlah\\_investor\\_saham\\_tembus\\_4\\_juta\\_20220725182203.pdf](https://www.ksei.co.id/files/uploads/press_releases/press_file/id-id/208_berita_pers_didominasi_milenial_dan_gen_z_jumlah_investor_saham_tembus_4_juta_20220725182203.pdf)
- [5] V. P. Ramadhan and F. Y. Pamuji, "Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI)," *J. Teknol. Dan Manaj. Inform.*, vol. 8, no. 1, pp. 39–45, Jul. 2022, doi: 10.26905/jtmi.v8i1.6092.
- [6] Saifuddin and A. Hermawan, "Prediksi Data Historis Saham PT.Bank Rakyat Indonesia Tbk (BBRI) Menggunakan Model Algoritma Artificial Neural Network," p. 6, 2019.
- [7] A. Santoso and S. Hansun, "Prediksi IHSG dengan Backpropagation Neural Network | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)," Aug. 2019, Accessed: Oct. 13, 2022. [Online]. Available: <http://www.jurnal.iaii.or.id/index.php/RESTI/article/view/887>
- [8] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [9] A. R. Wiranto, E. Setiawan, A. Nuryaman, and M. Usman, "Implementasi Metode Backpropagation Neural Network Dalam Meramalkan Tingkat Inflasi Di Indonesia,"

- MATHunesa J. Ilm. Mat.*, vol. 11, no. 1, pp. 8–16, May 2023, doi: 10.26740/mathunesa.v11n1.p8-16.
- [10] S. Setti and A. Wanto, “Analysis of Backpropagation Algorithm in Predicting the Most Number of Internet Users in the World,” *J. Online Inform.*, vol. 3, no. 2, p. 110, Feb. 2019, doi: 10.15575/join.v3i2.205.
- [11] F. Baskoro, Y. A. Suryatna, N. Kholis, A. I. Agung, and W. Aribowo, “Peramalan Beban Puncak Menggunakan Metode Feed Forward Backpropagation Dan Generalized Regression Neural Network,” *J. Tek. ELEKTRO*, vol. 10, no. 1, pp. 109–118, 2021, doi: 10.26740/jte.v10n1.p109-118.
- [12] D. Osinga, “Deep Learning Cookbook,” *O'REILLY*, Jun. 2018.
- [13] R. Meidyta, “Prakiraan Harga Saham Menggunakan Artificial Neural Network Dengan Algoritma Backpropagation (Studi Kasus: Harga Penutupan Saham Bulanan PT. Indofood Sukses Makmur (Tbk)),” p. 64, 2021.
- [14] Khoiri, “Cara Menghitung Mean Absolute Percentage Error (MAPE).” Accessed: Oct. 13, 2022. [Online]. Available: <https://www.khoiri.com/2020/12/pengertian-dan-cara-menghitung-mean-absolute-percentage-error-mape.html>
- [15] R-Stats, “Cara Menghitung MAE (Mean Absolute Error) di Excel dan R,” Rumus Statistik. Accessed: Oct. 11, 2022. [Online]. Available: <https://www.rumusstatistik.com/2021/05/cara-menghitung-mae.html>
- [16] I. D. G. Budiastawa, Iw. Santiyasa, and C. R. A. Pramatha, “Prediksi Dan Akurasi Nilai Tukar Mata Uang Rupiah Terhadap US Dolar Menggunakan Radial Basis Function Neural Network,” vol. 7, no. 4, 2019.
- [17] W. W. K. Wardani, “Prediksi Harga Saham Syariah Menggunakan Metode Reccurent Neural Network-Long Short Term Memory,” undergraduate, UIN Sunan Ampel Surabaya, 2021. Accessed: Oct. 13, 2022. [Online]. Available: <http://digilib.uinsby.ac.id/49542/>

*This page is intentionally left blank.*

# Klasifikasi Biji Jagung Berdasarkan Tekstur Dan Warna Menggunakan Metode *Backpropagation* Berbasis Web

M. Syafiih<sup>a1</sup>, Nadiyah<sup>a2</sup>

<sup>a1</sup>Teknologi Informasi, <sup>a2</sup>Teknik Informatika, Fakultas Teknik Universitas Nurul Jadid  
Probolinggo, Indonesia  
[1m.syafii@unuja.ac.id](mailto:1m.syafii@unuja.ac.id)  
[2nadiyah@unuja.ac.id](mailto:2nadiyah@unuja.ac.id)

## Abstract

*Indonesia is an agrarian country because most of its people depend on the agricultural sector. Paiton sub-district, Probolinggo district, East Java is one of the areas where the majority of people work as farmers who still cultivate corn crops. Corn is widely consumed by the surrounding community because it is rich in nutrients, corn can also be used as food and livestock. In previous studies there were less than optimal results in determining the quality of corn. Therefore, the quality of corn quality must be maintained in such a way because corn production is decreasing in productivity every year due to reduced planting land. Gray Level Co-occurrence Matric (GLCM), RGB (Red, Green, Blue) and Backpropagation methods. So that researchers will use these methods to classify the quality of corn kernels. It is hoped that this utilization can solve the problem of middlemen so as not to lose money when buying corn from farmers. The result of this research is the process of determining the quality of corn kernels based on color and texture features using the GLCM, RGB, and Backpropagation methods with a total of 150 images consisting of 120 training data and 30 test data. The classification system test results obtained an accuracy value of 75%. So that the backpropagation method can determine the quality of corn kernels based on images using a computer system so that it can be implemented.*

**Keywords:** Digital Image, RGB, Maize Seed quality classification, Backpropagation algorithm.

## 1. Pendahuluan

Indonesia merupakan negara yang mempunyai beragam bahan pangan lokal selain beras yang mengandung karbohidrat tinggi. [1] Jagung merupakan salah satu bahan pangan yang mudah ditemukan. Jagung biasa dikonsumsi oleh sebagian masyarakat sebagai makanan pokok. [2] Namun disayangkan, rendahnya tingkat konsumsi jagung dalam keluarga membuat sebagian besar masyarakat lebih memilih untuk mengonsumsi beras sebagai makanan pokok. [3] Belakangan tingkat konsumsi konsumsi beras semakin tinggi akibat jumlah penduduk yang semakin meningkat. [4] Parahnya mereka memiliki pemahaman yang salah bahwa makan jagung dikaitkan dengan kemiskinan. Sebagai dampaknya jagung semakin tidak termasuk dalam menu makanan keluarga. [5] Faktanya jagung memiliki kandungan nutrisi yang diperlukan untuk kesehatan. Kandungan karbohidrat dalam biji jagung sangat banyak, mencapai 80% dari keseluruhan bahan biji kering. Karbohidrat jagung lebih rendah dari beras. Untuk setiap 100 g jagung (dikupas, digiling) mengandung 361 hingga 366 kalori. Jagung memiliki kandungan serat dan karoten yang lebih tinggi dibandingkan beras [6]. Karoten merupakan zat penting dalam pembentukan vitamin A jagung dengan biji berwarna kuning mengandung lebih banyak protein dan vitamin A. Kecamatan Paiton Kabupaten Probolinggo Jawa Timur, sebagai salah satu wilayah yang mayoritas masyarakatnya bertani, bahkan sampai saat ini masih membudidayakan tanaman padi, kedelai, dan jagung. Jagung sendiri banyak dikonsumsi masyarakat sekitar karena memiliki kandungan gizi yang tinggi, jagung juga dapat digunakan sebagai bahan pangan dan ternak. Sehingga kualitas mutu jagung harus dijaga agar tetap baik karena produksi jagung semakin menurun produktivitasnya setiap tahun akibat berkurangnya lahan tanam. Masalah yang sering timbul di daerah sekitar ialah ketika petani menjual hasil panen jagung dengan menggunakan poster atau gambar yang diberikan kepada tengkulak. Ketika tengkulak melakukan pengecekan barang di lokasi panen jagung. Kondisi jagung berbeda dengan poster yang diberikan oleh petani, sehingga tengkulak perlu mengklasifikasikan kualitas biji jagung untuk menghindari ketidakstabilan harga. Tengkulak melakukan klasifikasi kualitas biji jagung untuk menghindari ketidakstabilan harga jagung. Hal ini menyebabkan pihak ketiga berperan dalam menentukan harga jagung. Kementerian Perdagangan telah membuat kebijakan harga untuk menjaga

stabilitas harga jagung bagi produsen dan konsumen. [7] Aplikasi ini dibuat untuk membantu dalam proses seleksi dan pengolahan biji jagung yang memiliki peran penting dalam penentuan kualitas dari jagung. Metode klasifikasi yang digunakan adalah Backpropagation yang merupakan teknik pemrosesan data yang terbukti efektif.

Proses evaluasi kualitas biji jagung untuk klasifikasi kualitas masih dilakukan secara manual melalui pengamatan visual. Pengamatan kualitas dengan cara tersebut memiliki beberapa kelemahan, antara lain membutuhkan waktu yang lama dan menghasilkan produk dengan kualitas yang tidak konsisten. Perbedaan ini disebabkan oleh keterbatasan visual manusia, kelelahan menyebabkan persepsi kualitas oleh setiap pengamat. [8] Klasifikasi ini untuk menentukan kualitas biji jagung berdasarkan tekstur dan warna berdasarkan pola citra digital dengan menggunakan metode *Backpropagation* untuk mendapatkan hasil yang akurat. kualitas biji jagung dengan memanfaatkan perbedaan indeks warna RGB dari berbagai kualitas biji jagung. Tujuan dari penelitian ini adalah untuk menghasilkan suatu sistem perangkat lunak berbasis backpropagation untuk mengklasifikasikan biji jagung ke dalam dua jenis kualitas biji jagung, yaitu biji jagung yang baik dan biji jagung yang tidak baik.

## 2. Metode Penelitian

Metode penelitian dapat mempermudah proses penelitian untuk memperoleh hasil penelitian dengan tahapan yang sistematis. Beberapa tahapan awal penelitian adalah studi literatur, identifikasi masalah, pengumpulan dataset, preproccesing, ekstraksi fitur, klasifikasi *backpropagation*, [9] implementasi sistem, pengujian, kesimpulan. Selanjutnya digambarkan pada gambar berikut :



Gambar 1. Kerangka Penelitian [10]

### 2.1. Identifikasi Masalah

Adapun permasalahan yang sering terjadi ialah penjual sering mengatakan jagung ini super. Atau kualitasnya bagus, bijinya tongkolnya berisi, klasifikasi ini sudah banyak dilakukan sebelumnya sehingga akan dilakukan pengembangan dari beberapa permasalahan yang terjadi. Dari hasil identifikasi masalah tersebut menjadi rujukan atau gambaran dalam pengembangan sistem ini, akan dibuat sebuah user interface pengklasifikasian jenis jagung berdasarkan fitur warna dan tekstur dengan menggunakan metode Backpropagation. Hal ini dapat mempermudah dan mempercepat dalam mengetahui kualitas baik atau buruk.

## 2.2. Studi Literatur

Merupakan bahan untuk mempelajari teori-teori dan pengetahuan yang berhubungan dengan masalah penelitian. Sebagai pembanding terhadap penelitian sebelumnya dan acuan dalam memperbaiki atau melengkapi penelitian yang digunakan sebagai solusi dari permasalahan yang telah diidentifikasi sebelumnya. Studi literatur berasal dari buku, jurnal dan penelitian terdahulu yaitu mengklasifikasikan jenis jagung berdasarkan fitur warna dan tekstur. Hal ini membutuhkan dataset berupa citra jagung yang berjumlah sekitar 75 untuk setiap jenisnya.

## 2.3. Pengumpulan Dataset

Pada tahap ini tempat yang menjadi objek adalah sawah yang ditanami jagung di daerah Paiton, desa Pandean. Pada saat mengajukan pertanyaan mengenai seputar jagung, jagung dipanen pada usia 99 - 106 hari. Dataset berisi sekumpulan citra biji jagung yang telah diambil dari petani sebanyak 150 data yang terdiri dari dua jenis yaitu 75 citra biji jagung yang jelek dan 75 citra biji jagung yang bagus. Adapun beberapa spesifikasi hardware yang digunakan dalam penelitian ini antara lain handphone vivo y22, *Random Acces Memory* (RAM) 4GB.

## 2.4. Preprocessing

*Pre-processing* yaitu bagian dari pengolahan data citra mentah yang telah didapatkan pada tahap pengumpulan data dengan tujuan untuk mendapatkan kualitas citra yang lebih baik untuk digunakan pada tahap selanjutnya. Tahapan yang dilakukan adalah *remove background* yang dilakukan secara manual, kemudian *resize* (pengecilan ukuran resolusi citra). Tahap pre-processing berguna untuk dapat meningkatkan kualitas citra yang diperoleh sehingga mudah untuk diolah pada tahap selanjutnya.

## 2.5. Ekstraksi Fitur

Pada tahap pertama ekstraksi fitur dalam penelitian ini adalah tahap ekstraksi fitur tekstur. Penelitian ini menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*) untuk menganalisa tekstur pada citra grayscale. Fitur yang digunakan pada penelitian ini adalah energi, kontras, homogenitas, dan korelasi. Langkah selanjutnya setelah mengekstraksi fitur tekstur dengan metode GLCM adalah tahap mengekstraksi fitur warna dengan menggunakan metode RGB (*Red, Green, Blue*) di dalam metode RGB dalam mengekstraksi gambar berwarna terdapat tiga komponen, yaitu:



Gambar 2 Ekraksi Fitur [11]

## 2.6. Klasifikasi *Backpropagation*

Pada proses klasifikasi, data yang telah memiliki label kelas digunakan dalam proses klasifikasi. Proses ini dilakukan untuk membentuk yang digunakan untuk mengklasifikasikan data baru, model yang digunakan untuk data testing adalah *Backpropagation*. Tujuan dari algoritma *Backpropagation* adalah untuk memodifikasi bobot untuk melatih jaringan syaraf agar dapat memetakan input yang berubah-ubah menjadi output dengan benar.

## 2.7. Implementasi Sisetem

Setelah tahap klasifikasi dengan algoritma *Backpropagation*, langkah selanjutnya adalah mengimplementasikan sistem dalam bentuk aplikasi web dengan menggunakan *web framework flask*. *Flask* merupakan *web framework* yang populer untuk mengembangkan aplikasi web dengan menggunakan bahasa *python*.

## 2.8. Pengujian

Tahap pengujian merupakan tahap yang sangat penting karena untuk mengetahui seberapa tepat keunggulan algoritma *Backpropagation* dalam memprediksi data baru. Pada pengujian ini menggunakan 20 data yang dibagi menjadi dua kelas, yaitu baik dan buruk Untuk menguji tingkat akurasi pada data testing dapat dilakukan dengan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{\sum \text{Data Benar}}{\sum \text{Data Uji}} \times 100 \quad (16)$$

Keterangan :

$\sum$  Data Benar = Banyak Data Benar

$\sum$  Data Uji = Banyak Data Uji

### 2.9. Kesimpulan

Hasil klasifikasi berupa informasi tentang kategori biji jagung misalnya kategori atau jenis tertentu dan tingkat akurasi model dalam prediksi. Tahapan penelitian yang dilakukan setelah uji coba selesai dilakukan terhadap sistem kemudian dianalisa mengenai keakuratan metode Backpropagation dalam mengklasifikasikan kualitas biji jagung. akhirnya didapatkan kesimpulan mengenai penelitian tentang klasifikasi kualitas biji jagung dengan pengolahan citra dari metode *Backpropagation*.

## 3. Hasil dan Pembahasan

### 3.1. Dataset

Dataset yang digunakan pada penelitian ini adalah citra biji jagung yang dikumpulkan sebanyak 150 biji jagung, dimana citra biji jagung tersebut merupakan salah satu jenis jagung yaitu jenis Bisi, masing-masing kelas memiliki 75 dataset yang baik dan 75 dataset yang buruk. Dari masing-masing kualitas biji jagung diambil 15 citra untuk dijadikan data testing. Jadi totalnya ada 30 sebagai data testing dan 120 sebagai data training. Pengambilan dataset menggunakan kamera handphone dengan ukuran 2.84 MB dengan tempat, kondisi dan waktu yang sama.



Biji Jagung Kualitas Bagus



Biji Jagung Kualitas Buruk

Gambar.3 Gambar Contoh Dataset

Pada gambar diatas merupakan gambar biji jagung yang belum melewati tahap pre-pocessing. Ketika proses pengambilan gambar biji jagung harus jelas, karena mempengaruhi ketika proses.

### 3.2. Analisis Data

Berdasarkan hasil penelitian yang dilakukan, didapat hasil pada setiap tahapan yang ada, Adapun hasil dari tahapan – tahapan yang mudah dilakukan pada penelitian ini, terdiri dari hasil pengumpulan dataset, pre-ppsessing, ekstrak fitur warna dan tekstur, normalisasi dan klasifikasi, berikut penjelasan yang lebih terperinci.

### 3.3. Pre`-processing

Selanjutnya data yang diperoleh diolah terlebih dahulu untuk mendapatkan kualitas citra yang baik sehingga memudahkan proses penentuan kualitas biji jagung berdasarkan fitur warna dan tekstur. Pada proses ini dilakukan app paint pada citra jagung dengan kualitas baik dan buruk yang diperoleh dari tahap pengambilan citra, dimana 120 citra biji jagung digunakan untuk data latih dan 30 citra untuk data uji. Tahap pra-pemrosesan dilakukan untuk background dengan warna putih secara manual dan mengubah gambar berwarna menjadi hitam dan putih menjadi abu-abu.

Pengubahan gambar asli menjadi grayscale bertujuan untuk mempermudah proses implementasi pada gambar grayscale. Adapun langkah gambar grayscale sebagai berikut:

a. *Remove background*

Remove Background dilakukan secara manual menggunakan google remove to remove, tujuannya agar lebih mudah saat melakukan preprocessing.

b. Proses Gambar *Grayscale*

Proses *Grayscale* Gambar menjadi hitam, putih dan abu-abu. Setiap gambar akan dikonversi menggunakan skrip python dari gambar berwarna menjadi abu-abu dengan format save \*.jpg.

```
rootPath = 'D:\Skripsi\DATASETFIX'  
folders = [os.path.join(rootPath, x) for x in (rootPath)]  
all_images = [img for folder in folders for img in (folder)]  
targetPath = "D:\Skripsi\ori2grayscale"  
# D:\Skripsi\thedata\ori2grayscale  
for currentPath in os.listdir(rootPath):  
    subPath = os.path.join(rootPath, currentPath)  
    subTargetPath = os.path.join(targetPath, currentPath)  
    if not os.path.isdir(subTargetPath):  
        os.mkdir(subTargetPath)  
    idx = 0  
    for filename in os.listdir(subPath):  
        filepath = os.path.join(subPath, filename)  
        img = cv2.imread(filepath, cv2.IMREAD_GRAYSCALE)  
        target_path = os.path.join(subTargetPath, "%s-%03d.jpg" % (currentPath, idx+1))  
        print(target_path)  
        cv2.imwrite(target_path, img)  
        idx += 1
```

Gambar 1.4 Segmen Program Grayscale



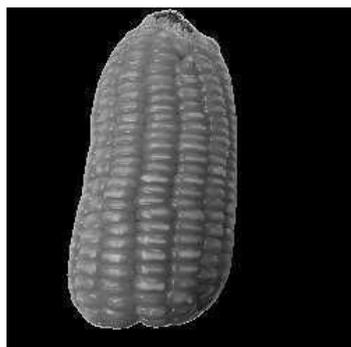
Gambar Asli



Gambar Hasil GrayScale

Gambar .5 Gambar asli dan grayscale

Gambar di atas merupakan dari gambar asli dan warna asli berubah menjadi menjadi warna abu-abu *setelah* dilakukan proses grayscale. Untuk contoh hasil grayscale dari setiap class terdapat pada gambar dibawah ini



Gambar .6 Hasil Grayscale Per Class

Selanjutnya melakukan Resize gambar dari ukuran 980 x 1280 pixel menjadi ukuran 250 x 250 pixel

```
class Glcm:
    def __init__(self):
        self.result = []
    def forOder0(self, a, b):
        size = 0
        for i in a[0]:
            size=size+1
        for i in range(len(a)):
            for j in range(len(a[i])-1):
                p=a[i][j]
                q=a[i][j+1]
                b[p][q]=b[p][q]+1
        matriksIterasi1=list(map(list,zip(*b)))
        for i in range(len(b)):
            for j in range(len(b)):
                b[i][j]=b[i][j]+matriksIterasi1[i][j]

        count=0
        for i in b:
            for j in i:
                count=count+j
        return self.normalisasi(b, count)
```

Gambar 7 Segmen Proses GLCM

menyiapkan kelas data yang akan disimpan ke file csv dari hasil ekstraksi GLCM. Dan fitur-fitur yang digunakan adalah fitur Entropi, Energi, Kontras, Korelasi, dan Homogenitas. Untuk source code lengkap nya dapat dilihat pada halaman lampiran.

```
!python object_detection_demo_flow-master\resize_images.py --raw-dir "D:\Skripsi\ori2grayscale\baik" --save-dir "D:\Skripsi\ori2grayscale2resize\baik"
75 files to resize from directory "D:\Skripsi\ori2grayscale\baik" to target size:(250, 250)
.....
Done resizing 75 files.
Saved to directory: "D:\Skripsi\ori2grayscale2resize\baik"

!python object_detection_demo_flow-master\resize_images.py --raw-dir "D:\Skripsi\ori2grayscale\buruk" --save-dir "D:\Skripsi\ori2grayscale2resize\buruk"
75 files to resize from directory "D:\Skripsi\ori2grayscale\buruk" to target size:(250, 250)
.....
Done resizing 75 files.
Saved to directory: "D:\Skripsi\ori2grayscale2resize\buruk"
```

Gambar .8 Hasil dari proses GLCM

Pada tahapan berikutnya ialah melakukan inialisasi untuk menyimpan fitur GLCM dimana hasil ekstraksi fitur akan otomatis tersimpan pada file csv dengan menjalankan fungsi insertoCsv() dengan nama filenya adalah 'FixGlcmFeatures.csv'.

```
class Data:
    def __init__(self, path):
        imageObject = Image.open(path)
        cropped = imageObject.crop((0,0,250,250))
        self.citra = np.array(cropped)
        self.citra = self.citra.transpose(2,0,1).reshape(3,-1)
        self.matriksIterasi = [[0 for i in range(256)] for j in range(256)]

    def insertoCsv(self,data):
        row = data;
        with open('FixGlcMFeatures.csv', 'a') as csvFile:
            writer = csv.writer(csvFile)
            writer.writerow(row)
        print("succes")
        csvFile.close()
```

Gambar.9 Segmen kelas Data Hasil Output dari GLCM

Merupakan inisialisasi untuk menyimpan fitur GLCM dimana hasil ekstraksi fitur akan otomatis tersimpan pada file csv dengan menjalankan fungsi insertoCsv() dengan nama filenya adalah 'FixGlcMFeatures.csv'. Adapun tabel hasil hasil ekstrak fitur GLCM sebagai berikut

Tabel 1 Hasil Ekstrak Fitur

Keterangan	Data 1	Data 2
Energi 0°	0.6132619368900873	0.6337320455609928
Contras 0°	0.6765349098000871	0.686295850062627
Homogeneity 0°	253.95759932159004	262.3981823709167
Corelation 0°	0.9588631444179109	0.9578719306985576
Energi 45°	0.9588631444179109	0.6337320455609295
Contras 45°	0.676534909800007	0.6862958500625519
Homogeneity 45°	253.95759932155875	262.39818237088673
Corelation 45°	0.9588631444179154	0.9578719306985594
Energi 90°	0.763077948320751	0.7786628834432208
Contras 90°	0.9999983559197761	0.9999984236819742
Homogeneity 90°	0.0012382125384342675	0.0012627954106108322

Corelation 90°	0.9999997240740344	0.9999997145212128
Energi 135°	0.6132631323488729	0.6337332022384019
Corelation 135°	0.6765374975283421	0.6862983597032368
Homogenity 135°	253.95556765442902	262.3960831787372

Corelation 135°	0.9588634084571864	0.9578721943478149
Energi Means	0.6132631323488729	0.6337332022384019
Corelation Means	0.6765374975283421	0.6862983597032368
Homegenity Means	253.95556765442902	262.3960831787372,
Corelation Means	0.9588634084571864	0.9578721943478149
Class	Baik	Buruk

Tabel diatas merupakan hasil dari ekstrak fitur Glcm dan hasil pengimputan data hasil ekstrak fitur. Hasil dari proses ekstrak fitur dari kualitas biji jagung.

```

for j in range(cols):
    k = image[i,j]
    if k[0] > k[1] and k[0] > k[2]:
        B_Color = B_Color + 1
        continue
    if k[1] > k[0] and k[1] > k[2]:
        G_Color = G_Color + 1
        continue
    if k[2] > k[0] and k[2] > k[1]:
        R_Color = R_Color + 1
        continue
    N_Color = N_Color + 1

totalImg = rows * cols

R.append(R_Color/totalImg)
G.append(G_Color/totalImg)
B.append(B_Color/totalImg)
print(" ")
print('Citra', filename, 'Blue:', B_Color/totalImg, 'Green:', G_Color / totalImg, 'Red:', R_Color / totalImg)
print(filepath, '\n')
simpanscv.append(filepath)

```

Gambar 10 Segmen Ekstrak fitur RGB

Adapun hasil dari proses tersebut seperti tabel di bawah ini

Tabel 2. Hasil Ekstrak fitur RGB

<b>B</b>	0.010384,	0.020016
----------	-----------	----------

Class	Baik	Buruk
-------	------	-------

### 3.4. Klasifikasi *Backpropagation*

Proses klasifikasi kualitas biji jagung berdasarkan fitur warna dan fitur tekstur menggunakan 120 data latih dan 30 data uji. Bertujuan untuk mentraining data pada data yang berguna untuk menentukan kualitas data yang akan diuji.

#### a. Split data training dan data testing

Berdasarkan data citra 150 citra biji jagung tersebut dibagi menjadi dua kelas, yaitu training 130 data dan 20 untuk data testing, sehingga dapat dilakukan pengujian performa model pada data tersebut. Pada pengujian ini, 20% data akan digunakan untuk data testing dan 80% untuk data training. Berikut adalah program data testing dan data training

```
x_train, x_val, y_train, y_val = train_test_split(X, y_train_encoded, test_size=0.2, random_state=0)

x_train.shape, x_val.shape, y_train.shape, y_val.shape
((120, 23), (30, 23), (120,), (30,))
```

Gambar 11. Segmen Split Data Training Dan Data Testing

#### b. Arsitektur Model

Pada proses ini terdapat beberapa lapisan dalam model. Pada baris pertama model disusun secara berurutan. Pada baris kedua, layer disusun dengan menggunakan 32 filter dan menggunakan relu aktivasi dan menentukan bentuk input sebanyak 23, pada baris terakhir, paramater output layer diisi sesuai dengan jumlah label yang tersedia.

```
}; import tensorflow as tf

}; import tensorflow as tf

model = tf.keras.Sequential([
    tf.keras.layers.Dense(32, activation='relu', input_shape=[23,]),
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(1, 'sigmoid')
])
```

Gambar 12. Segmen Arsitektur Model

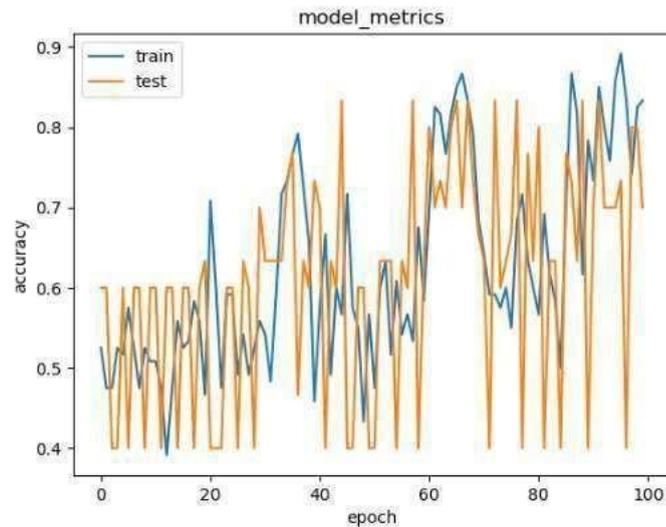
#### c. Training Algoritma

Pada tahapan pembelajaran model menggunakan data latih yakni `x_train`, `y_train` dengan fungsi `fit()`. Pembelajaran algoritma menggunakan 100 epochs atau iterasi dan validasi data juga diimplementasikan untuk memvalidasi proses pembelajaran terhadap data latih, dan terakhir callbacks yang akan memonitor proses pembelajaran model dimana akan memonitor accuracy dan validation accuracy berikut

```
results = model.fit(
    x_train, y_train,
    epochs=100,
    verbose=1,
    validation_data = (x_val, y_val),
    callbacks=[callbacks])
```

Gambar 13. Segmen Model

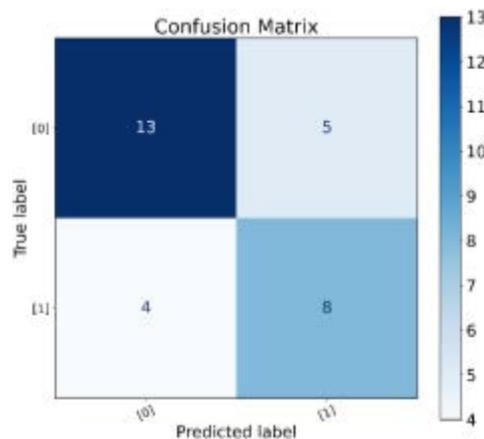
Pada table 1.12. dan table 1.13. merupakan visualisasi hasil proses pembelajaran selama 100 iterasi yang menunjukkan hasil akurasi dan nilai error (*loss*) pada model, dimana hasil proses pembelajaran model mendapatkan akurasi sebesar 89%



Gambar 14 Grafik Pembelajaran model

d. Hasil Prediksi

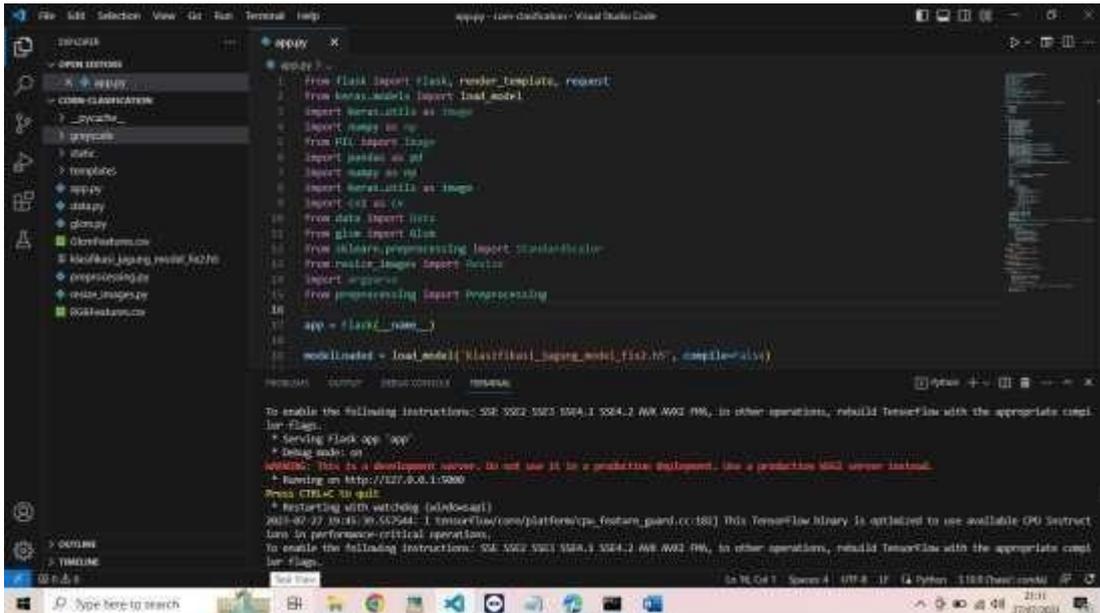
Tahapan hasil prediksi menjelaskan *confusion matrix* menampilkan hasil prediksi terhadap data validasi ada 13 data yang benar atau yang terdeteksi label 0, ada 5 data yang salah prediksi label 1, ada 4 data yang salah prediksi yang terdeteksi 0, ada 8 data yang benar atau terdeteksi label 1.



Gambar 15. Gambar Hasil Prediksi

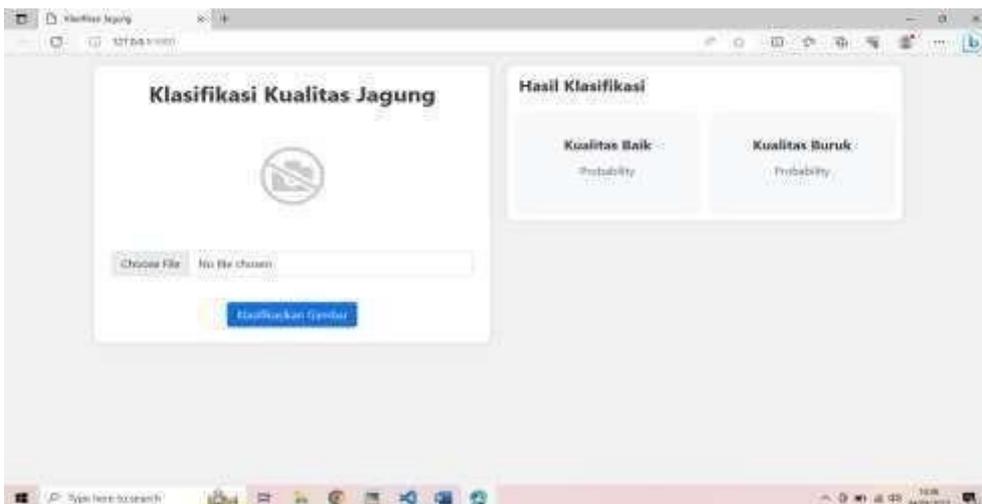
3.5. Implementasi sistem

Implementasi yang digunakan dalam penelitian ini menggunakan *website*. Aplikasi yang dibuat berfungsi untuk mengklasifikasikan gambar dengan menggunakan model yang telah dilatih dan disimpan sebelumnya. *Framework* yang digunakan adalah *Flask*. Pada implementasi ini menggunakan *VsCode (Visual Studio Code)* pada gambar 19.287 melakukan *proses run* dan cara menjalankan aplikasi *flask* dengan cara klik Terminal kemudian klik *Run python* file.



Gambar 16. Segmen Proses Run

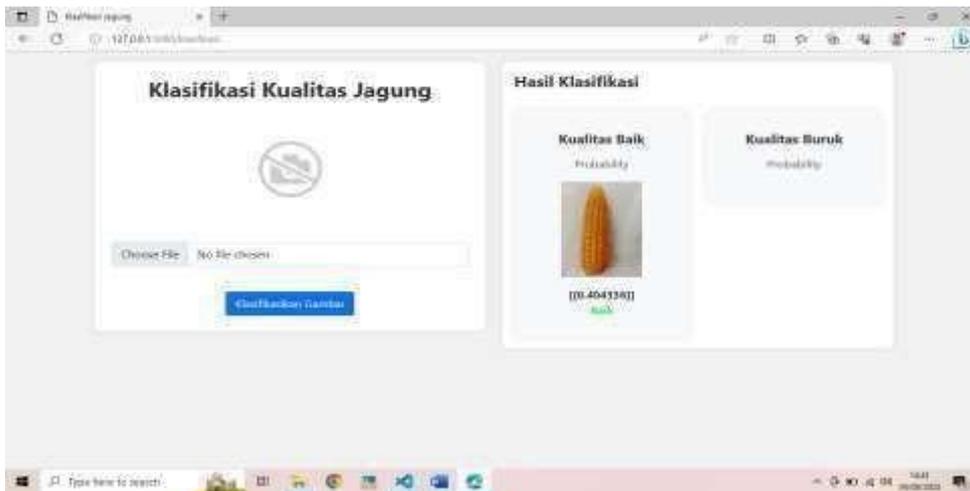
Setelah dijalankan akan mendapatkan server internal <http://127.0.0.1:5000/> klasifikasi dengan mengklik server kemudian akan muncul tampilan desain website Klasifikasi kualitas biji jagung seperti pada gambar 28 berikut :



Gambar 17. Desain Tampilan Website

Pada tampilan website, terdapat dua tombol, yaitu "pilih file" yang digunakan untuk mengambil gambar dan "klasifikasi gambar" yang digunakan untuk mengklasifikasikan gambar yang akan diklasifikasikan.

Langkah pertama klik choose file kemudian akan di arahkan ke directory untuk memilih gambar yang akan di klasifikasi. Langkah kedua yaitu klik "klasifikasi gambar" yang akan diklasifikasi kemudian akan muncul hasil dari klasifikasi jagung akan ditampilkan pada tabel disamping. Berikut adalah hasil dari klasifikasi kualitas biji jagung.



Gambar 18. Gambar Hasil Klasifikasi

### 3.6. Uji Coba

Pada tahapan uji coba dengan algoritma backpropagation yang sudah di implementasikan ke dalam website dengan menggunakan data baru sebanyak 20 data citra jagung terdiri dari 10 kualitas baik dan 10 kualitas buruk. Berikut adalah hasil uji coba website ditunjukkan pada tabel di bawah ini :

Tabel 3 Hasil Uji Coba

Nama Citra	Kelas Awal	Prediksi	Hasil
br1	Buruk	Buruk	Salah
br2	Buruk	Buruk	Benar
br3	Buruk	Buruk	Salah
br4	Buruk	Buruk	Salah
br5	Buruk	Buruk	Benar
br6	Buruk	Buruk	Benar
br7	Buruk	Buruk	Salah

br8	Buruk	Buruk	Benar
br9	Buruk	Buruk	Benar
br10	Buruk	Buruk	Salah
kb1	Baik	Baik	Benar
kb2	Baik	Baik	Benar
kb3	Baik	Baik	Salah
kb4	Baik	Baik	Benar
kb5	Baik	Baik	Benar
kb6	Baik	Baik	Benar
kb7	Baik	Baik	Benar
kb8	Baik	Baik	Salah
kb9	Baik	Baik	Benar
kb10	Baik	Baik	Benar

Selanjutnya pada tabel diatas akan dilihat seberapa besar akurasinya dengan rumus berikut :

$$Akurasi = \frac{\sum Data Benar}{\sum Data Uji} \times 100$$

Pada tabel diatas terdapat 20 data dengan 5 yang data salah sehingga :

$$Akurasi = \frac{15}{20} \times 100 = 75\%$$

Jadi pada pengujian klasifikasi kualitas biji jagung dengan metode backpropagation neural network berbasis web didapatkan akurasi sebesar 75%. Karena model Machine Learning Overfitting terlalu mempelajari data terlalu detail, sehingga yang ditangkap bukan hanya datanya saja, namun noise yang ada juga direkam. Model overfitting akan menghasilkan nilai akurasi yang sangat tinggi saat proses training, namun memiliki akurasi rendah Ketika tesing.

#### 4. Kesimpulan

Hasil penelitian menunjukkan bahwa aplikasi "Klasifikasi Benih Jagung Berdasarkan Tekstur dan Warna Menggunakan Metode Backpropagation Berbasis Web" berhasil mengklasifikasikan benih jagung dengan akurasi yang tinggi. Metode *Backpropagation* terbukti efektif dalam melakukan machine learning untuk mengenali pola tekstur dan warna, sehingga memberikan hasil klasifikasi yang handal. Interface web yang mudah digunakan memudahkan pengguna untuk mengunggah gambar biji jagung dan mendapatkan hasil klasifikasi dengan cepat. Keamanan data dan praktik privasi juga terjamin, sementara pemeliharaan rutin dan pembaruan model memastikan aplikasi tetap dioptimalkan dan responsif terhadap perubahan dataset atau kebutuhan pengguna.. Hasil dari penelitian klasifikasi kualitas biji jagung berdasarkan fitur warna dan tekstur dengan metode *Backpropagation* dapat disimpulkan bahwa proses menentukan kualitas biji jagung berdasarkan fitur warna dan tekstur dengan total 150 data dengan ukuran format gambar minimal 250 x 250 pixel.. Maka hasil pengujian sistem klasifikasi diperoleh nilai akurasinya 75% dengan menggunakan metode backpropagation dapat memprediksi kualitas biji jagung sehingga dapat diimplementasikan ke sistem *Web Flask*.

Daftar Pustaka

- [1] A. Wanto, "Prediksi Produktivitas Jagung Di Indonesia Sebagai Upaya Antisipasi Impor Menggunakan Jaringan Saraf Tiruan Backpropagation," *SINTECH (Science and Information Technology) Journal*, vol. 2, no. 1, pp. 53-62., 2019.
- [2] R. Aldillah, "Strategi pengembangan agribisnis jagung di Indonesia.," *Analisis Kebijakan Pertanian*, vol. 15, no. 1, pp. 43-66., 2017.
- [3] A. Wanto, D. Hartama, G. W. Bhawika, Z. Chikmawati, D. S. Hutauruk, P. H. Siregar and A. P. Windarto, "Model of Artificial Neural Networks in Predictions of Corn Productivity in an Effort to Overcome Imports in Indonesia," *In Journal of Physics: Conference Series*, vol. 1339, no. 1, p. 012057, 2019.
- [4] J. M. Maligan, D. D. Pratiwi and T. D. Widyaningsih, "Studi Preferensi Konsumen terhadap Nasi Putih dan Nasi Jagung Putih pada Pekerja Wanita di Kantor Pemerintah Kota Malang.," *Indonesian Journal of Human Nutrition*, vol. 6, no. 1, pp. 41-52, 2019.
- [5] D. Febrina, N. Khairunnisa and R. Febriyanti, "Pengaruh lama pemeraman dan metode pengolahan terhadap kualitas fisik dan kandungan nutrisi jerami jagung," *Jurnal Agripet*, vol. 20, no. 2, pp. 15-23, 2020.
- [6] A. R. Lapui, U. Nopriani and H. Mongi, "Analisis Kandungan Nutrisi Tepung Jagung (*Zea mays* Lam) dari Desa Uedele Kecamatan Tojo Kabupaten Tojo Una-Una untuk Pakan Ternak.," *Agropet*, vol. 18, no. 2, pp. 42-46, 2021.
- [7] M. I. Al-Qarazi, S. Sukardi and A. Anwar, "Analisis Peramalan Produksi, Konsumsi dan Harga Jagung Di Provinsi Nusa Tenggara Barat.," *Jurnal Agrimansion* , vol. 22, no. 1, pp. 49-60, 2021.
- [8] F. Sigaha, S. E. J. and S. Zainudin, "Evaluasi persentase karkas ayam kampung super dengan pemberian jermai jagung fermentasi.," *Jambura Journal of Animal Science*, vol. 2, no. 1, pp. 1-7, 2019.
- [9] G. Guntoro, L. Costaner and L. Lisnawita, "Prediksi Jumlah Kendaraan di Provinsi Riau Menggunakan Metode Backpropagation," *nformatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, vol. 14, no. 1, pp. 51-57., 2019.
- [10] A. Wanto, "Prediksi Produktivitas Jagung Di Indonesia Sebagai Upaya Antisipasi Impor Menggunakan Jaringan Saraf Tiruan Backpropagation," *SINTECH (Science and Information Technology) Journal*, vol. 2, no. 1, pp. 53-62, 2019.
- [11] A. Ciputra, E. H. Rachmawanto and A. Susanto, "Klasifikasi Tingkat Kematangan Buah Apel Manalagi Dengan Algoritma Naive Bayes Dan Ekstraksi Fitur Citra Digital," *Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, vol. 9, no. 1, pp. 465-472., 2018.

# Penerapan Metode MFCC dan LSTM untuk Speech Emotion Recognition

I Dewa Agung Adwitya Prawangsa, AAIN Eka Karyawati

Informatika, Universitas Udayana  
Jl. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
wawaca.waca@gmail.com  
eka.karyawati@unud.ac.id

## Abstract

*Speech recognition falls under the field of computational linguistics. This includes identification, recognition, and translation of speech detected into text by a computer. This research uses Mel Frequency Cepstral Coefficients (MFCC) and Recurrent Neural Networks (RNN) techniques as a form of Artificial Neural Network (ANN) architecture. The main objective of this research is to use speech recognition techniques to detect and identify various emotional voices within a person. The MFCC process will convert the voice signal into several vectors that help for the speech recognition process in this study. The results obtained are by determining two comparison parameters, namely a learning rate of 0.1 which results in an accuracy validation of 72% and a learning rate of 0.001 which results in 57%.*

**Keywords:** Feature Extraction, Speech recognition, MFCC, ANN, RNN

## 1. Pendahuluan

Emosi merupakan suatu kondisi mental seseorang yang dapat mendorongnya untuk melakukan suatu tindakan atau berekspresi yang dapat dipicu dari dalam atau luar dirinya. Dalam kehidupan sehari-hari sangat penting untuk memahami kondisi emosional seseorang dengan emosi tertentu. Emosi juga merupakan salah satu aspek penting bagi kehidupan [1]. Emosi seseorang dapat diketahui salah satunya dari ekspresi wajah, namun terkadang ekspresi wajah seseorang tidak sesuai dengan apa yang sedang dialaminya. Pada penelitian kali ini, saya akan membuat aplikasi guna sebagai penelitian yang dapat mengetahui emosi yang sedang dialami oleh seseorang, dibuatlah aplikasi untuk mendeteksi emosi seseorang berdasarkan suara menggunakan metode Mel Frequency Cepstral Coefficients (MFCC) dan Long short term memory network (LSTM).

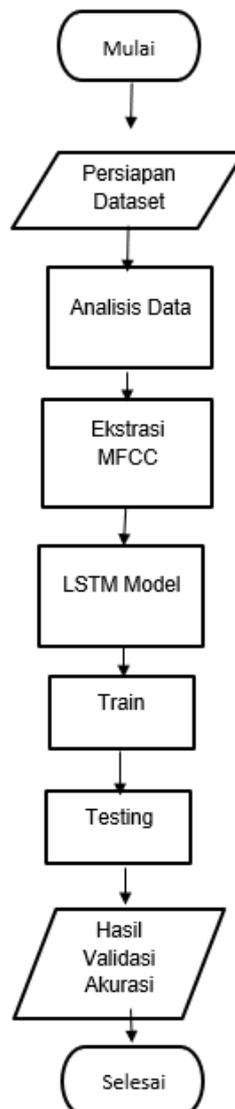
Penelitian ini juga berdasarkan dari beberapa penelitian terdahulu yang terkait yaitu pada jurnal pertama dari penulis Fatan Kasyidi, Ridwan Ilyas, Nida Muthi Annisa, dengan judul "Peningkatan Kemampuan Pengenalan Emosi melalui Suara dalam Bahasa Indonesia" Dalam review singkat pada jurnal ini berisikan tentang permasalahan tentang pengenalan emosi melalui suara terutama masalah korpus yang menjadi salah satu faktor yang menjadikan pengenalan emosi ini belum menghasilkan akurasi pengenalan yang optimal, khususnya berkaitan dengan imbalance data. Penelitian ini dilakukan untuk meningkatkan performa pengenalan emosi untuk mengenali lima kelas emosi yaitu senang, marah, sedih dan kepuasan serta netral menggunakan algoritma boosting. Selain itu, digunakan pula metode seperti CNN dan RNN untuk dapat dilakukan perbandingan serta penerapan SMOTE untuk korpusnya. Setelah eksperimen, dapat dihasilkan akurasi pengenalan mencapai 65% untuk akurasi untuk data tes berdasarkan konfigurasi 22050 Hz sebagai sampling rate, MFCCs dan oversampling SMOTE [1].

Lalu pada review jurnal penelitian yang kedua dengan penulisnya yang bernama Barlian Henryranu Prasetyo, Wijaya Kurniawan, Mochammad Hannats Hanafi Ichsan, yang berjudul "Pengenalan Emosi Berdasarkan Suara Menggunakan Algoritma HMM" dijelaskan bahwa penelitian ini memiliki tujuan untuk mengenali emosi seseorang melalui ucapan menggunakan algoritma HMM. Sistem dibangun dapat mengenali 3 jenis emosi yaitu marah, bahagia dan netral. Fitur yang digunakan dalam sistem ini adalah

pitch, energi dan formant. Database yang digunakan adalah suara dari rekaman film. Dari hasil obeservasi probabilitas emosi marah sebesar 0.196, bahagia 0.254 dan netral 0.045. Sistem memiliki tingkat akurasi rata-rata sebesar 86.66%. Rata waktu eksekusi sistem dalam mendeteksi dan mengklasifikasikan emosi sebesar 21.6ms [3].

## 2. Metodolgi Penelitian

### Flowchart Umum

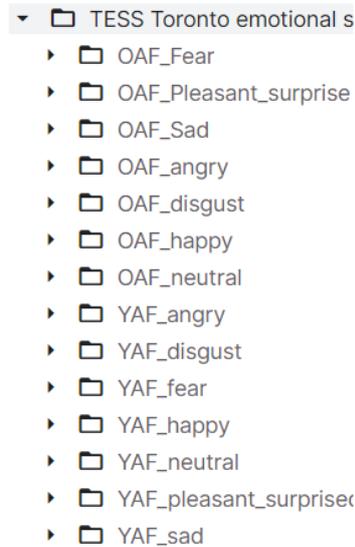


**Gambar 1.** Flowchart Alur Kerja

### 2.1 Persiapan Data Set

Mempersiapkan Data set yang diperlukan guna sebagai sample suara dalam melakukan pengenalan emosi dalam penelitian kali ini yaitu menggunakan Toronto emotional speech set (TESS) yang bersifat publik yang memudahkan dalam menggunakannya sebagai sample penelitian ini. Sumber dataset

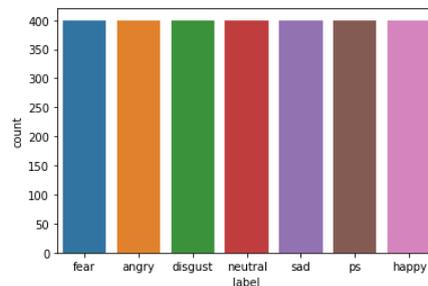
berapa pada tautan berikut ini : [www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tes](http://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tes). Ada satu set 200 kata target yang diucapkan dalam frasa pembawa "Ucapkan kata '\_' oleh dua aktris (berusia 26 dan 64 tahun) dan rekaman dibuat dari set yang menggambarkan masing-masing dari tujuh emosi (marah, jijik, takut, bahagia, kejutan, kesedihan, dan netral) Total ada 2800 titik data (file audio). Dataset diatur sedemikian rupa sehingga masing-masing dari dua aktor wanita dan emosi mereka terkandung dalam foldernya sendiri. Dan di dalamnya, semua file audio 200 kata target dapat ditemukan. Format file audio adalah format WAV.



**Gambar 2.** Toronto emotional speech set (TESS)

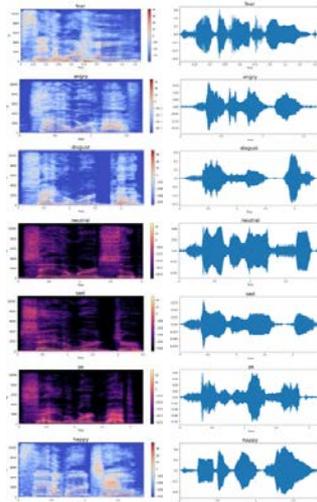
## 2.2 Analisis data

Dalam proses membuat data analisis, Saya akan membuat sebuah visual analisis yang pertama yaitu yang memuat label dari semua kelas dalam distribusi yang sama.



**Gambar 3.** Label semua kelas

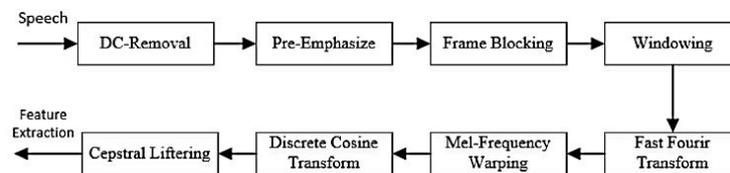
Lalu analisis yang kedua yaitu menampilkan waveform, spektogram serta menampilkan plot dari sample berkas yang digunakan.



**Gambar 4.** Spektogram & Waveform

### 2.3 Ekstraksi Fitur MFCC

Teknik Mel Frequency Cepstral Coefficients (MFCC) digunakan untuk ekstraksi ciri dari sinyal wicara dan membandingkan dengan penutur tak dikenal dengan penutur yang ada dalam database [2].

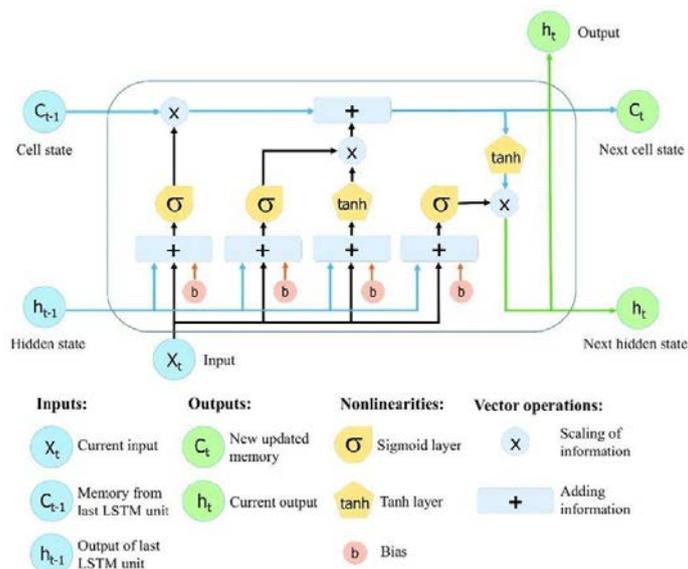


**Gambar 5.** Tahap proses Ekstraksi MFCC [6].

Durasi audio dibatasi hingga maksimal 3 detik untuk durasi ukuran file yang sama dan akan mengekstrak fitur koefisien cepstral frekuensi Mel (MFCC) dengan batas 40 dan mengambil rata-rata sebagai fitur akhir yang dimana berdasarkan data inputan 2800 sampel dan 7 kelas.

#### 1. Membuat LSTM Model

LSTM merupakan Long short term memory network (LSTM) yang dimana merupakan salah satu jenis dari Recurrent Neural Network (RNN) dimana dilakukan modifikasi dengan menambahkan memory cell yang dapat menyimpan informasi untuk jangka waktu yang lama [4]. LSTM diusulkan sebagai solusi untuk mengatasi terjadinya vanishing gradient pada RNN saat memproses data sequential yang panjang.



**Gambar 6.** Arsitektur LSTM [7]

Pada arsitektur LSTM ini diambil dari sampel hasil dari ekstrasi fitur MFCC yaitu 2800 sampel dan dimensi arraynya (40, 1 yang dimana akan ditaruh di layer pertama LSTM. Lalu layer selanjutnya yaitu Dense (64) dengan aktivasi relu untuk meminimalkan nilai kesalahan antara lapisan output dan kelas target lalu menambahkan dropout(0.2). Setelah itu layer dense dengan output 32 dan terakhir dengan output 7 dengan aktivasi softmax sebagai kategori pertama jika memiliki klasifikasi binary.

### 3. Hasil dan Pembahasan

Pada penelitian kali ini saya menggunakan Recurrent Neural Networks (RNN) sebagai salah satu bentuk arsitektur Artificial Neural Networks (ANN) yang dirancang khusus untuk memproses data yang bersambung / berurutan (sequential data) [5]. Pengenalan suara atau speech recognition adalah salah satu aplikasi yang termasuk dalam arsitektur RNN.

Layer (type)	Output Shape	Param #
lstm_30 (LSTM)	(None, 256)	264192
dropout_62 (Dropout)	(None, 256)	0
dense_72 (Dense)	(None, 128)	32896
dropout_63 (Dropout)	(None, 128)	0
dense_73 (Dense)	(None, 64)	8256
dropout_64 (Dropout)	(None, 64)	0
dense_74 (Dense)	(None, 7)	455

Total params: 305,799  
 Trainable params: 305,799  
 Non-trainable params: 0

**Gambar 7.** Model sequential

- Dense = lapisan linier dimensi tunggal dengan unit tersembunyi
- Dropout = digunakan untuk menambahkan regularisasi ke data, menghindari over fitting & drop out sebagian kecil dari data
- Loss='sparse\_categorical\_crossentropy' = menghitung kerugian lintas-entropi antara label yang sebenarnya dan label yang diprediksi.
- Optimizer='adam' = secara otomatis menyesuaikan tingkat pembelajaran untuk model selama jumlah epochs

Klasifikasi merupakan termasuk tahapan penting karena pada tahapan adalah penentuan hasil dari proses ekstraksi fitur yang telah dilakukan sebelumnya. Klasifikasi emosi pada penelitian ini menggunakan metode Jaringan Syaraf Tiruan: Multilayer Perceptron (MLP).

### 3.1 Data Uji

Pada tahap selanjutnya yaitu menentukan parameter-parameter yang terbaik, untuk digunakan ke tahap kesimpulan hasil akhir. Penentuan parameter tersebut berada pada tabel dibawah ini :

Parameter	Nilai
Learning rate	0.1
Epochs	50
Layer	7
Batch Size	64
Validation_split	0.2
Validasi Akurasi terbaik	72.32

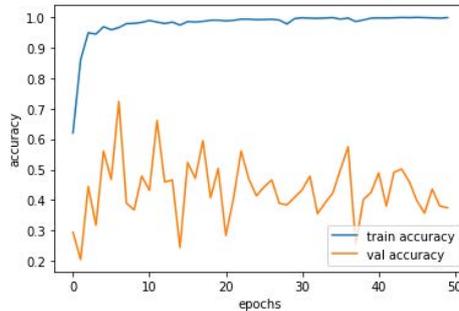
**Tabel 2.** Parameter Klasifikasi berdasarkan MLP

- batch\_size=64 = jumlah data yang akan diproses per langkah
- epochs=50 = nomor iterasi untuk melatih model
- validasi\_split=0.2 = melatih dan menguji persentase pemisahan
- Akurasi pelatihan dan akurasi validasi meningkatkan setiap iterasi
- akurasi validasi terbaik yang telah didapatkan yaitu 72.32 %

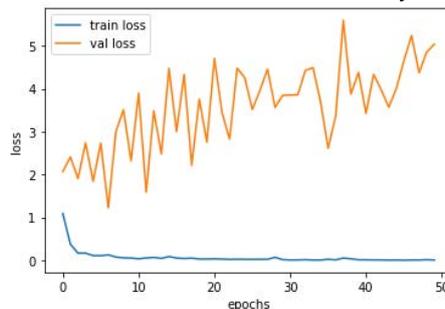
Jika diubah parameternya dalam menyesuaikan learning rate pada train model yang sudah dibuat dengan learning rate value nya yaitu 0.001 menjadikan hasil validasi akurasi hanya akan mendapatkan validasi akurasi dengan yang terbaik yaitu 57.14 % dan akan terus menurun seiring bertambahnya Epochs karena learning rate menjadi terlalu besar.

### Hasil melalui graphs plot

Pada tahap ini yaitu membuat plot hasil model yang dimana bagian yang dapat menyimpulkan hasil dari penelitian ini dengan menentukan validasi dari akurasi yang telah di uji sebelumnya.



**Gambar 8.** Hasil train acuracy



**Gambar 9.** Hasil train loss

Dalam penelitian ini didapatkan akurasi validasi akan menjadi sekitar 72,32% yang dimana lebih baik dalam konteks multiklasifikasi dasar. Pada penelitian ini dari perbedaan kalimat tidak dapat

mempengaruhi hasil akurasi melainkan beberapa faktor eksternal seperti noise dan emosi yang dibuat oleh sample suara. Akurasi akhir dari model pada data latih pada model yang dibuat dengan menyesuaikan parameter learning rate dengan value yang beda mempunyai perbedaan yang signifikan karena konvergensi menjadi cukup lambat sampai model mengalami overfitting. Hal tersebut disebabkan oleh tidak seimbangannya kelas emosi pada dataset sehingga membuat model akan cenderung memprediksi kelas yang labelnya lebih banyak. Selain itu, kurangnya heterogenitas dari dataset yang membuat karakter kelas emosi lebih berbeda dari yang lainnya sehingga mampu mengurangi bias pada model agar tidak membuat model overfitting.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan pada kesempatan kali ini, Dari akurasi tersebut yang mencakup semua kelas emosi yang telah ditentukan sebelumnya dengan menggunakan metode MFCC, Bisa disimpulkan permasalahan overfitting masih terjadi, Maka dari itu memerlukannya strategi lain. Hal ini akan ditangani di masa mendatang.

#### Daftar Pustaka

- [1] Fatan Kasyidi, Ridwan Ilyas, Nida Muthi Annisa, "Peningkatan Kemampuan Pengenalan Emosi melalui Suara dalam Bahasa Indonesia" *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, vol 6, no 2, Hal 194, 2021.
- [2] Raynaldy Arief, Nur Aviva Iriawan, Armin Lawi, "KLASIFIKASI AUDIO UCAPAN EMOSIONAL MENGGUNAKAN MODEL LSTM" Konferensi Nasional Ilmu Komputer (KONIK), Hal 524, 2021.
- [3] Barlian Henryranu Prasetio, Wijaya Kurniawan, Mochammad Hannats Hanafi Ichsan, "Pengenalan Emosi Berdasarkan Suara Menggunakan Algoritma HMM" (JTIK) *Jurnal Teknologi Informasi dan Ilmu Komputer*. Vol 4, No 3, 2017.
- [4] Yulistia Khoirotul Aini, Tri Budi Santoso, Titon Dutono, "Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia" (JKT) *Jurnal Komputer Terapan*, Vol 7, No 1, 2021.
- [5] Angga Anjaini Sundawa, Aji Gautama Putrada, S.T., M.T., Novian Anggis Suwastika, S.T., M.T., "Implementasi dan Analisis Simulasi Deteksi Emosi Melalui Pengenalan Suara Menggunakan Mel-Frequency Cepstrum Coefficient dan Hidden Markov Model Berbasis IOT", Universitas Telkom, 2019.
- [6] Irham Sidik Permana, Youllia Indrawaty Nurhasanah, Andriana Zulkarnain, "IMPLEMENTASI METODE MFCC DAN DTW UNTUK PENGENALAN JENIS SUARA PRIA DAN WANITA" *MIND Journal*, Vol 3, No 1, Hal 6, 2018.
- [7] Xuan Hien Le, Hung Viet Ho, Giha Lee, Sungho Jung, "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting" *MDPI Journal*, Department of Disaster Prevention and Environmental Engineering, Kyungpook National University, Korea.

*This page is intentionally left blank.*

# Perancangan Ulang Website Itinerary untuk Daerah Wisata Kabupaten Badung Menggunakan Metode User Centered Design

Putu Audy Cipta Pratiwi<sup>a1</sup>, Luh Arida Ayu Rahning Putri<sup>a2</sup>, I Wayan Santiyasa<sup>a3</sup>,  
Gst Ayu Vida Matrika Giri<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana  
Bukit Jimbaran, Badung, Bali, Indonesia

<sup>1</sup>audycipta02@gmail.com@email.com

<sup>2</sup>rahningputri@unud.ac.id

<sup>3</sup>santiyasa@unud.ac.id

<sup>4</sup>vida@unud.ac.id

## Abstract

Pariwisata menjadi sektor vital dalam perekonomian Indonesia, termasuk di Kabupaten Badung, Bali yang merupakan salah satu destinasi utama di Indonesia. Dengan peningkatan kunjungan wisatawan, diperlukan informasi wisata yang mudah diakses. Namun, pengalaman pengguna yang kurang optimal dalam menggunakan website Itinerary Travaa untuk merencanakan perjalanan wisata menunjukkan perlunya perancangan ulang. Penelitian ini menerapkan metode User Centered Design (UCD) dalam proses perancangan ulang website Itinerary Travaa, dengan tujuan untuk memastikan website tersebut dapat digunakan dengan mudah dan efektif oleh pengguna. Evaluasi Usability terhadap website Itinerary Travaa sebelum redesign menunjukkan beberapa kelemahan dalam dengan nilai rata-rata Usability mencapai 59,78%. Setelah dilakukan perancangan ulang dengan pendekatan User Centered Design, hasil evaluasi menunjukkan peningkatan yang signifikan yakni meningkat sebesar 48,21% dengan nilai rata-rata Usability mencapai 88,62%. Rata-rata nilai Usability tersebut termasuk kriteria "Sangat Baik", menunjukkan bahwa penerapan metode User Centered Design berhasil menghasilkan website Itinerary Travaa yang memenuhi kebutuhan dan preferensi pengguna, serta memberikan pengalaman pengguna yang optimal dan memuaskan.

**Keywords:** *Pariwisata, Redesign, User Centered Design, User Interface, User Experience*

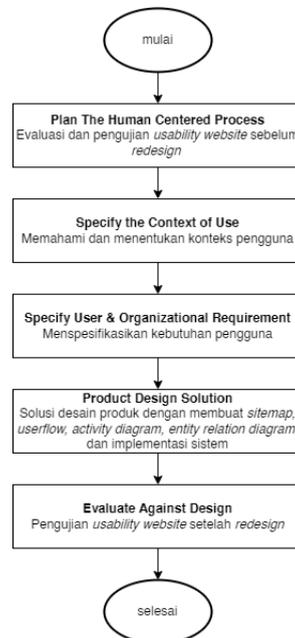
## 1. Pendahuluan

Menurut data Dinas Pariwisata Provinsi Bali tahun 2019, Badung menjadi salah satu destinasi paling populer di Bali, dengan kunjungan mencapai 2,9 juta orang, terdiri dari 1,3 juta wisatawan mancanegara dan 1,6 juta wisatawan domestik [1]. Hal ini menegaskan bahwa Badung memiliki potensi besar sebagai tujuan wisata di Bali, menarik bagi mereka yang ingin menikmati keindahan alam, pantai, dan budaya Bali. Dengan meningkatnya jumlah wisatawan ke Kota Badung, kebutuhan akan informasi tempat-tempat wisata yang mudah diakses semakin penting. Dengan meningkatnya jumlah wisatawan di Kota Badung, kebutuhan akan informasi akurat dan mudah diakses tentang tempat wisata semakin penting. Namun, banyaknya objek wisata menyulitkan wisatawan yang memiliki waktu terbatas untuk merencanakan kunjungan. Solusi umumnya adalah menggunakan Travel Agency, tetapi banyak wisatawan tidak sepenuhnya puas dengan pilihan perjalanan yang tersedia karena tidak sesuai dengan preferensi mereka [2].

Travaa adalah sebuah website itinerary yang membantu perencanaan perjalanan wisata, namun pengujian menunjukkan beberapa permasalahan seperti tampilan antarmuka yang kurang nyaman dan kesulitan dalam menemukan rekomendasi wisata. Evaluasi pengujian usability menggunakan 5 indikator yakni *Learnability*, *Memorability*, *Efficiency*, *Errors*, dan *Satisfaction* [5]. Pada pengujian tersebut mendapatkan nilai rata-rata usability sebesar 58,78%, termasuk dalam kriteria cukup baik namun perlu ditingkatkan.

Dalam penelitian ini, perancangan ulang website *Itinerary* Badung menggunakan metode *User Centered Design* (UCD) untuk memastikan bahwa website tersebut dapat digunakan dengan mudah dan efektif oleh pengguna karena konsep dari UCD adalah pengguna sebagai pusat dari proses pengembangan sistem, dan tujuan atau sifat-sifat, konteks dan lingkungan sistem semua didasarkan dari pengalaman pengguna [3].

## 2. Metodologi Penelitian



**Gambar 1.** Diagram Alur Metodologi Penelitian

Penelitian ini bertujuan untuk melakukan *redesign* dengan menggunakan *User Centered Design* (UCD) yang terdiri dari beberapa tahapan seperti yang ditunjukkan pada gambar 1, antara lain *Plan The Human Centered Process* dimana dilakukan evaluasi dan uji *usability website* sebelum *redesign*, *Specify the Context of Use* yakni memahami dan menentukan konteks pengguna, *Specify User and Organizational Requirement* yakni menspesifikasikan kebutuhan pengguna sesuai dengan perbaikan yang didapatkan dari permasalahan pada uji *usability*, *Product Design Solution* yakni pembuatan solusi desain produk seperti pembuatan *sitemap*, *redesign userflow* dan *activity diagram*, pembuatan database serta pengimplementasian sistem ke dalam coding, dan tahapan terakhir yakni *Evaluate Against Design* dengan melakukan uji *usability website* setelah *redesign*.

### 2.1. Plan The Human Centered Design

Pada tahapan ini, dilakukan pengumpulan data primer dengan melakukan pengujian *usability* untuk mengevaluasi website Travaa sebelum dilakukan perancangan ulang, dimana pengujian dilakukan kepada 30 responden yang terdiri dari wisatawan dan *UI/UX Designer* dengan penyebaran kuesioner dengan 5 skala likert dengan bobot nilai yakni (1) Sangat Tidak Setuju, (2) Tidak Setuju, (3) Netral (Setuju maupun Tidak Setuju), (4) Setuju, dan (5) Sangat Setuju [6]. Nilai yang diuji yakni *Learnability* (kemudahan dalam mempelajari), *Memorability* (kemudahan dalam mengingat), *Efficiency* (efisiensi), *Errors* (tingkat kesalahan), dan *Satisfaction* (kepuasan). Adapun indikator pertanyaan dari kuesioner yang ditunjukkan pada tabel 1 sebagai berikut:

**Tabel 1.** Indikator Pertanyaan *Usability*

No	Indikator	Kode	Pertanyaan pada Kuesioner
1	<i>Learnability</i> (A)	(A1) (A2) (A3) (A4)	Apakah website Travaa.com dapat dipelajari dengan mudah? Apakah anda dengan mudah dalam menavigasi antara halaman atau bagian berbeda pada website Travaa? Apakah anda mudah memahami isi dan konten informasi terkait wisata di Kabupaten Badung yang disajikan pada website Travaa? Apakah tanpa instruksi tertulis atau manual book, anda dapat mudah mempelajari penggunaan website Travaa?
2	<i>Memorability</i> (B)	(B1) (B2)	Apakah anda dapat dengan mudah mengingat penggunaan dari website Travaa? Apakah anda merasa mudah kapanpun menggunakan website Travaa?
3	<i>Efficiency</i> (C)	(C1) (C2) (C3)	Apakah anda dapat dengan cepat menyelesaikan tugas-tugas tertentu di Travaa, seperti mencari informasi rekomendasi wisata atau membuat agenda perjalanan wisata khususnya untuk Kabupaten Badung? Apakah anda dengan mudah dan cepat menerima informasi secara detail dan juga spesifik terkait wisata Kabupaten Badung pada website Travaa? Apakah anda dapat secara mudah langsung menemukan informasi terkait wisata Kabupaten Badung yang anda ingin cari dari awal membuka pada website Travaa?
4	<i>Errors</i> (D)	(D1) (D2)	Apakah anda tidak menemukan menu yang <i>Errors</i> atau tidak sesuai dengan fungsinya website Travaa? Apakah anda tidak mengalami kesalahan saat mencoba fitur pembuatan agenda perjalanan wisata <i>Itinerary</i> pada menu <i>My Trip</i> website Travaa?
5	<i>Satisfaction</i> (E)	(E1) (E2) (E3) (E4) (E5)	Apakah anda senang dengan desain antarmuka yang ada pada website Travaa.com Apakah anda merasa nyaman dalam menggunakan website Travaa? Paduan warna dan tata letak konten pada website Travaa.com tersebut nyaman untuk dilihat? Apakah website Travaa.com sesuai dengan ekspektasi anda, ketika anda melihat logo yang ada pada laman sistem tersebut? Apakah anda puas dengan pengalaman penggunaan situs web Travaa?

Sebelum kuesioner dibagikan kepada responden, dilakukan uji validitas dan reabilitas terlebih dahulu guna memastikan bahwa instrumen kuesioner yang digunakan dapat menghasilkan data yang akurat, konsisten, dan dapat diandalkan dalam mengukur persepsi dan pengalaman pengguna terhadap pengujian *website Itinerary* Travaa dengan cara membandingkan nilai dari *rhitung* dengan *rtabel*. Nilai *rtabel* pada *r-product moment* dengan taraf signifikan 0,05 untuk 30 responden yakni 0,361.

Setelah dilakukan uji validitas dan penyebaran terhadap kuesioner didapatkan umpan balik dari responden melalui uji *usability* berupa bobot nilai dari 5 skala likert, kemudian dilakukan perhitungan rata-rata dan presentase hasil *usability* dengan persamaan 1 dan 2

$$\text{Nilai Rata - rata} = \frac{\text{Jumlah Bobot Nilai}}{\text{Bobot Maksimum}} \quad (1)$$

$$\text{Persentase} = \frac{\text{Nilai Rata-Rata}}{\text{Bobot Maksimum}} \times 100\% \quad (2)$$

Hasil yang didapatkan melalui perhitungan persentase tersebut dapat dikategorikan seperti tabel 2

**Tabel 2.** Kriteria Interpretasi Skor *Usability*

% Jumlah Skor	Kriteria
20,00 % - 36,00%	Tidak Baik
36,01 % - 52,00%	Kurang Baik
52,01% - 68,00%	Cukup Baik
68,01% - 84,00%	Baik
84,01% - 100%	Sangat Baik

Tabel 3 menunjukkan hasil umpan balik yang didapatkan melalui pengujian *usability*, dimana rata-rata *usability* website *Itinerary Travaa* secara keseluruhan adalah 3.14 (62.95%), termasuk kriteria Cukup Baik, indikator *Learnability* mencapai 3.25 (65.00%), *Memorability* mencapai 3.27 (65.40%), *Efficiency* mencapai 3.04 (60.80%), *Errors* mencapai 3.23 (64.60%), dan *Satisfaction* mencapai 3.04 (60.80%). Dari tabel kriteria interpretasi skor, kelima indikator tersebut termasuk ke dalam kriteria "Cukup Baik".

**Tabel 3.** Hasil Nilai Pengujian Usability Website Sebelum *Redesign*

Indikator	Kode Pertanyaan	Presentase	Kriteria
<i>Learnability</i>	A1	65,33%	Baik
	A2	62,66%	Cukup Baik
	A3	62,66%	Cukup Baik
	A4	57,33%	Cukup Baik
<i>Memorability</i>	B1	61,33%	Cukup Baik
	B2	64%	Cukup Baik
<i>Efficiency</i>	C1	59,33%	Cukup Baik
	C2	57,33%	Cukup Baik
	C3	56%	Cukup Baik
<i>Erros</i>	D1	62,66%	Cukup Baik
	D2	60%	Cukup Baik
<i>Satisfaction</i>	E1	56%	Cukup Baik
	E2	58,66%	Cukup Baik
	E3	57,33%	Cukup Baik
	E4	56%	Cukup Baik
	E5	60%	Cukup Baik
Nilai Rata-Rata Keseluruhan		59,78%	Cukup Baik

## 2.2. Specify the Context of Use

Pada tahap ini dilakukan penentuan konteks user yaitu identifikasi user berguna untuk mengumpulkan data dari penggunaan sistem yang mencakup demografi, tujuan, kebutuhan untuk mengetahui tujuan dan dalam kondisi apa user akan menggunakan produk dimana pada penelitian ini melibatkan wisatawan yang sudah pernah menggunakan website *Itinerary Travaa* dan memiliki tujuan untuk menemukan informasi terkini tentang tempat wisata di Badung, merencanakan perjalanan dengan mudah serta kebutuhan mengenai informasi wisata lokal, kemudahan dalam merencanakan perjalanan, pengalaman yang nyaman dan *UI/UX Designer* yang merupakan pengguna baru website *Itinerary Travaa* yang memiliki tujuan untuk menganalisis tata letak dan navigasi situs web dan kebutuhan antarmuka pengguna yang intuitif, aplikasi website sesuai dengan kebutuhan pengguna, dan konsisten dari segi antarmuka.

## 2.3. Specify User and Organizational Requirement

### a. Data Kebutuhan Pengguna

Pada tahapan ini dilakukan pendeskripsian kebutuhan pengguna setelah analisis data dari evaluasi *usability* website sebelum *redesign* yang digunakan sebagai acuan dalam melakukan perbaikan dan perancangan ulang website *Itinerary Travaa* yang ditunjukkan pada Tabel 4.

**Tabel 4.** Fitur dan Temuan Masalah

Menu	Uraian Fitur Sebelum <i>Redesign</i>	Temuan Masalah
<i>Home</i>	Pada halaman <i>Home</i> page ini terdapat navigasi bar yang berisikan 3 menu yakni <i>Places</i> , <i>Blog</i> , dan <i>Itineraries</i> .	1. Pada indikator <i>Learnability</i> , <i>call to action</i> (CTA) " <i>Start</i> " pada <i>hero banner</i> tidak cukup jelas dan mungkin perlu

Menu	Uraian Fitur Sebelum <i>Redesign</i>	Temuan Masalah
	Selain navigasi bar terdapat <i>hero banner</i> yang berisi Call Action Button (CTA) “ <i>Start</i> ” yang akan mengarahkan user ke halaman <i>Itineraries</i> untuk membuat agenda atau rencana perjalanan.	diperjelas agar pengguna lebih mudah memahami langkah selanjutnya. 2. Pada indikator <i>Efficiency</i> Pengguna mengalami kesulitan dalam menavigasi atau menemukan informasi dengan cepat pada halaman <i>Home</i> .
<i>Places</i>	Pada halaman ini terdapat <i>list</i> tempat destinasi wisata yang dimana terdapat gambar serta deskripsi terkait tempat wisata tersebut.	Pada indikator <i>Satisfaction</i> beberapa responden tidak sepenuhnya puas dengan tata letak konten dan desain antarmuka pada halaman <i>Places</i> karena terlalu banyak <i>white space</i> dan terdapat <i>Errors</i> pada fitur <i>search bar</i> . Kemudian tidak terdapat tombol untuk menambahkan tempat wisata ke dalam <i>Itineraries</i> pengguna.
<i>Blog</i>	Pada halaman ini berisi artikel-artikel seputar tips dan cerita dari isi website ini.	Terlalu banyak <i>white space</i> pada awal menu <i>Blog</i> serta kurangnya warna pada <i>heading</i> membuat desain antarmuka menjadi kurang menarik sehingga perlu dilakukan perbaikan dan perubahan untuk meningkatkan indikator <i>Satisfaction</i> dan <i>Memorability</i> .
<i>Itineraries</i>	<i>Itineraries</i> merupakan fitur untuk membuat <i>trip planner</i> , atau melihat <i>trip planner</i> yang sudah user buat.	Pada indikator C3 <i>Efficiency</i> menunjukkan responden mengalami kesulitan dalam menemukan rencana perjalanan yang sesuai dengan keinginan mereka, serta penggunaan <i>copywriting</i> pada <i>button Itineraries</i> yang kurang tepat mempengaruhi tingkat <i>Learnability</i> pengguna.
<i>Login &amp; Register</i>	Pada menu ini dapat dilakukan dengan cara <i>Login</i> secara manual dengan akun <i>email</i> yang sudah didaftarkan sebelumnya dan <i>Login</i> secara otomatis dengan akun <i>Google</i> dan <i>Facebook</i> .	Responden mengeluhkan terdapat <i>Errors</i> yang mengakibatkan responden tidak dapat melakukan registrasi akun secara manual.
<i>My trip</i>	Menu ini merupakan menu yang berfungsi dalam pembuatan agenda perjalanan wisata yang dibuat oleh pengguna <i>Travaa</i> sendiri, dimana pada menu ini pengguna diwajibkan untuk melakukan <i>Login</i> terlebih dahulu. Pada menu ini pengguna dapat membuat agenda perjalanan baru atau menambahkan agenda melalui menu <i>Itineraries</i> .	Beberapa responden mengalami kesulitan dalam mengingat proses penggunaan <i>My Trip</i> sehingga mempengaruhi tingkat <i>Memorability</i> dari website <i>Itinerary Trava</i> ini, kemudian alur penggunaan dari menu ini tidak terlalu jelas serta responden mengalami kesulitan dalam menambahkan dan membuat <i>tripnya</i> sendiri.

#### 2.4. Product Design Solution

Pada tahap ini, dilakukan perancangan sistem dan aplikasi melalui pembuatan *User Flow*, *activity diagram*, desain basis data, *high-fidelity prototype design* website *Itinerary Travaa* dan implementasi sistem ke dalam kode program. Perancangan ulang *User Flow* disesuaikan dengan hasil evaluasi website *Itinerary Travaa* pada pengujian website sebelum *redesign*, dengan fokus meningkatkan *Learnability*, *Memorability*, *Efficiency*, dan *Errors* terutama pada menu *Home*, *Places*, dan *My Trip*, serta kepuasan pengguna secara keseluruhan. Perbaikan dan perancangan ulang desain antarmuka dilakukan untuk meningkatkan tingkat kepuasan pengguna website *Itinerary Trava* sesuai dengan kebutuhan dan umpan balik dari evaluasi usability website sebelum *redesign*.

## 2.5. Evaluate Against Design

Pada tahapan dilakukan pengujian dan mengevaluasi website yang telah dilakukan perancangan ulang dengan menggunakan metode yang sama pada saat proses pengumpulan data dari hasil evaluasi website *Itinerary Travaa* sebelum *redesign*. Kemudian dilakukan perbandingan *usability* antara hasil evaluasi pengujian website *Itinerary* sebelum dilakukan *redesign* dan website setelah dilakukan *redesign*.

## 3. Hasil dan Pembahasan

Hasil dan pembahasan menguraikan implementasi secara rinci mulai dari tahapan spesifikasi kebutuhan pengguna, solusi desain produk, implementasi sistem, serta evaluasi dan pengujian website setelah *redesign*.

### 3.1. Plan the Human Centered Design

Pada tahapan ini dilakukan uji validitas sebelum penyebaran kuesioner pengujian *usability*. Uji validitas pada penelitian ini dibantu menggunakan alat pengolahan data yaitu *Statistic SPSS* dengan menggunakan korelasi *Bivariate Pearson* (Produk Momen *Pearson*) yakni dengan cara mengkorelasikan masing-masing skor item dengan skor total [4]. Berikut ini merupakan hasil pengujian validitas yang ditunjukkan pada tabel 5:

**Tabel 5.** Hasil Nilai Uji Validitas Kuesioner

Indikator	Kode Pertanyaan	r-hitung	r-tabel	Keterangan
<i>Learnability</i>	A1	0,719	0,361	Valid
	A2	0,727	0,361	Valid
	A3	0,853	0,361	Valid
	A4	0,673	0,361	Valid
<i>Memorability</i>	B1	0,766	0,361	Valid
	B2	0,734	0,361	Valid
<i>Efficiency</i>	C1	0,546	0,361	Valid
	C2	0,813	0,361	Valid
	C3	0,596	0,361	Valid
<i>Erros</i>	D1	0,852	0,361	Valid
	D2	0,721	0,361	Valid
<i>Satisfaction</i>	E1	0,877	0,361	Valid
	E2	0,888	0,361	Valid
	E3	0,867	0,361	Valid
	E4	0,812	0,361	Valid
	E5	0,857	0,361	Valid

Validitas pengujian dianggap terpenuhi apabila  $r_{hitung} \geq r_{tabel}$  (uji 2 sisi dengan sig. 0,05) maka instrumen atau item-item pertanyaan berkorelasi signifikan terhadap skor total (dinyatakan valid). Berdasarkan tabel 5, hasil uji validitas menunjukkan bahwa 16 pertanyaan kuesioner memenuhi standar validitas yang ditetapkan. Selanjutnya, uji reliabilitas dilakukan menggunakan metode Cronbach's alpha dengan menggunakan SPSS.



Reliability Statistics	
Cronbach's Alpha	N of Items
.939	16

**Gambar 2.** Hasil Uji Reliabilitas

Hasil pada gambar 2 menunjukkan nilai alpha sebesar 0,939, menandakan bahwa seluruh item kuesioner memiliki tingkat reliabilitas yang tinggi, dengan nilai alpha yang melebihi 0,60, sehingga kuesioner dapat diandalkan untuk pengujian.

### 3.2. Specify User and Organizational Requirement

Setelah dilakukan pengujian *usability* pada *website Itinerary* Travaa sebelum *redesign*, didapatkan beberapa masukan dari responden pengujian *usability* yang ditunjukkan pada tabel 6 untuk meningkatkan indikator *usability* sesuai dengan temuan masalah pada tabel 4:

**Tabel 6.** Hasil Kuesioner Masukan Pengguna

Permasalahan	Insight Pengguna
Desain antarmuka yang kurang memadai	Perancangan ulang tata letak dan desain antarmuka serta atur jarak <i>white space</i> pada tampilan antarmuka
Tingkat kemudahan dalam mengingat yang rendah	Pemberian <i>highlight</i> pada judul dan perbaikan <i>copywriting</i> pada fitur.
<i>Efisiensi yang kurang pada fitur</i>	Penambahan <i>search engine</i> untuk menampilkan rekomendasi tempat wisata sesuai preferensi pengguna serta perancangan ulang alur navigasi pengguna.

Kemudian pada tahapan ini dilakukan spesifikasi kebutuhan pengguna dengan beberapa perbaikan fitur berdasarkan hasil pengujian *usability website* sebelum *redesign* yang ditunjukkan pada tabel 7 dan umpan balik yang diberikan pengguna

**Tabel 7.** Spesifikasi Perbaikan Kebutuhan Pengguna

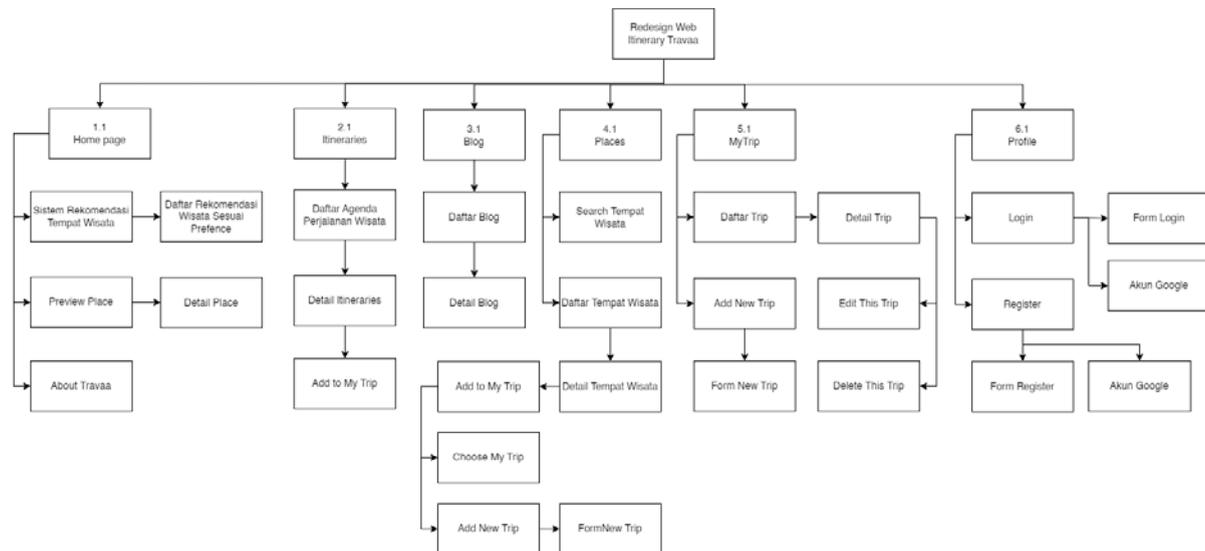
Menu	Perbaikan Fitur
<i>Home</i>	Pada pengujian <i>usability</i> , banyak pengguna mengalami kesulitan dalam memahami cara penggunaan <i>website</i> terutama karena menu navigasinya belum jelas. Oleh karena itu, dilakukan perubahan pada hero banner dengan menambahkan search bar yang memungkinkan pengguna mencari rekomendasi wisata sesuai preferensi pengguna dengan menampilkan perankingan destinasi wisata dari rate tertinggi-terendah, guna meningkatkan <i>Learnability</i> dan <i>Efficiency</i> . Selain itu, <i>redesign</i> antarmuka dilakukan untuk meningkatkan nilai <i>Satisfaction</i> . Konten juga diubah dengan menampilkan 'Trending Destination' agar informasi lebih mudah terbaca.
<i>Places</i>	Pada menu ini dilakukan perubahan terhadap tata letak dari <i>card</i> tempat wisata dan <i>search bar</i> yang jaraknya terlalu besar. Pada detail <i>place</i> juga tidak terdapat tombol untuk menambahkan tempat tersebut ke dalam agenda perjalanan ( <i>My Trip</i> ) sehingga perlu ditambahkan tombol tersebut dan perubahan tata letak serta penambahan mengenai informasi jam operasional, lokasi, dan testimoni dari pengguna yang diambil dari testimoni <i>Google Maps</i> .
<i>Blog</i>	Dilakukan perubahan terhadap tata letak dari <i>card</i> dan perubahan desain antarmuka guna meningkatkan indikator <i>Memorability</i> dan <i>Satisfaction</i> pengguna.
<i>Itineraries</i>	Menu ini berisi list <i>card</i> yang berisi kumpulan agenda tempat perjalanan wisata. Dilakukan beberapa perubahan, termasuk pada <i>detail Itineraries</i> , antarmuka <i>card Itineraries</i> , dan informasi tempat wisata yang ditampilkan untuk meningkatkan <i>Learnability</i> dan <i>Satisfaction</i> pengguna. Untuk meningkatkan <i>Memorability</i> , <i>Efficiency</i> , dan mengurangi <i>Errors</i> , dilakukan perancangan ulang terhadap <i>flow Itineraries</i> .
<i>Login / Register</i>	Pada menu ini dilakukan perbaikan sistem pada fitur <i>Login</i> secara manual dan perancangan ulang pada antarmuka.
<i>MyTrip</i>	Dilakukan perancangan ulang terhadap <i>flow</i> dan antarmuka.

### 3.3. Product Design Solution

Pada tahapan ini dilakukan pembuatan *site map*, *User Flow* terhadap *redesign* website *Itinerary Travaa*, *Activity diagram*, dan *High Fidelity Design Prototyping* serta pengimplementasian sistem ke dalam bentuk program sebagai solusi dalam perbaikan *website*.

#### a. Site Map

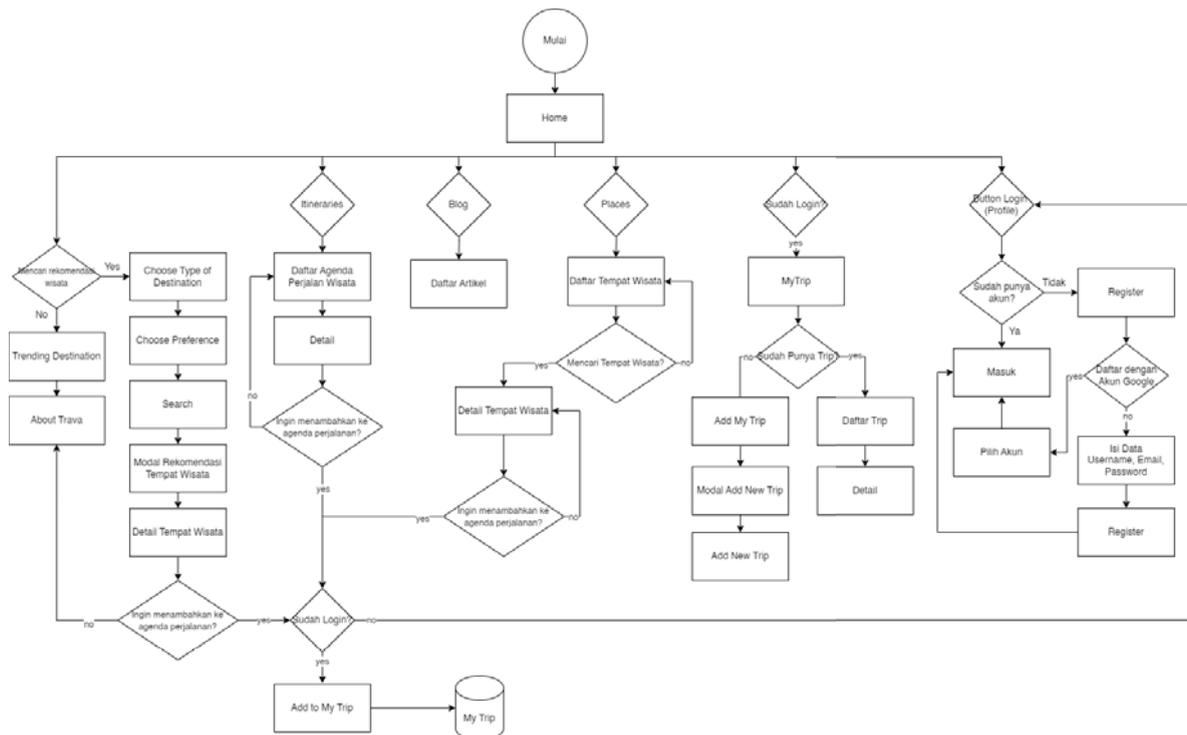
*Site map* merupakan daftar struktur hierarki halaman dalam sebuah situs web yang dapat membantu memberikan gambaran keseluruhan tentang hubungan antar halaman serta memastikan navigasi pengguna di situs web menjadi lebih efektif. Pada gambar 3 terlihat terdapat beberapa menu dan fitur seperti *Home Page*, *Itineraries*, *Blog*, *Places*, *My Trip*, dan *Profile* yang mana tidak terdapat penambahan atau pengurangan menu pada *redesign* website *Itinerary Travaa* sesuai dengan tabel kebutuhan pengguna pada tabel 7



Gambar 3. Site Map Website Itinerary Travaa

#### b. User Flow

*User Flow website* Itinerary Travaa yang dirancang ulang mencakup navigasi dari beranda ke fitur-fitur utama seperti rekomendasi pencarian destinasi, pembuatan *Itineraries*, dan interaksi dengan tempat wisata. Fokus utamanya adalah meningkatkan pengalaman pengguna dengan memastikan kemudahan navigasi, interaksi intuitif, dan keterhubungan yang jelas antara berbagai bagian aplikasi. Gambar 4 merupakan *redesign user flow* dari *website Itinerary Travaa* secara keseluruhan. Pada gambar menunjukkan saat pengguna memulai *website* pengguna akan masuk pada halaman *Home* yang terdapat menu-menu dan fitur seperti *Itineraries*, *Blog*, *Places*, *MyTrip* dan *Profile*. Pada halaman *Home* sendiri, jika pengguna ingin mencari rekomendasi tempat wisata pengguna dapat memilih 'type of destination' dan 'preference' lalu klik *button search*, kemudian sistem akan menampilkan rekomendasi tempat wisata berupa modal sesuai dengan preferensi pencarian pengguna. Jika pengguna ingin melihat daftar agenda perjalanan wisata, pengguna juga dapat memilih menu *Itineraries*, pengguna dapat melihat detail informasi mengenai agenda tersebut dan menambahkannya ke dalam *trip* pengguna. Selain itu pengguna juga dapat membaca artikel pada menu *Blog*. Lalu pada menu *Places*, pengguna dapat melihat, mencari, dan menambahkan destinasi wisata ke dalam agenda perjalanan pengguna. Namun untuk menambahkan ke dalam agenda perjalanan dan menggunakan fitur *Mytrip*, pengguna diharuskan untuk melakukan *login terlebih dahulu*. Lalu pada menu *Profile* pengguna dapat *Login* dan *Register* secara manual dengan menggunakan email atau pengguna juga dapat melakukannya dengan otomatis menggunakan akun Google yang terdapat pada perangkat yang digunakan pengguna.



Gambar 4. User Flow Redesign Website Itinerary Travaa

### c. High Fidelity Design Prototype dan Implementasi Sistem

Pembuatan *high-fidelity design* berdasarkan evaluasi dan pengujian *usability* sebelum *redesign*. Pada tahapan ini dilakukan pembuatan desain sistem dan *high fidelity design* yang kemudian diimplementasikan ke dalam sistem *website* agar dapat digunakan secara langsung oleh pengguna.

#### 1. Pembuatan Desain Sistem

Pembuatan desain sistem bertujuan untuk meningkatkan konsistensi dalam desain antarmuka, termasuk penggunaan warna dan *font*, efisiensi dalam pengembangan, dan pengalaman pengguna dengan menyediakan panduan dan aturan desain yang jelas. Pada *redesign website* Itinerary Traava, digunakan warna-warna seperti yang ditunjukkan pada gambar 5 yang mana warna oranye digunakan sebagai *primary color* dari *redesign website* ini, warna biru sebagai *secondary color*, dan ungu tua sebagai warna untuk *text heading* serta menggunakan *font* DM Sans agar memberikan kombinasi antara keterbacaan, modernitas, dan fleksibilitas.



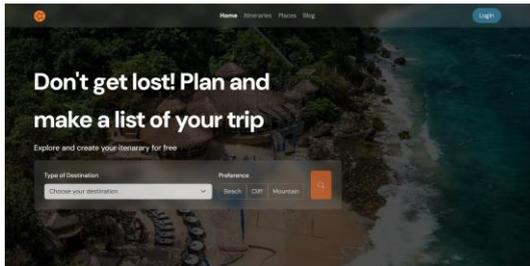
Gambar 5. Desain Sistem Warna dan Font Redesign Travaa

#### 2. Pembuatan High Fidelity Design dan Implementasi Sistem

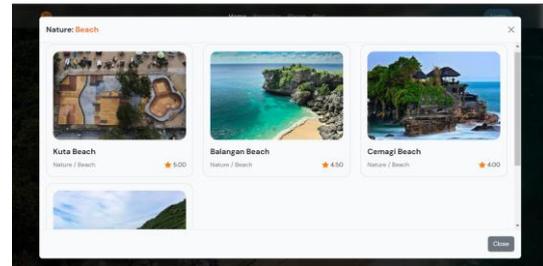
Pada tahapan ini dilakukan pembuatan *high fidelity design prototype* yang dirancang menggunakan tools *Figma*. Pada gambar 6, 7, 8 dan 9 merupakan tampilan antarmuka dari perancangan ulang beberapa menu dan pengembangan fitur sesuai dengan umpan balik pengguna pada tabel 6 dan spesifikasi kebutuhan pengguna pada tabel 7. Pada gambar 6 dilakukan penambahan fitur pencarian

Perancangan Ulang Website Itinerary untuk Daerah Wisata Kabupaten Badung Menggunakan Metode User Centered Design

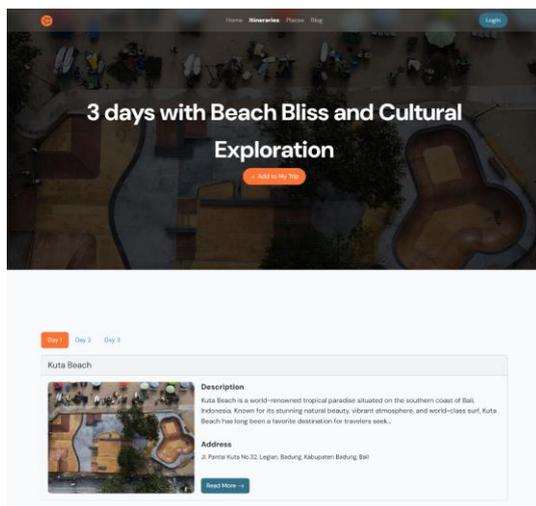
rekomendasi tempat wisata sesuai dengan preferensi pengguna yang akan menghasilkan rekomendasi wisata sesuai dengan perankingan rate terendah-tertinggi yang ditunjukkan pada gambar 7. Kemudian gambar 8 dan 9 merupakan tampilan dari detail menu *Itineraries* dan menu *Place*, dimana dilakukan perancangan ulang tampilan antarmuka dan *layout*. Pada gambar 8 dilakukan penambahan fitur 'add my trip' dan detail *trip* per harinya dan pada gambar 9 dilakukan penambahan informasi mengenai tempat wisata. Pada gambar10 yang merupakan detail dari menu Blog dilakukan perancangan ulang pada *font* dan pemberian warna pada judul agar pengguna lebih terfokus dan meminimalkan distraksi visual yang tidak perlu. Gambar 11 merupakan tampilan antarmuka menu detail *MyTrip*, dimana dilakukan penambahan fitur untuk menambahkan, mengedit, dan menghapus *trip* pengguna, serta tambahan agenda perjalanan wisata yang disajikan dalam bentuk kalender dan list tempat wisata per harinya.



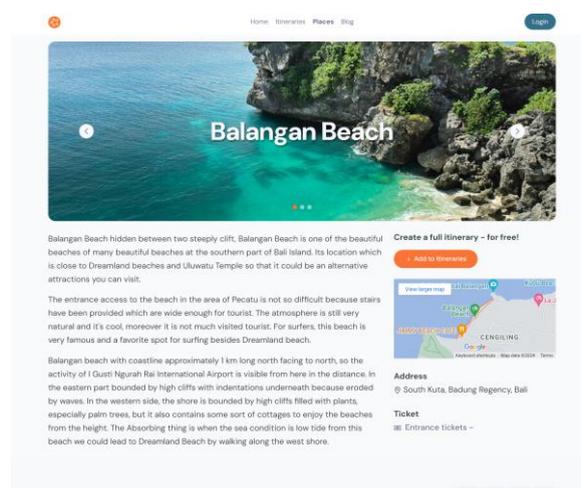
Gambar 6. Home Page



Gambar 7. Hasil Rekomendasi Tempat Wisata



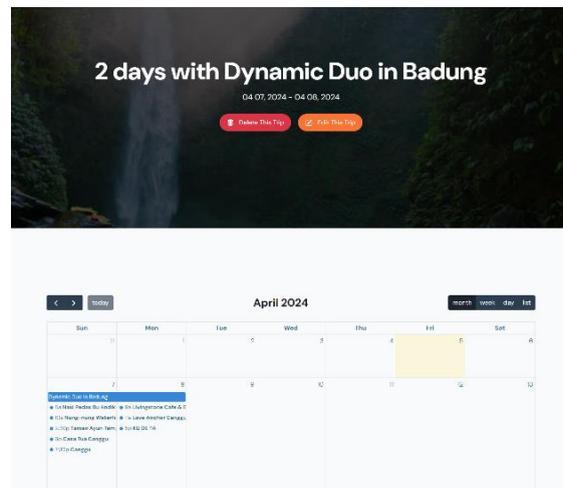
Gambar 8. Detail Itineraries



Gambar 9. Detail Place



Gambar 10. Detail Blog



Gambar 11. Detail My Trip

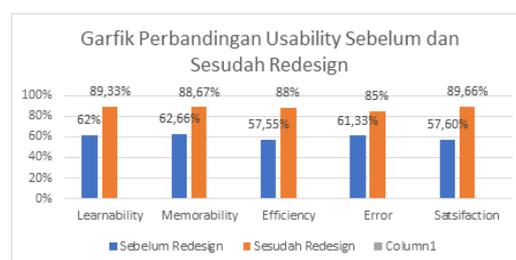
### 3.4. Evaluate Against Design

Pada tahapan ini dilakukan pengujian *usability website* setelah dilakukan *redesign*, dimana melibatkan 30 responden dan membagikan kuesioner yang sama dengan pengujian *website* sebelum *redesign* untuk membandingkan nilai *usability website Itinerary Travaa* sebelum dan sesudah perbaikan serta untuk mengetahui apakah *website* sudah memenuhi kebutuhan pengguna setelah dilakukan perbaikan dan perancangan ulang.

**Tabel 8.** Hasil Nilai Pengujian *Usability Website* Sesudah *Redesign*

Indikator	Kode Pertanyaan	Presentase	Kriteria
Learnability	A1	94,66%	Sangat Baik
	A2	86,66%	Sangat Baik
	A3	90,00%	Sangat Baik
	A4	86,00%	Sangat Baik
Memorability	B1	89,33%	Sangat Baik
	B2	88%	Sangat Baik
Efficiency	C1	89,33%	Sangat Baik
	C2	86,00%	Sangat Baik
	C3	88,66%	Sangat Baik
Erros	D1	84%	Baik
	D2	86,66%	Sangat Baik
Satisfaction	E1	91,33%	Sangat Baik
	E2	89%	Sangat Baik
	E3	90,66%	Sangat Baik
	E4	85,33%	Sangat Baik
	E5	92,00%	Sangat Baik
Nilai Rata-Rata Keseluruhan		88,62%	Sangat Baik

Dari hasil pengujian *usability* pada tabel 8, rata-rata nilai keseluruhan untuk 5 indikator Usability mencapai 88,62%, menunjukkan peningkatan yang signifikan. Indikator Learnability mendapatkan nilai antara 86,00% hingga 94,66%, dengan semua kriteria dikategorikan sebagai "Sangat Baik". Memorability juga mendapat nilai "Sangat Baik" dengan rentang nilai antara 88% hingga 89,33%. Efisiensi meningkat secara signifikan dengan nilai "Sangat Baik" antara 88,66% hingga 89,33%. Indikator Error juga mendapat nilai "Sangat Baik" dengan rentang antara 84% hingga 86,66%. Sedangkan indikator Satisfaction mencapai nilai antara 85,33% hingga 92,00%, juga dengan kriteria "Sangat Baik" untuk semua kriteria.



**Gambar 12.** Grafik Perbandingan *Usability* Sebelum dan Sesudah *Redesign*

Grafik perbandingan nilai rata-rata *Usability website* sebelum dan setelah *redesign* pada gambar 12 menunjukkan peningkatan yang signifikan pada tiap indikator *Usability* di antaranya *Learnability* meningkat sebanyak 44,24%, dari 62% menjadi 89,33%. *Memorability* meningkat 41,24%, dari 62,66% menjadi 89,33%. *Efficiency* meningkat 52,81%, dari 57,55% menjadi 88%. *Error* meningkat 38,48%, dari 61,33% menjadi 85%, dan *Satisfaction* meningkat 55,61%, dari 57,6% menjadi 89,66%. Peningkatan nilai *usability* setelah *redesign* sebesar 48,21% menunjukkan perubahan yang signifikan. Dari 59,78% menjadi 88,62%, hasil ini mencerminkan efektivitas dan keberhasilan perubahan yang dilakukan dalam proses *redesign*.

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

- a. Hasil evaluasi pengujian usability terhadap website Itinerary Travaa sebelum *redesign* menghasilkan nilai rata-rata secara keseluruhan sebesar 59,78% dengan kriteria "Cukup Baik". Hasil evaluasi menunjukkan, pada indikator *Learnability* beberapa pengguna mengalami kesulitan dalam mempelajari penggunaan situs; pada indikator *Memorability*, masih terdapat kekurangan penekanan informasi pada fitur-fitur kunci; pada indikator *Efficiency*, terdapat permasalahan pada alur navigasi dan pencarian informasi tidak optimal; evaluasi indikator Errors ditemukan beberapa pengguna mengalami kesalahan karena fitur dan menu tidak berfungsi; pada indikator *Satisfaction* beberapa pengguna tidak sepenuhnya puas dengan desain antarmuka dan alur navigasi.
- b. Perancangan ulang website *Itinerary Travaa* untuk daerah wisata di Kabupaten Badung, menggunakan metode *User Centered Design*, menghasilkan peningkatan signifikan dalam tingkat *Usability* menjadi "Sangat Baik", dengan rata-rata 88,62%. Kelima indikator *Usability* mencapai kriteria "Sangat Baik", menunjukkan keberhasilan metode *User Centered Design* dalam menghasilkan pengalaman pengguna yang optimal dan memuaskan dengan memprioritaskan kebutuhan dan preferensi pengguna dalam perancangan.

#### References

- [1] Dinas Pariwisata Provinsi Bali, "Data Statistik Wisata Bali Tahun 2019," 2019. [Online]. Available : <https://disparda.baliprov.go.id/category/data-statistik/>. [accessed on March , 17 2023]
- [2] Imanulhaq M.S, Andreswari Rachmadita, and Al Anshary F.M, "Perancangan Aplikasi Perencanaan Wisata (Travelkey) Pada Modul Itinerary Dengan Menggunakan Metode Iterative and Incremental," e-Proceeding Eng., vol. 8, no. 1, pp. 602–616, 2021.
- [3] I. U. Khasanah, M. Fachry, N. S. Adriani, N. Defiani, Y. Saputra, and A. Ibrahim, "Penerapan Metode User Centered Design dalam Menganalisis User Interface pada Website Universitas Sriwijaya," INTEGER J. Inf. Technol., vol. 3, no. 2, pp. 21–27, 2018, doi: 10.31284/j.integer.2018.v3i2.226.
- [4] J. F. Hair, W. C. Black, B. J. Babin, dan R. E. Anderson, *Multivariate Data Analysis*, edisi ke-8. Cengage Learning, 2019.
- [5] J. Nielsen, "Usability 101: Introduction to Usability," 2012. [Online]. Available : <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>. [accessed on March 23, 2023]
- [6] Sugiyono, *Metode Penelitian Kuantitatif dan R&D*. Bandung: Alfabeta, 2018.

# Implementasi Sistem Informasi E-Learning Berbasis Odoo Studi Kasus MI YPPI 1945 Babat

Ahmad Assrorul Abidin<sup>a1</sup>, Nufan Balafif<sup>a2</sup>, Eddy Kurniawan<sup>a3</sup>

<sup>a</sup>Program Studi Sistem Informasi  
Fakultas Sains dan Teknologi  
Universitas Pesantren Tinggi Darul Ulum Jombang

<sup>1</sup>[ahmadassrorulabidin23@email.com](mailto:ahmadassrorulabidin23@email.com)

<sup>2</sup>[nufanbalafif@ft.unipdu.ac.id](mailto:nufanbalafif@ft.unipdu.ac.id)

<sup>3</sup>[eddykurniawan@ft.unipdu.ac.id](mailto:eddykurniawan@ft.unipdu.ac.id)

## Abstrak

Penyampaian informasi materi sekolah menggunakan media buku paket membuat pembelajaran menjadi membosankan dan kurang daya tarik siswa mengikuti proses pembelajaran. Penilaian yang menghabiskan banyak biaya dengan menggunakan kertas dan sering hilangnya hasil kerja saat proses pengolahan nilai. Proses penilaian yang sering terjadi kesalahan dan membutuhkan waktu yang lama. Sehingga, informasi tidak cepat dan akurat. Analisis proses bisnis yang berjalan saat memberikan informasi materi, tugas dan penilaian hasil kerja menghasilkan proses bisnis yang diusulkan menggunakan teknologi. Aplikasi elearning odoo dapat menyampaikan materi secara online fleksibel, tugas yang praktis, efektif dan efisiensi, penilaian kinerja yang otomatis dan cepat serta target pembelajaran yang otomatis. Pengujian sistem dengan metode UAT (User Acceptance Testing) untuk melihat kualitas sistem yang berjalan. Penentuan sampel menggunakan rumus slovin untuk siswa dan sampel jenuh untuk guru. Hasil perhitungan rumus dengan populasi 73 menghasilkan 63 sampel siswa dengan signifikansi 5% dan 5 sampel guru. Sistem membantu proses pembelajaran efektif dan efisien disekolah dengan pengujian kualitas sistem elearning metode UAT. Pengujian sistem metode UAT dengan indikator Kelengkapan (Completeness), Konsisten (Consistency), Pelacakan (Traceability), Operabilitas (Operability), Pelatihan (Training), Keamanan (Security), Akurasi (Accuracy), Kesederhanaan (Simplicity), Kemudahan Eksekusi (Execution Efficiency) menghasilkan kualitas persentase 88% yang artinya sangat baik sistem e-learning yang berjalan di sekolah MI YPPI 1945 Babat

**Kata Kunci:** Informasi, Proses, Elearning, Odoo, UAT

## 1. Pendahuluan

Pendidikan merupakan pondasi untuk menciptakan seseorang yang berkualitas dan berkarakter sehingga dapat memiliki pandangan yang luas tentang kehidupan. Pendidikan memang wajib bagi setiap individu baik dari tingkat dasar maupun tingkat tinggi. Sekolah merupakan wadah untuk melakukan interaksi pembelajaran bagi setiap siswa untuk mendapatkan ilmu yang akan berguna di dunia. Proses pembelajaran disekolah diberikan oleh guru untuk memberikan isi materi kepada para siswa serta membimbing dan mengarahkan para siswanya dalam kegiatan belajarnya menggunakan metode penyampaian serta strategi pembelajaran yang tepat dan dengan dukungan dari sumber dan bahan yang tersedia [1]. Perkembangan teknologi didalam dunia pendidikan berjalan sangat pesat dan mudah diapatkan untuk mempermudah setiap aktifitas yang ada disekolah. Pemanfaatan sistem informasi yang mudah dan cepat menjadikan setiap aktifitas berjalan dengan efektif dan efisien. Elearning adalah suatu media pembelajaran digital yang sering dimanfaatkan pada dunia pendidikan sekolah untuk menunjang jalannya suatu proses penyampaian materi secara efektif [2]. Sistem elearning odoo merupakan aplikasi sistem yang digunakan untuk mempermudah dalam menyampaikan materi, penilaian kerja dan target belajar secara otomatis.

Lembaga pendidikan MI YPP 1945 Babat adalah lembaga pendidikan yang termasuk dalam Satuan Pendidikan Kementerian Agama. Banyaknya jumlah siswa kurang lebih 500 siswa menjadikan sekolah favorit dilingkungan sekitar. Ketersediaan fasilitas teknologi di lembaga yang cukup dapat dimanfaatkan untuk melakukan pembelajar berbasis teknologi. Kurangnya pemanfaatan teknologi di MI YPPI 1945 yang dilakukan guru, siswa maupun pihak yang bertanggungjawab di bidang teknologi menjadi

landasan untuk melakukan implementasi elearning di sekolah. Penyampaian informasi materi sekolah masih banyak menggunakan media buku paket yang membuat pembelajaran menjadi membosankan dan kurang daya tarik siswa mengikuti proses pembelajaran. Penilaian atau kuis yang diberikan pada siswa sering menghabiskan banyak biaya, karena masih menggunakan kertas dan terjadi hilangnya hasil kerja siswa saat melakukan proses pengolahan nilai. Proses penilaian kerja membutuhkan proses yang cukup lama dan sering terjadi kesalahan saat melakukan penilaian. Sehingga, informasi tidak berjalan dengan cepat dan akurat.

Dalam permasalahan yang terjadi di sekolah MI YPPI 1945 Babat. Implementasi elearning odoo diharapkan mampu menjadi solusi bagi pembelajaran sekolah. Dengan fitur yang sesuai kebutuhan sekolah terkait proses penyampaian materi, informasi tugas dan proses penilaian kerja serta memiliki tampilan yang dapat disesuaikan dengan kebutuhan akan membantu permasalahan yang ada di sekolah. Pengujian sistem menggunakan metode UAT (*User Acceptance Testing*) terhadap penerapan sistem yang berjalan diharapkan dapat melihat kualitas sistem dengan melakukan kuesioner responden terhadap pengguna. Sehingga, sistem yang berjalan terlihat kualitas sesuai atau tidaknya dengan kebutuhan yang ada di sekolah MI YPPI 1945 Babat Lamongan.

## **2. Tinjauan Pustaka**

### **2.1. Sistem Informasi**

Sistem informasi adalah bentuk dari komunikasi sistem yang dapat mewakili data serta diproses sebagai bentuk dari memori sosial. Sistem informasi dapat dianggap pula sebagai bahasa semi formal untuk mendukung manusia dalam mengambil keputusan atau tindakan. Pada hakikatnya sistem adalah sebuah sistem yang dilakukan dalam menyediakan informasi ketika hendak mengambil keputusan untuk manajemen dan dalam rangka menjalankan operasional dan prosedur yang terorganisir. Sistem informasi bertujuan untuk menjaga tercapainya pesan atau informasi dari seseorang atau kelompok kepada pihak lain [3]. Sistem informasi memiliki beberapa fungsi untuk mengumpulkan data yang telah diinput sebagai sumber utama jalannya sistem, menyimpan data yang sudah didapatkan saat melakukan pengumpulan data dan menghasilkan informasi yang dibutuhkan oleh sistem pengguna. Keunggulan sistem informasi untuk meningkatkan efisiensi operasional, inovasi dalam bisnis dan membangun informasi strategis.

### **2.2. E-learning**

E-learning adalah pembelajaran online dengan metode proses pembelajaran menggunakan teknologi informasi yang dilengkapi dengan jaringan internet. Pendekatan inovasi pembelajaran online yang mendistribusikan desain menarik untuk dijalankan dalam kegiatan belajar bagi setiap orang. Karakteristik E-learning memiliki aksesibilitas, mandiri, pengayaan, dan interaktivitas. Elearning merupakan pembelajaran yang berbasis teknologi untuk dijalankan secara fleksibel. Penggunaan pembelajaran teknologi yang tidak terbatas oleh ruang dan waktu baik secara bersama atau mandiri. Elearning meningkatkan ketertarikan seseorang untuk mendapatkan atau memberikan informasi dengan melalui media teknologi, seperti penggunaan website, video, audio ataupun media sosial lainnya. Dengan berbagai manfaat penggunaan elearning sebagai media pembelajaran di sekolah dapat meningkatkan mutu belajar guru dan siswa. Penggunaan elearning dapat mengontrol setiap aktivitas pembelajaran di sekolah. Baik itu, penyampaian informasi materi, kuis atau ujian online, dan penilaian kerja pengguna. Semakin menarik pembelajaran dengan elearning akan memudahkan dalam setiap interaksi yang terjadi antara sekolah, guru dan siswa.

### **2.3. Odoo**

Odoo adalah sebuah software ERP yang populer yang dapat digunakan serta dimanfaatkan oleh semua pihak baik perusahaan atau lembaga dalam menerapkan sistem ERP yang dibutuhkan. Odoo memiliki fleksibilitas yang dapat disesuaikan dengan berbagai kebutuhan perusahaan. Odoo yang dapat melakukan integrasi dan adaptasi keseluruhan perusahaan bahkan perusahaan kecil ataupun besar dengan permasalahan yang rumit apapun permasalahan yang ada di perusahaan. Aplikasi odoo merupakan aplikasi yang banyak memiliki kategori untuk digunakan perusahaan. Baik itu, kategori sales, kategori services, kategori accounting, kategori inventory, kategori manufacturing, kategori websites, kategori marketing, kategori human resources, kategori, productivity dan administrasi. Keuntungan yang didapatkan saat menggunakan aplikasi odoo. Antara lain, aplikasi gratis, mudah dioperasikan, tampilan menarik yang dapat disesuaikan, sistem fleksibel dan mudah dikembangkan.

## 2.4. Use Acceptance Testing Odoo

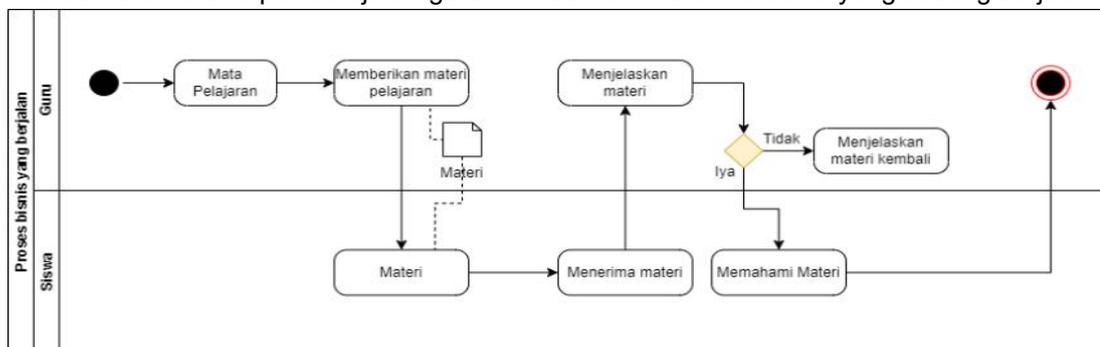
User Acceptance Testing (UAT) adalah pengujian sistem yang dilakukan oleh pengembang atau developer untuk melihat kualitas aplikasi yang berjalan. Dengan melakukan pengujian akan dapat melihat aplikasi sesuai dengan kebutuhan dan ekspektasi pengguna. *bug* merupakan masalah yang sering muncul saat sistem sudah berjalan. Dengan pengujian sistem, pengembang dapat melakukan identifikasi masalah *bug* dan dapat segera di perbaiki. Keuntungan yang didapatkan saat melakukan pengujian sistem dapat mengurangi resiko biaya yang akan dikeluarkan saat sistem sudah di pasaran atau diluncurkan. Pengujian sistem yang menghasilkan kualitas yang baik akan meningkatkan reputasi aplikasi tersebut. Pada umumnya UAT merupakan langkah penting untuk memvalidasi persyaratan pengguna, mendeteksi kesalahan, mengurangi risiko, dan memastikan kepuasan pengguna akhir. Dengan melibatkan pengguna dalam pengujian, UAT membantu membangun sistem yang sesuai dengan kebutuhan dan ekspektasi pengguna, meningkatkan keberhasilan implementasi, dan memberikan nilai tambah yang signifikan bagi organisasi atau entitas yang menggunakan sistem tersebut[4].

## 3. Metode Penelitian

### 3.1. Analisis Proses Bisnis

Proses bisnis masih berjalan di MI YPPI 1945 Babat menggunakan cara manual dan memakan waktu yang cukup lama untuk di jalan serta tidak siswa masih kesulitan dalam memahami materi yang terlewatkan sehingga pembelajaran kurang efektif dan efisiensi. Proses bisnis yang berjalan sebagai berikut :

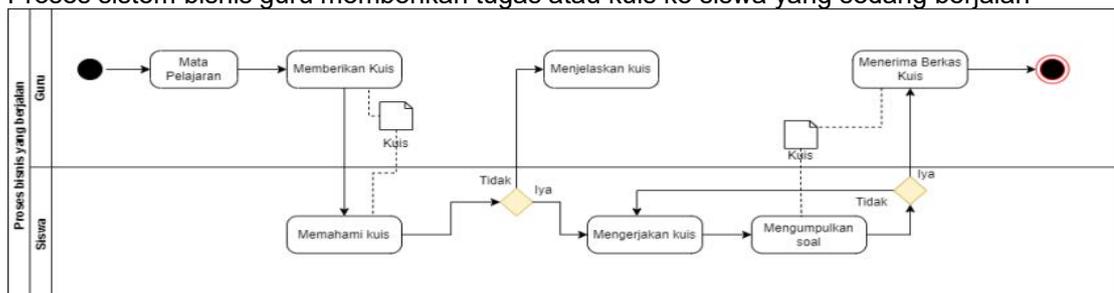
1. Proses sistem bisnis pembelajaran guru memberikan materi ke siswa yang sedang berjalan



Gambar 1. BPMN materi siswa yang berjalan

Pada gambar 1 diatas. dapat dijelaskan proses bisnis pemberian materi guru kepada siswa dilakukan secara tatap muka atau interaksi langsung. Memberian materi terbatas dengan adanya durasi waktu belajar sehingga pembelajaran siswa berjalan secara cepat dan sering terjadi ketidak fahaman siswa terhadap materi yang disampaikan. Disamping itu, pembelajaran masih belum menerapkan teknologi untuk menarik motivasi siswa dalam belajar

2. Proses sistem bisnis guru memberikan tugas atau kuis ke siswa yang sedang berjalan

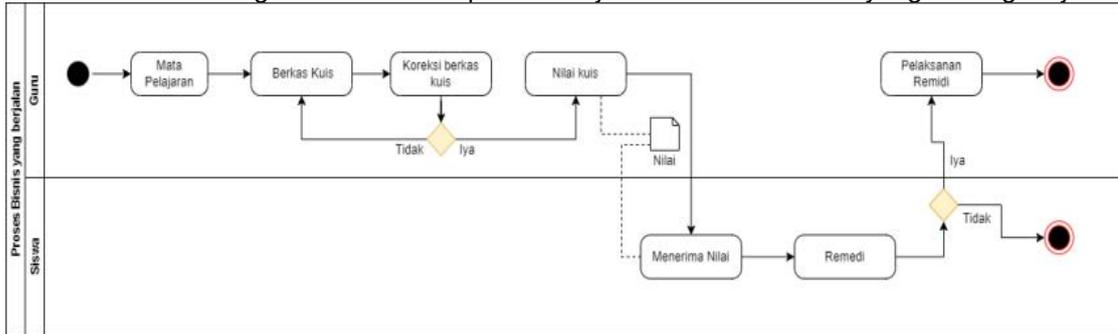


Gambar 2. BPMN tugas yang berjalan

Pada gambar 2 diatas. dapat dijelaskan bahwa proses bisnis pemberian ujian atau kuis pada siswa masih menggunakan kertas sebagai media alat kerja siswa. Media kertas masih

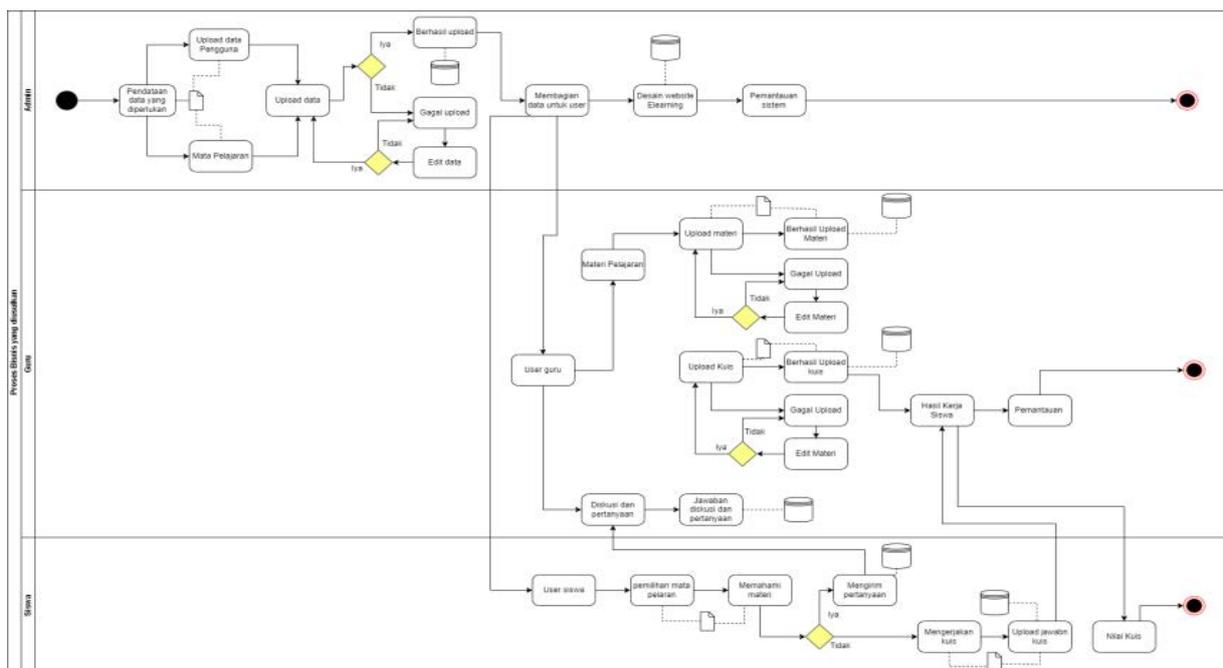
tergolong media yang mudah hilang dan mudah adanya kebocoran soal serta kurang menarik bagi siswa dalam melakukan kegiatan ujian.

3. Proses sistem bisnis guru memberikan penilaian ujian atau kuis ke siswa yang sedang berjalan



Gambar 3. BPMN penilaian tugas

3.2. Proses Bisnis yang diusulkan



Gambar 4. BPMN proses bisnis yang diusulkan

Pada gambar 4 diatas. dapat dijelaskan bahwa proses bisnis berjalan menggunakan teknologi dan data otomatis tersimpan saat melakukan proses upload data. Tampilan dan menu yang bisa disesuaikan dengan kebutuhan pembelajaran serta penilaian kinerja yang mudah dalam memantauan dan cepat.

Tabel 1. Analisis GAP sistem

No	Konten	YPPI 1945	Odoo	Keterangan
1	Penyampaian Materi	Buku	Fleksibel Online	Pembelajaran secara fleksibel online dari media dokumen, video, dan sumber lainnya cocok diterapkan di sekolah
2	Pemberian tugas atau kuis	Kertas manual	Online	Pemberian tugas secara praktis, efektif dan efisiensi dengan aplikasi odoo akan menghebat

				biaya operasional yang dikeluarkan
3	Penilaian Kinerja siswa	Perhitungan manual	Otomatis	Penilaian hasil kerja siswa yang otomatis akan mempercepat waktu dan mengurangi kesalahan saat penilaian
4	Data target pembelajaran	Manual	Otomatis	Target pencapaian sistem akan otomatis muncul di dalam sistem odoo
5	Pembelajaran secara tata muka langsung	V	X	Dilakukan secara tata muka langsung pada waktu yang ditentukan.

### 3.3. Metode UAT

Pengujian sistem e-learning dengan metode UAT digunakan untuk melihat seberapa kualitas sistem yang berjalan. Dengan pengambilan sampel populasi sebagai pengujian sistem e-learning yang berjalan. Populasi adalah keseluruhan objek atau subjek yang akan diteliti. Penentuan sampel siswa menggunakan rumus "Slovin" karena memiliki populasi yang banyak dan guru menggunakan sampel jenuh atau total karena populasi kurang dari 30 orang [5]. Rumus Slovin :

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Keterangan :

- n : Jumlah Sampel
- N : Jumlah Populasi
- e : Tingkat Signifikansi atau Batas Kesalahan

Metode UAT (*User Acceptance Testing*) terdapat rumus perhitungan yang digunakan untuk mengetahui tanggapan responden (pengguna) terhadap sistem yang dibuat. Rumus perhitungan tersebut membutuhkan bobot nilai untuk menghitung kuesioner yang akan digunakan [4].

*Tabel 2. Jawaban dan Bobot Nilai*

Jawaban	Bobot Nilai
Sangat Setuju (SS)	4
Setuju (S)	3
Tidak Setuju (TS)	2
Sangat Tidak Setuju (STS)	1

Proses perhitungan presentase dari hasil bobot nilai yang sudah dijawab responden menggunakan rumus :

- Mengitung jumlah nilai

$$(N1 \times 4) + (N2 \times 3) + (N3 \times 2) + (N4 \times 1) = \text{jumlah} \quad (2)$$

- Menghitung rata-rata nilai

$$\frac{\text{Jumlah Akhir}}{\text{Jumlah Pengguna}} = \text{Rata-rata} \quad (3)$$

- Menghitung presentase

$$\frac{\text{Rata-rata}}{\text{Nilai Bobot Maksimal}} \times 100 \% = \text{Presentase} \quad (4)$$

Untuk mempermudah dalam melakukan pengujian sistem diperlukan indikator penyusunan pernyataan yang nantinya diberikan kepada kuesioner. Dengan melakukan penyusunan pernyataan sesuai indikator pada sistem e-learning yang berjalan akan menemukan hasil kualitas produk yang telah dicoba oleh pengguna [6]. Terdapat 9 indikator untuk menguji kualitas sistem yang berjalan :

*Tabel 3. Indikator Pernyataan Siswa*

No	Indikator	Pernyataan
1	Kelengkapan (Completeness)	Sistem menampilkan setiap fitur menu
		Menampilkan kebutuhan sesuai fitur menu
		Sistem yang terupdate
2	Konsisten (Consistency)	Tampilan e-learning yang menarik
		Bahasa yang mudah di pahami
3	Pelacakan (Tracebility)	Dapat melakukan pencarian disetiap tampilan
		Dapat melihat kesalahan pengguna
4	Operabilitas (Operabilty)	Sistem mudah digunakan
		Menu yang mudah dipahami
		Informasi yang dicari mudah didapatkan
		Memberikan kenyamanan pengguna
5	Pelatihan (Training)	Memberikan motivasi belajar pada pengguna
		Memiliki arahan petunjuk bagi pengguna
		Ada fitur bantuan secara online
		Dapat mengirimkan kritik dan saran pada e-learning
6	Keamanan (Security)	Memberikan informasi bantuan secara jelas dan mudah dipahami
		Login mudah sesuai harapan
		Dapat melakukan pengaturan pada pengguna
7	Akurasi (Accuracy)	Keamanan materi dan soal sesuai harapan
		Menampilkan data sesuai perintah
		Memberikan data yang real
		Informasi akurat
8	Kesederhanaan (Simplicity)	Pengguna mendapatkan informasi sesuai kebutuhan
		Sistem informasi mudah dipahami
		Fitur menu mudah dipahami
9	Kemudahan Eksekusi (Execution Efficiency)	Tampilan yang sederhana namun menarik untuk pengguna
		Proses data yang cepat
		Fitur menu mudah dijalankan
		Jaringan yang lancar
		Fasilitas perangkat yang memadai

*Tabel 4. Indikator Pernyataan Guru*

No	Indikator	Pernyataan
	Kelengkapan (Completeness)	Sistem menampilkan fitur sesuai dengan kebutuhan
		Sistem yang terupdate
	Konsisten (Consistency)	Bahasa yang bisa disesuaikan dan mudah dipahami
		Tampilan menarik yang sesuai kebutuhan
	Pelacakan (Tracebility)	Dapat melakukan pencarian
		sistem yang bisa melakukan pengecekan pengguna
	Operabilitas (Operabilty)	Menu yang mudah dipahami
		Sistem yang memberikan kenyamanan dalam melakukan proses pembelajaran
		Informasi yang mudah didapatkan

Pelatihan (Training)	Terdapat petunjuk dalam sistem
	Bantuan bisa secara online
Keamanan (Security)	Privasi akun terjamin
	Dapat melakukan pengaturan akun
Akurasi (Accuracy)	Menampilkan data sesuai perintah pengguna
	Informasi akurat dan cepat
	Memberikan data yang sesuai keadaan
Kesederhanaan (Simplicity)	Tampilan yang simpel namun elegan
	Fitur sederhana sesuai kebutuhan
Kemudahan Eksekusi (Execution Efficiency)	Jaringan yang lancar
	Sistem yang mudah dan cepat
	Kemampuan fasilitas yang memadai

#### 4. Hasil dan Pembahasan

##### 4.1. Impelmentasi Sistem

Aplikasi odoo merupakan aplikasi berbasis web dengan bahasa pemrograman yaitu python, javascript, XML, sedangkan database yang digunakan yaitu postgresql. Sistem yang dapat di aplikasikan ke dalam pembelajaran yang ada di MI YPPI 1945 Babat Lamongan

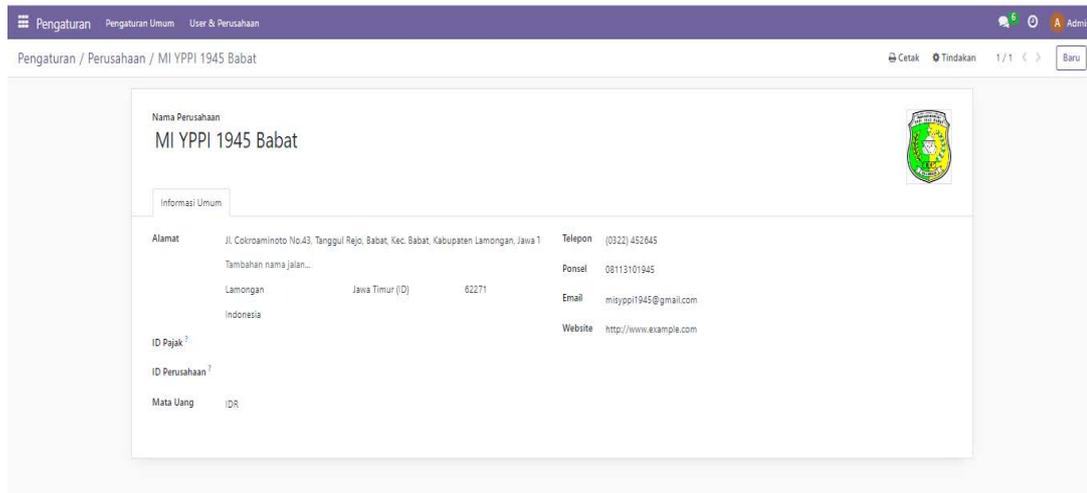
##### a. Instalasi Odoo



Gambar 5. Insall Aplikasi Odoo

Untuk mendapatkan software odoo melalui website resmi milik odoo. untuk menghindari malware yang berbahaya. Pada gambar 5 merupakan proses install aplikasi odoo yang akan digunakan untuk melakukan proses bisnis sesuai kebutuhan perusahaan.

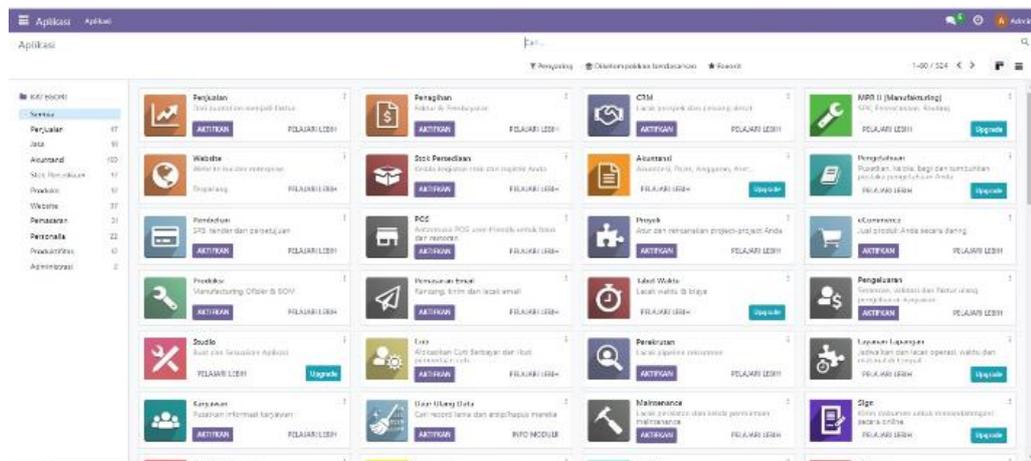
##### b. Konfigurasi Akun



Gambar 6. Konfigurasi Akun

Pada gambar 6 merupakan proses pengisian atau konfigurasi data perusahaan. Data yang diperlu berupa nama perusahaan, lokasi perusahaan, kontak person, dan id pajak perusahaan untuk memudahkan dalam proses bisnis.

### c. Halaman Apps Admin



Gambar 7. Halaman Apps Admin

Pada gambar 7 dapat melakukan pemilihan modul berdasarkan kebutuhan perusahaan. Dalam penerapan elearning odoo di sekolah MI YPPI 1945 Babat menggunakan 2 modul aplikasi.

### d. Aktivasi Modul





Gambar 8. Aktivasi Modul

Pemasangan modul aplikasi odoo implementasi elearning di MI YPPI 1945 babat menggunakan 2 modul. Pada gambar 8 merupakan modul website untuk menampilkan elearning dengan fitur menu sesuai kebutuhan. Modul elearning untuk mengisi informasi yang akan disampaikan kepada pengguna.

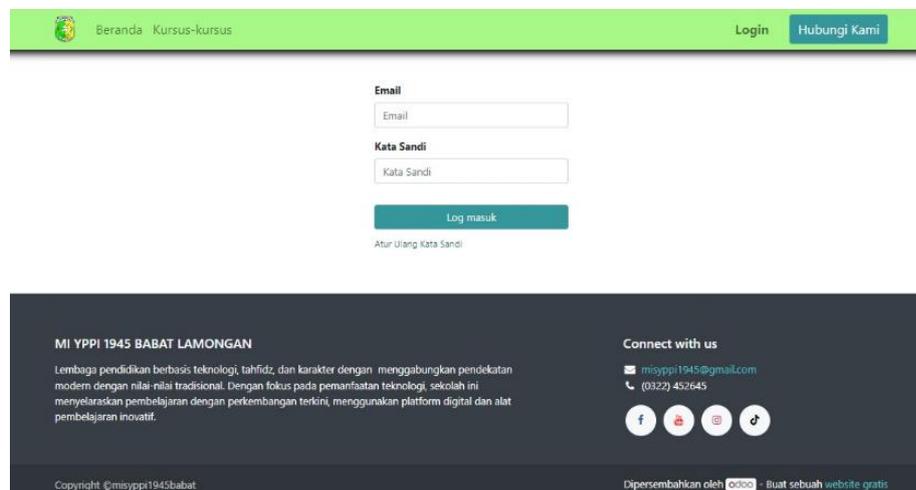
#### e. Tampilan Beranda



Gambar 9. Tampilan Beranda

Gambar 9 menampilkan halaman beranda dengan desain sesuai fitur menu yang sesuai dengan kebutuhan dan menampilkan informasi yang menarik bagi pengguna.

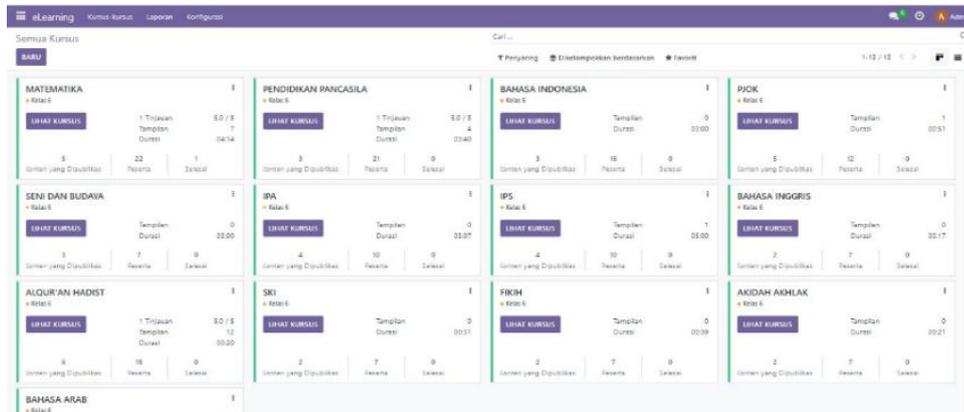
#### f. Halaman Login



Gambar 10. Halaman Login

Tampilan pada gambar 10 merupakan bagian penting untuk masuk kedalam informasi yang bersifat pribadi. Penggunaan email dan password sebagai keamanan informasi pada setiap akun.

### g. Dashboard Admin



Gambar 11. Halaman Dashboard Admin

Pada gambar 11 bagian dari dashboard admin saat berhasil login. Admin bias melihat aktivitas semua pembelajar yang berjalan dan dapat menambah konten sesuai hak akses yang dimiliki.

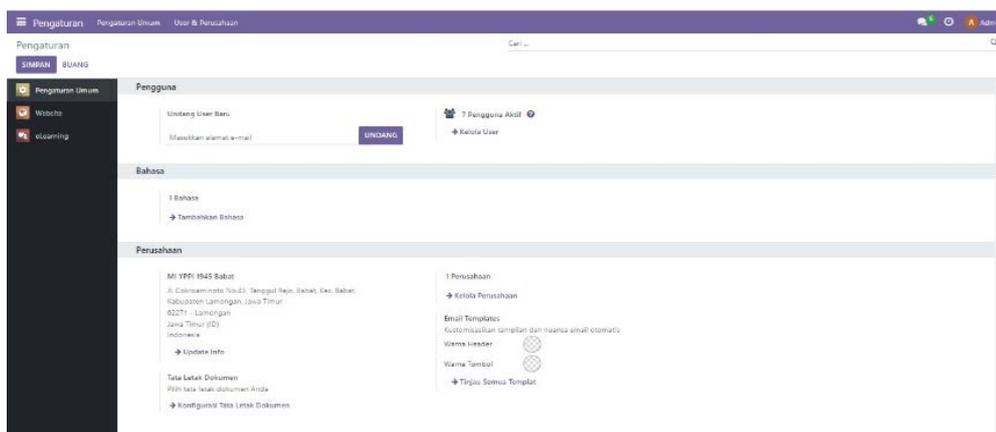
### h. Halaman Discus Admin



Gambar 12. Halaman Discus Admin

Halaman tersebut dapat melakukan komunikasi khusus setiap pengguna. Baik itu admin, guru dan siswa dengan hak ases yang telah di atur sebelumnya.

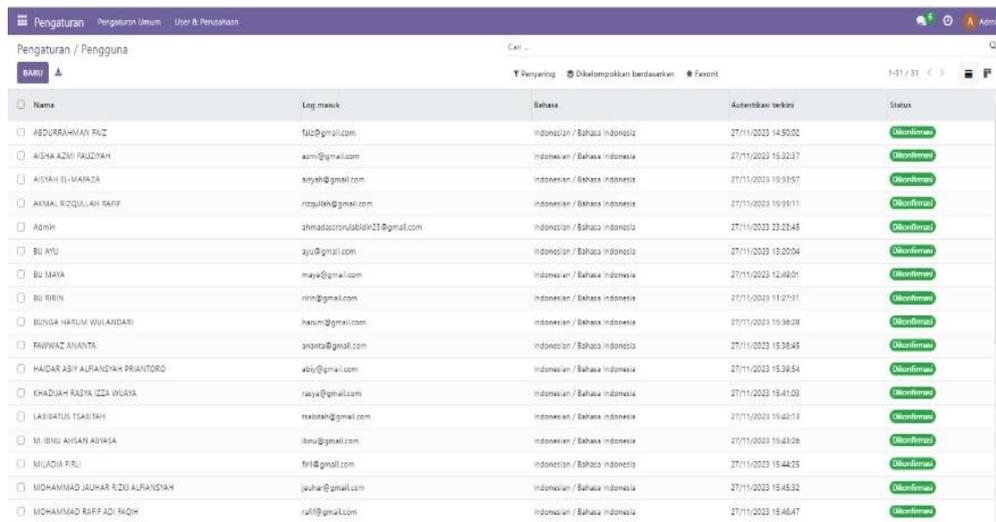
### i. Halaman Setting Admin



Gambar 13. Halaman Setting Admin

Gambar 13 merupakan pengaturan sistem yang disesuaikan dengan kebutuhan yang diinginkan. Terkait dengan hak pengguna, jumlah pengguna, dan sebagainya.

### j. Tampilan User

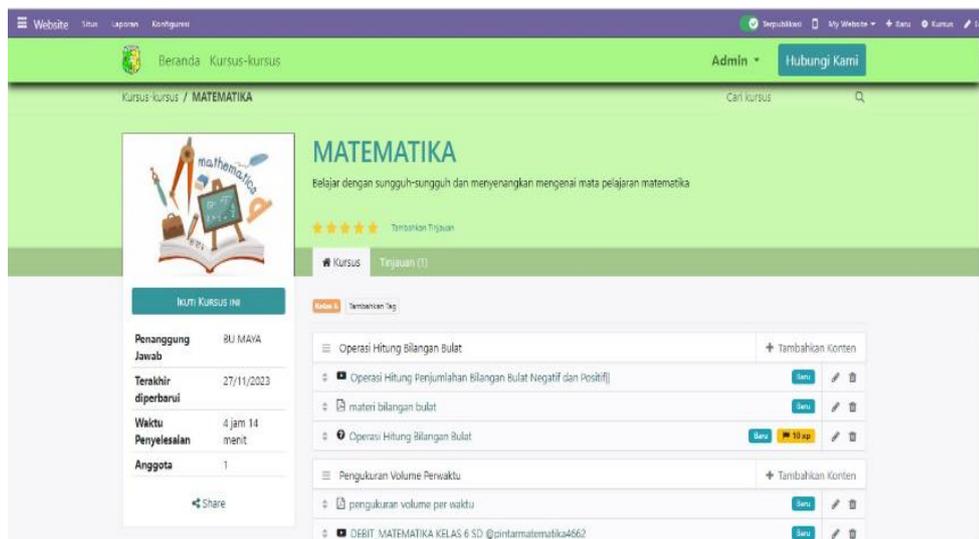


<input type="checkbox"/>	Nama	Log masuk	Bahasa	Autentikasi terakhir	Status
<input type="checkbox"/>	ABDURRAHMAN FAIZ	faiz@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 14:50:02	<span>Online</span>
<input type="checkbox"/>	AISNA AZMI RAUDYAH	am@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:22:37	<span>Online</span>
<input type="checkbox"/>	AISYAH EL-MAPAZA	aisyah@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 16:33:57	<span>Online</span>
<input type="checkbox"/>	AKMAL RIZQULAH RAFFI	rizqulah@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 16:36:11	<span>Online</span>
<input type="checkbox"/>	Admi	ahmadassorotubidin23@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 23:23:48	<span>Online</span>
<input type="checkbox"/>	BU AYU	ayu@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 19:20:04	<span>Online</span>
<input type="checkbox"/>	BU MAYA	maya@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 12:49:01	<span>Online</span>
<input type="checkbox"/>	BU RIRIN	ririn@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 11:27:51	<span>Online</span>
<input type="checkbox"/>	BUNGA HARUM WULANDARI	harum@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 19:38:28	<span>Online</span>
<input type="checkbox"/>	RAWAZ ANANTA	ananta@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:38:45	<span>Online</span>
<input type="checkbox"/>	HAIDAR ABY ALFANISAH PRIANTORO	aby@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:39:54	<span>Online</span>
<input type="checkbox"/>	KHADIAH RAZYA QIZA USANA	razya@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:41:03	<span>Online</span>
<input type="checkbox"/>	LABIBATUS TSABITAH	tsabtah@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:42:13	<span>Online</span>
<input type="checkbox"/>	M. IBNU AHSAN ABIRSA	ibnu@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:43:28	<span>Online</span>
<input type="checkbox"/>	MILADIA FIRLI	firli@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:44:25	<span>Online</span>
<input type="checkbox"/>	MOHAMMAD JAHRAR RIZQI ALFANISAH	jahrar@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:45:32	<span>Online</span>
<input type="checkbox"/>	MOHAMMAD RAFFADI RAQIH	rafi@gmail.com	Indonesia / Bahasa Indonesia	27/11/2023 15:46:47	<span>Online</span>

Gambar 14. Tampilan User

Tampilan seluruh pengguna (user) yang dapat dikelola oleh admin. Admin dapat melakukan penambahan user dan dapat menghapus data pengguna yang tidak dibutuhkan,

### k. Dashboard Guru



The screenshot shows a teacher's dashboard for a 'MATEMATIKA' course. The page includes a header with 'Beranda', 'Kursus-kursus', and 'Admin' options. The main content area features a course card for 'MATEMATIKA' with a description, a 5-star rating, and a 'Hubungi Kami' button. Below the course card, there is a section for 'KURSUS' with a list of lessons. The lessons listed are: 'Operasi Hitung Bilangan Bulat', 'Operasi Hitung Penjumlahan Bilangan Bulat Negatif dan Positif', 'materi bilangan bulat', and 'Operasi Hitung Bilangan Bulat'. Each lesson has a 'Setor' button and an edit/delete icon. The dashboard also shows course statistics: 'Penanggung Jawab: BU MAYA', 'Terakhir diperbarui: 27/11/2023', 'Waktu: 4 jam 14 menit', 'Penyelesaian: 1', and 'Anggota: 1'.

Gambar 15. Dasboar Guru

Dalam halaman ini, guru dapat membuat materi yang disampaikan dalam bentuk file, gambar, video, atau sumber lainnya sesuai dengan kebutuhan. Guru dapat melakukan kuis sebagai motivasi semangat belajar dengan mendapatkan poin nilai setiap pengerjaan soal atau aktivitas belajar.

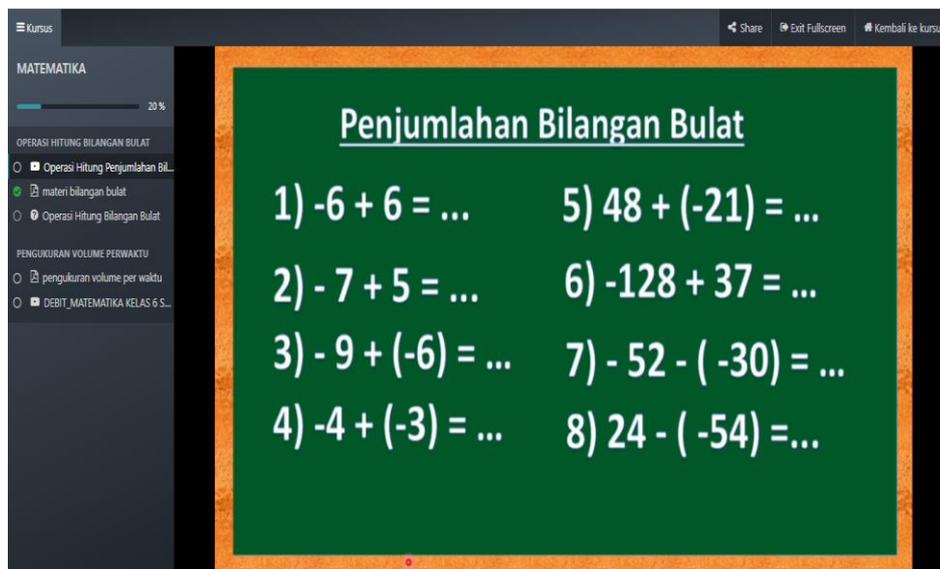
## I. Halaman View Kursus



Gambar 16. Halaman View Kursus

Halaman view kursur dapat memilih pelajaran yang sesuai dengan kebutuhan siswa. Halaman akan memunculkan poin nilai pengguna setelah melakan pembelajaran atau mengerjakan soal.

## m. Halaman Siswa



Gambar 17. Halaman Siswa

Halaman siswa menampilkan setiap informasi materi yang disampaikan oleh guru. Siswa juga dapat melakukan komentar atau penilaian terkait setiap materi yang sudah disampaikan.

## 4.2. Testing UAT

Proses pengumpulan data menggunakan google formulir dengan jumlah populasi 73 siswa kelas 6 dan 6 guru dengan memiliki perbedaan pernyataan kuesioner namun indikator yang sama. Dalam menentukan sampel kuesioner siswa menggunakan perhitungan sampel rumus slovin karena jumlah populasi yang banyak. Sedangkan perhitungan sampel guru menggunakan metode sampel jenuh atau klesuruhan karan populasi guru kurang dari 30 orang.

Rumus slovin pada populasi siswa yang berjumlah 73 sebagai berikut :

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

$$Sampel = \frac{73}{1 + 73 \cdot (0,05)^2}$$

$$Sampel = \frac{73}{1 + 73 \cdot (0,0025)}$$

$$Sampel = \frac{73}{1 + (0,2)}$$

$$Sampel = 62$$

Dari hasil perhitungan sampel menggunakan rumus slovin, maka hasil dari jumlah sampel yang nanti akan menjadi responden pada kuesioner berjumlah 62 siswa dari kelas 6 dengan tingkat signifikansi atau kegagalan 5%.

*Tabel 5. Hasil Presentase Indikator Kuesioner*

No	Indikator	Presentase Siswa	Presentase Guru	Jumlah Presentase
1	Kelengkapan (Completeness)	86%	94%	90%
2	Konsisten (Consistency)	82%	92%	87%
3	Pelacakan (Tracebility)	85%	88%	86%
4	Operabilitas (Operabilty)	84%	93%	88%
5	Pelatihan (Training)	87%	92%	89%
6	Keamanan (Security)	85%	90%	88%
7	Akurasi (Accuracy)	85%	92%	89%
8	Kesederhanaan (Simplicity)	85%	90%	87%
9	Kemudahan Eksekusi (Execution Efficiency)	81%	89%	85%

Pengukuran kualitas sistem informasi yang sudah diimplementasikan di sekolah MI YPPI 1945 Babat terhadap pengguna siswi dan guru dapat dilakukan dengan melihat hasil data presentase kuesioner dari responden.

*Tabel 6. Hasil Kualitas Sistem E-learning*

No	Kualitas	Presentase
1	Sangat Baik	81% - 100%
2	Baik	61% - 80%
3	Cukup	41% - 60%
4	Tidak Baik	21% - 40%
5	Sangat Tidak Baik	<21%

Berdasarkan tabel 5 diatas. dapat disimpulkan hasil dari kualitas sistem e-learning berbasis odoo yang ada di MI YPPI 1945 Babat Lamongan sebagai berikut :

- 1) Kelengkapan (Completeness) mendapatkan presentase kuesioner dari responden adalah 90%, yaitu dapat dimasukkan kedalam kategori kualitas "sangat baik"
- 2) Konsisten (Consistency) mendapatkan presentase kuesioner dari responden adalah 87%, yaitu dapat dimasukkan kedalam kategori kualitas "sangat baik"
- 3) Pelacakan (Tracebility) mendapatkan presentase kuesioner dari responden adalah 86%, yaitu dapat dimasukkan kedalam kategori kualitas "sangat baik"
- 4) Operabilitas (Operability) mendapatkan presentase kuesioner dari responden adalah 88%, yaitu dapat dimasukkan kedalam kategori kualitas "sangat baik"
- 5) Pelatihan (Training) mendapatkan presentase kuesioner dari responden adalah 89%, yaitu dapat dimasukkan kedalam kategori kualitas "sangat baik"

- 6) Keamanan (Security) mendapatkan presentase kuesioner dari responden adalah 88%, yaitu dapat dimasukkan kedalam kategori kualitas “sangat baik”
- 7) Akurasi (Accuracy) mendapatkan presentase kuesioner dari responden adalah 89%, yaitu dapat dimasukkan kedalam kategori kualitas “sangat baik”
- 8) Kesederhanaan (Simplicity) mendapatkan presentase kuesioner dari responden adalah 87%, yaitu dapat dimasukkan kedalam kategori kualitas “sangat baik”
- 9) Kemudahan Eksekusi (Execution Efficiency) mendapatkan presentase kuesioner dari responden adalah 85%, yaitu dapat dimasukkan kedalam kategori kualitas “sangat baik”.

## 5. Kesimpulan

Penggunaan odoo sebagai e-learning disekolah merupakan inovasi terbaru dalam penerapan teknologi pembelajaran yang ada. Sistem dapat melakukan proses penyampaian materi, pemberian tugas atau kuis, penilaian kinerja siswa dan data target pembelajaran sesuai yang diharapkan oleh sekolah. Sehingga, pembelajaran berjalan dengan efektif dan efisien. Sistem dapat membantu dalam proses pembelajaran yang ada disekolah dengan melakukan pengujian kualitas sistem e-learning dengan metode UAT. Pengujian sistem metode UAT dengan indikator Kelengkapan (Completeness), Konsisten (Consistency), Pelacakan (Traceability), Operabilitas (Operability), Pelatihan (Training), Keamanan (Security), Akurasi (Accuracy), Kesederhanaan (Simplicity), Kemudahan Eksekusi (Execution Efficiency) menghasilkan kualitas persentase 88% yang artinya sangat baik sistem e-learning yang berjalan di sekolah MI YPPI 1945 Babat.

## References

- [1] Y. Y. A. Anando and A. J. Gundo, “Pengaruh Antusiasme Belajar dan Media Belajar Website ‘Sekolah Digital SMKN 3 Salatiga’ Terhadap Prestasi Belajar Simulasi Digital,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 2, pp. 177–187, 2022, doi: 10.5281/zenodo.6133247.
- [2] N. H. Hidroes Abbrori, “Perancangan Learning Management System Menggunakan Odoo Dalam Pemanfaatan Media Pembelajaran Online di Pt. Gemilang Surya Kahyangan (GSK),” *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 6, no. 2, pp. 42–47, 2022, doi: 10.55886/infokom.v6i2.504.
- [3] A. R. Dewantara and L. G. Astuti, “Analisa dan Perancangan Sistem Informasi Survei Industri Besar dan Sedang (IBS) Bulanan Badan Pusat Statistik (BPS) Provinsi Bali,” *J. Elektron. Ilmu Komput. Udayana*, vol. 7, no. 2, pp. 94–106, 2018.
- [4] V. M. Anjasmara and A. Hadi Sumitro, “Pengembangan Sistem Informasi Masjid Darul Arham Menggunakan Metode V-Model dan UAT (User Acceptance Testing),” *Inf. Syst. Educ. Prof. J. Inf. Syst.*, vol. 8, no. 1, p. 47, 2023, doi: 10.51211/isbi.v8i1.2443.
- [5] N. F. Amin, S. Garancang, and K. Abunawas, “Konsep Umum Populasi dan Sampel dalam Penelitian,” *J. Pilar*, vol. 14, no. 1, pp. 15–31, 2023.
- [6] A. Z. Furgon, “Implementasi E-Learning Berbasis Odoo Untuk Mts Negeri 1 Nganjuk Menggunakan Metode Mccall,” *Skripsi Univ. Islam Negeri Maulana Malik Ibrahim Malang*, vol. 1, no. 2, pp. 1–90, 2022.

# Perbandingan Metode Ensemble Learning Random Forest Dan Adaboost Pada Pengenalan Chord Instrumen Piano Dan Gitar

I Dewa Agung Adwitya Prawangsa, AAIN Eka Karyawati

Informatika, Universitas Udayana  
Jl. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
prawangsa091@student.unud.ac.id  
eka.karyawati@unud.ac.id

## Abstract

*This paper presents a comprehensive analysis of Random Forest and Adaboost algorithms for classifying music chords, focusing on feature extraction and model optimization. The distribution of original data, including audio signal lengths and class distributions, is examined, revealing consistent characteristics across major and minor datasets. Mel Frequency Cepstral Coefficients (MFCC) are extracted with predefined parameters, ensuring feature extraction consistency. Subsequent feature normalization and oversampling using SMOTE address class imbalances. Evaluation metrics, including accuracy, precision, recall, and F1-score, are employed to assess model performance. The results demonstrate Random Forest's superiority in accuracy, precision, and recall over Adaboost. Furthermore, employing RandomizedSearchCV optimization enhances both models' performance, with Random Forest achieving an accuracy of 0.86 and Adaboost attaining 0.80. Confusion matrices illustrate Random Forest's higher prediction accuracy for both positive and negative classes compared to Adaboost. These findings underscore the effectiveness of Random Forest in accurately classifying music chords and highlight the significance of hyperparameter optimization in enhancing classification model performance.*

**Keywords:** Random Forest, Adaboost, Random Search, Confusion Matrix

## 1. Pendahuluan

Musik telah menjadi bagian tak terpisahkan dari kehidupan manusia selama berabad-abad, memberikan hiburan dan keindahan yang tak ternilai. Di era modern ini, perkembangan teknologi komputer dan kecerdasan buatan telah membuka peluang baru dalam pengenalan elemen-elemen musik secara otomatis, salah satunya adalah pengenalan Chord. Random forest merupakan metode pembelajaran terintegrasi penting berbasis bagging. Algoritma ensemble AdaBoost adalah algoritma iteratif berdasarkan multiple base learners dari jenis yang sama [1]. Penggunaan teknik ensemble learning seperti Random Forest dengan menggunakan Bootstrap Aggregating serta pemilihan acak untuk menghasilkan hasilnya [3] dan Adaboost yang berfokus pada pembangunan model yang lemah [4] telah menjadi sorotan dalam pengenalan pola suara, termasuk dalam konteks pengenalan Chord pada instrumen musik seperti piano dan gitar. Meskipun demikian, penelitian yang secara spesifik membandingkan kinerja kedua metode tersebut dalam pengenalan Chord masih terbatas. Sebagai elemen penting dalam teori musik, pengenalan Chord secara otomatis memperlihatkan tantangan kompleksitas tersendiri. Oleh karena itu, penelitian yang lebih mendalam perlu dilakukan untuk memahami dan membandingkan keunggulan serta keterbatasan masing-masing metode dalam konteks pengenalan Chord pada instrumen musik.

Studi sebelumnya, yang memiliki relevansi terhadap paper dibahas di dalam jurnal ilmiah yang berjudul A Mandarin Tone Recognition Algorithm Based on Random Forest and Feature Fusion [2] telah menunjukkan keberhasilan Ensemble Learning, terutama dengan metode Random Forest, dalam pengenalan pola suara pada konteks pengenalan nada dalam bahasa Mandarin. Meskipun fokusnya bukan pada pengenalan Chord dalam musik, namun penggunaan Random Forest dalam mengenali pola suara menunjukkan potensi yang relevan dengan konteks pengenalan Chord pada instrumen

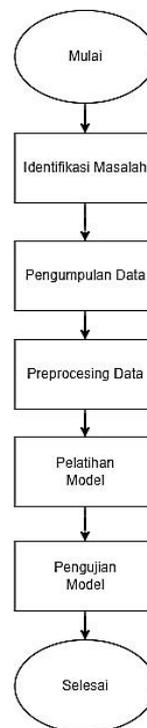
musik. Studi ini berhasil mencapai akurasi pengenalan nada yang tinggi di atas 83.57%, bahkan dengan jumlah sampel latihan yang kecil. Hasil ini memberikan pemahaman bahwa Random Forest memiliki kemampuan pembelajaran yang kuat dalam mengenali pola suara, yang dapat diterapkan dalam pengenalan Chord pada instrumen musik.

Dengan adanya permasalahan tersebut, serta urgensi untuk mengembangkan teknik pengenalan Chord secara otomatis yang lebih efektif, penelitian ini dilakukan. Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang kinerja masing-masing metode dalam konteks pengenalan Chord pada instrumen piano dan gitar. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi baru dalam pengembangan teknik pengenalan Chord secara otomatis, serta memberikan wawasan baru tentang aplikasi Ensemble Learning dalam bidang musik. Dengan demikian, penelitian ini memiliki urgensi yang kuat untuk dilakukan, karena akan memberikan pemahaman yang lebih dalam tentang teknik pengenalan Chord secara otomatis dan potensi penggunaan metode Ensemble Learning dalam konteks musik, yang dapat membawa dampak positif dalam pengembangan industri musik dan teknologi informasi.

## 2. Metodolgi Penelitian

### Flowchart Umum

Pada metodologi penelitian akan membahas langkah- langkah yang akan dilakukan dalam menjalankan penelitian ini. Langkah-langkah tersebut meliputi identifikasi permasalahan, pengumpulan data, Preprocessing Data, Perancangan model, serta Analisis Hasil. Visual dari flowchart berada pada gambar 1.



**Gambar 1.** Flowchart Alur Kerja

### 2.1 Persiapan Data Set

Pada tahap ini akan dilakukan pengumpulan data. Data yang digunakan dalam penelitian ini adalah Dataset open source yang berasal dari websie kaggle yang berisikan dua buah kategori folder yang bernamakan label Major dan Minor dan pada setiap folder memiliki banyak Sample Audio yang bertipekan file wav yang berisikan chord yang berasal dari instrument alat musik baik gitar maupun piano. Berdasarkan Tabel 1. Pada label Major memiliki jumlah sampel dengan total yaitu 502, Sementara pada label Minor memiliki jumlah sampel yakni 357 Data audio berformat wav, Sehingga

total dataset mencakupi semua label berjumlah 859. Dengan adanya perbedaan dalam perbandingan jumlah sampel dari antar label akan ditangani pada tahap selanjutnya.

**Tabel 1.** Informasi Data

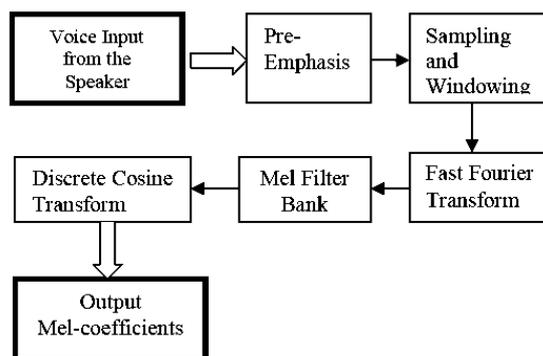
Parameter	Deskripsi	Tipe Data
Audio File	502 (Major) 357 (Minor) Total 859 Files	File (.wav)
Label	Label Chord (Mayor/Minor)	Kategori Folder
Distribusi Panjang Sinyal Audio (Major)	Rata-rata: 49360 Minimum: 48510 Maksimum: 50715	Hz
Distribusi Panjang Sinyal Audio (Minor)	Rata-rata: 49381 Minimum: 47408 Maksimum: 50715	

## 2.2 Preprocessing data

Pada tahapan ini, dilakukan proses preprocessing data untuk mempersiapkan dataset audio sebelum digunakan dalam model machine learning. Proses preprocessing ini meliputi beberapa langkah seperti di bawah ini :

### 2.2.1 Ekstraksi Fitur MFCC

Pertama, dilakukan ekstraksi fitur menggunakan metode Mel-frequency cepstral coefficients (MFCC) dari setiap file audio dalam dataset. MFCC adalah representasi fitur yang penting dalam pemrosesan suara, dan ekstraksi ini membantu menggambarkan karakteristik frekuensi dari sinyal audio [5].



**Gambar 2.** Diagram Alur MFCC [6].

Alur kerja berdasarkan pada gambar 2 dijelaskan flowchart dengan proses pengolahan sinyal suara. Pertama, sinyal suara masuk melalui input suara. Kemudian, dilakukan pre-emphasis untuk meningkatkan sinyal frekuensi tinggi. Selanjutnya, sinyal disampling dan di-windowing untuk membagi sinyal menjadi beberapa segmen yang lebih kecil. Proses fast Fourier transform (FFT) digunakan untuk mengonversi sinyal waktu ke domain frekuensi. Selanjutnya, mel filter bank digunakan untuk memperoleh spektrum energi sinyal dalam rentang frekuensi tertentu. Discrete cosine transform (DCT) digunakan untuk menghasilkan koefisien-koefisien mel yang merepresentasikan karakteristik sinyal suara dalam bentuk domain frekuensi. Output dari proses ini adalah mel-coefficients yang dapat digunakan untuk analisis dan pengolahan lanjutan dalam aplikasi pengenalan suara atau pengolahan sinyal suara lainnya.

### 2.2.2 Normalisasi (Z-score)

Setelah ekstraksi fitur MFCC, dilakukan normalisasi menggunakan Z-score normalization. Proses normalisasi ini dilakukan untuk memastikan agar distribusi nilai fitur seragam dan memiliki rata-rata nol serta standar deviasi satu.

### 2.2.3 Oversampling (SMOTE)

Selanjutnya, dilakukan proses oversampling pada label minoritas (kelas Minor) dalam dataset menggunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Oversampling ini bertujuan untuk menyeimbangkan distribusi kelas dalam dataset, sehingga mencegah bias dan meningkatkan kinerja model terhadap kelas minoritas. Setelah melakukan tahap smote, Data akan di bagi menjadi 80:20 untuk data latih dan data uji.

### 2.3 Random Search

Tuning hiperparameter untuk model hybrid yang ditujukan untuk klasifikasi Model dengan metode Random Search bertujuan sebagai untuk mencari kombinasi hiperparameter yang optimal yang dapat meningkatkan kinerja model dalam tugas klasifikasi.

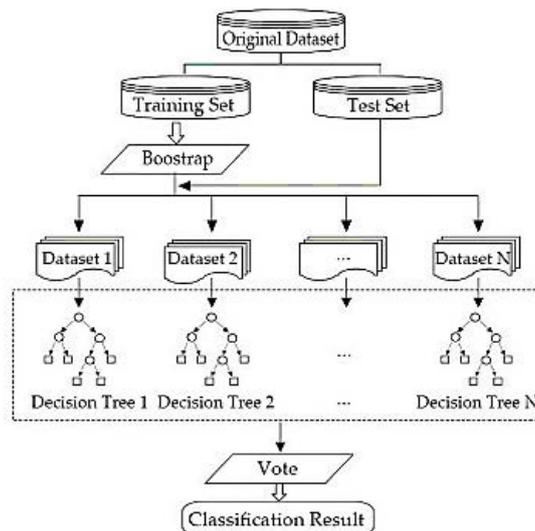
**Tabel 2.** Kombinasi nilai hyper-parameter yang digunakan

Model	Selected Hyperparameters	Range of Search
Random Forest	n_estimators, max_depth, min_samples_split, min_samples_leaf	[100, 200], [30, 50], [4, 7], [3, 5]
Adaboost	n_estimators, learning_rate	[50, 100, 200] , [0.01, 0.05, 0.1, 0.5, 1]

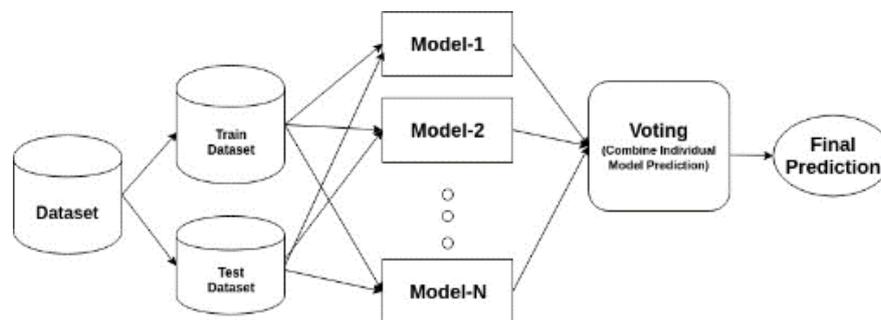
Berdasarkan tabel 2. Penentuan parameter untuk pencarian acak (Random Search) merupakan proses penting dalam pengembangan model machine learning. Dalam Tabel 3.3, parameter yang akan dioptimalkan untuk model Random Forest adalah n\_estimators (jumlah pohon dalam ensemble), max\_depth (kedalaman maksimum setiap pohon), min\_samples\_split (jumlah sampel minimum yang diperlukan untuk membagi node), dan min\_samples\_leaf (jumlah sampel minimum di setiap daun pohon). Rentang pencarian untuk masing-masing parameter didefinisikan sebagai [100, 200] untuk n\_estimators, [30, 50] untuk max\_depth, [4, 7] untuk min\_samples\_split, dan [3, 5] untuk min\_samples\_leaf. Sedangkan untuk model Adaboost, parameter yang akan dioptimalkan adalah n\_estimators (jumlah estimators dalam ensemble) dan learning\_rate (tingkat pembelajaran). Rentang pencarian untuk n\_estimators adalah [50, 100, 200], sedangkan untuk learning\_rate adalah [0.01, 0.05, 0.1, 0.5, 1]. Pencarian dilakukan dalam rentang ini untuk menemukan kombinasi parameter yang memberikan kinerja terbaik bagi model yang dibangun. Proses ini membantu dalam menemukan parameter optimal yang menghasilkan model yang memiliki kinerja prediksi yang lebih baik.

### 2.4 Perancangan Model

Random Forest adalah metode ensemble learning yang menggabungkan output dari beberapa pohon keputusan untuk membuat prediksi akhir. Lalu Adaboost (Adaptive Boosting) adalah algoritma ensemble learning yang digunakan untuk meningkatkan kinerja model dengan menggabungkan beberapa weak learners menjadi satu model yang kuat. Dengan pembangunan kedua model berdasarkan gambar 3 dan gambar 4.



**Gambar 3.** Arsitektur Model Random Forest



**Gambar 4.** Arsitektur Model Adaboost

Berdasarkan gambar 3, Arsitektur Random Forest melibatkan gabungan dari beberapa pohon keputusan yang dibangun dengan menggunakan dataset pelatihan yang dibagi menjadi dataset bootstrap. Setiap pohon keputusan dilatih dengan satu dataset bootstrap, dan hasil klasifikasi dari setiap pohon diambil melalui proses voting. Hasil akhir dari klasifikasi ditentukan oleh mayoritas suara dari semua pohon keputusan, yang menjadi output dari Random Forest. Di sisi lain, Pada gambar 4, pembangunan model Adaboost melibatkan inialisasi bobot sampel, iterasi untuk melatih pembelajaran lemah, dan menggabungkan hasil prediksi dari semua pembelajaran lemah menggunakan bobot model mereka untuk mendapatkan prediksi akhir.

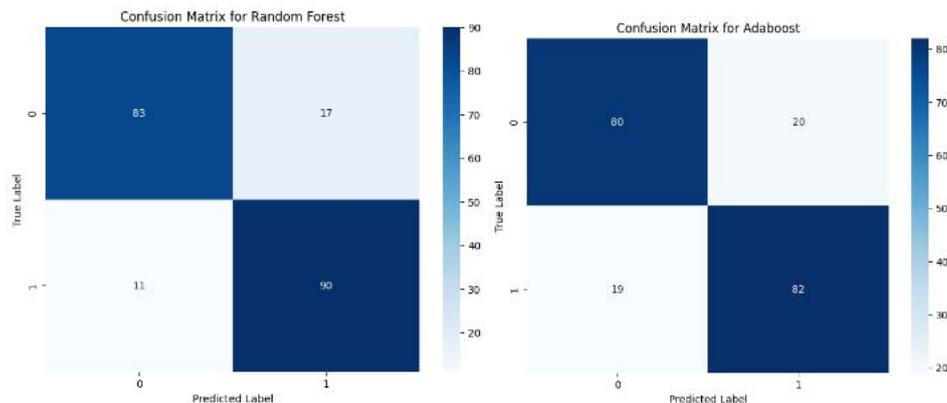
### 3. Hasil dan Pembahasan

Pada penelitian kali dilakukan perbandingan kinerja antar model ensemble learning Random Forest dengan Adaboost . Hasil eksperimen menunjukkan bahwa Random Forest berhasil mencapai tingkat akurasi sebesar 0.86, sementara Adaboost memperoleh nilai akurasi 0.80. Seperti terlihat pada Tabel 3.

**Tabel 3.** Hasil evaluasi model dengan menggunakan Random Search

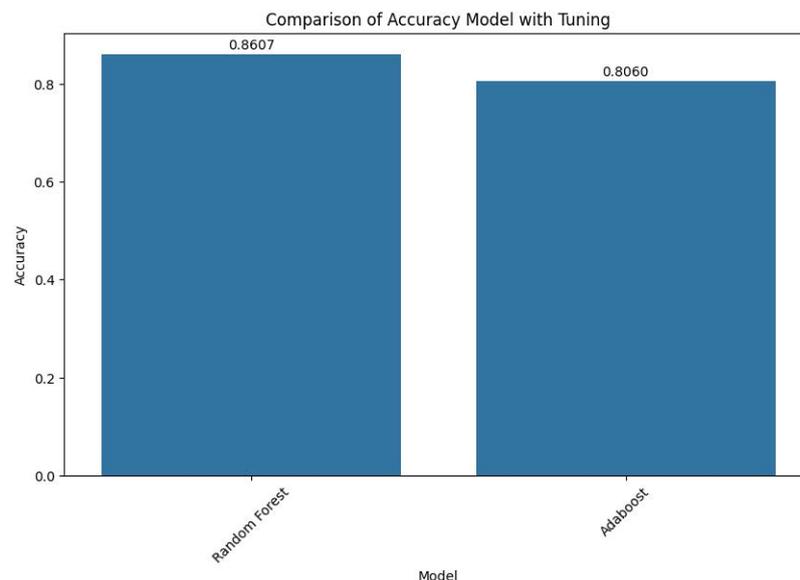
MetriK	Random Forest	Adaboost
Akurasi	0.86	0.80
Presisi	[0.833, 0.848]	[0.808, 0.803]
Recall	[0.85, 0.831]	[0.8, 0.811]
F1-score	[0.841, 0.84]	[0.804, 0.807]
Parameter	n_estimators: 100, min_samples_split: 7, min_samples_leaf: 3, max_depth: 30	n_estimators: 200, learning_rate: 1

Berdasarkan hasil dari tabel 3 menunjukkan bahwa penggunaan teknik RandomizedSearchCV telah berhasil meningkatkan kinerja keduanya model Random Forest dan Adaboost dalam tugas klasifikasi dengan teknik tuning hyperparameter.



**Gambar 5.** Confusion Matrix Random Forest & Adaboost

Berdasarkan gambar 5. menunjukkan bahwa Confusion Matrix untuk model Random Forest mendapatkan 82% prediksi yang benar untuk kelas positif dan 90% prediksi yang benar untuk kelas negatif, sementara model Adaboost menunjukkan tingkat prediksi yang lebih rendah dengan hanya 80% prediksi yang benar untuk kelas positif dan 82% prediksi yang benar untuk kelas negatif. Dalam konteks penelitian ini, Penulis menginterpretasikan hasil ini sebagai indikasi bahwa model Random Forest lebih mampu dalam mengklasifikasikan data dengan tepat dibandingkan dengan Adaboost. Tingkat prediksi yang lebih tinggi untuk kedua kelas dalam Confusion Matrix Random Forest menunjukkan bahwa model ini memiliki kemampuan yang lebih baik dalam mengidentifikasi kelas-kelas yang berbeda dalam data, memperkuat kesimpulan bahwa Random Forest mungkin merupakan pilihan yang lebih unggul dalam pengklasifikasian data untuk kasus yang diteliti.



**Gambar 6.** Perbandingan Akurasi Random Forest dengan Adaboost

Dalam konteks penelitian ini, Berdasarkan Gambar 6. Penulis menginterpretasikan hasil ini sebagai indikasi bahwa model Random Forest lebih mampu dalam mengklasifikasikan data dengan lebih baik dibandingkan dengan Adaboost.

#### 4. Kesimpulan

Berdasarkan evaluasi kinerja, Model Random Forest menunjukkan akurasi 81%, lebih tinggi daripada Model Adaboost yang mencapai 74%, menunjukkan perbedaan 7%. Keunggulan Random Forest dalam pengenalan Chord instrumen musik dijelaskan oleh kemampuannya mengenali pola kompleks

dalam data musik, berkat penggabungan banyak pohon keputusan dan adaptasi fitur-fitur informatif. Dalam pengenalan Chord, akurasi klasifikasi sangat penting, sehingga keunggulan Random Forest menjadi faktor penentu pilihan model yang optimal. Penerapan pendekatan Random Search untuk menemukan parameter optimal telah terbukti signifikan dalam meningkatkan kinerja model secara keseluruhan. Dengan menggunakan parameter yang dioptimalkan, akurasi Model Random Forest meningkat menjadi 86%, mewakili kenaikan sebesar 5% dari sebelumnya. Begitu juga, penggunaan teknik pencarian parameter optimal meningkatkan akurasi Model Adaboost menjadi 80%, mencerminkan peningkatan sebesar 6% dari sebelumnya. Ini menegaskan bahwa penyesuaian parameter yang tepat secara substansial meningkatkan kinerja model klasifikasi, menunjukkan dampak positif yang signifikan terhadap akurasi dan kemampuan prediktifnya.

#### Daftar Pustaka

- [1] Runchuan Li, Wenzhi Zhang, Shengya Shen, Jinliang Yao, Bicao Li, Bing Zhou, Gang Chen, Zongmin Wang. An Intelligent Heartbeat Classification System Based on Attributable Features with AdaBoost & Random Forest Algorithm. *Journal of Healthcare Engineering*. 2021.
- [2] Yan J, Meng Q, Tian L, Wang X, Liu J, Li M, Zeng M, Xu H. A Mandarin Tone Recognition Algorithm Based on Random Forest and Feature Fusion †. *Mathematics*. 2023.
- [3] Ochim Gold, Agaji Iorshase. Heart failure prediction framework using random forest and J48 with Adaboost algorithms . *Science World Journal*. 2023.
- [4] F. Z. SOLAK, "COVID-19 DETECTION USING VARIATIONAL MODE DECOMPOSITION OF COUGH SOUNDS", *KONJES*, vol. 11, no. 2, pp. 354–369, 2023.
- [5] Shen Y, Cheng W. A Tree-Based Machine Learning Method for Pipeline Leakage Detection. *Water*. 2022.
- [6] Ali, Shalbbya, et al. "Mel frequency cepstral coefficient: a review." *ICIDSSD* 2020.

*This page is intentionally left blank.*

# Analisis Keranjang Belanja Untuk Meningkatkan Penjualan Fashion Di E-Commerce Australia

Azzahra Putri Maharani<sup>a1</sup>, Apriade Voutama<sup>a2</sup>

<sup>a</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, Jln. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

<sup>1</sup>[azzahramaharani183@gmail.com](mailto:azzahramaharani183@gmail.com)

<sup>2</sup>[apriade.voutama@staff.unsika.ac.id](mailto:apriade.voutama@staff.unsika.ac.id)

## Abstract

*Analisis keranjang belanja dalam e-commerce fashion di Australia menjadi krusial dalam meningkatkan penjualan produk. Penelitian ini bertujuan untuk mengeksplorasi data keranjang belanja untuk mendapatkan wawasan yang berharga dalam meningkatkan penjualan. Melalui metode analisis seperti Root Cause Analysis (RCA), Exploratory Data Analysis (EDA), dan visual data mining, dataset dari platform e-commerce Australia periode Januari hingga Oktober 2021 dieksplorasi. Hasil analisis menunjukkan fluktuasi pendapatan bulanan dari \$85,597 hingga \$131,364 selama periode tersebut. Faktor-faktor seperti karakteristik produk dan perubahan dalam basis pelanggan diidentifikasi sebagai penyebab utama fluktuasi penjualan. Insight penting ditemukan, termasuk preferensi produk berdasarkan segmentasi usia dan tren pelanggan baru dan setia. Rekomendasi disarankan untuk meningkatkan pendapatan bulanan, seperti mengoptimalkan penjualan di wilayah tertentu, segmentasi pelanggan dan produk, promosi produk dengan penjualan rendah, penyempurnaan produk, dan analisis kinerja serta evaluasi berkala. Penelitian ini memberikan pandangan berharga bagi perusahaan e-commerce fashion di Australia untuk meningkatkan strategi penjualan dan mencapai pertumbuhan bisnis yang berkelanjutan.*

**Keywords:** E-Commerce Fashion, Analisis Data, Keranjang Belanja, Rekomendasi, Insight

## 1. Pendahuluan

Pada zaman digital ini, sektor *e-commerce fashion* telah mencapai tingkat dinamisme yang luar biasa di Australia. Perusahaan-perusahaan fashion terus mengembangkan inovasi baru untuk memperkuat persaingan dalam industri yang sangat kompetitif [1]. Kehadiran industri *fashion* memiliki signifikansi yang besar karena meraih pangsa pasar yang luas di berbagai wilayah dunia [2]. *E-commerce* adalah bentuk perdagangan daring yang dilakukan melalui internet [3]. Perusahaan *fashion online* dihadapkan pada tantangan untuk mempertahankan kompetitivitas dan meningkatkan penjualan di tengah persaingan yang sengit. Oleh karena itu, analisis data menjadi krusial sebagai alat strategis untuk mendukung pengambilan keputusan yang efektif. Analisis data adalah proses untuk mengumpulkan dan mengatur data dari pengamatan, wawancara, dan sumber informasi lainnya untuk mendalami pengetahuan peneliti tentang topik yang sedang diselidiki [4].

Fokus utama proyek ini adalah mengatasi masalah bisnis terkait penjualan produk dengan target penjualan rata-rata \$100,000 dalam 6 bulan ke depan. Analisis data keranjang belanja menjadi krusial untuk mencapai target ini. Dengan memahami pola pembelian, preferensi produk, dan variabel lain yang mempengaruhi keputusan pembelian konsumen, perusahaan bisa merencanakan strategi penjualan yang lebih optimal. Dengan memahami secara komprehensif tingkah laku konsumen serta tren pasar, perusahaan dapat menyesuaikan produk dan layanan mereka untuk meningkatkan retensi pelanggan dan menarik pelanggan baru.

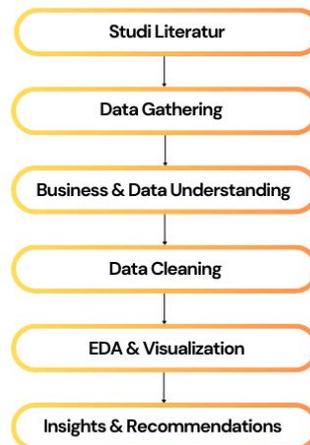
Penelitian ini bertujuan untuk melakukan analisis mendalam terhadap dataset keranjang belanja dari *platform e-commerce* di Australia, khususnya produk-produk *fashion*. Metode analisis yang digunakan dalam penelitian ini mencakup Metode RCA (*Root Cause Analysis*), EDA (*Exploratory Data Analysis*), dan *visual data mining*. Melalui analisis pola pembelian yang komprehensif, penelitian ini bertujuan untuk mengidentifikasi tren konsumen, preferensi produk, dan berbagai faktor lain yang berdampak pada pembelian yang dilakukan oleh konsumen di ranah *fashion*

*e-commerce*. Dataset ini diperoleh dari pengguna situs Kaggle bernama Ruchi Bhatia dengan judul "*Shopping Cart Database*", berisi data penjualan produk *fashion e-commerce* dari Australia periode Januari hingga Oktober 2021. Keberagaman informasi dalam dataset ini menjadi faktor penentu pemilihan data untuk analisis ini.

Diharapkan penelitian ini akan memberikan pandangan yang berharga bagi perusahaan dan pelaku *e-commerce fashion* lainnya di Australia, membantu dalam meningkatkan strategi penjualan, mengoptimalkan pengalaman pelanggan, serta mencapai pertumbuhan bisnis yang berkelanjutan.

## 2. Metode Penelitian

Alur penelitian yang diterapkan oleh peneliti melibatkan serangkaian langkah-langkah sistematis yang dirancang untuk mencapai tujuan penelitian dengan efektif. Tahapan penelitian merupakan langkah-langkah atau prosedur yang dilakukan oleh penulis untuk mencapai hasil yang diharapkan [5]. Untuk informasi lebih lanjut mengenai tahapan analisis, dapat dilihat pada Gambar 1.



**Gambar 1.** Skema Alur Penelitian

### 2.1. Studi Literatur

Analisis data merupakan proses untuk mengumpulkan dan mengatur data dari pengamatan, wawancara, dan sumber informasi lainnya untuk mendalami pengetahuan peneliti tentang topik yang sedang diselidiki [4]. Beberapa Teknik yang umum digunakan dalam analisis data termasuk *Root Cause Analysis* (RCA), *Exploratory Data Analysis* (EDA), dan *Visual Data Mining*.

#### 2.1.1. Root Cause Analysis (RCA)

*Root Cause Analysis* (RCA) merupakan metode yang diterapkan untuk menemukan akar masalah dalam suatu sistem atau perangkat mekanis. Pendekatan ini melibatkan penelusuran terhadap penyebab kegagalan hingga ke tingkat yang paling fundamental, lalu mengelompokkan setiap penyebab untuk memfasilitasi identifikasi penyebab utama dari kegagalan tersebut. [6].

#### 2.1.2. Exploratory Data Analysis (EDA)

*Exploratory Data Analysis* (EDA) adalah suatu teknik yang dipergunakan untuk mengungkap susunan, kerangka, dan hubungan dalam data dengan memvisualisasikannya dalam bentuk grafik, sehingga memudahkan peneliti dalam memahami data tersebut. EDA adalah bagian dari bidang statistik dan analisis data yang melibatkan eksplorasi data menggunakan teknik statistik deskriptif dan visualisasi tanpa melakukan asumsi sebelumnya. Maksudnya adalah untuk mencari pola dalam data, menganalisis ciri-ciri data, dan membangun model dari data tanpa memiliki prasangka sebelumnya [7].

#### 2.1.3. Visual Data Mining (VDM)

Metode *Visual Data Mining* (VDM) adalah sebuah pendekatan yang diterapkan untuk menginterpretasikan visualisasi data dengan tujuan mengurangi kesalahan dalam komunikasi pemahaman data. Langkah-langkah yang ada dalam pendekatan ini meliputi *Project Planning Phase*, *Data Preparation Phase*, dan *Data Analysis Phase* [8].

## 2.2. Data Gathering

Pada tahap *data gathering* dalam alur penelitian, kegiatan yang dilakukan adalah mengumpulkan dataset dari sumber yang telah ditentukan, yaitu kaggle.com. Dataset adalah

sebuah himpunan data, yang umumnya terdiri dari satu tabel dalam basis data, atau dapat juga berupa matriks data di mana setiap kolom menggambarkan suatu variabel tertentu, dan setiap baris mewakili sejumlah data [9].

Data yang didapat terdiri dari 4 tabel, yaitu table customers, products, sales, dan orders. Berikut adalah deskripsi dari masing-masing tabel:

**Tabel 1. Customers**

Atribut	Jumlah Data	Tipe Data	Deskripsi
customer_id	1000	Int64	Nomor unik pelanggan
customer_name	1000	Object	Nama pelanggan
gender	1000	Object	Jenis kelamin pelanggan
age	1000	Int64	Usia pelanggan
home_address	1000	Object	Alamat tempat tinggal pelanggan
zip_code	1000	Int64	Kode pos tempat tinggal pelanggan
city	1000	Object	Kota tempat tinggal pelanggan
state	1000	Object	Nama negara bagian pelanggan
country	1000	Object	Nama negara tempat tinggal pelanggan

**Tabel 2. Orders**

Atribut	Jumlah Data	Tipe Data	Deskripsi
order_id	1000	Int64	Nomor unik untuk Setiap pelanggan
customer_id	1000	Int64	Nomor unik untuk Setiap pelanggan yang melakukan pesanan
payment	1000	Int64	Pembayaran yang dilakukan pelanggan
order_date	1000	Datetime64	Tanggal pesanan
delivery_date	1000	Object	Tanggal pesanan diantarakan
order_month	1000	Int32	Bulan ketika pesanan ditempatkan

**Tabel 3. Products**

Atribut	Jumlah Data	Tipe Data	Deskripsi
product_ID	1260	Int64	Nomor unik Setiap produk
product_type	1260	Object	Kategori produk
product_name	1260	Object	Nama produk
size	1260	Object	Ukuran produk
colour	1260	Object	Warna produk
price	1260	Int64	Harga produk per unit
quantity	1260	Int64	Jumlah produk
description	1260	Object	Deskripsi singkat tentang produk

**Tabel 4. Sales**

Atribut	Jumlah Data	Tipe Data	Deskripsi
sales_id	5000	Int64	Nomor unik Setiap transaksi penjualan
order_id	5000	Int64	Nomor ID yang menghubungkan transaksi penjualan dengan pesanan tertentu
product_id	5000	Int64	Nomor ID yang mengidentifikasi produk yang dijual dalam transaksi
price_per_unit	5000	Int64	Harga per unit untuk produk yang dijual
quantity	5000	Int64	Jumlah unit yang terjual
total_price	5000	Int64	Total harga untuk semua produk dalam satu transaksi penjualan

### 2.3. Business & Data Understanding

*Business & Data Understanding* adalah langkah awal penting dalam proyek analisis data. Tahapan ini memerlukan pemahaman tentang konteks bisnis, struktur data, dan permasalahan yang ingin dipecahkan [10]. *Business understanding* melibatkan identifikasi tujuan bisnis dan merumuskan tujuan penelitian menjadi permasalahan yang harus diselesaikan. Sementara itu, *data understanding* melibatkan pengumpulan dan analisis awal data untuk memahami karakteristiknya [11].

Penerapan metode *Root Cause Analysis (RCA)* membantu memahami akar permasalahan yang ingin diselesaikan. RCA adalah proses penyelesaian masalah yang menyelidiki peristiwa atau ketidaksesuaian [12]. Dengan menerapkan RCA dalam *Business & Data Understanding*, peneliti dapat memahami permasalahan yang ingin dipecahkan melalui analisis data dengan lebih mendalam, memastikan bahwa upaya analisis akan relevan dan efektif.

Pada tahap ini, penelitian menggunakan Google Spreadsheet dan BigQuery SQL sebagai alat utama untuk melakukan analisis awal terhadap data.

### 2.4. Data Cleaning

*Data cleaning* merupakan tahapan penting dalam penelitian. Pada langkah ini, data dibersihkan untuk memastikan keakuratan dan konsistensi datanya, serta meningkatkan kualitasnya. Proses *data cleaning* dilakukan dengan menggunakan *Google Spreadsheet* untuk memudahkan identifikasi dan perbaikan data yang tidak sesuai atau tidak lengkap. Proses ini bertujuan untuk mengidentifikasi dan menghapus kesalahan serta inkonsistensi yang terdapat dalam data [13]. Pada proses ini yang dilakukan yaitu mencakup identifikasi dan penanganan data duplikat, nilai yang hilang, pengubahan format data, pemeriksaan *whitespace*, dan deteksi *outlier*. Selain itu, dibuat tabel tambahan '*scope\_data*' yang berisi data yang akan dianalisis.

### 2.5. EDA & Visualization

Dalam fase ini, data yang telah melalui proses pembersihan digunakan untuk melakukan EDA. Ditahap ini, penelitian menggunakan *tools BigQuery* dan *Looker Studio* untuk menganalisis lebih lanjut pola dan keterkaitan dalam dataset yang ada. *Exploratory Data Analysis (EDA)* merupakan suatu teknik yang berguna untuk menemukan pola, susunan, dan keterkaitan dalam data tersebut dengan menerapkannya dalam bentuk grafs. Tujuannya adalah untuk mempersembahkan data dengan cara yang lebih sederhana agar dipahami oleh peneliti [7]. EDA merujuk kepada tahap yang penting dalam proses penyelidikan awal terhadap data, di mana ringkasan statistik dan visualisasi grafis digunakan untuk menemukan pola, anomali, serta untuk menguji hipotesis dan asumsi [14]. Visualisasi data menggambarkan informasi dan data yang membantu dalam eksplorasi visual data, yang seringkali terhubung dengan EDA [15]. Proses

*data mining* melibatkan tiga tahap, yakni tahap perencanaan proyek, tahap persiapan data, dan tahap analisis data. *Data mining* adalah proses analisis yang menghasilkan informasi berharga dan menarik dengan memanfaatkan pola-pola yang terdapat dalam Kumpulan data yang besar [16]. Terkadang, istilah penambangan data juga merujuk pada proses menemukan pengetahuan dan informasi yang tersembunyi dalam data yang tidak diketahui sebelumnya secara eksplisit, dengan potensi untuk memberikan manfaat [17].

Pada tahapan perencanaan proyek, langkah pertama adalah mengidentifikasi tujuan dan menetapkan batasan masalah bisnis untuk memahami serta memenuhi kebutuhan subjek dengan baik. Pada tahap persiapan data melibatkan seleksi data yang relevan serta menyiapkannya untuk keperluan analisis lebih lanjut, serta melakukan verifikasi untuk mendeteksi kesalahan data. Terakhir, dalam tahap analisis data, melibatkan penggunaan pertanyaan penelitian yang telah ditetapkan sebelumnya untuk melakukan analisis, dan juga melibatkan analisis visual atau pembuatan model penambangan data untuk memvalidasi hasil dan menunjukkan informasi yang ditemukan melalui visualisasi.

### 2.6. Insight & Recommendation

*Insight & Recommendation* merupakan tahap penting dalam alur penelitian untuk mendapatkan pemahaman mendalam tentang hasil analisis data dan menyusun rekomendasi relevan. Peneliti menafsirkan hasil analisis data dari *EDA & Visualization*, mencari pola, tren, dan hubungan yang terungkap. Wawasan berharga ditemukan dan rekomendasi disusun, termasuk saran untuk perbaikan proses bisnis, pengambilan keputusan strategis, atau langkah-langkah berikutnya. Rekomendasi didasarkan pada pemahaman yang mendalam tentang konteks bisnis dan tujuan yang ingin dicapai. Dengan demikian, rekomendasi ini menjadi panduan strategis bagi perusahaan dalam mengoptimalkan penjualan produk *fashion* di *platform e-commerce* Australia.

## 3. Hasil dan Pembahasan

### 3.1. Data Gathering

Dataset yang dianalisis dalam penelitian ini adalah "*Shopping Cart Database*", yang dapat diakses melalui *platform* Kaggle. Sebagai dataset yang tersedia secara terbuka, informasi yang terdapat di dalamnya dapat dimanfaatkan untuk berbagai keperluan penelitian. Dataset ini terdiri dari empat file yang disimpan dalam format CSV (*Comma Separated Values*), yaitu *customers.csv*, *orders.csv*, *products.csv*, dan *sales.csv*.

Tampilan data dari *customers.csv* dapat dilihat pada Gambar 2 di bawah ini. Gambar tersebut menampilkan data *customers.csv*, mencakup atribut seperti *customer\_id*, *customer\_name*, *gender*, *age*, *home\_address*, *zip\_code*, *city*, *state*, dan *country*.

customer_id	customer_name	gender	age	home_address	zip_code	city	state	country
1	Leanne Bosson	Female	30	8506 Victoria Terrace Suite 559	5484	Johnstonhaven	Northern Territory	Australia
2	Zabrina Harrows	Genderfluid	69	8327 Kilde Summit Apt. 401	8223	New Zacharyfort	South Australia	Australia
3	Shina Dullaghan	Polygender	59	269 Gemma Summit Suite 189	5661	Alburgh	Australian Capital Territory	Australia
4	Hevet McVie	Bigender	67	743 Bailey Grove Suite 141	1729	South Justinhav	Queensland	Australia
5	Rubia Ashleigh	Polygender	30	48 Hyatt Manor Suite 375	4032	Griffithshire	Queensland	Australia
6	Conley Tolcher	Genderfluid	40	7118 Micolough Square Suite 639	9996	Stahaven	New South Wales	Australia
7	Windsor Erbam	Bigender	76	52 Hills Station St Apt 683	793	Masonfurt	Queensland	Australia
8	Marlowe Wynn	Agender	76	383 Muller Summit Suite 809	7681	Samside	Northern Territory	Australia
9	Brittaney Gordie	Male	51	57 Greenfelder Hill Apt. 077	2	Beleport	Northern Territory	Australia
10	Susantha Wilah	Bigender	70	615 Hayley Knoll Suite 454	2116	Jobelburgh	Western Australia	Australia
11	Michaeline Kuhn	Agender	39	96 Danier Place Suite 898	7409	Georgeland	Northern Territory	Australia
12	Fedora Dmych	Bigender	76	66 Kayla Mews Suite 951	6334	Taylorburgh	South Australia	Australia
13	Marabel Swinfor	Male	42	16 Kuhn Loop Suite 442	6170	Maddisonmouth	Queensland	Australia
14	Christia Wackly	Bigender	36	403 Doherty Run Suite 497	2050	North Benjamin	New South Wales	Australia
15	Avril Rossiter	Genderfluid	34	254 An Ridge Apt. 350	491	Kiehnburgh	Western Australia	Australia
16	Gabbie Johnson	Male	79	424 Mason Place Apt. 151	6436	New Kal	Victoria	Australia
17	Devonna Cuttix	Genderfluid	32	29 Imogen Crest Suite 275	8309	Lake Graceside	Australian Capital Territory	Australia
18	Chan Duchesne	Male	79	13 Bailey Manor Apt. 848	7171	Walterland	New South Wales	Australia
19	Chadwick Crudd	Agender	41	802 Beau Plaza Apt. 306	7053	West Bailey	Victoria	Australia

Gambar 2. Data Customers

Selanjutnya, pada Gambar 3, ditampilkan data dari *orders.csv*. Gambar ini memuat beberapa data *orders*, termasuk detail seperti *order\_id*, *customer\_id*, *payment*, *order\_date*, dan *delivery\_date*.

	A	B	C	D	E
1	order_id	customer_id	payment	order_date	delivery_date
2		722	401 \$ 20,524.00	10/24/2021	11/20/2021
3		227	406 \$ 36,809.00	10/23/2021	11/19/2021
4		153	727 \$ 49,677.00	10/21/2021	11/17/2021
5		89	448 \$ 29,268.00	10/22/2021	11/16/2021
6		196	630 \$ 48,990.00	10/22/2021	11/16/2021
7		980	84 \$ 57,728.00	10/19/2021	11/14/2021
8		258	742 \$ 38,266.00	10/21/2021	11/13/2021
9		885	16 \$ 27,561.00	10/18/2021	11/13/2021
10		247	68 \$ 39,513.00	10/19/2021	11/12/2021
11		607	648 \$ 36,692.00	10/19/2021	11/12/2021
12		715	225 \$ 51,678.00	10/19/2021	11/12/2021
13		93	908 \$ 52,379.00	10/22/2021	11/11/2021
14		708	687 \$ 56,307.00	10/24/2021	11/10/2021
15		458	757 \$ 13,142.00	10/21/2021	11/10/2021
16		419	745 \$ 19,518.00	10/20/2021	11/10/2021
17		449	163 \$ 56,294.00	10/23/2021	11/9/2021
18		333	367 \$ 21,091.00	10/18/2021	11/9/2021
19		128	936 \$ 17,661.00	10/13/2021	11/9/2021
20		212	62 \$ 29,887.00	10/14/2021	11/8/2021

Gambar 3. Data Orders

Data dari *products.csv* dapat dilihat pada Gambar 4 di bawah ini. Gambar tersebut menampilkan data products yang dijual, termasuk atribut seperti *product\_ID*, *product\_type*, *product\_name*, *size*, *colour*, *price*, *quantity*, dan *description*.

	A	B	C	D	E	F	G	H
1	product_id	product_type	product_name	size	colour	price	quantity	description
2	0	Shirt	Oxford Cloth	XS	red	\$ 114.00	66	A red coloured, XS sized, Oxford Cloth Shirt
3	1	Shirt	Oxford Cloth	S	red	\$ 114.00	53	A red coloured, S sized, Oxford Cloth Shirt
4	2	Shirt	Oxford Cloth	M	red	\$ 114.00	54	A red coloured, M sized, Oxford Cloth Shirt
5	3	Shirt	Oxford Cloth	L	red	\$ 114.00	69	A red coloured, L sized, Oxford Cloth Shirt
6	4	Shirt	Oxford Cloth	XL	red	\$ 114.00	47	A red coloured, XL sized, Oxford Cloth Shirt
7	5	Shirt	Oxford Cloth	XS	orange	\$ 114.00	45	A orange coloured, XS sized, Oxford Cloth Shirt
8	6	Shirt	Oxford Cloth	S	orange	\$ 114.00	72	A orange coloured, S sized, Oxford Cloth Shirt
9	7	Shirt	Oxford Cloth	M	orange	\$ 114.00	77	A orange coloured, M sized, Oxford Cloth Shirt
10	8	Shirt	Oxford Cloth	L	orange	\$ 114.00	46	A orange coloured, L sized, Oxford Cloth Shirt
11	9	Shirt	Oxford Cloth	XL	orange	\$ 114.00	43	A orange coloured, XL sized, Oxford Cloth Shirt
12	10	Shirt	Oxford Cloth	XS	yellow	\$ 114.00	72	A yellow coloured, XS sized, Oxford Cloth Shirt
13	11	Shirt	Oxford Cloth	S	yellow	\$ 114.00	78	A yellow coloured, S sized, Oxford Cloth Shirt
14	12	Shirt	Oxford Cloth	M	yellow	\$ 114.00	56	A yellow coloured, M sized, Oxford Cloth Shirt
15	13	Shirt	Oxford Cloth	L	yellow	\$ 114.00	75	A yellow coloured, L sized, Oxford Cloth Shirt
16	14	Shirt	Oxford Cloth	XL	yellow	\$ 114.00	50	A yellow coloured, XL sized, Oxford Cloth Shirt
17	15	Shirt	Oxford Cloth	XS	green	\$ 114.00	68	A green coloured, XS sized, Oxford Cloth Shirt
18	16	Shirt	Oxford Cloth	S	green	\$ 114.00	56	A green coloured, S sized, Oxford Cloth Shirt
19	17	Shirt	Oxford Cloth	M	green	\$ 114.00	60	A green coloured, M sized, Oxford Cloth Shirt
20	18	Shirt	Oxford Cloth	L	green	\$ 114.00	52	A green coloured, L sized, Oxford Cloth Shirt

Gambar 4. Data Products

Terakhir, pada Gambar 5, terlihat data dari *sales.csv*. Gambar ini menunjukkan beberapa data *sales*, termasuk informasi seperti *sales\_id*, *order\_id*, *product\_id*, *price\_per\_unit*, *quantity*, dan *total\_price*.

	A	B	C	D	E	F
1	sales_id	order_id	product_id	price_per_unit	quantity	total_price
2	0	1	1	218 \$	106.00	2 \$ 212.00
3	1	1	1	481 \$	118.00	1 \$ 118.00
4	2	1	1	2 \$	96.00	3 \$ 288.00
5	3	1	1	1002 \$	106.00	2 \$ 212.00
6	4	1	1	691 \$	113.00	3 \$ 339.00
7	5	1	1	981 \$	106.00	3 \$ 318.00
8	6	2	2	915 \$	96.00	1 \$ 96.00
9	7	2	2	686 \$	113.00	1 \$ 113.00
10	8	2	2	1091 \$	115.00	3 \$ 345.00
11	9	2	2	1196 \$	105.00	1 \$ 105.00
12	10	2	2	157 \$	91.00	3 \$ 273.00
13	11	2	2	1174 \$	99.00	2 \$ 198.00
14	12	3	3	54 \$	104.00	1 \$ 104.00
15	13	3	3	6 \$	96.00	2 \$ 192.00
16	14	3	3	1010 \$	106.00	2 \$ 212.00
17	15	4	4	1033 \$	90.00	1 \$ 90.00
18	16	4	4	1184 \$	99.00	3 \$ 297.00
19	17	4	4	769 \$	109.00	1 \$ 109.00
20	18	4	4	923 \$	96.00	3 \$ 288.00

Gambar 5. Data Sales

### 3.2. Business & Data Understanding

Setelah mempersiapkan data, langkah selanjutnya adalah memahami masalah-masalah bisnis yang dihadapi. Ini penting untuk pengambilan keputusan efektif. Dengan pendekatan ini, perusahaan dapat mengidentifikasi titik-titik lemah dalam proses bisnisnya. Penggunaan analisis *pivot table* memungkinkan perusahaan untuk dengan cepat memahami pola penurunan penjualan yang mungkin terjadi.

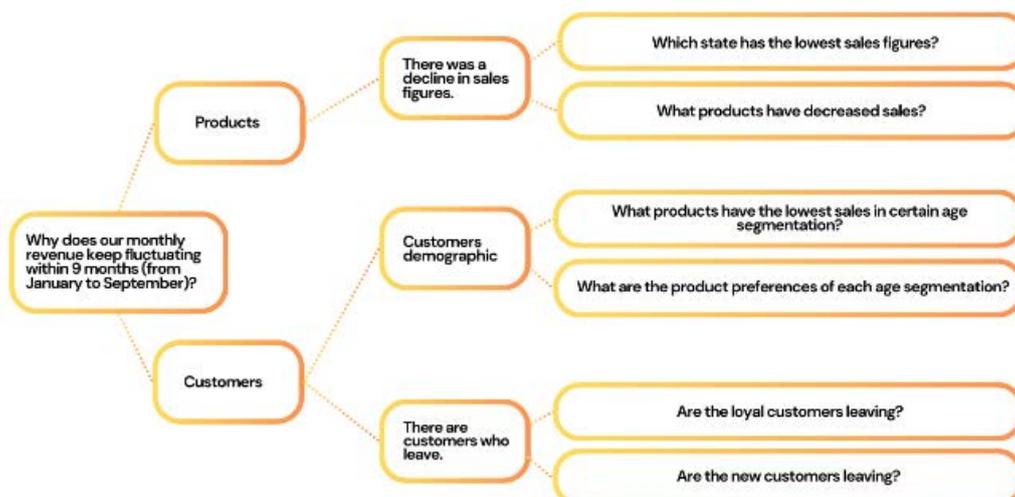
Month	SUM of Revenue
01	\$ 119,333.00
02	\$ 94,894.00
03	\$ 131,364.00
04	\$ 97,530.00
05	\$ 85,597.00
06	\$ 101,458.00
07	\$ 116,081.00
08	\$ 104,565.00
09	\$ 96,526.00

Gambar 6. Revenue

Dari hasil analisis *pivot table*, terlihat dinamika penjualan bulanan. Penjualan mencapai puncaknya pada Januari (\$119,333.00), namun turun signifikan pada Februari (\$95,080.00). Meskipun mengalami kenaikan pada Maret (\$131,364.00), tren penurunan terjadi pada April (\$97,530.00) dan Mei (\$85,597.00). Bulan Juni mencatat kenaikan (\$101,458.00), tetapi turun lagi pada Juli (\$116,081.00) dan Agustus (\$104,565.00), mencapai titik terendah di September (\$96,526.00). Dengan memahami dinamika penjualan ini, perusahaan dapat mengidentifikasi penyebab fluktuasi.

Pendapatan bulanan fluktuasi antara \$85.597 dan \$131.364 selama 9 bulan terakhir (Januari hingga September). Dengan menetapkan pernyataan masalah "Bagaimana meningkatkan pendapatan bulanan hingga rata-rata lebih dari \$100.000 dalam 6 bulan ke depan?", tujuan jelas ditetapkan untuk mencari solusi. Analisis data ini memberikan dasar untuk merumuskan strategi meningkatkan pendapatan perusahaan dalam periode yang ditentukan.

Dari permasalahan yang diidentifikasi, maka selanjutnya menentukan *Root Cause Analysis* (RCA). Berikut hasil dari *root cause analysis*:



Gambar 7. Root Cause Analysis

Fluktuasi pendapatan bulanan pada produk fashion dalam e-commerce di Australia dapat dipicu oleh dua faktor utama, yaitu faktor *products* (produk) dan faktor *customers* (pelanggan). Di bawah faktor *products* (produk), terdapat satu subfaktor yang menyebabkan penurunan penjualan, yaitu Penurunan dalam angka penjualan (*Decline in Sales Figures*), yang menjelaskan bahwa terjadi penurunan dalam angka penjualan. Sedangkan di bawah faktor *customers* (pelanggan), terdapat dua subfaktor yang menyebabkan penurunan penjualan, yaitu Demografi Pelanggan (*Customers Demographic*) yang menjelaskan bagaimana karakteristik demografis pelanggan mempengaruhi penjualan, dan Pelanggan yang Meninggalkan (*There are Customers Who Leave*) yang mengarah pada perubahan dalam basis pelanggan.

Meskipun beberapa faktor lainnya juga dapat memengaruhi penjualan, tetapi tidak dapat dipastikan dari data yang ada. Karena itu, dari ketiga subfactor tersebut, muncul beberapa *business questions* yang akan dianalisis. Beberapa di antaranya adalah:

- 1) Negara bagian mana yang memiliki angka penjualan terendah?
- 2) Produk mana yang mengalami penurunan penjualan?
- 3) Produk apa yang memiliki penjualan terendah dalam segmentasi usia tertentu?
- 4) Preferensi produk apa yang dimiliki oleh setiap segmen usia?
- 5) Apakah pelanggan setia pergi?
- 6) Apakah pelanggan baru pergi?

Untuk menjawab beberapa *business questions* diatas, maka akan digunakan data dengan cakupan sebagai berikut:

- 1) Kumpulan data mencakup wilayah Australia
- 2) Data order mencakup periode 1 Januari 2021 hingga 30 September 2021
- 3) Ada tiga jenis produk: *shirts*, *jackets*, dan *trousers*

### 3.3. Data Cleaning

Setelah menetapkan pertanyaan bisnis dan memahami cakupan data, langkah berikutnya adalah *data cleaning*. Proses ini dilakukan secara manual menggunakan *Microsoft Excel*. Berikut langkah-langkah dalam proses *data cleaning*:

- 1) Menghapus Baris Duplikat: Pada tabel *customers*, *orders*, *products*, dan *sales*, tidak ditemukan adanya baris duplikat setelah dilakukan pemeriksaan.
- 2) Pemeriksaan *Missing Values*: Setelah pemeriksaan, tidak terdapat *missing values* pada tabel *customers*, *orders*, *products*, dan *sales*. Hal ini menunjukkan bahwa data lengkap dan tidak ada nilai yang hilang.
- 3) Pengubahan Format Data: Beberapa kolom diubah formatnya untuk konsistensi dan kemudahan pemrosesan data. Seperti, kolom *payment*, *price*, *price\_per\_unit*, dan *total\_price* diubah formatnya menjadi format mata uang. Sedangkan kolom *order\_date* dan *delivery\_date* diubah formatnya menjadi format tanggal.
- 4) Pemeriksaan *Whitespace*: Tidak terdapat *whitespace* yang tidak diinginkan pada tabel *customers*, *orders*, *products*, dan *sales* setelah dilakukan pemeriksaan.
- 5) Pemeriksaan *Outlier*: Untuk memeriksa *outlier*, kolom *total\_price* pada tabel *sales* ditambahkan, yang merupakan hasil perkalian antara *price\_per\_unit* dengan *quantity*. Selain itu, kolom *z\_score* juga ditambahkan untuk menilai seberapa jauh suatu nilai *total\_price* dari rata-rata populasi dalam satuan standar deviasi. Hasil dari penambahan kolom ini menunjukkan bahwa tidak terdapat *outlier* dalam data.

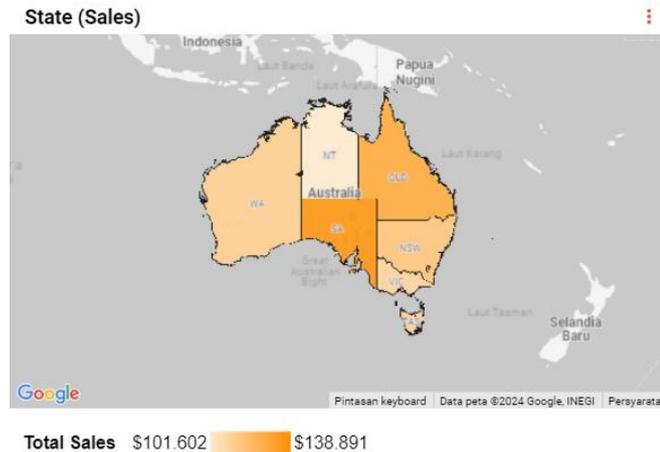
### 3.4. EDA & Visualization

Setelah data dibersihkan, selanjutnya yaitu analisis data yang dilakukan menggunakan BigQuery, yang mampu mengeksekusi *SQL queries* pada dataset besar [18]. Visualisasi hasil analisis menggunakan *Looker Studio*, platform yang memungkinkan pengguna untuk menggambarkan data melalui visualisasi [19]. Analisis mendalam terhadap penjualan produk *fashion* di *e-commerce* Australia selama Januari hingga September 2021 menghasilkan beberapa insight penting:

- 1) Negara dengan Penjualan Terendah



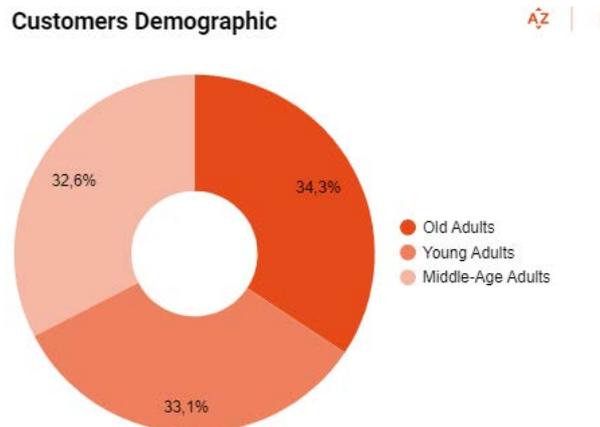
Gambar 8. Penjualan Terendah di Negara-negara – Diagram



**Gambar 9.** Penjualan Terendah di Negara-negara – Peta

*Northern Territory* menunjukkan angka penjualan terendah sebesar \$101.602, sementara Australia Selatan mencatat angka penjualan tertinggi sebesar \$138.891. Perbedaan antara penjualan tertinggi dan terendah ini mencapai sekitar \$37.000, menyoroti disparitas yang signifikan dalam performa penjualan di berbagai wilayah. Total penjualan selama periode tersebut mencapai \$947.534.

## 2) Demografi Pelanggan



**Gambar 10.** Demografi Pelanggan

Segmentasi usia "*Old-Adults*" merupakan kelompok terbesar dengan 343 pelanggan, yang menyumbang 34,3% dari total pelanggan. Sementara itu, segmentasi usia "*Middle-Age Adults*" merupakan kelompok dengan jumlah pelanggan terendah, hanya memiliki 326 pelanggan, atau sekitar 32,6% dari total pelanggan.

## 3) Produk yang Mengalami Penurunan Penjualan

**Products (Decreased Sales)**

Product	Quantity Sold...	Price/Unit	Total Sales
1. Pullover	255	\$91	\$23.205
2. Cords	209	\$94	\$19.646
3. Mandarin Collar	205	\$109	\$22.345

**Gambar 11.** Penurunan Penjualan Produk

*Insight* yang didapat dari data tersebut menunjukkan bahwa produk-produk yang mengalami penurunan penjualan selama periode Januari hingga September adalah *Cords*, *Mandarin Collar*, dan *Pullover*. Dari ketiga produk tersebut, *Cords* merupakan produk dengan penjualan terendah, hanya mencapai total penjualan sebesar \$19.646. Menariknya, meskipun *Pullover* memiliki harga satuan terendah yaitu \$91, namun produk ini mampu mencatat penjualan tertinggi di antara ketiga produk tersebut, mencapai \$23,2 ribu.

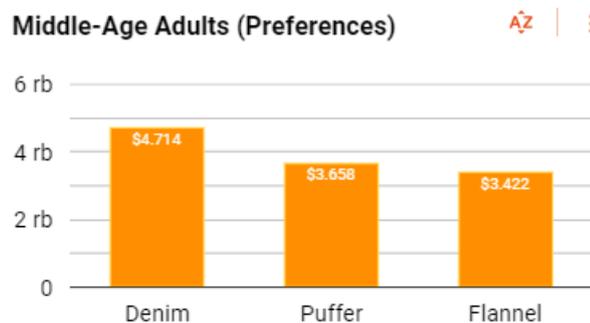
4) Preferensi Produk untuk Usia Muda (*Young-Adults*)



**Gambar 12.** Preferensi Produk untuk Usia Muda

Tiga produk dengan penjualan tertinggi untuk "*Young-Adults*" dari Januari hingga September adalah *Denim* (\$4,8K), *Pleated* (\$3,6K), dan *Tracksuit Bottom* (\$3,2K). Dari data ini, kita dapat melihat bahwa *Denim* menjadi produk yang paling diminati oleh kelompok usia muda (*Young-Adults*), dengan total penjualan mencapai sekitar \$4.806. Hal ini menunjukkan bahwa *Denim* memiliki daya tarik yang kuat di kalangan dewasa muda (*Young-Adults*) dan dapat menjadi fokus dalam strategi pemasaran lebih lanjut untuk segmen ini.

5) Preferensi Produk untuk Usia Tengah (*Middle-Aged Adults*)

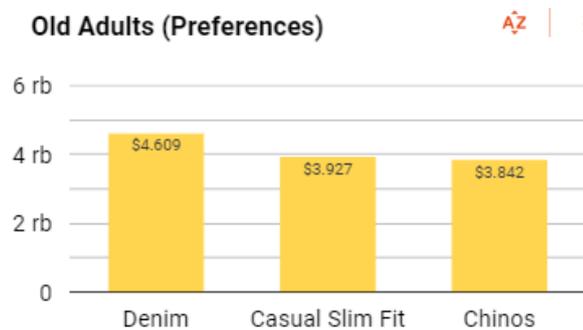


**Gambar 13.** Preferensi Produk untuk Usia Tengah

*Insight* yang diperoleh dari data tersebut adalah bahwa terdapat perbedaan preferensi produk antara dua kelompok usia, yaitu usia tengah (*Middle-Aged Adults*) dan usia muda (*Young-*

*Adults*). Untuk usia Tengah (*Middle-Aged Adults*), tiga produk dengan penjualan tertinggi dari Januari hingga September adalah *Denim* (\$4,7K), *Puffer* (\$3,6K), dan *Flanel* (\$3,4K). Dari data tersebut, kita dapat melihat bahwa *Denim* menjadi produk favorit dengan total penjualan yang signifikan, sekitar \$4.714, untuk kategori usia Tengah (*Middle-Aged Adults*). Hal ini menunjukkan bahwa *Denim* memiliki popularitas yang tinggi di kalangan dewasa usia menengah (*Middle-Aged Adults*).

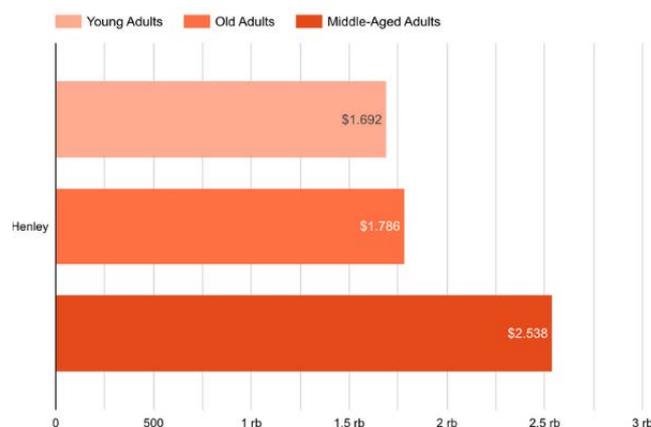
6) Preferensi Produk untuk Usia Lanjut (*Old-Adults*)



Gambar 14. Preferensi Produk untuk Usia Lanjut

Data menunjukkan bahwa dari Januari hingga September, terdapat tiga produk dengan penjualan tertinggi untuk kelompok usia tua (*Old-Adults*), yaitu *Denim* (\$4,6K), *Casual Slim Fit* (\$3,9K), dan *Chinos* (\$3,8K). Insight ini menyoroti pentingnya fokus pada produk-produk ini dalam upaya meningkatkan pendapatan dari segmen pasar yang lebih tua. Khususnya, *Denim* menonjol sebagai produk paling diminati dengan total penjualan mencapai \$4.609, menunjukkan potensi besar untuk pertumbuhan bisnis dalam segmen ini.

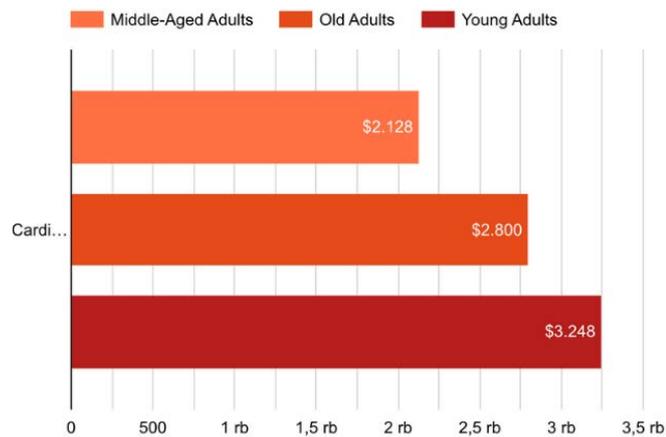
7) Produk dengan Penjualan Terendah untuk Usia Muda (*Young-Adults*)



Gambar 15. Produk Penjualan Terendah untuk Usia Muda

Data menunjukkan bahwa meskipun produk *Henley* memiliki penjualan terendah untuk segmen "*Young-Adults*" dengan total penjualan sebesar \$1.692, namun terdapat perbedaan yang signifikan dalam penjualan antara segmen "*Young-Adults*" dan "*Middle-Aged Adults*" sebesar \$846. Hal ini menunjukkan bahwa produk yang memiliki kinerja penjualan rendah pada kelompok "*Young-Adults*" belum tentu memiliki kinerja yang sama pada segmen usia yang berbeda, terutama pada segmen "*Middle-Aged Adults*".

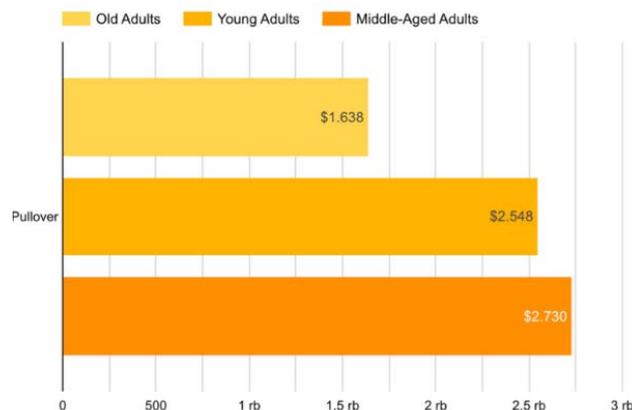
8) Produk dengan Penjualan Terendah untuk Usia Tengah (*Middle-Aged Adults*)



**Gambar 16.** Produk Penjualan Terendah untuk Usia Tengah

*Cardigan* merupakan produk dengan penjualan terendah untuk "*Middle-Aged Adults*" dengan total penjualan \$2.128. Perbedaan hasil penjualan *Cardigan* di setiap segmentasinya cukup terlihat. Perbedaan hasil penjualan *Cardigan* pada masing-masing segmentasi cukup mencolok, dengan selisih penjualan sebesar \$672 untuk "*Old-Adults*", dan \$1.120 untuk "*Young Adults*". Berdasarkan data tersebut, kita dapat menyimpulkan bahwa "*Middle-Aged Adults*" kurang tertarik membeli *Cardigan* jika dibandingkan dengan segmen usia lainnya.

9) Produk dengan Penjualan Terendah untuk Usia Lanjut (*Old-Adults*)



**Gambar 17.** Produk Penjualan Terendah untuk Usia Lanjut

Analisis menunjukkan bahwa pullover merupakan produk dengan penjualan terendah untuk segmen "*Old-Adults*", dengan total penjualan sebesar \$1.638.

Terdapat perbedaan yang signifikan dalam hasil penjualan *pullover* di setiap segmen usia. Penjualan *pullover* untuk segmen "*Young Adults*" memiliki selisih sebesar \$910, sementara untuk segmen "*Middle-Aged Adults*" selisihnya mencapai \$1.092. *Insight* yang didapat dari data tersebut adalah bahwa "*Old-Adults*" cenderung kurang tertarik untuk membeli *pullover* dibandingkan dengan segmen usia lainnya.

10) Apakah Pelanggan Setia dan pelanggan baru Meninggalkan?

		different_time / total_users									
cohort_...	first_pu...	0	1	2	3	4	5	6	7	8	9
1 Jan 2...	104	104	13	11	12	10	10	15	10	6	10
1 Feb 2...	77	77	14	9	11	5	10	6	6	8	-
1 Mar 2...	90	90	5	5	11	12	12	11	11	-	-
1 Apr 2...	70	70	4	9	8	6	5	7	-	-	-
1 Mei 2...	48	48	7	1	1	6	3	-	-	-	-
1 Jun 2...	57	57	8	1	6	5	-	-	-	-	-
1 Jul 20...	46	46	7	4	7	-	-	-	-	-	-
1 Agu 2...	58	58	4	2	-	-	-	-	-	-	-
1 Sep 2...	44	44	3	-	-	-	-	-	-	-	-

Gambar 18. Analisis Pelanggan Setia dan Pelanggan Baru (1)

		different_time / total_users									
cohort_...	first_pu...	0	1	2	3	4	5	6	7	8	9
1 Jan 2...	104	100%	12,5%	10,58%	11,54%	9,62%	9,62%	14,42%	9,62%	5,77%	9,62%
1 Feb 2...	77	74,04%	13,46%	8,65%	10,58%	4,81%	9,62%	5,77%	5,77%	7,69%	-
1 Mar 2...	90	86,54%	4,81%	4,81%	10,58%	11,54%	11,54%	10,58%	10,58%	-	-
1 Apr 2...	70	67,31%	3,85%	8,65%	7,69%	5,77%	4,81%	6,73%	-	-	-
1 Mei 2...	48	46,15%	6,73%	0,96%	0,96%	5,77%	2,88%	-	-	-	-
1 Jun 2...	57	54,81%	7,69%	0,96%	5,77%	4,81%	-	-	-	-	-
1 Jul 20...	46	44,23%	6,73%	3,85%	6,73%	-	-	-	-	-	-
1 Agu 2...	58	55,77%	3,85%	1,92%	-	-	-	-	-	-	-
1 Sep 2...	44	42,31%	2,88%	-	-	-	-	-	-	-	-

Gambar 19. Analisis Pelanggan Setia dan Pelanggan Baru (2)

Jumlah pelanggan baru menunjukkan tren penurunan seiring berjalannya waktu. Pada tanggal 1 Januari, terdapat 104 pelanggan yang melakukan pembelian pertama, namun berakhir dengan hanya 44 pelanggan pada tanggal 1 September. Dari data ini, terlihat adanya penurunan signifikan dalam jumlah pelanggan baru selama periode tersebut.

Selain itu, insight menarik lainnya adalah mengenai persentase pelanggan setia. Pada bulan Juni, tercatat persentase pelanggan setia tertinggi sebesar 14,4% dari total pelanggan baru yang melakukan pembelian pertama pada bulan Januari. Hal ini menunjukkan bahwa meskipun jumlah pelanggan baru menurun, persentase pelanggan setia masih relatif tinggi pada bulan tersebut.

Namun, pada bulan Juni dan Juli, terjadi penurunan drastis dalam persentase pelanggan setia menjadi hanya 0,96% dari total pelanggan baru yang berbelanja pada bulan Mei.

### 3.5. Insight & Recommendation

Berikut adalah *key insight* dan rekomendasi untuk meningkatkan pendapatan bulanan hingga rata-rata lebih dari \$100.000 dalam 6 bulan ke depan:

- 1) Optimalkan Penjualan di *Northern Territory*: Fokuskan pemasaran dan promosi di wilayah ini. Identifikasi faktor-faktor yang mempengaruhi penjualan rendah dan kembangkan strategi untuk menarik pelanggan. Dengan alokasi sumber daya yang cerdas, potensi pangsa pasar dapat dimanfaatkan.
- 2) Segmentasi Pelanggan dan Produk: Perhatikan preferensi produk berdasarkan usia. Perluas penawaran produk yang diminati oleh setiap kelompok usia. Misalnya, produk seperti *Denim*, *Pleated*, dan *Tracksuit Bottom* diminati oleh "Young-Adults".
- 3) Promosikan Produk dengan Penjualan Rendah: Identifikasi produk dengan penjualan terendah dan kembangkan strategi promosi khusus. Misalnya, *produk Cords*, *Mandarin Collar*, dan *Pullover* bisa mendapatkan dorongan melalui diskon khusus.
- 4) Penyempurnaan Produk: Lakukan peninjauan produk dengan penjualan rendah. Pastikan harga, kualitas, dan relevansi produk sesuai dengan preferensi pelanggan.
- 5) Analisis Kinerja dan Evaluasi Berkala: Tinjau kembali data penjualan dan respons pelanggan untuk mengidentifikasi tren dan peluang baru, serta menyesuaikan strategi yang ada agar lebih efektif.

#### 4. Kesimpulan

Dari hasil analisis yang dilakukan, fluktuasi pendapatan bulanan produk *fashion* dalam *e-commerce* di Australia menyoroti pentingnya memahami faktor-faktor yang memengaruhi pola pembelian konsumen. Metode analisis RCA, EDA, dan visual data mining telah membantu mengidentifikasi penyebab utama fluktuasi penjualan, termasuk karakteristik produk dan perubahan dalam basis pelanggan. Insight penting ditemukan, seperti preferensi produk berdasarkan segmentasi usia dan tren pelanggan baru dan setia. Rekomendasi yang diusulkan mencakup strategi pengoptimalan penjualan di wilayah tertentu, segmentasi produk dan pelanggan yang lebih tepat, promosi produk dengan penjualan rendah, serta penyempurnaan produk dan evaluasi kinerja berkala. Kontribusi penelitian ini terletak pada pemahaman mendalam terhadap perilaku konsumen dan rekomendasi strategis yang diberikan, yang dapat membantu perusahaan *e-commerce fashion* di Australia meningkatkan strategi penjualan dan mencapai pertumbuhan bisnis yang berkelanjutan.

#### References

- [1] T. N. Muthmainnah and A. Voutama, "Volume 6 ; Nomor 2," *Juli*, vol. 6, pp. 463–471, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [2] B. Berliandika, D. Isfianadewi, and A. Priyono, "Strategi Sustainable Business Model Innovation pada Industri Fashion Digital Ilustrasi," *J. Samudra Ekon. dan Bisnis*, vol. 13, no. 1, pp. 46–60, 2021, doi: 10.33059/jseb.v13i1.3798.
- [3] A. P. Kehista *et al.*, "Analisis Keamanan Data Pribadi pada Pengguna E-Commerce: Ancaman, Risiko, Strategi Kemanan (Literature Review)," *J. Ilmu Manaj. Terap.*, vol. 4, no. 5, pp. 625–632, 2023, [Online]. Available: <https://dinastirev.org/JIMT/article/view/1541>
- [4] I. A. Siregar, "Analisis Dan Interpretasi Data Kuantitatif," *ALACRITY J. Educ.*, vol. 1, no. 2, pp. 39–48, 2021, doi: 10.52121/alacrity.v1i2.25.
- [5] A. V. Dica Parameswari Syifa Dewi, "Application of K-Means Algorithm on HIV Cases in West Java for Clustering Based on The Level of Spread in Each District / City Using RapidMiner," *Jtsi*, vol. 4, no. 2, pp. 289–300, 2023.
- [6] I. S. Haq and M. A. Purba, "Kajian Penyebab Kerusakan Door Packing pada Tabung Sterilizer Menggunakan Metode Root Cause Analysis (RCA) di Sungai Kupang Mill," *J. Vokasi Teknol. Ind.*, vol. 2, no. 2, 2020, doi: 10.36870/jvti.v2i2.177.
- [7] D. Leni, F. Earnestly, R. Sumiati, A. Adriansyah, and Y. P. Kusuma, "Evaluasi sifat mekanik baja paduan rendah berdasarkan komposisi kimia dan suhu perlakuan panas menggunakan teknik exploratory data analysis (EDA)," *Din. Tek. Mesin*, vol. 13, no. 1, p. 74, 2023, doi: 10.29303/dtm.v13i1.624.
- [8] G. Yoshua, U. Apsiswanto, S. Dharma Wacana Metro, I. Darmajaya Lampung, J. Kenangan No, and M. Kota Metro, "Visualisasi Persebaran Obat Menggunakan Visual Data Mining (Vdm) Studi Kasus Pada Pt. Bernofarm Wilayah Lampung," *J. Inf. dan Komput.*, vol. 11, no. 1, pp. 165–170, 2023.
- [9] S. Faisal, "Klasifikasi Data Mining Menggunakan Algoritma C4.5 Terhadap Kepuasan Pelanggan Sewa Kamera Cikarang," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2019, doi: 10.36805/technoexplorev.v4i1.541.
- [10] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [11] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.
- [12] F. R. Zani and H. Supriyanto, "Analisis Perbaikan Proses Pengemasan Menggunakan Metode Root Cause Analysis Dan Failure Mode and Effect Analysis Dalam Upaya Meningkatkan Kualitas Produk Pada Cv. Xyz," *Semin. Nas. Sains dan Teknol. Terap. IX*, pp. 140–146, 2021.
- [13] M. Riyyan and H. Firdaus, "PERBANDINGAN ALGORITME NAA• VE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022, doi: 10.36595/jire.v5i1.547.
- [14] D. T. Husni *et al.*, "Analisis Big Data Penjualan Video Games Menggunakan Eda," *J. Tek. Inf. dan Komput.*, vol. 5, no. 1, p. 43, 2022, doi: 10.37600/tekinkom.v5i1.517.
- [15] E. D. Madyatmadja, M. N. Ridho, A. R. Pratama, M. Fajri, and L. Novianto, "Penerapan

- Visualisasi Data Terhadap Klasifikasi Tindak Kriminal Di Indonesia,” *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 61–68, 2022, doi: 10.37365/jti.v8i1.127.
- [16] A. Yoga Pratama *et al.*, “Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja),” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [17] A. Ade Christian, Riki Supriyadi, Fattya Ariani, “PEMBELIAN PADA TOKO UMBAH SEPATU,” *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 2, pp. 162–169, 2023, doi: 10.36595/jire.v6i2.
- [18] E. Riana, “Implementasi Cloud Computing Technology dan Dampaknya Terhadap Kelangsungan Bisnis Perusahaan Dengan Menggunakan Metode Agile dan Studi Literatur,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 3, p. 439, 2020, doi: 10.30865/jurikom.v7i3.2192.
- [19] Y. Setiani, N. Rachmah, and I. Purnama, “Visualisasi Data Malnutrisi Anak Di Asia Menggunakan Looker Studio Serta Analisis Data Dengan Metode ANOVA,” *J. Ilm. Sist. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 188–212, 2023, [Online]. Available: <https://doi.org/10.55606/juisik.v3i3.701>

*This page is intentionally left blank.*

# Sistem Pendukung Keputusan Gerai Di Pusat Perbelanjaan *Modern* Provinsi Bali: Metode ANP

Andien Rachma Fadillah<sup>1</sup>, I Komang Ari Mogi<sup>2</sup>, Cokorda Rai Adi Pramatha<sup>3</sup>,  
I Gusti Ngurah Anom Cahyadi Putra<sup>4</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus Unud, Jimbaran, Bali, 80361, Indonesia

<sup>1</sup>andienrf06@gmail.com

<sup>2</sup>arimogi@unud.ac.id

<sup>3</sup>cokorda@unud.ac.id

<sup>4</sup>anom.cp@unud.ac.id

## Abstract

The rapid competition in the business world due to the Era of Revolution 4.0 has resulted in the need for the business world to maintain competitiveness in front of consumers. Expanding the market by establishing a presence in easily accessible locations such as modern shopping centers is one way to achieve this. Choosing the best location among the many modern shopping centers in Bali for opening an outlet is a challenge for business people, who in this research only used case studies in Bali Province of 16 Modern Shopping Centers. The aim of this research is to develop a web-based decision support system to facilitate the location selection process for business actors. This research will utilize the Analytic Network Process (ANP) method for decision making regarding outlet selection in modern shopping centers. The development of a web-based decision support system enables businesses to effectively select the most suitable locations within modern shopping centers to open new outlets, thereby maximizing their market presence. This system will help businesses make decisions on selecting locations for modern shopping centers in Bali Province.

**Keywords:** *Decision Support System, Analytic Network Process, Location Selection, Modern Shopping Centers, Websites*

## 1. Pendahuluan

Pada saat ini dapat dikatakan bahwa Era Revolusi 4.0 telah menjadi salah satu alasan persaingan industri sangat pesat dan membuat persaingan bisnis di sektor ekonomi menjadi ketat. Dengan begitu, para pelaku bisnis yang bergelut di sektor ekonomi diharuskan memutar otak untuk mempertahankan keunggulan serta tetap diingat oleh konsumen di pasaran. Dengan menguasai pasar yang luas dapat menjadi salah satu caranya. Perluasan pasar dapat membantu suatu produk memiliki nama yang dapat dikenal oleh masyarakat luas. Untuk mewujudkan hal tersebut suatu perusahaan ataupun *brand* diperlukan memiliki lokasi yang mudah dijangkau sehingga dapat menyebarkan produk ke masyarakat luas dengan lebih mudah.

Dari banyaknya lokasi lahan yang tersedia, pusat perbelanjaan modern menjadi salah satu pilihan bagi para pelaku bisnis. Sebagai elemen besar dalam kota, pusat perbelanjaan mempunyai pengaruh yang penting dalam suatu lingkungan perkotaan, oleh karena itu bangunan seperti ini tidak boleh dianggap sebagai satu elemen arsitektur tunggal tanpa dikaitkan dengan keadaan disekitarnya. Dengan perkembangan global kebutuhan dan juga gaya hidup masyarakat, definisi pusat perbelanjaan menjadi semakin kompleks dan berubah sesuai dengan tuntutan penggunaannya. Shopping Mall merupakan one stop leisure dimana pengunjung mendapatkan berbagai macam jenis leisure. Pada umumnya, shopping mall

merupakan pusat perbelanjaan dimana terdapat retail-retail yang menyuguhkan berbagai macam kebutuhan para pengunjung. Sehingga, shopping mall merupakan sarana hedonic dan utilitarian yang dapat menjadi tempat untuk mendapatkan kebutuhan sekaligus mendapatkan kesenangan[1].

Dari banyaknya aktivitas yang dapat dilakukan, membuat masyarakat tertarik untuk menjadikan pusat perbelanjaan modern menjadi salah satu list tempat berpergian. Ketua Pengurus Pusat Asosiasi Pusat Belanja Indonesia (APPBI) menyatakan bahwa pada tahun 2020 rata-rata pengunjung pusat perbelanjaan *modern* sebesar 50% dari kondisi normal (sebelum pandemi), tahun 2021 sebesar 60% dan tahun 2022 diperkirakan dapat mencapai 70 hingga 80%[2]. Dari hal tersebut, pembukaan gerai baru di pusat perbelanjaan modern bagi pengusaha bisnis dapat menjadi salah satu solusi yang tepat untuk memasarkan produknya ke pasar yang lebih luas.

Namun, dilihat dari banyaknya pusat perbelanjaan modern di Provinsi Bali dapat menimbulkan kebingungan dalam pemilihan gerai yang tepat. Berdasarkan hal tersebut, penulis ingin mengembangkan sebuah aplikasi berbasis web menggunakan metode Analytic Network Process untuk membantu dan mempermudah pengambilan keputusan dan pemilihan lokasi pembukaan gerai di dalam Pusat Perbelanjaan Modern (Mall) di Bali dan diharapkan dapat menyelesaikan masalah tersebut.

## 2. Metode Penelitian

Pada tahap ini dijelaskan rancangan metodologi penelitian yang berisi proses penelitian secara umum, metode penelitian yang digunakan sebagai berikut.

### 2.1 Sistem Pendukung Keputusan (SPK)

Sistem Pendukung Keputusan atau SPK merupakan sistem yang membantu pengambilan keputusan dengan memanfaatkan metode dan data untuk menyelesaikan masalah tidak terstruktur ataupun semi terstruktur, hal tersebut guna mencapai tujuan yang diinginkan dimana sistem harus sederhana, mudah dikelola, dapat beradaptasi sepenuhnya terhadap isu-isu utama, dan mudah dikomunikasikan.

### 2.2 Analytic Network Process (ANP)

Metode *Analytic Network Process* (ANP) merupakan pengembangan lebih lanjut dari metode *Analytical Hierarchy Process* (AHP). Metode ANP dapat membenahi perbedaan struktur AHP dalam hal kemampuan beradaptasi terhadap hubungan antar kriteria atau alternatif.

### 2.3 Data Penelitian

Sumber data untuk penelitian ini adalah sebagai berikut:

#### a. Data Primer

Data primer menggunakan data dummy untuk harga dan ukura gerai, sedangkan gambar dan jumlah pesaing gerai didapatkan dari pengamatan langsung di lapangan. Dari hasil penelitian yang dilakukan didapatkan 3 kriteria penilaian yaitu Ukuran Gerai, Harga Gerai, dan Jumlah Pesaing Gerai. Pusat perbelanjaan *modern* yang digunakan berjumlah 16 diantaranya adalah Bali Collection, Samasta Lifestyle Village, Sidewalk Jimbaran, Park 23, Lippo Mall Kuta, Discovery Shopping Mall, Beachwalk Shopping Centre, Lippo Plaza Sunset, Mall Bali Galeria, Seminyak Village, Seminyak Square, Trans Studio Mall Bali, Level 21 Mall, Ramayana Bali Mall, Lippo Plaza Sunset dan Living World Denpasar.

#### b. Data Sekunder

Data sekunder sendiri berasal dari sumber literatur sebelumnya yang terkait dengan objek penelitian.

Untuk lebih jelasnya dapat dilihat pada Tabel 1. Contoh Data.

**Tabel 1. Contoh Data**

No.	Nama Pusat Perbelanjaan Modern	Ukuran Gerai	Harga Gerai	Lokasi Gerai
1	Bali Collection	70 m <sup>2</sup>	Rp 24.000.000	Kab. Badung
2	Sidewalk Jimbaran	35 m <sup>2</sup>	Rp 18.000.000	Kab. Badung
3	Level 21 Mall	30 m <sup>2</sup>	Rp 21.000.000	Kota Denpasar

## 2.4 Teknik Pengumpulan Data

Data diperoleh menggunakan data dummy untuk harga dan ukuran lokasi gerai, sedangkan jumlah pesaing dan gambar gerai dilakukan dengan pengamatan langsung di lokasi, selain itu dilakukan pula wawancara langsung kepada pengusaha bisnis terkait kriteria-kriteria yang diperlukan dalam sistem.

## 2.5 Teknik Analisis Data

Metode analisis data di sini berbentuk ANP (Analytic Network Process) yang melibatkan penalaran logis dan nilai-nilai yang mengandalkan imajinasi, pengalaman, dan pengetahuan. Di sisi lain, metode ini membantu dalam pengambilan keputusan dan pemecahan masalah.

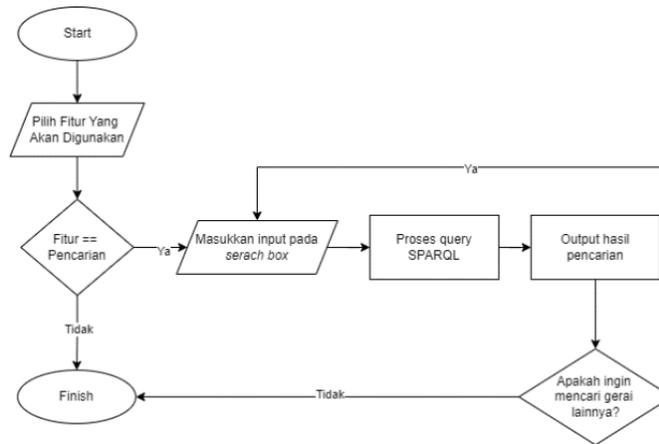
Terdapat langkah-langkah yang harus dilakukan guna mendapatkan hasil Keputusan akhir dalam hal pemilihan gerai. Pertama, user harus menentukan alternatif/mall apa yang akan dibandingkan sesuai kriteria ukuran, harga dan jumlah pesaing gerai. Kemudian, masing-masing kriteria dan alternatif akan diberikan nilai bobot sesuai persepsi atau keinginan user sesuai data yang dimiliki. Jika, nilai akhir dari masing-masing bobot kriteria dan alternatif bersifat Konsisten dapat dikatakan bahwa nilai bobot yang diinputkan sudah sesuai antar satu sama lainnya. Selanjutnya, user akan mendapatkan nilai hasil Unweighted dan Weighted Supermatrix sesuai dengan hasil nilai eigen yang didapatkan dari perhitungan nilai bobot. Terakhir hasil limiting supermatrix didapat dari hasil kali baris dan kolom pada nilai Unweighted dan Weighted Supermatrix hingga memiliki nilai yang stabil per baris nya.

## 2.6 Perancangan Sistem

Pada tahap ini menampilkan gambaran secara umum terhadap user mengenai sistem yang diciptakan.

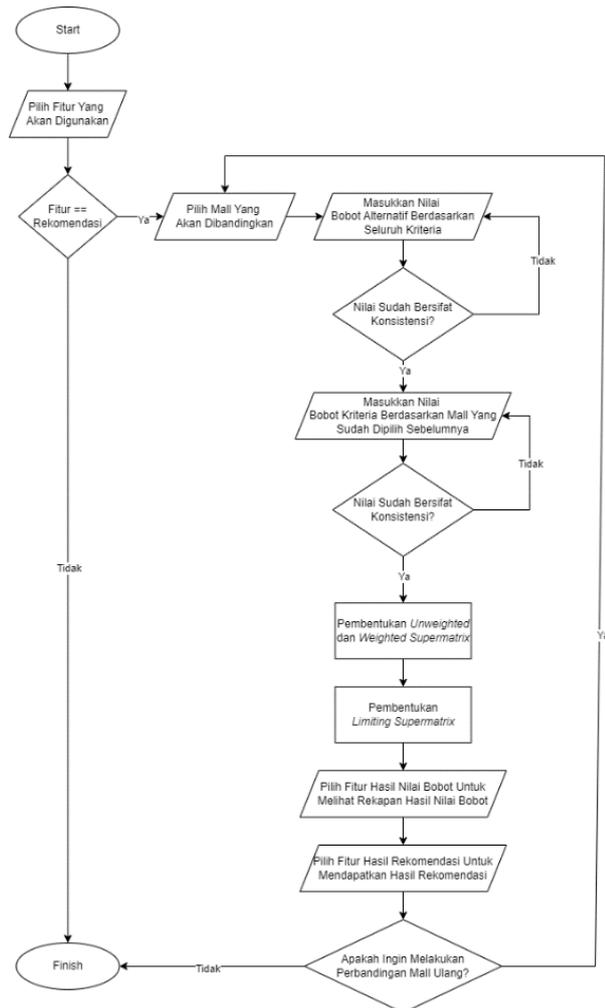
### 2.6.1 Flowchart

Pada sistem ini terdapat fitur pencarian, yang bertujuan mendapatkan list atau daftar gerai yang tersedia untuk pengusaha bisnis. Dapat dilihat pada Gambar 1. fitur pencarian diawali dengan memasukkan kata kunci berupa nama pusat perbelanjaan *modern* atau lokasi berupa kabupaten/kota dari pusat perbelanjaan *modern*. Kemudian, query SPARQL akan mengeksekusi atau melakukan proses pencarian sehingga Menampilkan hasil pencarian berdasarkan kata kunci yang diberikan.



Gambar 1. Flowchart Pencarian

Selain fitur pencarian, terdapat fitur rekomendasi yang akan memberikan hasil akhir berupa peringkat pusat perbelanjaan *modern* yang paling sesuai dengan penilaian user. Untuk mendapatkan hasil tersebut, dapat dilihat pada Gambar 2. Fitur rekomendasi diawali dengan memilih pusat perbelanjaan *modern* yang ingin dibandingkan satu sama lain. Kemudian, memilih fitur nilai bobot alternatif untuk seluruh kriteria guna memasukkan nilai bobot pada masing-masing kriteria sesuai dengan pusat perbelanjaan yang telah dipilih sebelumnya. Terdapat syarat dalam pengisian nilai bobot tersebut yaitu harus bernilai lebih dari 0,1 pada akhir nilai konsistensi. Setelah itu, dilanjutkan dengan pengisian nilai bobot kriteria pada alternatif sesuai dengan pusat perbelanjaan *modern* yang dipilih. Cara pengisian sama dengan pengisian nilai bobot alternatif dan memiliki syarat yang sama. Ketika syarat sudah terpenuhi user dapat memilih fitur Nilai Bobot untuk melihat pembentukan nilai *unweighted* dan *weighted* supermatriks yang diambil dari nilai eigen pada masing-masing nilai bobot yang telah dibentuk. Terakhir, untuk mendapat rekomendasi user dapat memilih fitur Hasil Rekomendasi yang akan menampilkan hasil *limiting* supermatrix dari perkalian baris dan kolom nilai *unweighted* dan *weighted* supermatriks, serta ditampilkan pula hasil peringkat pusat perbelanjaan *modern* sebagai bentuk rekomendasi yang diberikan.

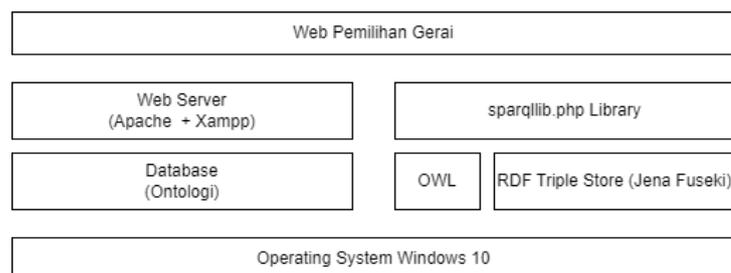


**Gambar 2.** Flowchart Rekomendasi

Pada Gambar 3. dapat dilihat bahwa terdapat langkah dalam pengambilan Keputusan

**2.6.2 Arsitektur Sistem**

Pada bagian arsitektur sistem dapat dilihat sesuai dengan Gambar 3. Arsitektur Sistem. Pengembangan sistem pemilihan gerai peneliti menggunakan Apache dan Xampp sebagai server web, untuk database sendiri menggunakan ontologi, dan Apache Jena sebagai triple store RDF. Terdapat library yang digunakan untuk menghubungkan PHP dengan query SPARQL yaitu sparqlib.php.



**Gambar 3.** Arsitektur Sistem

## 2.7 Pengujian Sistem

Pada tahap pengujian sistem menggunakan 2 pengujian yaitu Black-Box dan pengujian pemahaman pengguna terhadap sistem.

### 2.7.1 Pengujian *Black-Box*

Pengujian *black-box* digunakan untuk melakukan pengujian fungsional untuk keseluruhan sistem. Untuk fitur yang akan diuji sendiri adalah fitur pencarian dan fitur rekomendasi. Untuk fitur pencarian sendiri akan diuji keseluruhan fitur untuk melihat apakah hasil yang diharapkan sesuai dengan hasil yang diinginkan oleh peneliti. Begitu pula untuk pengujian fitur rekomendasi, seluruh fitur yang dijelaskan sebelumnya pada flowchart rekomendasi akan diuji oleh user untuk melihat apakah sesuai dengan hasil yang diharapkan.

### 2.7.2 Pengujian Pemahaman Sistem

Pengujian pemahaman terhadap sistem ini dilakukan dengan pemberian kuisioner kepada responden yang akan melakukan pengujian terhadap pemahaman sistem. Sebelumnya, peneliti sudah memberikan gambaran secara umum terkait bagaimana sistem bekerja dikarenakan responden akan diberikan pertanyaan untuk menguji masing-masing fitur. Hal ini dilakukan untuk melihat apakah user cukup memahami cara penggunaan sistem dengan melihat jawaban dari pertanyaan yang diajukan peneliti.

## 3. Hasil dan Pembahasan

### 3.1 Proses Metode ANP

Pertama yang harus dilakukan adalah dengan menentukan kriteria, alternatif dan nilai bobot kepentingan kriteria yang akan ditentukan oleh user saat menggunakan sistem. Banyaknya inputan sesuai dengan jumlah alternatif yang ingin dibandingkan dalam hal ini menggunakan 3 alternatif/mall. Setelah nilai diinputkan maka akan disesuaikan pada matriks perbandingan berpasangan (*pairwise matrix*), dan dilanjutkan dengan menentukan perhitungan eigen vektor pada setiap kriteria dan alternatif. Detail perhitungan dapat dilihat berikut ini.

**Tabel 2.** Perhitungan Nilai Bobot Kriteria Ukuran Terhadap Alternatif

	A04	A05	A06	Normalisasi			Eigen
A04	1,00000	0,14286	0,20000	0,07692	0,09677	0,04762	0,07377
A05	7,00000	1,00000	3,00000	0,53846	0,67742	0,71429	0,64339
A06	5,00000	0,33333	1,00000	0,38462	0,22581	0,23810	0,28284
Total	13,00000	1,47619	4,20000	1,00000	1,00000	1,00000	1,00000

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak Nilai Lamda didapatkan melalui perhitungan berikut:

$$\begin{aligned} \lambda_{Max} &= (0,07377 \times 13,00000) + (0,64339 \times 1,47619) + (0,28284 \times 4,20000) \\ &= 3,09670 \end{aligned}$$

Selanjutnya, untuk Consistency Index didapatkan dari perhitungan berikut:

$$CI = \frac{3,09670 - 3}{3 - 1} = 0,04835$$

Dan untuk mendapatkan nilai Consistency Ratio dapat melalui perhitungan berikut:

$$CR = \frac{0,04835}{0,58} = 0,08336$$

Nilai 0,58 didapatkan dari nilai random consistency index (RI) yang disesuaikan dengan jumlah elemen mall yaitu 3. Dengan nilai Consistency Ratio adalah 0,08336 maka pembobotan pada setiap alternatif bernilai konsisten dikarenakan nilai  $CR < 0,1$ .

**Tabel 3.** Perhitungan Manual Nilai Bobot Kriteria Harga Terhadap Alternatif

	A04	A05	A06	Normalisasi			Eigen
<b>A04</b>	1,00000	3,00000	3,00000	0,600000	0,60000	0,60000	0,60000
<b>A05</b>	0,33333	1,00000	1,00000	0,20000	0,20000	0,20000	0,20000
<b>A06</b>	0,33333	1,00000	1,00000	0,20000	0,20000	0,20000	0,20000
<b>Total</b>	1,66667	5,00000	5,00000	1,00000	1,00000	1,00000	1,00000

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak melalui perhitungan seperti sebelumnya. Nilai Lamda didapatkan 3,00000, kemudian untuk Consistency Index mendapatkan nilai 0 dan Consistency Ratio bernilai 0. Dengan nilai Consistency Ratio adalah 0 maka pembobotan pada setiap alternatif bernilai konsisten dikarenakan nilai  $CR < 0,1$ .

**Tabel 4.** Perhitungan Manual Nilai Bobot Kriteria Pesaing Terhadap Alternatif

	A04	A05	A06	Normalisasi			Eigen
<b>A04</b>	1,00000	3,00000	7,00000	0,67742	0,71429	0,53846	0,64339
<b>A05</b>	0,33333	1,00000	5,00000	0,22581	0,23810	0,38462	0,28284
<b>A06</b>	0,14286	0,20000	1,00000	0,09677	0,04762	0,07692	0,07377
<b>Total</b>	1,47619	4,20000	13,00000	1,00000	1,00000	1,00000	1,00000

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak melalui perhitungan seperti sebelumnya. Nilai Lamda didapatkan 3,09673, kemudian untuk Consistency Index mendapatkan nilai 0,04836 dan Consistency Ratio bernilai 0,08338. Dengan nilai Consistency Ratio adalah 0,08338 maka pembobotan pada setiap alternatif bernilai konsisten dikarenakan nilai  $CR < 0,1$ .

**Tabel 5.** Perhitungan Manual Nilai Bobot Alternatif 1 Terhadap Kriteria

	K01	K02	K03	Normalisasi			Eigen
<b>K01</b>	1,00000	0,33333	0,20000	0,11111	0,07692	0,13043	0,10616
<b>K02</b>	3,00000	1,00000	0,33333	0,33333	0,23077	0,21739	0,26050
<b>K03</b>	5,00000	3,00000	1,00000	0,55556	0,69231	0,65217	0,63335

<b>Total</b>	9,00000	4,33333	1,53333	1,00000	1,00000	1,00000	1,00000
--------------	---------	---------	---------	---------	---------	---------	---------

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak melalui perhitungan seperti sebelumnya. Nilai Lamda didapatkan 3,05536, kemudian untuk Consistency Index mendapatkan nilai 0,02768 dan Consistency Ratio bernilai 0,04772. Dengan nilai Consistency Ratio adalah 0,04772 maka pembobotan pada setiap kriteria bernilai konsiststen dikarenakan nilai CR < 0,1.

**Tabel 6.** Perhitungan Manual Nilai Bobot Alternatif 2 Terhadap Kriteria

	<b>K01</b>	<b>K02</b>	<b>K03</b>	<b>Normalisasi</b>			<b>Eigen</b>
<b>K01</b>	1,00000	3,00000	0,33333	0,23077	0,33333	0,21739	0,26050
<b>K02</b>	0,33333	1,00000	0,20000	0,07692	0,11111	0,13043	0,10616
<b>K03</b>	3,00000	5,00000	1,00000	0,69231	0,55556	0,65217	0,63335
<b>Total</b>	4,33333	9,00000	1,53333	1,00000	1,00000	1,00000	1,00000

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak melalui perhitungan seperti sebelumnya. Nilai Lamda didapatkan 3,05536, kemudian untuk Consistency Index mendapatkan nilai 0,02768 dan Consistency Ratio bernilai 0,04772. Dengan nilai Consistency Ratio adalah 0,04772 maka pembobotan pada setiap kriteria bernilai konsiststen dikarenakan nilai CR < 0,1.

**Tabel 7.** Perhitungan Manual Nilai Bobot Alternatif 3 Terhadap Kriteria

	<b>K01</b>	<b>K02</b>	<b>K03</b>	<b>Normalisasi</b>			<b>Eigen</b>
<b>K01</b>	1,00000	0,33333	3,00000	0,23077	0,21739	0,33333	0,26050
<b>K02</b>	3,00000	1,00000	5,00000	0,69231	0,65217	0,55556	0,63335
<b>K03</b>	0,33333	0,20000	1,00000	0,07692	0,13043	0,11111	0,10616
<b>Total</b>	4,33333	1,53333	9,00000	1,00000	1,00000	1,00000	1,00000

Kemudian, dapat dihitung nilai dari lamda hingga mengetahui apakah pembobotan nilai dapat dikatakan konsisten atau tidak melalui perhitungan seperti sebelumnya. Nilai Lamda didapatkan 3,05536, kemudian untuk Consistency Index mendapatkan nilai 0,02768 dan Consistency Ratio bernilai 0,04772. Dengan nilai Consistency Ratio adalah 0,04772 maka pembobotan pada setiap kriteria bernilai konsiststen dikarenakan nilai CR < 0,1.

Selanjutnya dilakukan perhitungan supermatrix yang dimana terdiri dari Unweighted Supermatrix yang berisi penggabungan nilai eigen vector kriteria dan alternatif. Kemudian, ada pula Weighted Supermatrix yang dimana pada kasus ini hanya terdapat dua cluster yakni kriteria dan alternatif dan kedua cluster tersebut tidak diperbandingkan, sehingga matriks cluster secara default seperti tabel di bawah ini:

**Tabel 8.** Perbandingan Cluster

<i>Cluster</i>	<i>Alternatif</i>	<i>Kriteria</i>
<b>Alternatif</b>	0	1
<b>Kriteria</b>	1	0

**Tabel 9.** Hasil Unweighted dan Weighted Supermatrix

	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	K01	K02	K03
A01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,10616	0,26050	0,63335
A05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,26050	0,10616	0,63335
A06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,26050	0,63335	0,10616
A07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K01	0	0	0	0,07377	0,64339	0,28284	0	0	0	0	0	0	0	0	0	0	0	0	0
K02	0	0	0	0,60000	0,20000	0,20000	0	0	0	0	0	0	0	0	0	0	0	0	0
K03	0	0	0	0,64339	0,28284	0,07377	0	0	0	0	0	0	0	0	0	0	0	0	0

Untuk perhitungan Limit Supermatrix didapatkan melalui perhitungan dengan memangkatkan nilai pada Weighted Supermatrix secara terus menerus hingga mendapatkan nilai yang stabil. Perhitungan Unweighted Supermatrix, Weighted Supermatrix dan Limiting Supermatrix.

**Tabel 10.** Hasil Limiting Supermatrix

	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	K01	K02	K03
A01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,17867	0,26414	0,55718
A05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,17867	0,26414	0,55718
A06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,17867	0,26414	0,55718
A07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K01	0	0	0	0,53015	0,32538	0,14447	0	0	0	0	0	0	0	0	0	0	0	0	0
K02	0	0	0	0,53015	0,32538	0,14447	0	0	0	0	0	0	0	0	0	0	0	0	0
K03	0	0	0	0,53015	0,32538	0,14447	0	0	0	0	0	0	0	0	0	0	0	0	0

Terakhir, hasil rekomendasi dilihat pada nilai setiap baris Alternatif, yang dimana jika memiliki nilai tertinggi maka akan menjadi rekomendasi pertama untuk pusat perbelanjaan modern.

**Tabel 11.** Hasil Perangkingan Pusat Perbelanjaan Modern

<b>Peringkat</b>	<b>Kode Pusat Perbelanjaan Modern</b>	<b>Nama Pusat Perbelanjaan Modern</b>
1	A04	Park 23
2	A05	Lippo Mall Kuta

3	A06	Discovery Shopping Mall
---	-----	-------------------------

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan terdapat beberapa kesimpulan diantaranya:

1. Metode *Analytic Network Process* (ANP) digunakan dengan menyertai kriteria analisa yakni Ukuran Gerai, Harga Gerai dan Jumlah Pesaing Gerai
2. Sistem berhasil memberikan rekomendasi terkait lokasi gerai Pusat Perbelanjaan *Modern* Provinsi Bali yang tepat kepada user sesuai dengan kriteria yang digunakan dan alternatif yang dibandingkan.
3. Hasil merupakan nilai akhir dari perkalian baris dan kolom pada nilai Unweighted dan Weighted Supermatrix hingga menemukan nilai yang stabil (sama per barisnya). Kemudian, hasil tersebut akan dilihat sesuai alternatif yang memiliki nilai tertinggi.
4. Jumlah alternatif atau pusat perbelanjaan modern yang dibandingkan dengan perhitungan algoritma memiliki jumlah maksimal hingga 5 alternatif pada sistem. Hal tersebut bertujuan meminimalisir kesulitan user dalam menentukan nilai bobot Perbandingan. Sistem ini hanya berfokus pada rekomendasi pemilihan lokasi gerai di Pusat Perbelanjaan *Modern* Provinsi Bali.

#### Referensi

- [1] J. U. Hasanah, "Inclusive Shopping Mall", Elibrary Unikom, 2019. Retrieved from <https://elibrary.unikom.ac.id/id/eprint/1901/8/08%20BAB%20II%20DESKRIPSI%20PROYEK.pdf>
- [2] S. Ambarwati, "APPBI Targetkan Kunjungan Mall Pada 2023 Mencapai 100 Persen", Antara Kantor Berita Indonesia, 2023. Retrieved from <https://www.antaraneews.com/berita/3410256/appbi-targetkan-kunjungan-mal-pada-2023-capai-100-persen>
- [3] A. A., G. B. Prasetyo, & D. W. Wibowo, "Sistem Pendukung Keputusan Berbasis Ontologi Untuk Pemilihan Pemain Sesuai Role Pada Tim PUBG Mobile Menggunakan Metode ANP," Repository JTI Politeknik Negeri Malang, Skripsi, 2020. Retrieved from <http://repota.jti.polinema.ac.id/692/>
- [4] A. A. Larasati, A. S. F. Utami, & F. Prasetyo, "Sistem Pendukung Keputusan Dalam Pemilihan Belanja Online Marketplace Menggunakan Analytic Network Process (ANP)," *Informatics For Educators and Professionals*, vol. 4, no. 2, p. 133-142, 2020. doi: <https://doi.org/10.51211/itbi.v4i2.1310>
- [5] A. B. Nasution, "Penerapan Sistem Pendukung Keputusan Dalam Memilih Rumah Sakit Bersalin Dengan Metode ANP," *Jurnal Sistem Informasi Kaputama*, vol. 2, no. 1, p. 73-83, 2018. doi: <https://doi.org/10.1234/jsik.v2i1.85>
- [6] A. Febriyanti, A. L. N. Afifah, & H. Sharafina, "Penerapan Analytical Network Process Dalam Pemilihan Lokasi Cabang Arosah Fashion," *Seminar Nasional INAECO*, p. 256-263, 2019. Retrieved from <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/10708>
- [7] A. J. Olanta, M. E. Sianto, & I. Gunawan, "Perbandingan Metode ANP dan AHP Dalam Pemilihan Jasa Kurir Logistik Oleh Penjual Gadget Online," *Scientific Journal Widya Tenik*, vol. 18, no. 2, p. 96-101, 2019. doi: <https://doi.org/10.33508/wt.v18i2.2275>
- [8] B. Kurniawan, O. Wahyudi, Marini, & A. P. Windarto, "Analisa Metode ANP Pada Pemilihan Alat Cukur Rambut," *Kajian Ilmiah Informatika dan Komputer*, vol. 1, no. 6, p. 248-252, 2021. Retrieved from <http://djournals.com/klik/article/view/203>
- [9] C. Pramarta, J. G. Davis, & K. K. Y. Kuan, "Digital Preservation of Cultural Heritage: An Ontology-Based Approach," *Australian Conference on Information Systems*, 2017. Retrieved from [https://www.researchgate.net/publication/321527990\\_Digital\\_Preservation\\_of\\_Cultural\\_Heritage\\_An\\_Ontology-Based\\_Approach](https://www.researchgate.net/publication/321527990_Digital_Preservation_of_Cultural_Heritage_An_Ontology-Based_Approach)

- [10] J. E. Leal, "AHP-express: A simplified version of the analytical hierarchy process method", Pontifical Catholic University of Rio de Janeiro. 2019.
- [11] Juliana & H. Gunawan, "Sistem Pendukung Keputusan Perpanjangan Kontrak Kerja Dengan Menggunakan Metode Analytic Network Process (ANP) Berbasis Web," *Jurnal Information System and Data Science*, vol. 1, no. 2, p. 59-68, 2023. Retrieved from <https://ejournal.cip.or.id/index.php/Inseds/article/view/65>
- [12] M. Abdillah, Ilhamsyah, & R. Hidayati, "Penerapan Metode Analytic Network Process (ANP) Berbasis Android Sebagai Sistem Pendukung Keputusan Dalam Pemilihan Tempat Kos," *Jurnal Coding, Rekayasa Sistem Komputer Untan*, vol. 6, no. 3, p. 12-22, 2018. doi: <http://dx.doi.org/10.26418/coding.v6i3.27437>

*This page is intentionally left blank.*

# Ekstraksi Fitur Dengan *Convolutional Neural Network* Dan Rekomendasi *Fashion* Menggunakan Algoritma *K-Nearest Neighbours*

I Gede Teguh Permana<sup>a1</sup>, Ida Bagus Gede Dwidasmara<sup>a2</sup>,  
Made Agung Raharja<sup>b3</sup>, I Wayan Santiyasa<sup>b4</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Udayana, Kuta Selatan, Badung Bali, Indonesia

<sup>1</sup>teguhpermana096@gmail.com

<sup>2</sup>dwidasmara@unud.ac.id

<sup>3</sup>made.agung@unud.ac.id

<sup>4</sup>santiyasa@unud.ac.id

## Abstract

Pesatnya pertumbuhan industri *fashion* pada *platform e-commerce* sehingga *fashion* dapat diperoleh dengan mudah oleh berbagai segmentasi konsumen. Segmentasi konsumen dapat direpresentasikan disetiap *search* jenis *fashion* yang di inginkan, namun *search* jenis *fashion* pada *e-commerce* dilakukan dengan *search* berbasis kata kunci *string* sehingga segmentasi konsumen terhadap karakteristik *fashion* sulit dilakukan. *Fashion* merupakan *object* yang mudah dikenali secara *visual* sehingga *search* berbasis gambar sangat diperlukan pada *platform e-commerce* untuk memilih *fashion* berbasis segmentasi konsumen. Implementasi *search* berbasis gambar dapat dilakukan dengan rekomendasi *fashion* berbasis *content* dengan *k-nearest neighbour* (KNN) untuk melakukan pendekatan antara *feature fashion* terhadap *input image fashion* oleh konsumen dengan setiap *feature* data dilakukan ekstraksi *feature* kedalam *convolution layer* pada model *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG) dapat dievaluasi dengan *top-n accuracy* terhadap model ResNet, GoogLeNet, VGG, dan HOG dengan masing-masing performa model tersebut dibandingkan sehingga dapat diperoleh *accuracy* sebesar 93% pada GoogLeNet dengan KNN sebagai model terbaik dalam rekomendasi *fashion*. Adapun pendekatan antara *feature fashion* dilakukan berbasis hasil label dari proses *classification* ke dalam *convolution* dan *fully connected layer* pada *convolutional neural network* (CNN) dapat dievaluasi dengan *evaluation matrices* terhadap model ResNet, GoogLeNet, VGG dengan masing-masing performa model tersebut dibandingkan sehingga dapat diperoleh nilai *accuracy* sebesar 99%, *precision* sebesar 100%, *recall* 99%, *f1-score* 99% pada VGG sebagai model terbaik untuk identifikasi jenis *fashion*.

**Keywords:** *Fashion, Ekstraksi Feature, Sistem Rekomendasi, Arsitektur CNN, HOG, KNN, Evaluation Matrices, Top-n accuracy*

## 1. Introduction

Pesatnya pertumbuhan *fashion* industri pada *platform e-commerce* sehingga *fashion* dapat diperoleh dengan mudah oleh berbagai segmentasi konsumen. Hal tersebut dapat sebagai keuntungan bagi konsumen, konsumen diberikan berbagai macam pilihan *fashion* tetapi kerugian dapat ditimbulkan dari beragamnya segmentasi konsumen. Saat konsumen *search fashion* yang diinginkan, umumnya konsumen kesulitan dalam *search* kata kunci pencarian *fashion* yang diinginkan. Proses *search* kata kunci berbasis *string* dapat menyebabkan kebingungan pada konsumen, *search* berdasarkan *image* sangat dibutuhkan disebabkan *fashion* bersifat *visual* daripada tekstual sehingga konsumen dapat melakukan proses pencarian berbasis gambar *fashion* yang diinginkan. Adapun sebagai pemahaman segmentasi gaya *fashion* yang beragam diperlukan layanan identifikasi jenis *fashion*, ekstraksi fitur gambar *fashion* dan sistem rekomendasi *fashion* yang disesuaikan dalam mengatasi keterbatasan untuk pemahaman segmentasi gaya *fashion* [1].

Ekstraksi fitur dengan *convolutional neural network* (CNN) sebagai salah satu metode dalam *computer vision* mampu mengatasi kompleksitas fitur yang terdapat pada *image* untuk

mengekspresikan gambar dengan lebih detail, mempelajari fitur lebih spesifik dan mengatasi batasan yang dialami oleh ekstraksi fitur dengan metode konvensional [2]. Salah satu ekstraksi fitur dengan metode konvensional adalah *histogram oriented gradient* (HOG) yang hanya dapat didekomposisi menjadi beberapa sel dengan skala yang berbeda dan mengekstraksi titik fitur pada besaran *pixel* dalam sel data sehingga dapat memunculkan fenomena yang disebabkan oleh deteksi *histogram feature* yang berskala tunggal di bawah skala yang berbeda sehingga *histogram* yang dihasilkan dapat digabungkan untuk mewakili deskriptor *image* [3]. Pada penelitian [4] mengusulkan *image feature matching* berbasis *convolutional neural network* (CNN) dengan memperhatikan *patch* gambar, dalam pencocokan titik fitur gambar, memperoleh fitur dengan *convolutional layer* yang parameternya dicapai dengan proses *learning* sehingga dapat dengan kuat mengekspresikan beberapa fitur pada gambar. Pada hasil penelitian menunjukkan metode yang diusulkan memiliki evaluasi yang lebih baik dibandingkan *baseline model*.

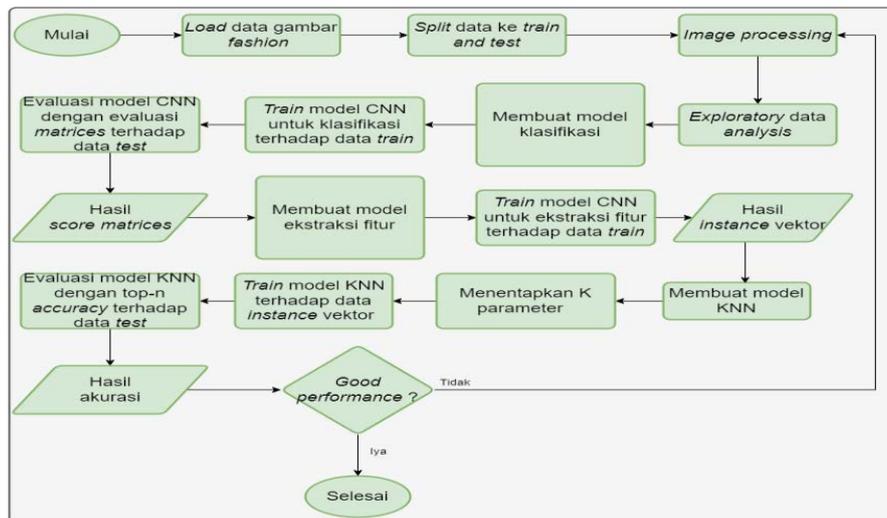
Identifikasi jenis *fashion* dengan *convolutional neural network* (CNN) sebagai salah satu metode campuran ekstraksi fitur dengan *convolution layer* dan *classification* sebagai penetapan jenis *fashion* dengan *fully connected layer*. *Convolutional neural network* (CNN) terdiri dari beragam perkembangan arsitektur untuk mengatasi semakin kompleksnya sebuah permasalahan saat ini [5]. Pada penelitian sebelumnya [6] melakukan perbandingan arsitektur *convolutional neural network* (CNN) dalam penetapan rekomendasi jenis *american sign language* (ASL) terhadap 1000 data per huruf, didapatkan bahwa performa arsitektur LeNet menghasilkan akurasi sebesar 92% dan pada arsitektur AlexNet menghasilkan akurasi sebesar 91% sehingga model arsitektur LeNet memiliki performa yang lebih baik daripada AlexNet. Sistem rekomendasi *Fashion* adalah sebuah sistem yang dapat membantu konsumen dalam *search fashion* yang diinginkan. Pada implementasi sistem rekomendasi terdiri atas beberapa jenis metode, salah satu jenisnya adalah *content based system* sebagai karakteristik informasi yang menyediakan data terhadap asosiasi produk *feature* [7]. Pada penelitian sebelumnya mengusulkan sistem rekomendasi hotel dengan *content based filtering* menggunakan *K-Nearest Neighbor* (KNN) dan *haversine formula*, *K-Nearest Neighbor* (KNN) dan *haversine formula* sebagai *association* setiap *instance* pada ruang vektor matriks untuk mencari kedekatan antara fitur [8].

Pada penelitian saat ini menggunakan metode *classification* dan ekstraksi *feature* dengan CNN untuk identifikasi jenis *fashion* dengan setiap jenis *image fashion* yang berkaitan diekstraksi menggunakan arsitektur CNN dengan *convolutional layer* sehingga diperoleh *instance* vektor dengan setiap *instance* vektor mewakili *image fashion*. Setiap *instance* vektor dilakukan proses rekomendasi berbasis *content based system* dengan *similarity measure* antar *feature instance vector* terhadap *initial vector*, sehingga didapatkan indeks hasil *similarity* dari berbagai *feature* gambar *fashion* sebagai rekomendasi *fashion* terhadap data *input* konsumen yang diproses ke dalam *initial* vektor. Diharapkan dengan identifikasi dan ekstraksi fitur gambar *fashion* dengan *convolutional neural network* (CNN) terhadap implementasi *fashion recommendation system* dapat berguna bagi pihak konsumen dan *platform e-commerce* dalam memberikan rekomendasi gambar *fashion* yang sesuai berdasarkan *input* gambar *fashion* konsumen terhadap *platform e-commerce*.

## 2. Research Methods

Pada tahap ini dapat dilihat pada Gambar 1. Mengenai alur secara umum dari penelitian yang akan dilakukan. Proses pertama ditunjukkan pada *load data fashion image* yang berasal dari situs web kaggle *fashion product images dataset* dengan setiap data yang didapatkan di lakukan proses *split* data menjadi data *train* dan *test*, 80% untuk data *train* dan 20% untuk data *test*. Pembuatan model *convolutional neural network* (CNN) kedalam tiga bentuk model yang terdiri atas: ResNet, VGG, GoogLeNet untuk proses ekstraksi fitur. Setiap model tersebut dilakukan proses *training* dengan data *train* untuk menjalankan proses ekstraksi fitur dengan menggunakan *convolutional layer* dari data gambar untuk menghasilkan sebuah ruang vektor, *output* yang diharapkan berupa satu *instance* terkandung beberapa bobot fitur yang sudah diekstrak di setiap satu gambar. Pembuatan model *convolutional neural network* (CNN) kedalam tiga bentuk model yang terdiri atas: ResNet, VGG, GoogLeNet untuk identifikasi jenis *fashion*. Setiap model tersebut dilakukan proses *training* dengan data *training* untuk menjalankan proses ekstraksi fitur pada *convolutional layer* terhadap data *image fashion* dan dilanjutkan melakukan proses *classification* pada *fully connected layer* terhadap sebelas label jenis *fashion*.

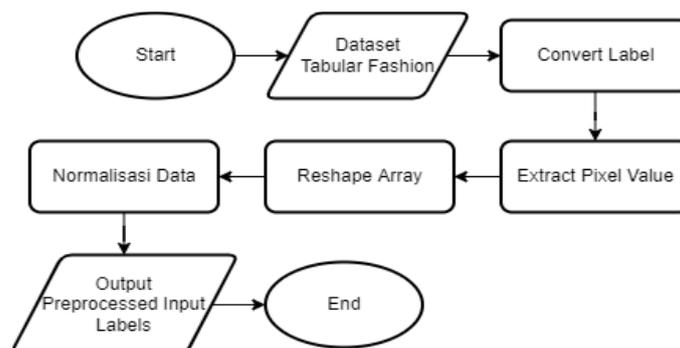
Proses selanjutnya setelah proses ekstraksi fitur dengan *convolutional neural network* (CNN) yakni membuat model *K-nearest neighbour* (KNN) yang diawali dengan memetakan parameter K sebagai jumlah *instance* data yang dipetakan ke dalam ruang vektor untuk didekatkan dengan *initial* data yang diinputkan *user* untuk melakukan proses rekomendasi *fashion* yang sesuai. Proses evaluasi dilakukan pada model *K-nearest neighbour* (KNN) dengan top-n *accuracy* yang dapat direpresentasikan terhadap hasil proses *association* pada data *test* untuk mengukur performa dari kinerja *K-nearest neighbours* sebagai algoritma sistem rekomendasi terhadap model *convolutional neural network* (CNN) untuk ekstraksi fitur *fashion image*. Proses evaluasi model yang dilakukan pada model *convolutional neural network* (CNN) dengan evaluasi *matrices* terhadap hasil proses *classification* pada data *test*, evaluasi *matrices* dapat direpresentasikan untuk mengukur performa model CNN terhadap *accuracy*, *precision*, *recall* dan *f1-score* dalam identifikasi jenis *fashion*.



Gambar 1. Desain Penelitian

### 2.1. Image Processing

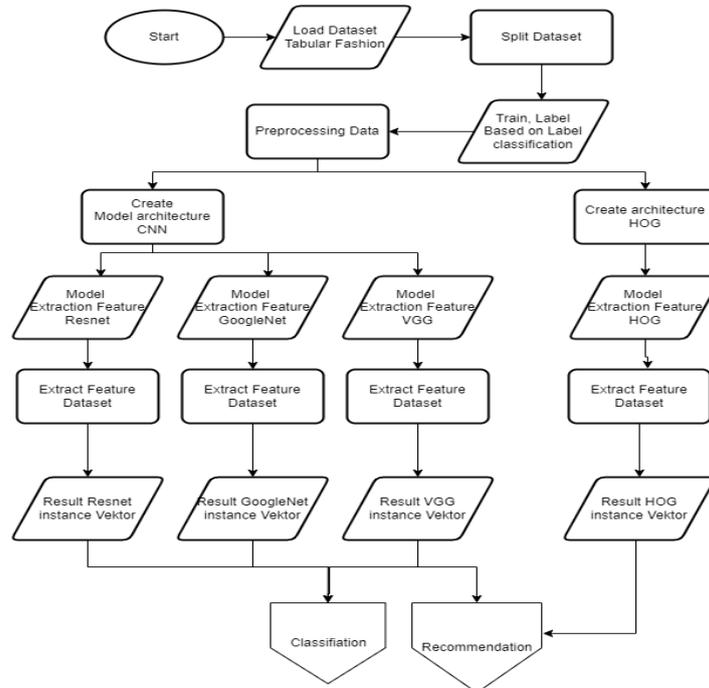
*Image preprocessing* data adalah sebuah metode yang digunakan untuk memproses sebuah data *image* sehingga model mudah untuk mengolah sebuah data sebagai bahan *input* dalam mengenali sebuah *pattern* yang dapat dilihat pada Gambar 2. Pengenalan sebuah *pattern* disesuaikan berdasarkan kegunaan pada model *output*. *Image preprocessing* data diawali dengan *load* dataset dalam bentuk tabular yang mana setiap label *feature* di *convert* dalam bentuk *categorical* data sebagai label *output*. Masing-masing nilai *pixel* di ekstrak dalam bentuk *preprocessed input* yang hanya mengambil nilai pada *feature pixel* dengan tidak mengambil *feature* awal sebagai label *feature*. Nilai setiap *preprocessed input* dilakukan proses *reshape* ukuran *array* sebesar 28 x 28 *pixel* dengan masing-masing nilai di normalisasi untuk membuat data lebih terpusat dan mengurangi penyebaran data, sehingga didapatkan sebuah *output* data berupa *preprocessed input* dan label data.



Gambar 2. Preprocessing Data

## 2.2. Ekstraksi *Feature*

Pada tahap ini. Data gambar yang diperoleh dapat diekstraksi sebelum memasuki model yang sesuai dengan Gambar 3. Tahap ini dilakukan untuk memperoleh data yang lebih sederhana dengan membuang bagian yang tidak terlalu penting namun bobot fitur yang penting tetap dipertahankan, pada tahapan ini menggunakan metode *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG), setiap dataset tabular *fashion* dalam bentuk matriks diekstraksi dengan setiap arsitektur model. Pada penggunaan arsitektur *convolutional neural network* (CNN) dilakukan proses *tuning* arsitektur model yang terbaik sehingga dataset *fashion* dilakukan proses transformasi menjadi *instance* vektor, setiap *instance* vektor yang bertujuan untuk merapatkan sebuah data yang tersebar ke titik pusat digunakan untuk mempermudah semua proses *training* dan pengenalan pola dari suatu data yang diimplementasikan kedalam rekomendasi *fashion* berdasarkan *content* dan klasifikasi sebagai identifikasi jenis *fashion*.



Gambar 3. *Schema Ekstraksi Feature*

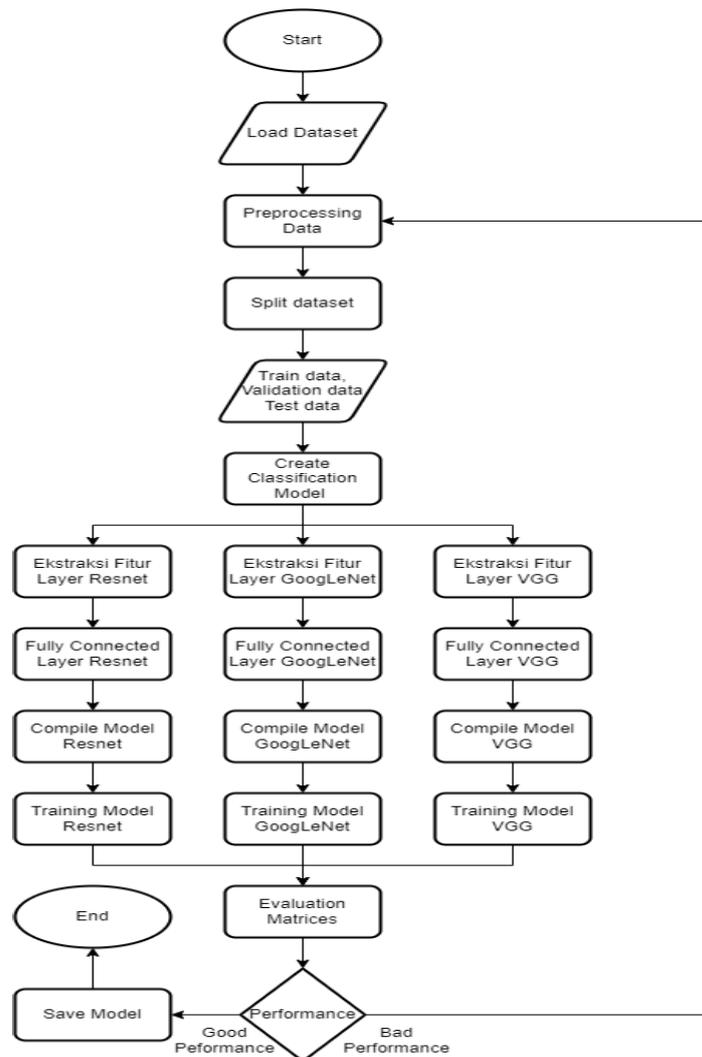
Adapun arsitektur dari *convolutional neural network* dapat disegmentasi sebagai ResNet (*residual neural network*), VGG (*Visual geometry group*), GoogLeNet. Penentuan *tuning* model arsitektur ditentukan berdasarkan nilai evaluasi yang diperoleh dari setiap model arsitektur tersebut. Setiap hasil ekstraksi fitur dengan arsitektur *convolutional neural network* sebagai ruang vektor yang terdiri satu *instance* data di setiap satu gambar yang mengandung setiap fitur gambar yang mengandung bobot fitur yang penting, jumlah bobot fitur yang dikandung setiap hasil ekstraksi dipengaruhi oleh struktur arsitektur *convolutional neural network* yang digunakan. Pada penggunaan arsitektur *histogram oriented gradient* (HOG) dilakukan proses ekstraksi fitur untuk menghasilkan data dalam bentuk *instance* vektor yang dapat digunakan untuk proses rekomendasi *fashion*. Arsitektur HOG sebagai metode konvensional dalam proses ekstraksi fitur yang dibandingkan dengan performa pada model arsitektur CNN untuk model ekstraksi fitur sebagai data masukan ke dalam model rekomendasi *fashion* dengan setiap fitur didekatkan berdasarkan *content*.

## 2.3. Klasifikasi

Pada tahap klasifikasi dilakukan untuk identifikasi jenis *fashion* yang dimasukkan kedalam sebuah model *convolutional neural network*, hal ini bertujuan untuk meningkatkan optimalisasi hasil rekomendasi yang mana pada proses klasifikasi dapat menghasilkan *output* berupa label *fashion* yang terdiri atas: *ankle boot, bag, coat, dress, hat, sandal, shirt, sneaker, trouser, t shirt, pullover*, sehingga pada proses *similarity* dengan *euclidean distance* hanya *image*

*fashion* yang sesuai dari label *output* proses klasifikasi. Tahap implementasi proses klasifikasi ditujukan pada Gambar 4. Dengan menggunakan tiga jenis arsitektur CNN, diantaranya: ResNet, GoogLeNet, dan VGG. Setiap model tersebut melakukan proses ekstraksi fitur pada *convolutional layer* dan proses klasifikasi pada *fully connected layer*. Setiap model arsitektur dilakukan proses *training* data terhadap data *training* dan data *validation* dengan jenis *optimizer* adalah *Adaptive Moment Estimation (Adam)*, *learning rate* sebesar 0.001, kalkulasi *loss function* dengan *categorical cross entropy*, *matrices calculation* dengan *accuracy* di setiap *epoch* terhadap data *validation* dan data *train*.

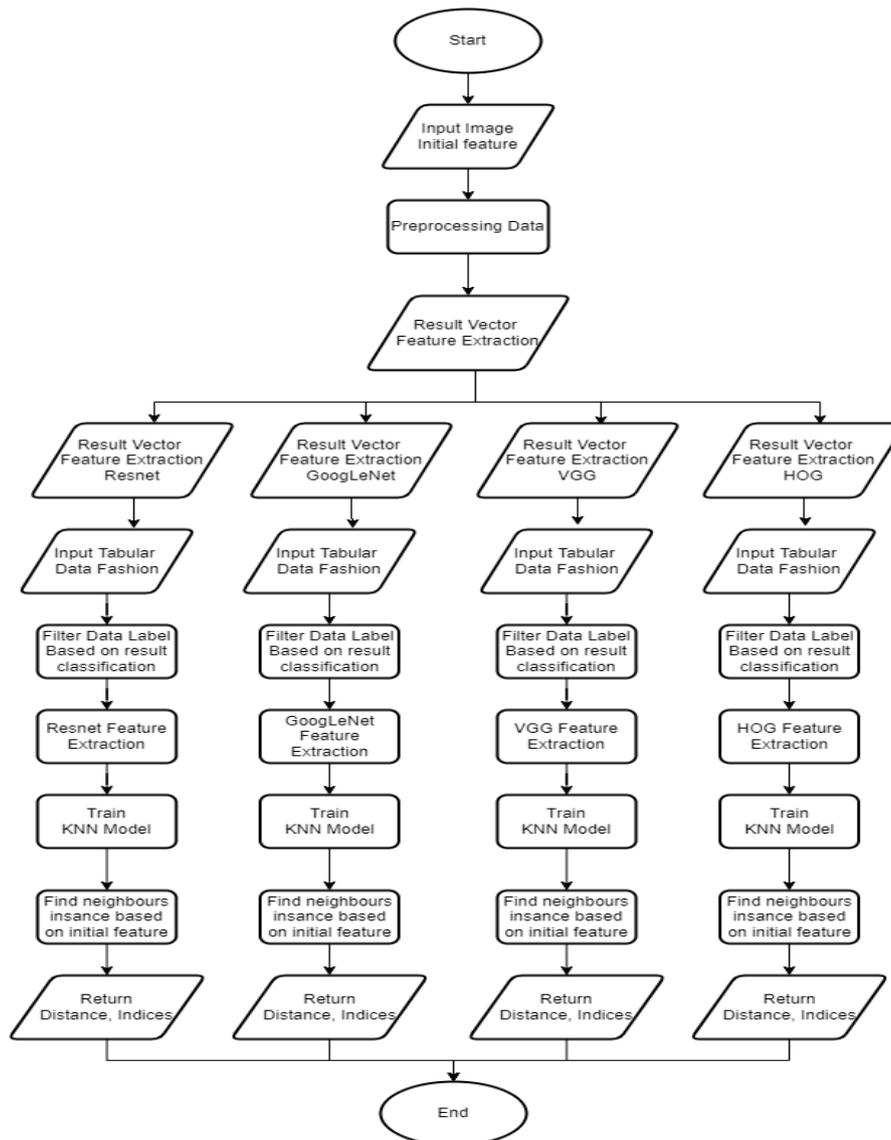
Pada masing-masing model di *training* di setiap 8 batch size data dengan menggunakan 20 *epoch* dengan hasil *training* di setiap *epoch* berupa nilai dari *loss function* dan *accuracy*. Hasil yang direpresentasikan pada data *training* dan data *testing* dapat sebagai rujukan untuk melihat kualitas model di setiap *epoch* selama proses *training* data. Pada proses klasifikasi didasari pada metode komputasi *backpropagation*, *backpropagation* sebagai metode komputasi *neural network* berdasarkan *forward pass* dan *backward pass*. Pada proses *forward pass* sebagai proses komputasi dari *input layer* untuk mendapatkan hasil klasifikasi di setiap label pada *output layer*, pada proses *backward pass* sebagai proses komputasi dari *output layer* untuk mendapatkan hasil optimasi setiap parameter proses *training* sehingga didapatkan proses parameter yang lebih baik di setiap *epoch* yang ditunjukkan dengan perhitungan *loss* yang menurun dan perhitungan *matrices accuracy* yang meningkat di setiap *epoch*.



Gambar 4. *Schema* Klasifikasi

#### 2.4. Sistem Rekomendasi

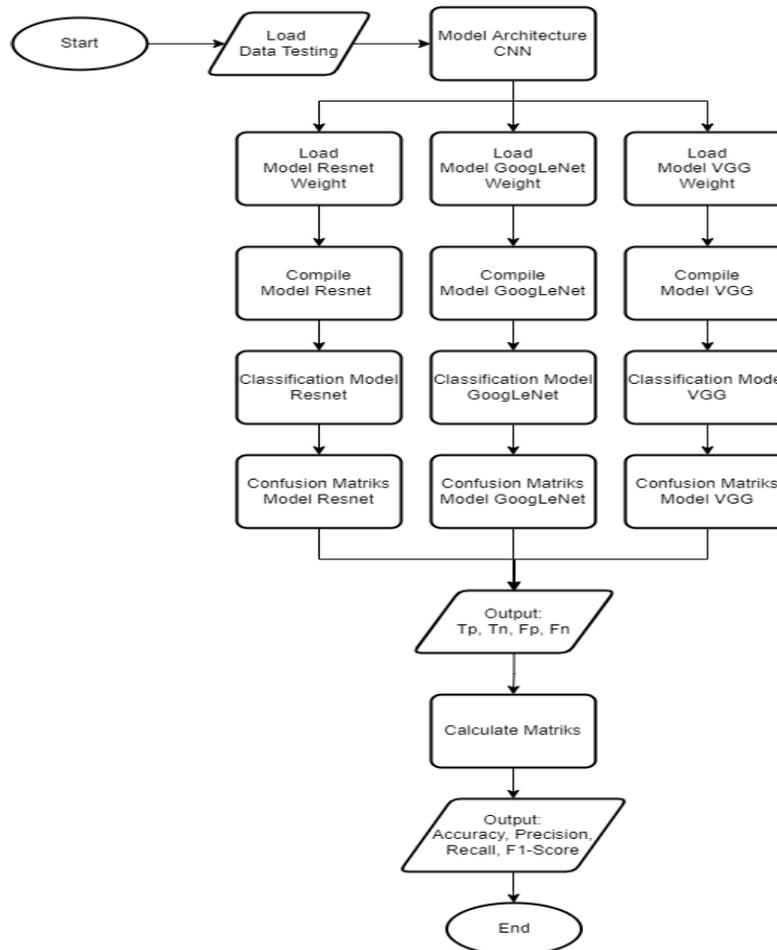
*K-Nearest Neighbors* digunakan sebagai algoritma untuk mengoperasikan sistem rekomendasi yang ditujukan pada Gambar 5. Pada pengoperasiannya, *K-Nearest Neighbors* membutuhkan inputan dari hasil pada proses ekstraksi fitur untuk diproses dengan sistem rekomendasi berdasarkan *content*, yang mana setiap *instance* dari ekstraksi fitur yang mewakili setiap data citra pada dataset akan dipetakan dalam bentuk ruang vektor yang terdiri beberapa *instance* data citra, setiap *initial* data dapat dikorelasikan kedekatannya berdasarkan *euclidean distance* yang berbasis kedekatan antara fitur sesuai dengan kesamaan fitur yang dimilikinya. Sistem rekomendasi bergantung pada hasil vektor dalam ekstraksi fitur dengan *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG). Masing-masing hasil vektor tersebut dikorelasikan terhadap *initial feature* dengan *K-Nearest Neighbors* berdasarkan kedekatan jarak setiap *feature* yang dikalkulasi dengan *euclidean distance*. Setiap hasil rekomendasi dapat *return* berupa *distance* dan *indices*, *distance* merepresentasikan jarak *instance* tersebut terhadap *initial feature*, sedangkan *indices* merepresentasikan identifikasi suatu data *image fashion*.

Gambar 5. *Schema Recommendation*

## 2.5. Evaluasi *Matrices*

Evaluasi *matrices* dilakukan untuk mengetahui sebuah performa model *convolutional neural network* (CNN) dalam melakukan *classification* untuk identifikasi jenis *fashion* yang ditujukan dengan Gambar 6. Evaluasi *matrices* dilakukan berdasarkan hasil setiap model

arsitektur *convolutional neural network* (CNN) yang terdiri atas ResNet, GoogLeNet, VGG. Setiap model tersebut di *training* dengan setiap hasil *training* disimpan kedalam *weight*, setiap model *weight* di load terhadap struktur masing-masing model CNN. Setiap model di *compile* untuk mengkonfigurasi parameter yang digunakan model yang sudah di *training* untuk melakukan proses *classification*. Hasil *classification* berupa label *fashion* yang terdiri sebelas jenis label *fashion* terhadap setiap data *image fashion*. Setiap label yang dihasilkan di kalkulasi dengan *confusion matriks* terhadap *true* label *fashion* sehingga menghasilkan *output* berupa; *true positif* (Tp), *true negatif* (Tn), *false positif* (Fp), *false negatif* (Fn). Setiap *output confusion matrices* dikalkulasi terhadap kalkulasi *matrices* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, *f1-score* yang merepresentasikan kualitas performa model *convolutional neural network* (CNN) untuk identifikasi jenis *fashion*.

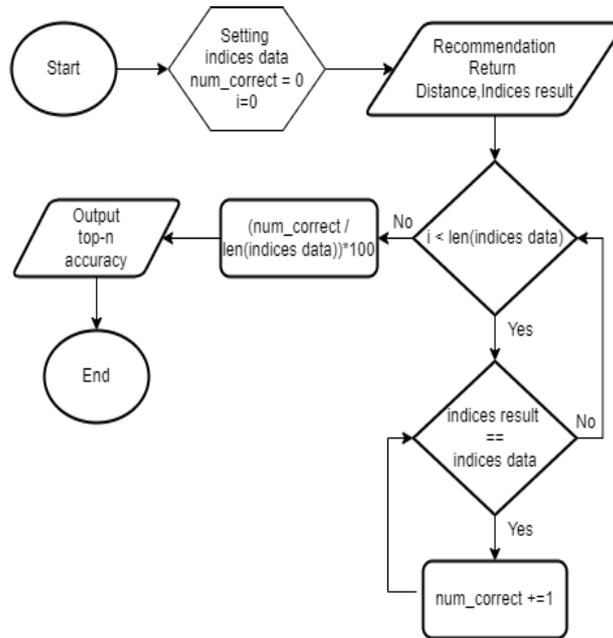


Gambar 6. Evaluasi *Matrices*

## 2.6. Top-n Accuracy

Evaluasi dengan top-n *accuracy* dilakukan untuk mengetahui bagaimana performa model yang telah dibuat sebelumnya yang ditujukan pada Gambar 7. Mengevaluasi model *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG) terhadap performanya dalam ekstraksi fitur untuk sistem rekomendasi dengan *k-nearest neighbours*. Penelitian ini menggunakan teknik evaluasi top-n *accuracy*. Top-n *accuracy* dapat mengevaluasi setiap n banyak data citra *fashion* yang serupa terhadap setiap proses rekomendasi dengan berapa banyak n yang serupa terhadap *initial image* sebagai *testing*, top-n *accuracy* direpresentasikan untuk mengetahui performa dalam suatu model yang dipengaruhi oleh *distance* dan *indices result*. Kalkulasi top-n *accuracy* di kalkulasi berdasarkan *neighbour* yang didekatkan pada model *k-Nearest Neighbors* sehingga setiap *indices* yang dihasilkan dapat dibandingkan terhadap *indices* data *fashion* dengan setiap hasil disimpan sebagai *num\_correct*. Pada *num\_correct* di *increment* di setiap proses perbandingan dengan *indices* data terhadap *indices result* sehingga

hasil yang diperoleh dapat dibagi dengan banyaknya *indices* data sebagai *accuracy*. Nilai *accuracy* tersebut merepresentasikan kemampuan model dalam melakukan rekomendasi berbasis *content*.



Gambar 7. Top-n Accuracy

### 3. Result and Discussion

#### 3.1. Hasil Ekstraksi Feature

Hasil ekstraksi feature dapat dilihat berdasarkan Tabel 1. Pada tabel tersebut merepresentasikan nilai *feature* yang dapat dihasilkan di setiap model ekstraksi *feature* dengan *convolutional neural network* (CNN) yang dapat disegmentasikan ke dalam model ResNet, GoogLeNet, VGG dan *histogram oriented gradient* (HOG) sebagai metode konvensional dalam ekstraksi *feature*. Pada model VGG menghasilkan 2304 *feature* sehingga *pixel* yang dapat diperoleh lebih kompleks dibandingkan model ResNet dan GoogLeNet yang hanya memperoleh 512 *feature* dari hasil proses komputasi ekstraksi *feature* di *convolution layer*. Pada model konvensional dalam ekstraksi *feature* dengan *histogram oriented gradient* (HOG) diperoleh 144 *feature* sehingga *pixel* yang dapat diperoleh lebih sederhana dibandingkan dengan menggunakan arsitektur model *convolutional neural network* (CNN).

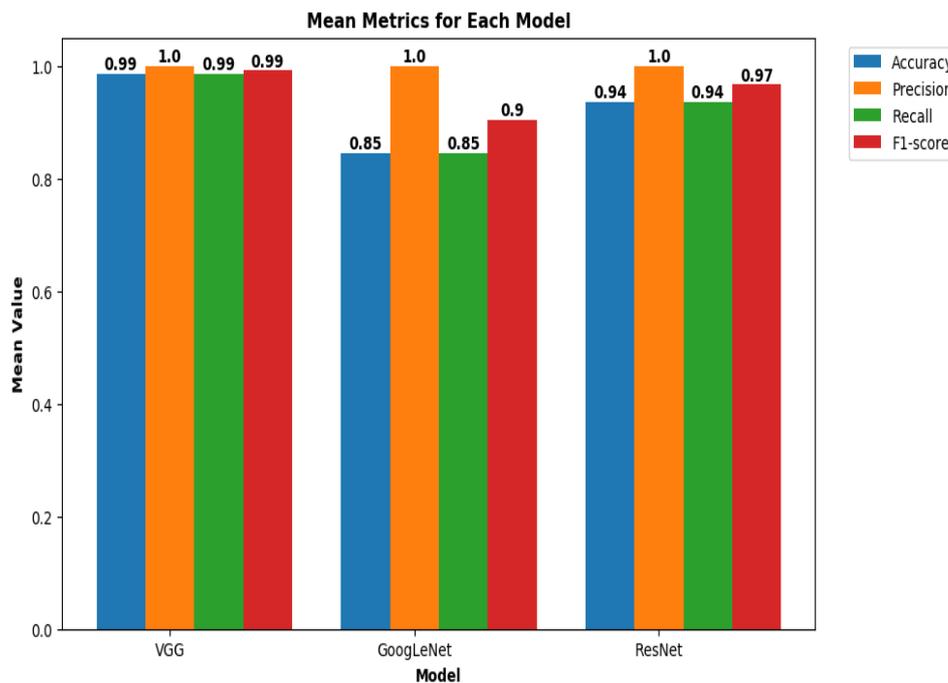
<i>Model</i>	<i>Feature</i>
ResNet	512
GoogLeNet	512
VGG	2304
HOG	144

Tabel 1. Hasil Ekstraksi *Feature*

#### 3.2. Hasil Evaluasi *Matrices Model Classification*

Hasil *evaluation matrices* model *classification* dengan arsitektur model *convolutional neural network* (CNN) dapat dilihat berdasarkan Gambar 8. Pada gambar tersebut didapatkan sebuah hasil berupa performa model sebagai identifikasi jenis *fashion* dengan menggunakan metode *classification*. Performa yang diperoleh dari *evaluation matrices* adalah *accuracy*, *precision*, *recall*, *f1-score*. Model VGG memperoleh performa model 99% pada *accuracy*, 100% pada *precision*, 99% pada *recall*, 99% pada *f1-score*. Model GoogLeNet memperoleh performa model 85% pada *accuracy*, 100% pada *precision*, 85% pada *recall*, 90% pada *f1-score*. Model ResNet memperoleh performa model 94% pada *accuracy*, 100% pada *precision*, 94% pada *recall*, 97% pada *f1-score*. Berdasarkan hasil yang diperoleh maka performa model VGG sebagai model terbaik untuk identifikasi jenis *image fashion* dibandingkan dengan model ResNet dan GoogLeNet terhadap data *testing*. Pada model GoogLeNet diperoleh hasil tidak cukup baik untuk identifikasi jenis *image fashion* dibandingkan dengan model ResNet yang memiliki performa *good fit*.

Setiap hasil dari masing-masing arsitektur model *convolutional neural network* (CNN) diperoleh model VGG sebagai model terbaik yang mana model VGG memiliki arsitektur dengan berfokus pada peningkatan dua kali nilai *neuron* pada tumpukan *layer* dalam *convolutional layer* dan terjadi pengurangan setengah nilai *neuron* pada tumpukan *layer* dalam *fully connected layer*. Pada model GoogLeNet sebagai model yang tidak cukup baik terhadap arsitektur *layer* yang berfokus pada penggunaan *inception layer* dalam *convolution layer* dan tanpa penggunaan *dense layer* dalam *fully connected layer*. Pada model Resnet sebagai model yang memiliki performa secara keseluruhan cukup baik terhadap arsitektur *layer* yang berfokus pada penggunaan metode *merge skip layer* dan *deep layer* berdasarkan *identity block* serta *convolutional block* dalam *convolutional layer* dan penggunaan 512 *neuron* pada penggunaan *dense layer* yang disesuaikan berdasarkan *output* dari *convolutional layer* dalam *fully connected layer*.

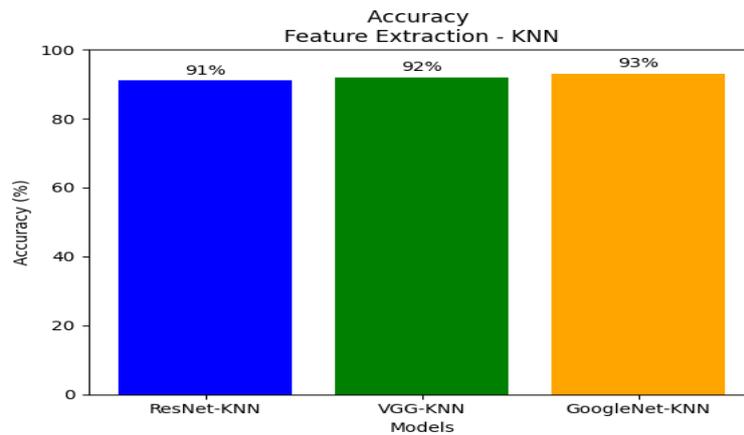


Gambar 8. Hasil *Evaluation Matrices*

### 3.3. Hasil Evaluasi Top-N Accuracy Sistem Rekomendasi

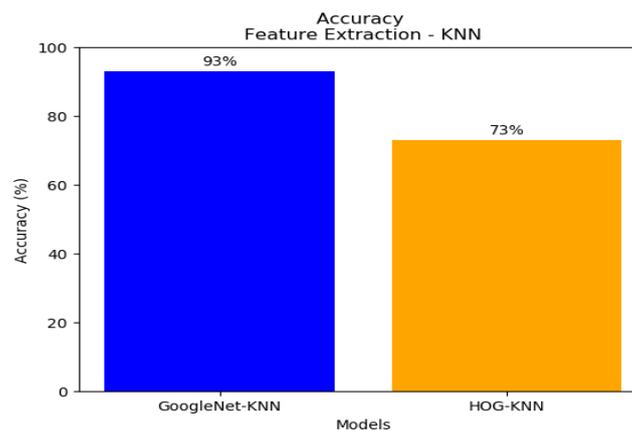
Hasil *evaluation top-n accuracy* pada sistem rekomendasi dapat dilihat pada Gambar 9. Pada hasil tersebut diperoleh performa model *k-nearest neighbor* (KNN) dalam mencari *distance* pada *instance* terdekat terhadap *initial* data yang di *input* user. Performa *k-nearest neighbor* sangat berpengaruh terhadap arsitektur model arsitektur *feature* dengan *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG). Pada performa sistem rekomendasi

dengan *k-nearest neighbor* terhadap ekstraksi *feature* dengan model diperoleh bahwa model GoogLeNet dengan KNN memiliki performa yang terbaik dalam rekomendasi *fashion* yang sesuai terhadap *user* yang memiliki akurasi sebesar 93%, namun performa hasil yang diperoleh berdasarkan hasil perhitungan *evaluation top-n accuracy* tidak terlalu berbeda signifikan terhadap model ResNet dengan KNN yang memiliki akurasi sebesar 91% dan model VGG dengan KNN yang memiliki akurasi sebesar 92%.



Gambar 9. Hasil *Evaluation* Sistem Rekomendasi

Hasil yang terdapat pada Gambar 10. Hasil *evaluation* sistem rekomendasi memperoleh sebuah performa model yang tidak terlalu berbeda signifikan satu sama lain, maka dari itu dilakukan perbandingan performa model konvensional untuk ekstraksi *feature* dengan HOG dengan KNN untuk sistem rekomendasi terhadap model terbaik CNN dengan KNN. Perbandingan CNN terhadap metode konvensional, diperoleh bahwa model CNN dengan KNN memiliki performa yang lebih baik sebesar 93% dibandingkan HOG dengan KNN yang hanya sebesar 73%.

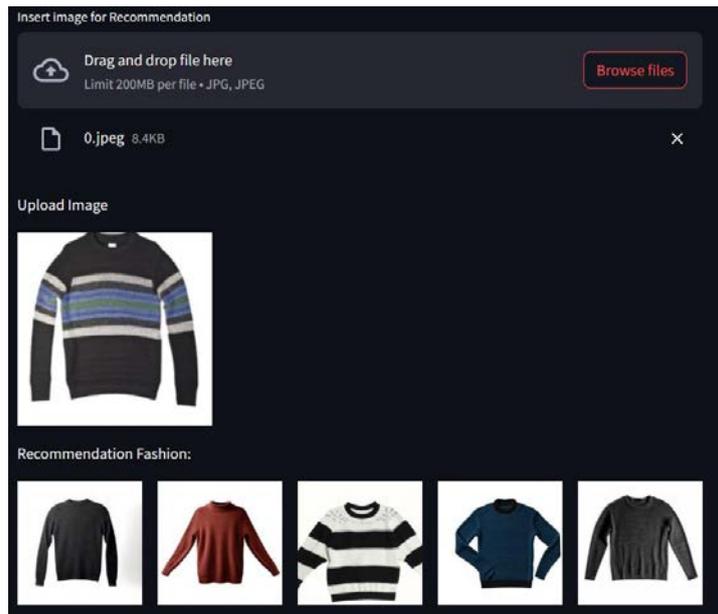


Gambar 10. Hasil *Evaluation* Sistem Rekomendasi

### 3.4. *Prototype* Sistem Rekomendasi

Model terbaik yang diperoleh dari hasil ekstraksi *feature*, identifikasi jenis *fashion*, dan rekomendasi *fashion* dapat digunakan untuk *prototype*. *Prototype* sistem rekomendasi di *deployment* dengan menggunakan *framework streamlit*, adapun tampilan dari *prototype* sistem rekomendasi dapat dilihat pada Gambar 11. Proses penggunaan *prototype* diawali dengan *input* data *image fashion* oleh *user* sebagai *initial* data yang di ekstrak dan di klasifikasi dengan menggunakan arsitektur model VGG untuk mengetahui identifikasi jenis *fashion* berdasarkan *initial* data sehingga menghasilkan label jenis *fashion*. Ekstraksi *feature* berdasarkan label jenis *fashion* di ekstrak sebagai *instance* data dengan model GoogLeNet yang di dekatkan setiap *feature* berdekatan terhadap *initial* data berbasiskan perhitungan *euclidean distance* pada KNN

sehingga *user* memperoleh rekomendasi hasil gambar yang sesuai berdasarkan *input image fashion*.



Gambar 11. Hasil *Prototype* Sistem Rekomendasi

#### 4. Conclusion

Berdasarkan paparan penelitian yang telah dilakukan sebelumnya, maka simpulan dalam penelitian ini diantaranya:

1. Perbandingan Performa model arsitektur *convolutional neural network* (CNN) dalam ekstraksi *feature* dan *classification* untuk identifikasi jenis *fashion* memperoleh hasil bahwa *visual geometry group* (VGG) sebagai model yang memiliki performa terbaik berdasarkan kalkulasi *evaluation matrices* terhadap nilai *accuracy*, *precision*, *recall*, *f1-score*, memperoleh hasil sebesar 99% pada *accuracy*, 100% pada *precision*, 99% pada *recall*, 99% pada *f1-score* dibandingkan terhadap performa model ResNet dengan hasil sebesar 94% pada *accuracy*, 100% pada *precision*, 94% pada *recall*, 97% pada *f1-score*, dan dibandingkan terhadap performa model GoogLeNet dengan hasil sebesar 85% pada *accuracy*, 100% pada *precision*, 85% pada *recall*, 90% pada *f1-score*.
2. Performa sistem rekomendasi *fashion* yang menggunakan *K-Nearest Neighbor* (KNN) terhadap arsitektur *convolutional neural network* (CNN) dan *histogram oriented gradient* (HOG) untuk melakukan ekstraksi fitur diperoleh hasil bahwa kombinasi antara model GoogLeNet dalam ekstraksi *feature* dengan *K-Nearest Neighbor* (KNN) dalam rekomendasi *fashion* sebagai model terbaik berdasarkan hasil perhitungan top-n *accuracy* yang memperoleh hasil sebesar 93%, dibandingkan dengan kombinasi antara model VGG dan *K-Nearest Neighbor* (KNN) memperoleh *accuracy* sebesar 92%, serta dibandingkan dengan kombinasi antara model ResNet dan *K-Nearest Neighbor* (KNN) memperoleh *accuracy* sebesar 91%. Jika di bandingkan antara arsitektur model *convolutional neural network* (CNN) diperoleh hasil *accuracy* yang tidak terlalu berbeda secara signifikan, namun jika dibandingkan dengan metode ekstraksi *feature* secara konvensional dengan menggunakan *histogram oriented gradient* (HOG) terjadi perbedaan signifikan dengan *accuracy* sebesar 73%.

## References

- [1] Jo, J. , Lee, S. , Lee, C. , Lee, D. and Lim, H. , 2020. Development of *fashion* product retrieval and *recommendations* model based on *deep learning*. *Electronics*, 9(3), p. 508
- [2] DeTone, D. , Malisiewicz, T. and Rabinovich, A. , 2018. *Superpoint: Self-supervised interest point detection and description*. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 224-236).
- [3] Tanjung, J. P. & Muhathir, 2020. *Classification* of facial expressions using SVM and HOG. *JITE (Journal of Informatics and Telecommunication Engineering)*, 3(2), pp. 210-215.
- [4] Y. Liu, X. Xu and F. Li, "*Image* Feature Matching Based on *Deep learning*," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 2018, pp. 1752-1756, doi: 10. 1109/CompComm. 2018. 8780936.
- [5] Jumaryadi, Y., Ihsan, A.M. and Priambodo, B., 2023. Klasifikasi Jenis Buah-Buahan Menggunakan Citra Digital Dengan Metode *Convolutional neural networks*. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(3), pp.1737-1746.
- [6] Al Rivani, M.E. and Riyadi, A.G., 2021. Perbandingan arsitektur lenet dan alexnet pada metode *convolutional neural network* untuk pengenalan american sign language. *Jurnal Komputer Terapan*, 7(1), pp.53-61.
- [7] P. Kumar and R. S. Thakur, "*Recommendation* sistem *techniques and related issues: a survey*," *Int. j. inf. technol.* , vol. 10, no. 4, pp. 495–501, 2018, doi: 10. 1007/s41870-018-0138-8.
- [8] Muliawan, A. , Badriyah, T. and Syarif, I. , 2022. Membangun Sistem Rekomendasi Hotel dengan *Content Based Filtering* Menggunakan *K-Nearest Neighbor* dan *Haversine Formula*. *Technomedia Journal*, 7(2 October), pp. 231-247.

# Penerapan SVM dengan Seleksi Fitur Mutual Information untuk Memprediksi Sentimen PEMILU 2024

I Gusti Bgs Darmika Putra<sup>a1</sup>, Cokorda Pramatha<sup>a2</sup>,  
Anak Agung Istri Ngurah Eka Karyawati<sup>a3</sup>, Made Agung Raharja<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, Universitas Udayana

<sup>b</sup>Center for Interdisciplinary Research on the Humanities and Social Sciences, Universitas Udayana  
Bali, Indonesia

<sup>1</sup>gstbgdarmika@gmail.com

<sup>2</sup>cokorda@unud.ac.id

<sup>3</sup>eka.karyawati@unud.ac.id

<sup>4</sup>made.agung@unud.ac.id

## Abstract

A wealth of information on the 2024 Indonesian Election floods Twitter, from campaign schedules to candidate profiles and the latest survey results indicating candidate popularity. This information overload poses challenges in discerning comments' sentiment. Manual classification is feasible but time-consuming. Hence, this study aims to streamline data analysis for the 2024 Election. It employs a dataset of 1000 entries categorized as positive or negative. Support Vector Machine (SVM) with Mutual Information feature selection is utilized for classification. Results reveal that Mutual Information feature selection enhances SVM performance. Without it, SVM achieves 88% accuracy and 87.9% f-measure using the rbf kernel ( $C=1$ ,  $\gamma=2$ ), computed in about 0.07 seconds. With feature selection, SVM's accuracy improves to 90%, and f-measure to 89.9% with 60% features, using rbf kernel ( $C=10$ ,  $\gamma=0.5$ ), reducing computation time to 0.02 seconds, optimizing both performance and efficiency. The website system scored 88.63 in usability, higher than the global average of 68, based on a SUS questionnaire with 10 questions and 20 respondents. This indicates excellent performance and user satisfaction, as evaluated from the web system.

**Keywords:** Sentiment Analysis, Classification, Support Vector Machine, Mutual Information, Indonesian Election 2024

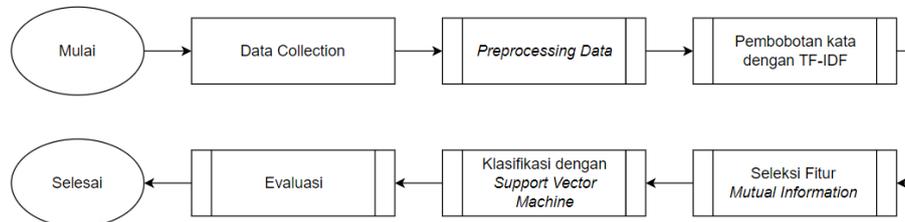
## 1. Pendahuluan

Pemilihan umum merupakan suatu proses dalam sistem demokrasi yang melibatkan partisipasi aktif masyarakat dalam menentukan wakil-wakil para perwakilannya di lembaga-lembaga pemerintahan [1]. Banyak informasi mengenai Pemilu 2024 tersebar di *Twitter*, mulai dari jadwal kampanye, profil pasangan calon, hingga hasil survei terbaru yang menunjukkan seberapa populer masing-masing kandidat [2]. Hal ini menyulitkan untuk menentukan apakah suatu komentar bersifat positif atau negatif, meskipun dapat dilakukan secara manual, namun memerlukan waktu dan usaha yang lebih. Analisis sentimen hadir sebagai solusi yang mempermudah dalam memahami reaksi orang terhadap suatu topik, dan dengan demikian dapat mendukung pengambilan keputusan yang lebih tepat. Dalam analisis sentimen, data biasanya dikelompokkan menjadi positif, negatif, atau netral untuk memudahkan pemahaman mengenai opini atau tanggapan yang terkandung dalam teks [3]. Sebagai contoh, dalam penelitian mengenai analisis sentimen tweet terkait program kartu prakerja [4]. Hasil eksperimen tersebut menunjukkan bahwa SVM dapat digunakan secara efektif untuk mengklasifikasikan tweet dengan Tingkat akurasi mencapai 98.67%, *precision* 98%, *recall* 99%, dan *f-measure* 98%. Terdapat juga penelitian yang melakukan analisis sentimen terhadap ulasan produk berdasarkan aspek-aspeknya, menggunakan pemilihan fitur *Mutual Information* (MI) dengan pendekatan *K-Nearest Neighbor* (KNN) oleh [5]. Berdasarkan hasil eksperimen, ditemukan bahwa penggunaan MI memberikan pengaruh yang relatif baik terhadap peningkatan akurasi pada aspek harga 39,77%, aspek kemasan 7,31%, dan 9,18% pada aspek aroma. Berdasarkan uraian dan referensi penelitian sebelumnya, penulis akan menganalisis sentimen pada data *tweet* terkait pemilu 2024 menggunakan metode SVM dengan seleksi fitur *Mutual Information* (MI). Penulis memilih kasus pemilu dalam

penelitian ini karena pemilu merupakan proses demokrasi yang penting, mencerminkan partisipasi masyarakat dalam menentukan pemimpin dan kebijakan publik. Dengan menggunakan kombinasi ini, diharapkan penelitian ini bisa menjadi panduan bagi perbandingan pilihan metode pendekatan dalam menganalisis sentimen terhadap opini di *Twitter*.

## 2. Metode Penelitian

Pada bagian metodologi penelitian, penulis menerapkan serangkaian pendekatan atau metode untuk melakukan analisis sentimen. Langkah-langkah yang diambil oleh penulis dalam penyusunan penelitian ini mencakup pencarian dan pengumpulan data, preprocessing, pelabelan data, pembobotan data dan klasifikasi metode *Support Vector Machine* (SVM) dengan seleksi fitur *Mutual Information* (MI). Alur penelitian dapat dilihat pada Gambar 1 berikut:



Gambar 1. Diagram Alir Penelitian

### 2.1. Pengumpulan Data

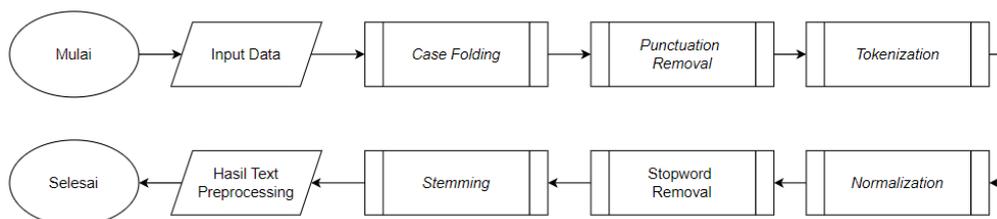
Penelitian ini menggunakan data sekunder dari media sosial *Twitter* berbahasa Indonesia dengan kata kunci "pemilu 2024". Total data yang digunakan adalah 1000 *tweet* yang terbagi menjadi dua kelas, yaitu positif dan negatif, dengan masing-masing 500 *tweet*. Pengumpulan data dilakukan mulai bulan Mei hingga Desember 2024. Proses pelabelan data dilakukan secara manual dengan melibatkan ahli bahasa Indonesia dari Fakultas Ilmu Budaya Universitas Udayana. Contoh data untuk setiap kelas dapat dilihat pada Tabel 1 berikut:

Tabel 1. Contoh Data Tweet

Tweet	Label
Pemilu 2024 adalah kesempatan bagi kita semua untuk membuktikan bahwa demokrasi Indonesia bisa menjadi contoh bagi dunia @hidupdamaiiiii Ganjar Mahfud Menang <a href="https://t.co/QOKmnZBMQX">https://t.co/QOKmnZBMQX</a>	Positif
#TolakKetuaKPUCurang ... tolak ketua KPU CURANG...semoga cepat mati sebelum pemilu 2024.	Negatif

### 2.2. Text Preprocessing

*Text preprocessing* adalah tahap awal dalam proses klasifikasi teks yang bertujuan untuk mengubah data teks asli yang tidak terstruktur menjadi data yang terstruktur, serta memastikan data yang berkualitas atau inputan yang baik agar dapat dilanjutkan ke proses selanjutnya dengan efisien [6]. Pada penelitian ini, dilakukan beberapa tahap *text preprocessing* yang dijelaskan pada Gambar 2:



Gambar 2. Tahapan *Text Preprocessing*

Tahapan pertama dalam *text preprocessing* adalah *case folding*, yaitu mengubah semua karakter menjadi huruf kecil. Dilanjutkan dengan *punctuation removal* untuk penghapusan tanda baca yang diterapkan dalam suatu teks yang bertujuan untuk membedakan antara kalimat dan bagian penyusunnya, serta untuk menjelaskan maknanya dengan lebih jelas. Kemudian, teks dibagi menjadi token dengan proses *tokenization*, diikuti oleh *normalization* untuk mengonversi variasi bentuk kata menjadi bentuk standar. Setelah itu, langkah selanjutnya adalah *stopword removal* untuk

menghilangkan kata-kata umum yang kurang relevan. Terakhir, tahap *stemming* dilakukan proses mengubah kata menjadi bentuk dasarnya dengan menghapus akhiran kata. Ini membantu mengurangi variasi dasar antara kata yang serupa [7].

### 2.3. Term Frequency – Inverse Document Frequency

Pembobotan kata merupakan suatu proses yang bertujuan untuk mengurangi dampak dari kata-kata yang umum atau tidak relevan dalam sebuah dokumen, sementara pada saat yang sama memberikan bobot yang lebih besar untuk kata-kata yang dianggap penting atau memiliki makna yang lebih spesifik. Salah satu algoritma yang umum digunakan adalah TF-IDF [8]. Adapun langkah perhitungannya sebagai berikut:

- a. Menghitung *term b frequency*

$$TF_{(t,d)} = \frac{\text{jumlah kata } t \text{ dalam dokumen } d}{\text{jumlah kata total dalam dokumen } d} \quad (1)$$

- b. Menghitung *inverse document b frequency*

$$IDF_{(t)} = \log \frac{N}{DF_{(t)}} \quad (2)$$

- c. Menghitung bobot TF-IDF

$$TF - IDF_{(t,d)} = TF(t,d) \times IDF(t) \quad (3)$$

Keterangan:

d = dokumen yang sedang dihitung.

t = Kata yang diukur frekuensinya.

N = Jumlah total dokumen.

IDF(t) = frekuensi dokumen dari term t.

TF(t,d) = Nilai *Term Frequency* kata (t) dalam dokumen (d).

IDF(t) = Nilai IDF kata (t) dalam kumpulan dokumen.

### 2.4. Mutual Information

*Mutual Information* dalam analisis sentimen membantu dalam memahami keterkaitan antara fitur-fitur yang muncul dalam teks dengan sentimen yang diungkapkan [9]. Hal ini memungkinkan untuk membangun model analisis sentimen yang lebih akurat dan dapat memberikan wawasan yang lebih baik tentang sentimen yang terkandung dalam teks. Untuk mencari nilai *Mutual Information* pada setiap kelas kata dalam kategori, menggunakan rumus berikut:

$$I(U, C) = \sum_{et \in \{1,0\}, ec \in \{1,0\}} P(U = et, C = ec) \log_2 \frac{P(U = et, C = ec)}{P(U = et) P(C = ec)} \quad (4)$$

Persamaan (4) dapat dijabarkan kembali menjadi persamaan (5). Berikut merupakan persamaan dari *Mutual Information* yang dijabarkan kembali.

$$I(U, C) = \frac{N_{11}}{N} \log_2 \frac{N \cdot N_{11}}{N_1 \cdot N_1} + \frac{N_{01}}{N} \log_2 \frac{N \cdot N_{01}}{N_0 \cdot N_1} + \frac{N_{10}}{N} \log_2 \frac{N \cdot N_{10}}{N_1 \cdot N_0} + \frac{N_{00}}{N} \log_2 \frac{N \cdot N_{00}}{N_0 \cdot N_0} \quad (5)$$

Keterangan:

N = Total dokumen yang memiliki et dan ec.

N<sub>1</sub> = Total dokumen yang memiliki et.

N<sub>0</sub> = Total dokumen yang memiliki ec.

N<sub>0</sub> = Total dokumen yang memiliki tidak memiliki et.

N<sub>0</sub> = Total dokumen yang memiliki tidak memiliki et.

### 2.5. Support Vector Machine

*Support Vector Machine* (SVM) adalah algoritma klasifikasi data yang termasuk dalam kategori *supervised learning*. Dalam penggunaannya, SVM memerlukan data yang telah dilabeli sebelumnya sebagai masukan. SVM bertujuan memisahkan data ke dalam kategori berbeda dengan menemukan

garis pemisah optimal [10]. SVM menggunakan konsep kernel dalam dimensi tinggi untuk memecahkan masalah *non-linear*. Konsep ini memungkinkan SVM untuk menemukan hubungan yang kompleks antara fitur dalam data, sehingga meningkatkan kemampuan SVM dalam menangani berbagai jenis data yang tidak dapat dipisahkan secara *linear* [11]. Formulasi optimasi SVM dibagi menjadi klasifikasi linear dan non-linear. Berikut merupakan rumus SVM:

- a. Menghitung garis *hyperplane* menggunakan rumus persamaan

$$w_i \cdot x + b = 0 \quad (6)$$

- b. Jika data termasuk kelas positif menggunakan rumus persamaan

$$w_i \cdot x + b > 0 \quad (7)$$

- c. Jika data termasuk kelas negatif menggunakan rumus persamaan

$$w_i \cdot x + b < 0 \quad (8)$$

- d. Dengan adanya dua garis pemisah tersebut maka dapat menghasilkan persamaan

$$y_i((\vec{x}_i) \cdot \vec{w} + b - 1 \geq 0) \quad (9)$$

Keterangan:

$w$  = Vektor bobot (*weight vector*) yang merupakan vektor normal terhadap garis *hyperplane*.

$x$  = Vektor fitur *input* yang merupakan data yang ingin diklasifikasikan.

$b$  = Bias atau pergeseran *hyperplane*.

$y_i$  = Label kelas dari data latih.

$\vec{x}_i$  = Vektor fitur *input* dari data latih.

$\vec{w}$  = Vektor bobot yang digunakan untuk memisahkan dua kelas data.

## 2.6. System Usability Scale

SUS merupakan alat pengukuran standar yang digunakan untuk mengevaluasi tingkat kegunaan (*usability*) suatu sistem [12]. Salah satu keunggulan SUS adalah kemampuannya untuk memberikan hasil yang cukup signifikan dengan jumlah sampel yang relatif sedikit, sehingga dapat mengurangi biaya pengujian *usability* secara keseluruhan. Berikut merupakan rumus *System Usability Scale* (SUS):

$$\text{Skor SUS} = ((P1 - 1) + (5 - P2) + (P3 - 1) + (5 - P4) + (P5 - 1) + (5 - P6) + (P7 - 1) + (5 - P8) + (P9 - 1) + (5 - P10)) * 2.5 \quad (10)$$

## 2.7. Uji Validitas

Uji validitas adalah langkah untuk memeriksa seberapa baik alat yang digunakan dalam penelitian bisa mengukur variabel yang seharusnya diukur dengan benar dan tepat [13]. Dengan kata lain, validitas membantu memastikan bahwa alat tersebut benar-benar mengukur apa yang seharusnya diukur dengan akurat. Uji validitas melibatkan dua metode utama: *Expert Review* dan uji *Product-Moment Coefficient*. *Expert Review* adalah evaluasi oleh ahli dalam bidang terkait, yang diminta untuk menilai apakah instrumen yang telah dibuat sudah valid atau belum [13].

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{(n(\sum x_i^2) - (\sum x_i)^2)(n(\sum y_i^2) - (\sum y_i)^2)}} \quad (11)$$

Keterangan:

$r$  = koefisien korelasi.

$n$  = jumlah responden.

$x_i$  = skor setiap item pada instrument.

$y_i$  = skor setiap item pada kriteria.

Setelah menemukan koefisien korelasi, langkah selanjutnya untuk mengetahui validitas dapat melakukannya dengan membandingkan nilai  $t_{hitung}$  dengan nilai  $t_{tabel}$ . Nilai  $t_{tabel}$  didapatkan sesuai dengan tingkat signifikansi penelitian yang telah ditentukan sebelumnya, dalam kasus ini adalah 5% atau 0.05. Sebuah butir dianggap valid jika nilai  $t_{hitung}$  lebih besar dari nilai  $t_{tabel}$ . Sedangkan nilai  $t_{hitung}$  didapatkan dengan menggunakan rumus [13].

$$r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (12)$$

Keterangan:

r = koefisien korelasi

n = jumlah responden

t = nilai  $t_{hitung}$

## 2.8. Uji Reliabilitas

Uji reliabilitas adalah proses menilai seberapa konsisten suatu metode atau instrumen pengukuran dalam menghasilkan hasil yang serupa dalam situasi yang berbeda [13]. Dalam konteks kuesioner, uji reliabilitas penting untuk memastikan konsistensi pertanyaan dalam mengukur variabel yang sama. Ini memungkinkan hasil kuesioner tetap dapat diandalkan dan konsisten dalam berbagai situasi, mendukung analisis yang akurat dalam penelitian. Nilai yang diperoleh dari uji *Cronbach's Alfa* biasanya berada dalam rentang antara lebih dari 0,7 dan kurang dari 0,9. Semakin tinggi nilai *Cronbach's Alfa*, semakin reliabel instrumen penelitian tersebut dianggap [13]. Pengujian reliabilitas menggunakan rumus *Cronbach's Alfa* dengan rumus:

$$r_i = \frac{k}{(k - 1)} \left( 1 - \frac{\sum s_i^2}{s_t^2} \right) \quad (13)$$

Keterangan:

$r_i$  = koefisien reliabilitas *Cronbach's Alfa*.

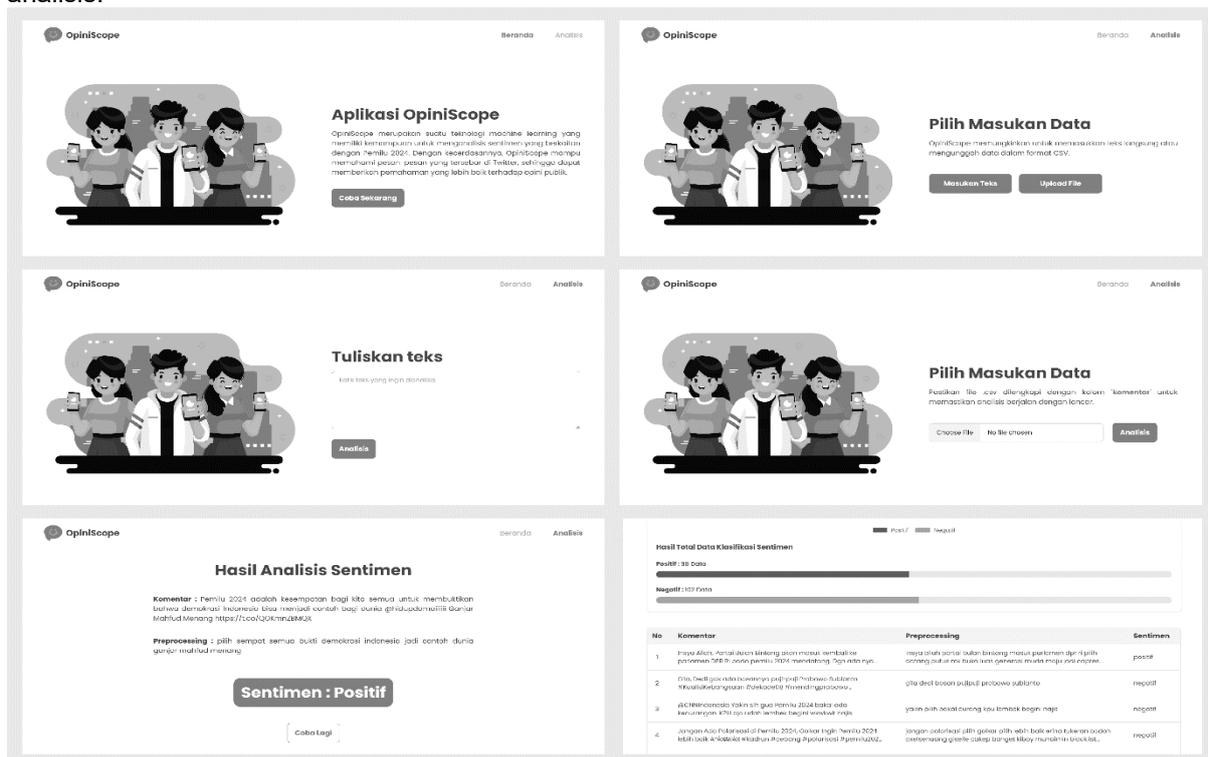
k = jumlah item soal.

$s_i^2$  = varians tiap item.

$s_t^2$  = varians total.

## 2.9. Rancangan Antarmuka

Rancangan antarmuka (*interface design*) merupakan proses perencanaan dan pengembangan tampilan program aplikasi yang dapat dilihat dan digunakan oleh pengguna [14]. Antarmuka sistem pada penelitian ini dibangun berbasis website menggunakan *Python* dengan *framework Flask* dan *Preline UI* sebagai *frontend*, yang menggunakan *Tailwind CSS component library*. Gambar 3 adalah hasil dari proses rancangan antarmuka sistem halaman beranda, halaman analisis, dan halaman hasil analisis.



Gambar 3. Rancangan Antarmuka

## 2.10. Skenario Pengujian dan Evaluasi

Penelitian ini menggunakan *K-Fold Cross-Validation* untuk pengujian model klasifikasi. Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Kemudian, data latih dilakukan iterasi sebanyak  $k$  bagian atau folds, setiap bagian dibagi menjadi data latih dan data validasi secara bergantian. Tujuan pengujian ini adalah untuk menemukan parameter rasio seleksi fitur yang optimal dengan memperhatikan rata-rata nilai metrik seperti akurasi dan *f-measure* dari seluruh iterasi terhadap berbagai rasio seleksi fitur yang diujikan [15]. Selanjutnya, kinerja parameter rasio seleksi fitur terbaik akan dibandingkan saat diterapkan pada data uji, baik dengan menggunakan seleksi fitur maupun tanpa penggunaan seleksi fitur. Sedangkan untuk evaluasi sistem website menggunakan kuesioner SUS dengan skala *Likert*. Kuesioner ini dirancang untuk mendapatkan persepsi responden terhadap *usability* (kemudahan penggunaan) website. Kegunaan diukur dari seberapa percaya seseorang bahwa penggunaan sistem tersebut akan meningkatkan kinerja kerjanya. Kemudahan penggunaan diukur dari seberapa percaya seseorang bahwa menggunakan sistem tersebut akan mudah tanpa memerlukan usaha [16].

## 3. Hasil dan Pembahasan

### 3.1. Pembangunan Model Klasifikasi

Terdapat 2 skenario klasifikasi, yaitu dengan menggunakan seleksi fitur dan tanpa seleksi fitur. Terdapat 3 *tuning hyperparameter* yang digunakan yaitu: kernel,  $C$ , dan  $\gamma$  ( $\gamma$ ). Adapun kernel yang diuji meliputi *linear*, *rbf*, dan *polynomial*. Rentang nilai yang dievaluasi untuk parameter  $C$  adalah 0.1, 1, dan 10, sementara untuk  $\gamma$  ( $\gamma$ ) adalah 0.5, 1, dan 2. Pengujian dilakukan dengan menggunakan metode *K-fold validation*, di mana nilai *K-fold validation* yang digunakan adalah 5. Berikut adalah tabel perbandingan akurasi dan *f-measure* antara penggunaan metode seleksi fitur dan tanpa seleksi fitur dalam klasifikasi SVM. Hasil evaluasi klasifikasi tanpa seleksi fitur ditunjukkan pada Tabel 2 sedangkan hasil evaluasi dengan seleksi fitur ditunjukkan pada Tabel 3.

**Tabel 2.** Hasil Evaluasi Klasifikasi Tanpa Seleksi Fitur

Kernel	C	$\gamma$	Jumlah Fitur	Accuracy	F-measure	Waktu Komputasi
Linear	1	-	1235	0.865	0.864	0.04
Rbf	1	2	1235	0.880	0.879	0.07
Polynomial	10	0.5	1235	0.835	0.839	0.06

**Tabel 3.** Hasil Evaluasi Klasifikasi dengan Seleksi Fitur

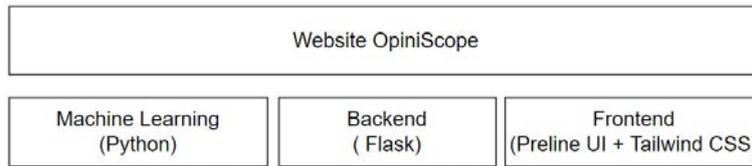
Kernel	Nilai Threshold	C	$\gamma$	Jumlah Fitur	Accuracy	F-measure	Waktu Komputasi
Linear	60%	10	-	741	0.890	0.889	0.02
Rbf	60%	10	0.5	741	0.900	0.899	0.02
Polynomial	80%	10	1	984	0.875	0.873	0.05

Pada hasil klasifikasi tanpa menggunakan seleksi fitur, ditemukan bahwa nilai tertinggi untuk akurasi dan *f-measure* tercapai saat menggunakan kernel *rbf* dengan parameter  $C = 1$  dan  $\gamma = 2$ , dengan total fitur sebanyak 1235. Hasilnya menunjukkan akurasi sebesar 88% dan *f-measure* sebesar 87,9%, dengan waktu komputasi yang mencapai 0,07 detik. Sementara itu, dalam pengujian klasifikasi sentimen dengan menerapkan seleksi *Mutual Information* (MI), nilai akurasi dan *f-measure* tertinggi tercapai dengan menerapkan kernel *rbf* dengan seleksi fitur sebesar 60%, parameter  $C = 10$ ,  $\gamma = 0.5$  dengan jumlah fitur yang digunakan sebanyak 741 fitur yang menghasilkan akurasi sebesar 90% dan *f-measure* sebesar 89,9%, dengan waktu komputasi yang lebih efisien, yaitu 0,02 detik. Pengaruh signifikan dari seleksi fitur dalam klasifikasi menggunakan SVM terlihat jelas, di mana penerapan seleksi fitur mampu meningkatkan performa klasifikasi dengan mengurangi jumlah fitur yang digunakan. Ini tidak hanya mengoptimalkan waktu komputasi, tetapi juga dapat meningkatkan akurasi dan nilai *f-measure* secara keseluruhan.

### 3.2. Arsitektur Sistem

Arsitektur Sistem website yang dikembangkan mencakup penggunaan *Python* versi 3.12.1 sebagai dasar utamanya untuk membangun *machine learning* klasifikasi guna menganalisis data opini. *Flask* versi 3.0.2 digunakan di bagian *backend* aplikasi untuk menangani permintaan dan pengelolaan data di sisi server, sementara tampilan depannya diperkuat dengan *Preline UI* versi 2.0.0 yang

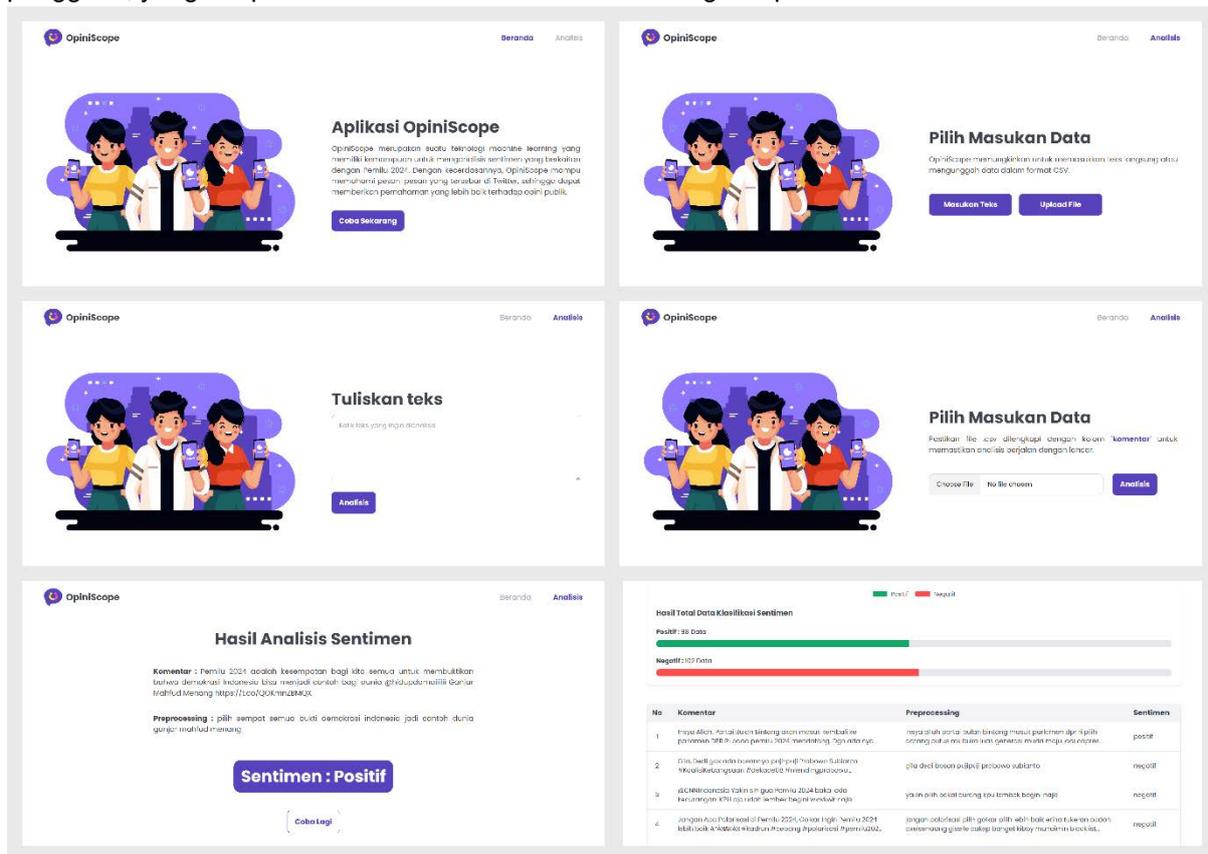
menggunakan *Tailwind CSS component library* untuk tampilan yang responsif dan menarik. Arsitektur sistem dapat dilihat pada Gambar 4 berikut:



Gambar 4. Arsitektur Sistem

### 3.3. Implementasi Antarmuka

Implementasi sistem ini akan dijelaskan sesuai dengan urutan tahapan yang telah ditetapkan sebelumnya. Sistem ini dirancang untuk memberikan fleksibilitas dalam hal jenis dan jumlah *input* komentar yang dapat digunakan serta memberikan hasil yang relevan dan mudah dimengerti bagi pengguna, yang direpresentasikan melalui Gambar 5 sebagai implementasi antarmuka sistem.



Gambar 5. Implementasi Antarmuka Sistem

### 3.4. Pengujian System Usability Scale (SUS)

Proses evaluasi website menggunakan kuesioner SUS dengan skala *Likert*. Kuesioner ini dirancang untuk mendapatkan persepsi responden terhadap *usability* website. Responden diminta untuk menilai sejumlah pernyataan yang diberikan dengan menggunakan skala yang ditentukan. Selanjutnya, hasil tanggapan responden akan diuji validitas dan reliabilitasnya. Langkah-langkah ini penting untuk memastikan bahwa evaluasi *usability* website didasarkan pada data yang valid dan dapat dipercaya. Berikut adalah sepuluh pernyataan yang digunakan untuk menguji sistem *OpiniScope* menggunakan metode SUS, yang tertera pada Tabel 4.

Tabel 4. Kuisisioner System Usability Scale

No	Pernyataan
P1	Saya akan sering mengakses website <i>OpiniScope</i> .
P2	Saya merasa website <i>OpiniScope</i> sulit untuk digunakan.

P3	Saya merasa website <i>OpiniScope</i> mudah digunakan.
P4	Saya butuh bantuan dari orang lain atau teknisi untuk menggunakan website <i>OpiniScope</i> .
P5	Saya merasa bahwa fitur-fitur situs website <i>OpiniScope</i> berjalan dengan baik.
P6	Saya merasa ada banyak hal yang tidak konsisten di website <i>OpiniScope</i> .
P7	Saya yakin orang lain cepat memahami cara menggunakan website <i>OpiniScope</i> .
P8	Saya merasakan kebingungan dalam menggunakan website <i>OpiniScope</i> .
P9	Saya merasa tidak mengalami kesulitan dalam menggunakan website <i>OpiniScope</i> .
P10	Saya merasa perlu membiasakan diri sebelum menggunakan website <i>OpiniScope</i> .

Evaluasi sistem pada penelitian ini melibatkan 20 responden sebagai partisipan. Di bawah ini adalah hasil evaluasi skor SUS yang diperoleh dari responden, yang tercantum pada Tabel 5:

**Tabel 5.** Hasil Perhitungan Skor SUS

Resp.	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Jumlah	Skor SUS
1	0	3	3	3	4	3	4	3	3	3	29	72,5
2	4	4	3	3	4	3	4	4	4	3	36	90
3	4	4	4	4	4	4	4	4	4	4	40	100
4	4	4	4	4	4	4	4	4	4	4	40	100
5	3	4	4	4	4	4	4	4	4	3	38	95
6	4	4	3	4	4	3	4	4	3	4	37	92,5
7	3	4	4	4	4	4	4	4	4	4	39	97,5
8	3	3	3	3	4	3	3	3	3	4	32	80
9	2	4	4	4	4	4	3	4	4	3	36	90
10	3	4	4	3	3	4	4	4	3	3	35	87,5
11	3	3	4	4	4	4	4	4	4	4	38	95
12	3	4	4	3	4	3	4	4	4	3	36	90
13	2	3	4	4	3	3	3	4	3	3	32	80
14	4	4	4	4	4	3	3	4	4	3	37	92,5
15	3	4	3	3	3	3	3	3	3	2	30	75
16	2	4	4	3	3	4	3	3	3	3	32	80
17	4	4	4	4	4	4	4	4	4	4	40	100
18	3	3	3	3	3	2	3	3	3	0	26	65
19	4	4	4	4	2	2	4	4	4	4	36	90
20	4	4	4	4	4	4	4	4	4	4	40	100
<b>Skor Rata-Rata SUS</b>											<b>88,63</b>	

Berdasarkan hasil perhitungan, website *OpiniScope* memperoleh nilai SUS sebesar 88,63 yang menunjukkan bahwa pengguna menganggap sistem ini dapat diterima dalam penggunaannya. Tingkat kegunaan sistem tersebut dinilai sebagai "Acceptable" berdasarkan *acceptability ranges*. Pada *grade scale*, website *OpiniScope* mendapatkan peringkat B, menunjukkan bahwa secara keseluruhan, tingkat kegunaan sistem tersebut berada pada tingkat yang baik. Peringkat B menunjukkan bahwa secara keseluruhan, website memiliki kualitas yang baik dan mudah digunakan. Sedangkan untuk *adjective rating* termasuk ke kelompok "Excellent" menunjukkan bahwa website memiliki tingkat kualitas yang sangat baik dalam hal usability dan pengalaman pengguna.

### 3.5. Hasil Uji Validitas

Dalam penelitian ini, validitas diukur menggunakan koefisien korelasi *Pearson* (2 tail) terhadap skor total dari 10 item pernyataan dalam kuesioner SUS. Untuk menentukan validitas, penulis melakukan perbandingan antara nilai  $t_{hitung}$  dengan nilai  $t_{tabel}$  pada tingkat signifikansi 0,05. Dalam kasus ini,

$t_{tabel}$  memiliki nilai 0,444 karena jumlah responden adalah 20. Setelah melakukan pengukuran, didapatkan bahwa nilai  $t_{hitung}$  lebih besar dari  $t_{tabel}$ . Hal ini menunjukkan bahwa kuesioner memiliki validitas yang baik, karena hubungan antara setiap item dalam kuesioner dengan skor totalnya signifikan secara statistik. Berdasarkan hasil uji validitas, pernyataan dalam kuesioner dianggap valid karena nilai  $t_{hitung}$  yang dihasilkan lebih besar daripada  $t_{tabel}$  yang ditentukan. Rincian perhitungan lebih lanjut dapat ditemukan pada Tabel 6, yang memberikan penjelasan tentang hasil uji validitas data kuesioner.

**Tabel 6.** Hasil Uji Validitas Data Kuesioner

	$t_{hitung}$	$t_{tabel}$	Keterangan
P1	0.613	0.444	Valid
P2	0.593	0.444	Valid
P3	0.628	0.444	Valid
P4	0.714	0.444	Valid
P5	0.469	0.444	Valid
P6	0.619	0.444	Valid
P7	0.615	0.444	Valid
P8	0.827	0.444	Valid
P9	0.792	0.444	Valid
P10	0.764	0.444	Valid

### 3.6. Hasil Uji Reliabilitas

Hasil uji reliabilitas yang terdapat pada Tabel 7, menunjukkan bahwa kuesioner SUS memiliki tingkat reliabilitas yang cukup tinggi yaitu 0.830. Standar umum menyatakan bahwa nilai *Cronbach's Alpha* yang berada di antara 0.7 dan 0.9 menunjukkan tingkat reliabilitas yang baik. Dengan nilai 0.830 yang berada dalam rentang ini, dapat disimpulkan bahwa kuesioner tersebut dianggap reliabel. Ini berarti pertanyaan pada kuesioner dapat diandalkan untuk mengukur apa yang dimaksudkan dengan konsistensi yang baik.

**Tabel 7.** Hasil Uji Reliabilitas Data Kuesioner

Nilai Cronbach's Alfa	Jumlah Item	Keterangan
0.830	10	Reliabel

## 4. Kesimpulan

Hasil Penelitian ini, menunjukkan bahwa seleksi fitur *Mutual Information* (MI) dapat meningkatkan kinerja metode SVM dibandingkan tanpa seleksi fitur. Pada pengujian tanpa seleksi fitur, kinerja SVM mencapai akurasi 88% dan *f-measure* 87,9% dengan kernel *rbf* menggunakan nilai  $C = 1$  dan nilai  $\gamma = 2$  dengan waktu komputasi sekitar 0,07 detik. Saat seleksi fitur diterapkan, terjadi peningkatan kinerja SVM yang cukup signifikan. Akurasi meningkat menjadi 90%, sedangkan *f-measure* mencapai 89,9% dengan pemanfaatan 60% fitur, kernel *rbf*, dan parameter  $C = 10$  serta  $\gamma = 0,5$ . Selain mengoptimalkan kinerja, penerapan seleksi fitur juga memberikan dampak positif pada efisiensi waktu komputasi, yang turun drastis menjadi hanya 0,02 detik.

## References

- [1] Imam Fahrur Rozi, Yushintia Pramitarini, dan Novia Puspitasari, "Analisis Mengenai Calon Presiden Indonesia 2019 Di Twitter Menggunakan Metode Backpropagation," *J. Inform. Polinema*, vol. 6, no. 2, hal. 27–31, 2020, doi: 10.33795/jip.v6i2.306.
- [2] R. Ardiansyah, "Analisis Sentimen Calon Presiden Dan Wakil Presiden Periode 2019-2024 Pasca Debat Pilpres Di Twitter," *Sci. Comput. Sci. Informatics J.*, vol. 2, no. 1, hal. 21, 2019, doi: 10.22487/j26204118.2019.v2.i1.13068.
- [3] R. Tineges, A. Triayudi, dan I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, hal. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [4] N. Hendrastuty, A. Rahman Isnain, A. Yanti Rahmadhani, dan S. Styawati, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector

- Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, hal. 150–155, 2021, [Daring]. Tersedia pada: <http://situs.com>.
- [5] A. W. Wilantapoera, “Analisis Sentimen Kategori Aspek Pada Ulasan Produk Menggunakan Metode KNN Dengan Seleksi Fitur Mutual Information Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung Lembar Persetujuan Analisis Sentimen Kategori Aspek Pada U,” vol. 10, no. 2, hal. 1673–1681, 2022.
- [6] P. Widyantara dkk., “Analisis Sentimen pada Teks Berbahasa Bali Menggunakan Metode Multinomial Naive Bayes dengan TF-IDF dan BoW,” *Jnatia*, vol. 2, no. 1, hal. 37–46, 2023.
- [7] Merinda Lestandy, Abdurrahim Abdurrahim, dan Lailis Syafa’ah, “Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [8] T. F. T. Hidayat, G. Garno, dan A. A. Ridha, “Analisis Sentimen Opini Pemindahan Ibu Kota Pada Twitter Dengan Metode Support Vector Machine,” *J. Ilmu Komput.*, vol. 14, no. 1, hal. 49, 2021, doi: 10.24843/jik.2021.v14.i01.p06.
- [9] S. Al Qodrin dkk., “Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma,” vol. 14, no. 2, hal. 44–52, 2022.
- [10] P. Arsi dan R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, hal. 147, 2021, doi: 10.25126/jtiik.0813944.
- [11] P. W. A. Wibawa dan C. Pramatha, “Systematic Literature Review: Machine Learning Methods in Emotion Classification in Textual Data,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, hal. 425–433, 2023, doi: 10.32736/sisfokom.v12i3.1787.
- [12] F. Rosyad, D. Pramono, dan K. C. Brata, “Analisis dan Perbaikan Usability Pada Aplikasi Ker Menggunakan Metode Usability Testing dan System Usability Scale (SUS),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 7, hal. 2261–2268, 2020.
- [13] M. L. L. Usman dan M. A. Gustalika, “Pengujian Validitas dan Reliabilitas System Usability Scale (SUS) Untuk Perangkat Smartphone,” *J. Ecotipe (Electronic, Control. Telecommun. Information, Power Eng.)*, vol. 9, no. 1, hal. 19–24, 2022, doi: 10.33019/jurnalecotipe.v9i1.2805.
- [14] A. S. ABADI dan P. F. DEWI, “Multimedia Mobile Application of National Heroes History Learning for Children’s Character Education,” *Telematika*, vol. 18, no. 3, hal. 308, 2021, doi: 10.31315/telematika.v18i3.5542.
- [15] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, dan B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM,” vol. 7, no. 1, hal. 11–18, 2022.
- [16] C. Pramatha, J. G. Davis, dan K. K. Y. Kuan, “A Semantically-Enriched Digital Portal for the Digital Preservation of Cultural Heritage with Community Participation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11196 LNCS, no. October, hal. 560–571, 2018, doi: 10.1007/978-3-030-01762-0\_49.

# Perancangan Sistem Informasi Pemasok Barang Rongsokan Dengan Pendekatan User Centered Design

Tri Adi Ningsih<sup>a1</sup>, I Made Widiartha<sup>a2</sup>, Dr. Ir. I Ketut Gede Suhartana<sup>a3</sup>, I Gusti Ngurah Anom Cahyadi Putra<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, FMIPA, Universitas Udayana  
Badung, Indonesia

<sup>1</sup>[triadiningsih65@gmail.com](mailto:triadiningsih65@gmail.com)

<sup>2</sup>[madewidiartha@unud.ac.id](mailto:madewidiartha@unud.ac.id)

<sup>3</sup>[ikg.suhartana@unud.ac.id](mailto:ikg.suhartana@unud.ac.id)

<sup>4</sup>[anom.cp@unud.ac.id](mailto:anom.cp@unud.ac.id)

## Abstract

*Information systems play a crucial role in facilitating business processes, including in the scrap recycling industry. In this context, this research aims to develop an information system that enables efficient management of the procurement process of scrap materials from suppliers. The User Centered Design (UCD) method is adopted to ensure that the developed system truly meets the needs and preferences of users. The system implementation is carried out using web-based technologies with PHP and MySQL as the database backend. End users, including scrap material suppliers, can access the system through a responsive web interface. System evaluation is conducted using the Single Ease Question (SEQ) and System Usability Scale (SUS) methods to measure user responses and system usability. The development results show that the scrap material supplier information system has been successfully developed. Users responded positively to the user-friendly interface and provided functionality. The testing involved 10 respondents, yielding a SEQ score calculation of 88%, indicating success. Additionally, SUS testing resulted in a high satisfaction level of 87.25% among users regarding system usability. Thus, this research contributes to the development of information systems that integrate UCD principles to support efficiency and effectiveness in managing scrap material suppliers in the recycling industry.*

**Keywords:** Information System, Waste, User Centered Design, Single Ease Question, Testing

## 1. Pendahuluan

Permasalahan akan sampah dari tahun ke tahun belum juga dapat diatasi oleh pemerintah hal ini dikarenakan kurangnya kesadaran dan kepedulian kita terhadap lingkungan. Masih banyak masyarakat yang membuang sampah di tempat-tempat yang seharusnya bukan tempat pembuangan sampah. Masalah sampah diperkotaan salah satunya timbulan sampah yang makin ke tahun makin menumpuk. Data yang diambil bersumber dari Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) menunjukkan bahwa wilayah Provinsi Bali menghasilkan jumlah sampah yang cukup besar setiap tahunnya. Timbulan sampah di Bali mencapai 525,962.04 ton sepanjang tahun 2022. Dari data tersebut, terlihat bahwa timbulan sampah di Bali sangat signifikan dan menjadi masalah serius yang perlu segera diatasi. Permasalahan timbulan sampah yang terjadi di Bali memberikan dampak negatif terhadap lingkungan dan keindahan pulau ini. Timbulan sampah yang tidak dikelola dengan baik dapat mencemari lingkungan, merusak ekosistem laut, serta mengganggu keindahan alam dan kesehatan masyarakat sekitar. Dalam konteks ini, penanganan sampah menjadi sebuah isu penting yang perlu mendapat perhatian serius. Salah satu cara yang dapat dilakukan untuk mengurangi timbulan sampah adalah dengan memanfaatkan barang rongsokan. Penjualan barang rongsokan dapat menjadi solusi untuk mengurangi jumlah sampah di Bali dengan mengolah kembali barang-barang yang sudah tidak terpakai menjadi bahan yang dapat digunakan lagi.

Menurut penelitian yang dilakukan oleh User Interface Engineering, Inc. pada tahun 2004, 60% waktu terbuang karena orang kesulitan menemukan informasi yang mereka butuhkan. Hal ini

tidak hanya mengakibatkan penurunan produktivitas, tetapi juga meningkatkan tingkat frustrasi[1]. Pengalaman seperti ini dapat mempengaruhi persepsi pengguna saat mengunjungi sebuah situs web, bahkan mungkin membuat mereka menilai situs tersebut tidak layak dikunjungi. Oleh karena itu, banyak situs web yang ditinggalkan oleh pengguna karena tidak berhasil mencapai tujuan awal pembuatannya. Metode User Centered Design (UCD) memperkenalkan pendekatan baru dalam proses perancangan sistem berbasis web, dengan mengedepankan pengguna sebagai pusat dari seluruh proses pengembangan. Konsep UCD menekankan bahwa tujuan, sifat-sifat, konteks, dan lingkungan sistem harus didasarkan pada pengalaman pengguna. UCD merupakan proses interaktif yang melibatkan langkah-langkah perancangan dan evaluasi mulai dari awal proyek hingga implementasi[2]. Pendekatan ini tidak hanya merancang antarmuka yang menarik perhatian pengguna, tetapi juga memastikan bahwa antarmuka tersebut dapat memberikan pengalaman yang optimal. Berdasarkan uraian tersebut, penulis mengajukan judul “Perancangan Sistem Informasi Pemasok Barang Rongsokan Menggunakan Metode User Centered Design”.

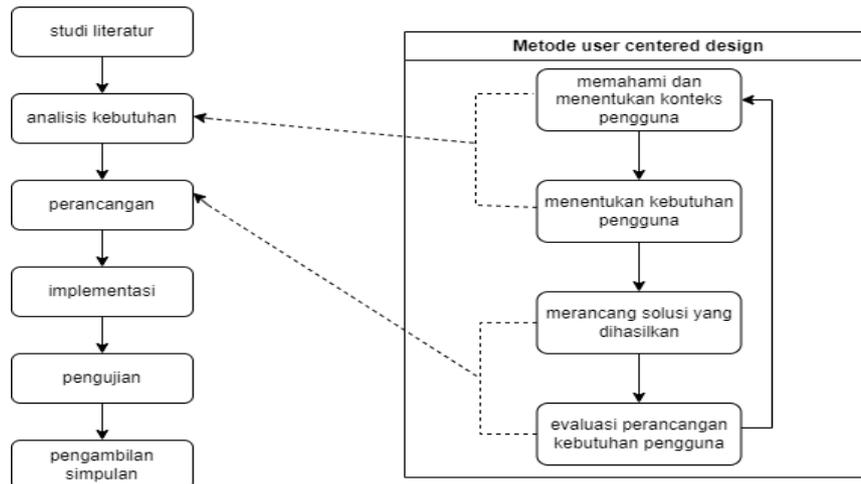
Dalam konteks pengembangan sistem informasi untuk pemasok barang rongsokan, penelitian sebelumnya dilakukan oleh Ade Pratama dalam jurnal yang berjudul “Sistem Informasi Jual Beli Barang Rongsokan Menggunakan Fitur Location Based Service “telah mengungkapkan sejumlah permasalahan yang mendesak. Salah satunya adalah kesenjangan yang terjadi antara kebutuhan pengguna yang semakin kompleks dan kemampuan sistem yang ada untuk memenuhi kebutuhan tersebut. Penelitian-penelitian terdahulu juga menyoroti kurangnya integrasi masukan pengguna secara efektif dalam proses pengembangan, yang berujung pada ketidakcocokan antara fitur-fitur yang disediakan oleh sistem dan kebutuhan sebenarnya pengguna. Ditemukan bahwa banyak sistem informasi yang sudah ada kurang mengadopsi pendekatan user-centered design, yang pada akhirnya menyebabkan kesulitan pengguna dalam penggunaan, kurangnya efisiensi operasional, dan tingkat ketidakpuasan pengguna yang signifikan[3]. Oleh karena itu, dalam konteks penelitian ini, pentingnya adopsi metode user-centered design sebagai pendekatan pengembangan yang relevan dan sesuai dengan kebutuhan pengguna telah diakui. Langkah ini diambil untuk mengatasi permasalahan-permasalahan yang diidentifikasi dan memastikan bahwa sistem informasi yang dikembangkan dapat lebih efektif memenuhi kebutuhan dan harapan pengguna yang beragam.

Sistem informasi pemasok barang rongsokan ini nantinya akan memudahkan masyarakat untuk menjual barang rongsokan yang sudah masyarakat kumpulkan tersebut agar bisa di daur ulang kembali. Masyarakat nantinya dapat menggunakan aplikasi ini untuk dapat menjualkan rongsokannya ke pemasok dan memilih lokasi keberadaan agen pemasok barang bekas disekitar mereka. . Dengan adanya sistem informasi ini diharapkan dapat membantu menanggulangi permasalahan sampah khususnya di Provinsi Bali. Dalam proses perancangan, website ini tentunya akan di rancang dengan mengikuti serangkaian tahap evaluasi dan pengujian sebelum digunakan. Tahap evaluasi ini akan melibatkan penggunaan metode SEQ (Single Ease Questions) dan SUS (Software Usability Scale) untuk memperoleh pemahaman yang komprehensif tentang kemudahan penggunaan serta tingkat kepuasan pengguna terhadap website ini. Melalui proses ini, peneliti akan dapat mengevaluasi performa website secara menyeluruh, memperbaiki potensi masalah, dan memastikan bahwa website siap untuk digunakan.

## **2. Metode Penelitian**

Dalam bab metode penelitian jurnal ini, peneliti menerapkan pendekatan kuantitatif dengan metode tertentu. Data diperoleh melalui wawancara mendalam dengan responden yang dipilih secara purposif, serta observasi partisipatif di lapangan. Teknik analisis data yang digunakan meliputi pengkodean tematik dan pembentukan pola. Pemilihan metode penelitian ini bertujuan untuk memahami secara mendalam dinamika dan konteks yang terlibat dalam fenomena yang diteliti. Tahapan penelitian ini dijelaskan secara detail dalam gambar 1 :

:



**Gambar 1.** Tahapan Penelitian

Gambar di atas menjelaskan alur penelitian yang dilakukan oleh penulis dimulai dari tahap studi literatur sampai dengan tahap pengambilan simpulan. Penjelasan mengenai tahapan penelitian ini akan dibahas pada bab 2.1 sampai dengan 2.3.

## 2.1 Studi Literatur

Studi literatur adalah proses sistematis untuk meninjau, mengevaluasi, dan menyintesis literatur yang relevan terhadap topik penelitian yang sedang diteliti, yang bertujuan untuk memperoleh pemahaman yang mendalam tentang kerangka teoritis yang ada serta mengidentifikasi kesenjangan pengetahuan yang dapat diisi oleh penelitian baru[4]. Pada tahap pertama ini penulis melakukan pemetaan terhadap teori-teori yang ada untuk mengidentifikasi kesenjangan pengetahuan atau konflik dalam literatur yang dapat dijadikan dasar untuk penelitian lebih lanjut. Selain itu, tinjauan literatur kami memperhatikan berbagai pendekatan metodologis yang digunakan dalam penelitian terdahulu, sehingga memungkinkan penulis untuk mengevaluasi kekuatan dan kelemahan masing-masing metode tersebut dalam konteks penelitian saya. Hasil dari analisis literatur ini memberikan dasar yang kuat untuk merumuskan kerangka teoritis yang mendukung penelitian ini, serta memandu desain metodologi yang tepat untuk mengumpulkan dan menganalisis data.

## 2.2 Analisis Kebutuhan

Dalam tahap ini, tujuannya adalah Untuk mengidentifikasi kebutuhan sistem informasi pemasok barang rongsokan, penelitian ini mengadopsi metode UCD (User Centered Design) yang mencakup dua tahapan analisis, yaitu menentukan konteks penggunaan dan menentukan kebutuhan pengguna serta organisasi. Pada tahap menentukan konteks penggunaan, dilakukan identifikasi terhadap berbagai pihak yang mungkin akan menggunakan sistem informasi ini. Langkah berikutnya adalah menjalankan tahapan menentukan kebutuhan pengguna dan organisasi setelah berhasil mengidentifikasi para aktor, untuk mengetahui kebutuhan dari calon pengguna. Metode yang digunakan pada tahapan ini adalah wawancara dengan 5 calon pengguna yang telah ditentukan sebelumnya. Adapun kebutuhan yang dihasilkan setelah melakukan wawancara dapat dilihat pada tabel 1 berikut ini :

**Tabel 1. Analisis Kebutuhan Pengguna**

No.	Kebutuhan Pengguna	Kebutuhan Fungsional Dalam Sistem
1.	Pengguna ingin dapat mendaftar sebagai penjual rongsok di platform	Sistem harus menyediakan fitur pendaftaran akun yang memungkinkan pengguna untuk membuat akun.
2.	Pengguna ingin dapat mengakses informasi tentang harga jual rongsokan terkini.	Sistem harus memiliki antarmuka yang responsive yang menampilkan harga jual rongsokan yang terbaru.
3.	Pengguna ingin bisa menemukan lokasi agen pemasok barang rongsokan terdekat.	Sistem harus memiliki fitur pemilihan lokasi berdasarkan kota atau kabupaten yang memungkinkan pengguna untuk menemukan lokasi agen pemasok barang rongsokan terdekat.
4.	Pengguna ingin memiliki riwayat transaksi jual-beli yang dilakukan.	Sistem harus menyediakan fitur riwayat transaksi yang memungkinkan pengguna untuk melihat riwayat transaksi yang telah dilakukan sebelumnya.
5.	Pengguna memerlukan penggunaan peta atau aplikasi penunjuk arah dapat membantu mempercepat pemasok dalam proses pengambilan barang bekas	Sistem harus menyediakan fitur peta interaktif yang menunjukkan lokasi pemasok untuk proses pengambilan barang bekas

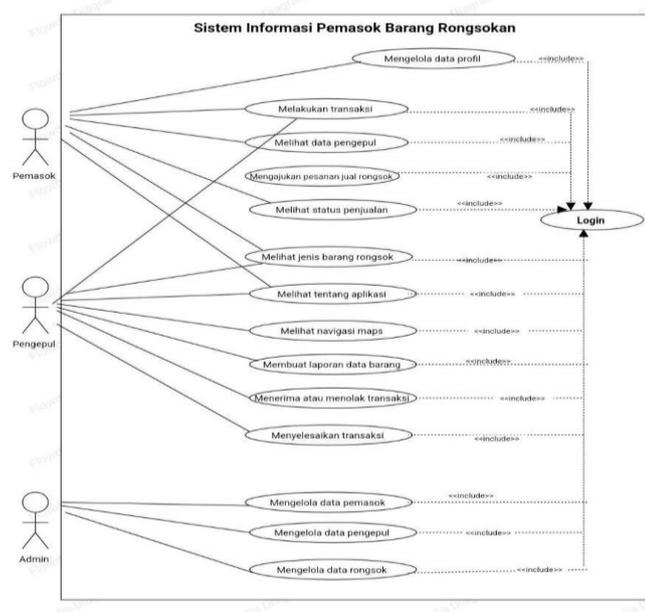
Berdasarkan tabel kebutuhan pengguna yang diperoleh melalui wawancara, Beberapa aspek penting yang perlu diperhatikan dalam pengembangan sistem informasi pemasok barang rongsokan ini disajikan dalam tabel ini, yang tentunya akan menjadi acuan utama dalam proses pengembangan sistem informasi yang akan dibangun.

### 2.3 Perancangan

Tahap berikutnya adalah perancangan, yang terbagi dua tahapan lagi pada metode Ucer Centered Design yaitu *produce design solution dan evaluate design*. Tahap perancangan adalah proses penentuan langkah-langkah dan kerangka kerja yang akan digunakan dalam penelitian atau proyek tertentu, yang melibatkan identifikasi tujuan, pemilihan metode, serta pengembangan strategi untuk mencapai hasil yang diinginkan[5]. Dimulai dari perancangan perancangan use case diagram, perancangan basis data dan diagram aktivitas dan perancangan wireframe. Setelah desain solusi dibuat, langkah berikutnya adalah menyerahkannya kepada calon pengguna untuk dievaluasi, baik dari segi tampilan maupun fitur yang ada. Lalu evaluasi tersebut digunakan untuk memperbaiki sistem. Evaluasi ini dilakukan dengan mewawancarai 5 pengguna, yang kemudian memberikan evaluasi terhadap hasil perancangan pada tahap ini.

#### 1. Use Case Diagram

Use Case Diagram adalah jenis dari UML (Unified Modeling Language), yang menjadi bahasa pemodelan standar untuk merancang dan mendokumentasikan perangkat lunak. Diagram Use Case mengilustrasikan interaksi antara aktor-aktor yang terlibat dalam sistem dengan proses-proses yang ada di dalam sistem. Use case juga berfungsi sebagai deskripsi sistematis tentang interaksi antara sistem yang sedang dikembangkan dan pihak-pihak yang terlibat di dalamnya, yang mencakup berbagai skenario atau situasi yang mungkin terjadi selama penggunaan sistem tersebut. Use case digunakan untuk memodelkan fungsionalitas sistem secara terperinci dan membantu dalam merancang solusi yang sesuai dengan kebutuhan pengguna[6]. Use Case Diagram untuk Sistem Informasi Pemasok Barang Rongsokan lebih jelas lagi akan disajikan pada gambar 2. di bawah ini :

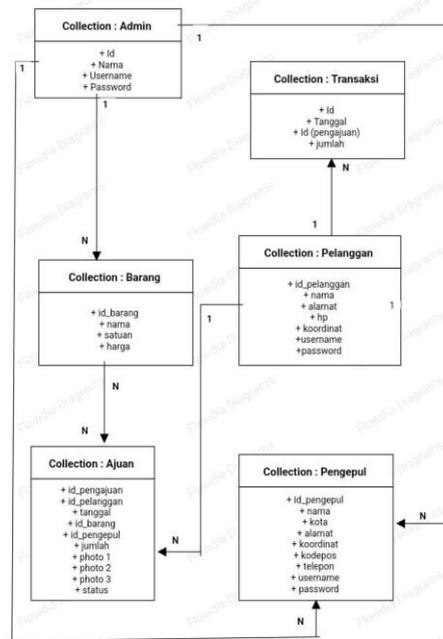


**Gambar 2.** Use Case Diagram

Pada Gambar di atas terdapat 3 aktor yang berperan pada sistem serta 15 fitur utama yang akan dikembangkan berdasarkan kriteria dari calon pengguna yang sudah diwawancarai sebelumnya.

## 2. Perancangan Basis Data

Rancangan basis data adalah proses merencanakan struktur dan organisasi dari basis data, termasuk entitas, atribut, hubungan antarentitas, serta aturan integritas data yang diperlukan. Rancangan basis data mencakup pemodelan konseptual, logis, dan fisik, yang bertujuan untuk menghasilkan basis data yang efisien, mudah dipahami, dan dapat diakses dengan baik[7]. Dalam merancang basis data, penulis memilih menggunakan sistem manajemen basis data MySQL. MySQL menggunakan bahasa kueri SQL yang bertugas mengelola dan mengakses data dalam basis data. Desain basis data untuk Sistem Informasi Pemasok Barang Rongsok dapat dilihat lebih detail pada gambar 3 di bawah ini:

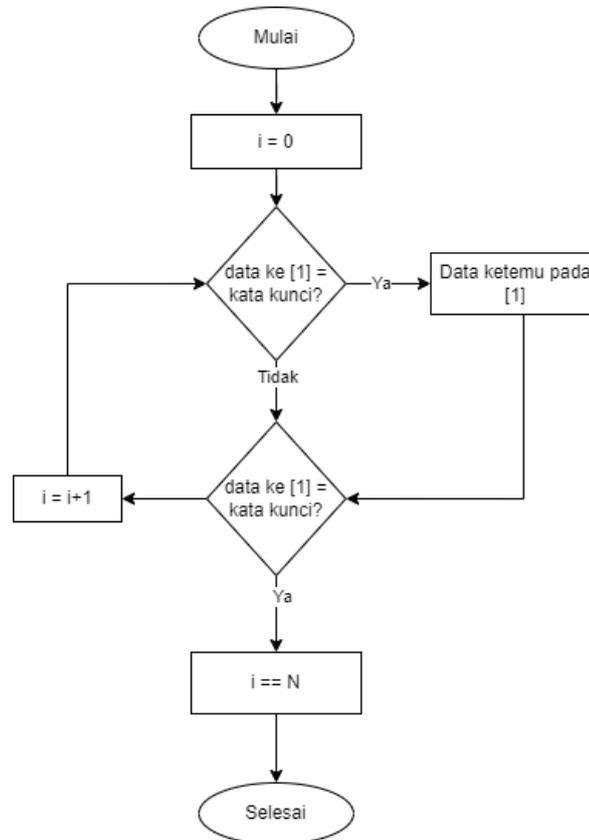


**Gambar 3.** Rancangan Basis Data Sistem

Pada Gambar di atas terdapat daftar tabel dan nama kolom yang akan dikembangkan untuk membantu proses penyimpanan data yang disimpan ke dalam basis data mysql.

### 3. Implementasi Kode Program

Implementasi kode program adalah proses mengubah desain atau rancangan program yang telah dibuat menjadi bentuk nyata atau aplikasi yang berfungsi sepenuhnya. Implementasi mencakup penulisan, pengujian, dan debug kode program sesuai dengan spesifikasi yang telah ditentukan[8]. Pada implementasi sistem yang akan dibangun ini penulis menggunakan salah satu algoritma sequential search. Algoritma sequential search merupakan sebuah metode pencarian sederhana yang digunakan untuk menemukan suatu elemen dalam kumpulan data dengan cara menelusuri secara berurutan. Algoritma ini bekerja dengan cara memeriksa setiap elemen dalam urutan, satu per satu, hingga elemen yang dicari ditemukan atau sampai akhir dari kumpulan data tercapai[9]. Adapun alur kerja dari algoritma sequential search pada sistem yang akan digunakan dapat dilihat lebih jelas pada gambar 4. sebagai berikut :

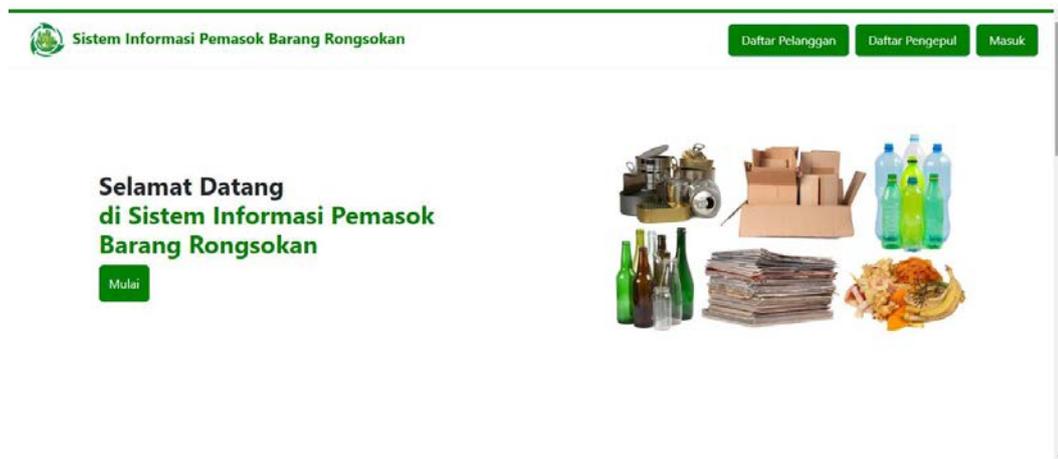


**Gambar 4.** Flowchart Sequential Search

Algoritma Sequential Search ini akan memulai pencarian dari elemen pertama, sampai pada elemen berikutnya. Karena elemen saat ini sama dengan elemen yang dicari, algoritma akan mengembalikan indeks elemen tersebut.

#### 4. Implementasi User Interface Sistem

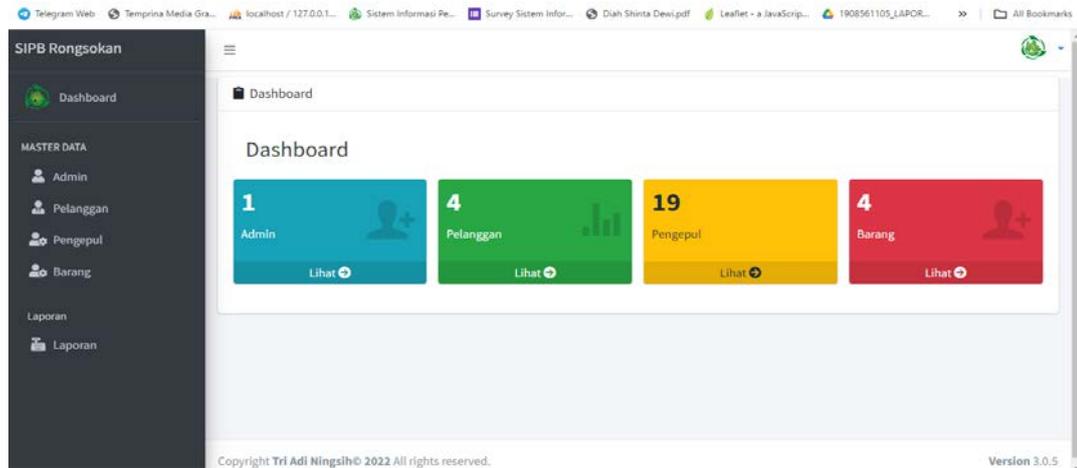
Setelah melakukan proses pengkodean yang lumayan lama lalu terbentuklah user interface dari sistem informasi pemasok barang rongsokan yang dibangun. Adapun hasil interface yang dibangun dapat dilihat pada gambar 5 sampai dengan gambar 7. sebagai berikut :



**Gambar 5.** Homepage

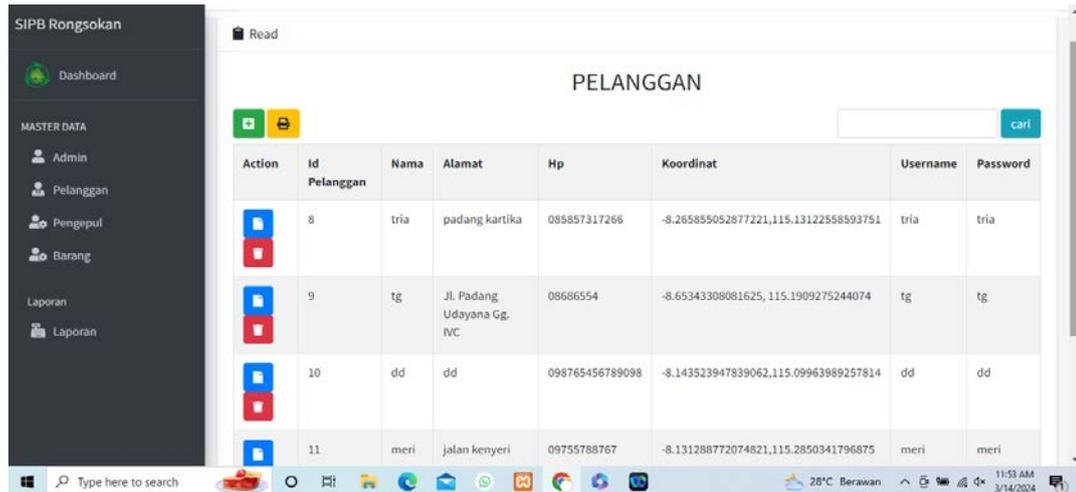
Gambar diatas merupakan tampilan homepage dari sistem yang dirancang. Halaman ini yang pertama kali akan muncul saat aktor admin, pengepul dan pemasok saat mengakases

sistem informasi pemasok barang rongsokan. Pada halaman ini juga nantinya aktor akan diarahkan ke halaman login ataupun register. Di halaman ini juga aktor dapat membaca beberapa informasi yang disediakan seperti fitur yang ditawarkan dan juga informasi terkait sistem informasi yang dibangun



**Gambar 6. Dashboard Halaman Admin**

Gambar di atas ini adalah halaman untuk menampilkan dashboard Admin. Pada halaman ini admin akan disajikan beberapa pilihan menu untuk dapat dikelola seperti halaman pelanggan yang menyajikan data pelanggan, halaman pengepul yang menyajikan data pengepul, halaman barang yang menyajikan data jenis barang dan halaman laporan yang menyajikan data penjualan harian.



**Gambar 7. Daftar Pelanggan**

Gambar diatas merupakan halaman untuk data pelanggan. Pada halaman ini admin dapat memperbaiki data pelanggan seperti menambahkan, mengubah atau menghapus data pelanggan

### 3. Hasil dan Pembahasan

#### 3.1 Pengujian

Tahapan selanjutnya yakni tahapan pengujian yaitu pada tahap ini akan dilakukan evaluasi dan juga pengujian rancangan solusi yang telah diberikan apakah sudah sesuai dengan keinginan pengguna ataupun kebutuhannya. Hal ini tentunya akan menjadi tolak ukur

keberhasilan penerimaan aplikasi oleh pengguna terkait. Pengujian ini bertujuan untuk mendapatkan hasil ukur tingkat kepuasan pengguna pada sistem yang telah dibangun. Pada penelitian ini, terdapat 2 tahap testing, yaitu kuisioner SEQ dan SUS. Dalam tahap usability testing, partisipan diminta untuk berinteraksi dengan desain sistem dengan menjalankan tugas sesuai dengan peran mereka masing-masing. Setelah menyelesaikan tugas, partisipan diminta untuk menilai tingkat kesulitan atau kemudahan tugas tersebut menggunakan kuesioner Single Ease Question (SEQ). Selain itu, mereka juga diminta untuk menilai tingkat kepuasan mereka terhadap rancangan sistem menggunakan kuisioner Single Usability Scale (SUS). Jumlah responden yang dibutuhkan untuk mengisi kuisioner usability testing adalah sebanyak 10 partisipan.

Kuisoner Single Ease Question (SEQ) yang digunakan dalam penelitian ini memiliki beberapa fungsi yang akan diuji, di mana responden akan diberikan beberapa tugas berdasarkan fungsi yang telah ditetapkan. Penilaian dilakukan untuk mengevaluasi seberapa mudah tugas-tugas tersebut dengan menggunakan tingkatan skala Likert tujuh poin, yang terdiri dari sangat sulit (1), sulit (2), tidak mudah (3), cukup (4), tidak sulit (5), mudah (6), hingga sangat mudah (7). Selain itu, responden juga diminta untuk memberikan keterangan Ya/Tidak saat menyelesaikan setiap tugas yang diberikan. Setelah menyelesaikan semua tugas, responden akan menilai tugas secara keseluruhan. Dalam penelitian ini, terdapat 5 fungsi yang akan diuji oleh pemasok dan 5 fungsi yang akan diuji oleh pengepul. Responden diminta untuk menyelesaikan beberapa tugas berdasarkan fungsi-fungsi tersebut. Hasil analisis dari kuisioner single ease question akan disajikan dalam tabel 2 dan tabel 3.

**Tabel 2. Tabel Pengujian SEQ (Pemasok)**

No.	Skenario	Skor						
		1	2	3	4	5	6	7
1	Daftar Akun					2	3	5
2	Login						3	7
3	Lihat Daftar Barang						6	4
4	Menambahkan Pengajuan Penjualan Rongsok						4	6
5	Melihat Status Pengajuan					5	3	2
Jumlah vote		0	0	0	0	7	19	24

Dari tabel di atas dapat dilakukan perhitungan sebagai berikut :

$$\text{Effectiveness} = \frac{\text{number of tasks completed successfully}}{\text{total number of tasks undertake}} \times 100$$

total number of task undertaken = 5 questions x 10 user = 50  
 Total suara nilai 7 = 24 / 50 x 100% = 48%  
 Total suara nilai 6 = 19 / 50 x 100% = 38%  
 Total = 86%

**Tabel 3. Tabel Pengujian SEQ (Pengepul)**

No.	Skenario	Skor						
		1	2	3	4	5	6	7
1	Daftar Akun					3	2	6
2	Login						3	7
3	Lihat Daftar Barang					1	3	6
4	Melihat Data Pengajuan Penjualan Barang						3	7
5	Melihat Navigasi					2	4	4
Jumlah vote		0	0	0	0	6	14	30

Dari tabel di atas dapat dilakukan perhitungan sebagai berikut :

$$\text{Effectiveness} = \frac{\text{a number of tasks completed successfully}}{\text{total number of tasks undertake}} \times 100$$

total number of task undertaken = 5 questions x 10 user = 50

Total suara nilai 7 =  $30 / 50 \times 100\% = 60\%$

Total suara nilai 6 =  $14 / 50 \times 100\% = 28\%$

Total = 88%

Kuisisioner Single Usability Scale (SUS) terdiri dari 10 item pertanyaan yang akan diajukan kepada responden. Kuisisioner ini menggunakan skala likert 5 poin, di mana responden akan diminta untuk memberikan penilaian dengan keterangan Sangat Tidak Setuju (STS), Tidak Setuju (TS), Ragu-Ragu (RG), Setuju (S), dan Sangat Setuju (SS). Hasil pengujian sistem menggunakan kuisisioner usability scale menghasilkan hasil sebagai berikut:

**Tabel 4. Tabel Pengujian SUS (Pemasok)**

No. Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
R1	5	2	5	1	5	1	5	1	4	1	95
R2	4	2	4	2	3	1	5	1	3	2	77.5
R3	4	2	5	1	5	2	4	2	4	1	85
R4	5	3	5	1	5	3	5	1	5	2	87.5
R5	4	2	5	2	5	1	5	3	3	3	77.5
R6	5	3	5	1	5	2	4	1	4	2	85
R7	5	2	4	1	4	3	5	1	4	2	82.5
R8	5	2	5	1	5	1	5	1	3	2	90
R9	5	2	5	1	5	1	5	1	3	1	92.5
R10	4	2	5	1	5	1	4	1	5	1	92.5

Nilai SUS	<b>86.5</b>
-----------	-------------

**Tabel 5.** Tabel Pengujian SUS (Pengepul)

No. Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
R1	4	1	5	1	5	1	5	2	4	2	90
R2	5	2	4	2	3	1	4	1	5	1	85
R3	5	2	4	1	4	2	5	1	4	2	85
R4	5	1	4	1	5	3	5	1	5	1	92.5
R5	4	2	4	2	4	1	5	1	4	1	85
R6	5	1	5	1	3	2	4	1	4	2	85
R7	4	2	4	1	3	3	5	1	4	1	80
R8	5	1	5	1	5	1	5	1	4	2	95
R9	4	1	4	1	5	1	5	1	3	1	90
R10	5	2	4	1	5	1	4	1	5	1	92.5
Nilai SUS											<b>88</b>

#### 4. Kesimpulan

Dari hasil pengukuran menggunakan metode single ease question (SEQ), didapatkan nilai sebesar 88%, yang menunjukkan bahwa sebagian besar pengguna merasa sistem informasi yang dirancang dengan metode UCD mudah digunakan. Selain itu, pengukuran usability sistem juga menunjukkan nilai sebesar 87.25%, mengindikasikan tingkat kepuasan yang tinggi dari pengguna terhadap usability sistem yang telah dirancang. Hasil pengukuran ini memberikan gambaran yang positif tentang efektivitas metode UCD dalam merancang sistem informasi yang memperhatikan kebutuhan dan preferensi pengguna secara optimal. Kemudahan pengguna yang tinggi, sebagaimana tercermin dari nilai SEQ dan pengukuran usability sistem, merupakan indikator bahwa desain sistem telah berhasil mengakomodasi pengguna dengan baik. Dengan demikian, hasil pengukuran ini memberikan dukungan yang kuat terhadap keberhasilan penerapan metode UCD dalam merancang sistem informasi pemasok barang rongsokan. Selain sebagai indikator kepuasan pengguna, hasil ini juga dapat menjadi dasar untuk implementasi perbaikan dan peningkatan lanjutan pada usability sistem di masa depan, demi terus meningkatkan pengalaman pengguna yang lebih baik lagi

#### Referensi

- [1] User Interface Engineering, Inc. ([www.usability.gov](http://www.usability.gov), 2004), "The Role of User-Centered Design in Improving Software Usability: A Review of Current Practices and Future Directions."
- [2] R. Ones and T. Smith, "Metode Penelitian dalam Ilmu Sosial," Jurnal Penelitian Sosial, vol. 10, pp. 25-39

- [3] Pratamaa, A., Sholvaa, Y., & Irwansyah, M. A. (2023). Aplikasi Jual Beli Barang Rongsok Berbasis Jarak Menggunakan Fitur Location Based Service. *Juristi*,
- [4] J. Smith and L. Johnson, "Metodologi Penelitian: Panduan Praktis," *Jurnal Metodologi Penelitian*, vol. 5, pp. 68-82
- [5] A. Smith and B. Johnson, "Understanding Use Cases: A Practical Guide," *Journal of Software Engineering*, vol. 8, pp. 45-58
- [6] C. Brown and M. Davis, "Database Design Principles: A Comprehensive Approach," *Journal of Database Management*, vol. 12, pp. 102-115
- [7] T. Johnson and D. Smith, "Understanding Program Implementation: Best Practices and Guidelines," *Journal of Software Development*, vol. 15, pp. 78-91
- [7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms," 3rd ed. MIT Press, 2009.
- [8] User Interface Engineering, Inc. ([www.usability.gov](http://www.usability.gov), 2004), "The Role of User-Centered Design in Improving Software Usability: A Review of Current Practices and Future Directions."
- [9] User Interface Engineering, Inc. ([www.usability.gov](http://www.usability.gov), 2004), "The Role of User-Centered Design in Improving Software Usability: A Review of Current Practices and Future Directions."

# Analisis Dan Visualisasi Data Penjualan Pada NYC Property Menggunakan EDA

Meliana Endang Nyimas Lisna<sup>a1</sup>, Apriade Voutama<sup>a2</sup>

<sup>a</sup>Universitas Singaperbangsa Karawang; Jl. HS. Ronggowaluyo, Telukjambe Timur, 41363, Jawa Barat, Indonesia

<sup>1</sup>[melianaendang25@gmail.com](mailto:melianaendang25@gmail.com)

<sup>2</sup>[apriade.voutama@staff.unsika.ac.id](mailto:apriade.voutama@staff.unsika.ac.id)

## Abstract

Di tengah perkembangan pesat teknologi digital, nilai data penjualan semakin meningkat bagi pelaku bisnis, terutama dalam sektor industri real estate. Penjualan properti di New York City (NYC) menjanjikan peluang keuntungan besar namun juga menimbulkan risiko yang signifikan. Penelitian ini bertujuan untuk menerapkan Metode *Exploratory Data Analysis* (EDA) guna mendalami tren penjualan properti di NYC pada periode 2016-2017. Data yang digunakan diambil dari dataset Kaggle yang mencakup beragam variabel seperti lokasi, jenis properti, harga, dan tanggal penjualan. Proses analisis meliputi langkah-langkah pemahaman bisnis, pembersihan data, EDA, serta penyusunan insight dan rekomendasi. Temuan analisis menunjukkan adanya penurunan signifikan dalam penjualan pada bulan Juni 2017 sebesar 20,4%. Properti *residential* lebih diminati daripada properti *commercial* di berbagai wilayah, dengan properti *residential* unit dalam kelas pajak 2 menunjukkan peningkatan signifikan dalam penjualan. Analisis juga menyoroti fluktuasi pasar properti yang mencolok di berbagai wilayah Kota New York serta penurunan signifikan dalam penjualan unit *commercial* dari bulan ke bulan. Rekomendasi bisnis yang dihasilkan termasuk peningkatan strategi pemasaran, peninjauan ulang kebijakan perpajakan properti, dan fokus pada pengembangan apartemen besar yang sesuai dengan preferensi pasar.

**Keywords:** *Exploratory Data Analysis, Property Industri, Sale*

## 1. Pendahuluan

Dalam era digital yang terus berkembang, data penjualan menjadi semakin berharga sebagai sumber daya yang tak ternilai bagi bisnis. Informasi tentang barang yang terjual, klien yang dilayani, pola penjualan, dan faktor-faktor lainnya memberikan wawasan yang mendalam tentang kinerja bisnis. Dengan banyaknya data yang tersedia, bisnis dari berbagai industri berusaha untuk memanfaatkannya guna meraih keunggulan kompetitif. Diprediksi bahwa jumlah data yang dikonsumsi secara global akan terus meningkat, mencapai lebih dari 180 zettabytes pada tahun 2025.[1] Anraeni

Di bidang real estate, di mana penjualan properti menjadi fokus utama, analisis data pasar menjadi krusial. Kesuksesan penjualan properti tidak hanya memengaruhi profitabilitas, tetapi juga nilai bisnis secara keseluruhan. Di kota seperti New York City (NYC), di mana pasar real estate sangat dinamis, pemahaman yang mendalam tentang tren penjualan, preferensi konsumen, dan faktor-faktor lainnya sangat penting.[2]

Dalam menjawab tantangan ini, metode *Exploratory Data Analysis* (EDA) muncul sebagai solusi yang menjanjikan. EDA memungkinkan analisis yang mendalam terhadap data penjualan, membantu mengidentifikasi tren penjualan tertinggi dan terendah, serta memahami faktor-faktor yang memengaruhi pola penjualan. Namun, sementara EDA telah terbukti efektif dalam berbagai konteks, penting untuk menyesuaikan pendekatan ini dengan kebutuhan dan karakteristik pasar *real estate* yang unik.[3]

Penelitian ini bertujuan untuk mengisi kesenjangan pengetahuan dalam analisis penjualan properti di NYC dengan menerapkan EDA pada data penjualan properti di lima wilayah di kota tersebut dari tahun 2016 hingga 2017. Dengan fokus pada peningkatan penjualan, penelitian ini akan mengidentifikasi tren penjualan yang signifikan, mempertimbangkan berbagai faktor seperti jenis properti, wilayah, bulan penjualan, dan harga.

Melalui pendekatan yang cermat terhadap *Big Data*, penelitian ini berharap dapat memberikan wawasan yang berharga bagi praktisi industri *real estate*, membantu mereka membuat keputusan yang lebih baik dalam strategi pemasaran, alokasi sumber daya, dan pengembangan properti. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan terhadap pemahaman kita tentang dinamika pasar real estate dan memperluas cakupan pengetahuan dalam analisis data penjualan properti.

## 2. Metode Penelitian

Metode penelitian yang diterapkan dalam penelitian ini adalah metode *Exploratory Data Analysis* (EDA), yang mana metode ini berfokus pada pengumpulan data dan analisis mendalam. Dalam penelitian ini, kami akan menggali lebih dalam tentang data penjualan properti di New York City (NYC) menggunakan berbagai tahapan analisis.



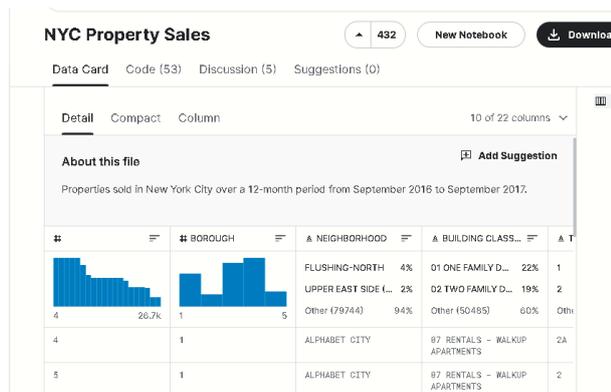
**Gambar 1** Alur Penelitian

### 2.1. Studi Literatur

Studi literatur merupakan bagian integral dari proses penelitian yang melibatkan penggunaan informasi dari berbagai sumber dan teknik pengumpulan data yang tersedia di perpustakaan. Ini mencakup referensi dari buku, hasil penelitian sebelumnya dalam bidang yang sama, artikel, catatan, dan berbagai jurnal yang relevan dengan permasalahan yang akan diteliti. Melalui studi literatur, peneliti dapat memperoleh pemahaman yang mendalam tentang lanskap pengetahuan terkait topik penelitian mereka, memperluas wawasan, dan mendukung pengembangan metodologi penelitian yang tepat.[3]

### 2.2. Pengumpulan Data

Metode pengumpulan data dalam penelitian ini menggunakan dataset *Kaggle NYC Property Sales* yang berisi informasi tentang penjualan properti di berbagai wilayah di New York City selama tahun 2016-2017. Dataset ini mencakup berbagai variabel, termasuk lokasi geografis, jenis properti, kategori bangunan, harga, tanggal penjualan, tahun dibangun, alamat, tingkat pajak dll.



Gambar 2 Datasets NYC Property Sales Pada Website Kaggle

### 2.3. Data Cleaning

Proses pembersihan data bertujuan untuk menghilangkan *noise* dan mengisi nilai yang hilang. Dalam konteks ini, pendekatan yang umum digunakan adalah dengan menghapus catatan yang mengandung *noise* atau nilai yang hilang.[3] Tahap ini meliputi pemilihan dataset, yang bertujuan untuk memilih data yang sesuai dengan kebutuhan penelitian, transformasi dataset, yang bertujuan untuk mengubah data melalui proses seperti pembersihan dan penyaringan, sehingga dapat divisualisasikan, dan verifikasi dataset, yang bertujuan untuk memastikan bahwa data tersebut akurat, bebas dari kesalahan, dan tidak menghasilkan bias dalam kumpulan data yang digunakan.[4]

### 2.4. Exploratory Data Analysis (EDA)

*Exploratory Data Analysis* (EDA) bertujuan untuk mendeteksi kesalahan, menemukan data yang sesuai, memeriksa asumsi, dan menentukan korelasi di antara variabel-variabel penjelas. Dalam konteks ini, EDA dianggap sebagai analisis data yang tidak termasuk inferensi dan pemodelan statistik. Analitika merupakan teknik penting bagi setiap profesi karena dapat meramalkan masa depan dan pola-pola tersembunyi.[5] Penggunaan EDA merupakan aspek kritis dari proses ilmiah dan satu area di mana EDA dapat bermanfaat adalah analisis akar penyebab (RCA), di mana data dieksplorasi secara grafis untuk memperoleh wawasan tentang kemungkinan akar penyebab. Wawasan-wawasan ini kemudian dievaluasi secara detail. Hipotesis-hipotesis terkait penyebab dan efek berdasarkan EDA dapat disebabkan oleh pengamatan celah dalam data, keberadaan *outlier*, atau pola-pola.[6].

### 2.5. Visualisasi Data

Visualisasi data adalah teknik yang digunakan untuk mempresentasikan informasi dalam bentuk visual dengan maksud meningkatkan pemahaman atau penalaran. Dengan memanfaatkan kemampuan persepsi visual, data yang kompleks dan banyak dapat diubah menjadi bentuk yang lebih mudah dipahami dan dianalisis. Tujuan utama dari visualisasi data adalah untuk mengidentifikasi pola dan tren, membuat data lebih mudah dipahami, melakukan analisis data yang lebih efektif, meningkatkan penjualan bisnis, dan mengkomunikasikan informasi dengan lebih efisien.[7]

## 3. Hasil dan Pembahasan.

### 3.1. Data Cleaning

Setelah mengimpor dataset NYC *Property Sales* ke dalam spreadsheet, langkah pertama dalam proses pembersihan data adalah melakukan pemeriksaan terhadap duplikasi data. Setelah dilakukan, hasilnya menunjukkan bahwa tidak ada duplikasi data dalam dataset tersebut. Langkah selanjutnya adalah melakukan pemeriksaan terhadap nilai-nilai yang hilang atau *missing values*.

Setelah identifikasi, langkah yang diambil adalah menghapus atau mengabaikan nilai-nilai yang hilang tersebut, sehingga memastikan konsistensi data yang digunakan dalam analisis.

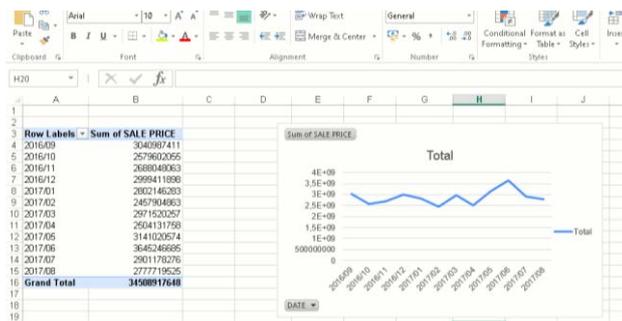
*Data cleaning* yang dilakukan juga mencakup perubahan tipe data tanggal penjualan agar sesuai dengan format yang diinginkan. Setelah penyesuaian, data tanggal penjualan telah berhasil diubah menjadi format yang sistematis, memudahkan analisis data secara lebih efektif.

BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASEMENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT
1	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40			311 EAST 13RD STREET, 17	
4	CHELSEA	10 COOPS - ELEVATOR APARTMENTS	2	748	1		04	321 WEST 24TH STREET, 30	
5	CHELSEA	10 COOPS - ELEVATOR APARTMENTS	2	751	1		04	290 6TH AVENUE, 33	
5	CHELSEA	10 COOPS - ELEVATOR APARTMENTS	2	751	1		04	290 6TH AVENUE, 14B	
11	CHELSEA	10 COOPS - ELEVATOR APARTMENTS	2	762	1		04	385 WEST 20TH STREET, 11D	
13	CIVIC CENTER	13 CONDOS - ELEVATOR APARTMENTS	2	173	183		R4	63 WORTH STREET	
19	GRAMERCY	10 COOPS - ELEVATOR APARTMENTS	2	875	25		04	82 RIVING PLACE, 3B	
20	GRAMERCY	17 CONDO COOPS	2	903	1127		R9	235 EAST 22ND STREET HALLW	
21	GRAMERCY	17 CONDO COOPS	2	903	1127		R9	235 EAST 22ND STREET 14K	
22	GREENWICH VILLAGE-CENTRAL	10 COOPS - ELEVATOR APARTMENTS	2	963	37		00	67 EAST 11TH STREET, 003	
23	GREENWICH VILLAGE-CENTRAL	10 COOPS - ELEVATOR APARTMENTS	2	968	19		04	20 EAST 19TH STREET, 19D	
26	GREENWICH VILLAGE-CENTRAL	17 CONDO COOPS	2	960	1203		R9	80 EAST 10TH STREET, 837	
30	HARLEM-CENTRAL	10 COOPS - ELEVATOR APARTMENTS	2	124	33		04	1 WEST 10TH STREET, 52	
36	HARLEM-CENTRAL	13 CONDOS - ELEVATOR APARTMENTS	2	2045	1180		R4	68 BRADHURST AVENUE	104
37	HARLEM-CENTRAL	17 CONDO COOPS	2	2045	1501		R9	130 BRADHURST AVENUE, 1205	
40	HARLEM-EAST	13 CONDOS - ELEVATOR APARTMENTS	2	1028	1238		R4	1815 3 AVENUE	10
41	HARLEM-EAST	15 CONDOS - 2 TO UNIT RESIDENTIAL	2C	1710	1208		R1	400 EAST 117 STREET	PH-0A
44	HARLEM-EAST	31 COMMERCIAL VACANT LAND		1835	127		V1	185 EAST 107 STREET	
45	INWOOD	10 COOPS - ELEVATOR APARTMENTS	2	2238	17		04	680 WEST 200TH STREET, 1F	
46	INWOOD	10 COOPS - ELEVATOR APARTMENTS	2	2242	322		04	50 PARK TERRACE EAST, 3L	
47	KIP'S BAY	13 CONDOS - ELEVATOR APARTMENTS	2	124	1248		R4	330 EAST 38TH STREET	204
48	LOWER EAST SIDE	13 CONDOS - ELEVATOR APARTMENTS	2	900	1014		R4	20 CLINTON STREET	0A
50	MANHATTAN VALLEY	13 CONDOS - ELEVATOR APARTMENTS	2	1944	1005		R4	478 CENTRAL PARK WEST	2B
51	MANHATTAN VALLEY	13 CONDOS - ELEVATOR APARTMENTS	2	1944	1005		R4	478 CENTRAL PARK WEST	2B
52	MIDTOWN-CBN	10 COOPS - ELEVATOR APARTMENTS	2	1212	39		04	185 EAST 57TH STREET, 18B	2B
53	MIDTOWN EAST	10 COOPS - ELEVATOR APARTMENTS	2	1323	1		04	306 EAST 40TH STREET, 3B	
54	MIDTOWN EAST	10 COOPS - ELEVATOR APARTMENTS	2	1361	30		04	12 BEEKMAN PLACE, 12E	
55	MIDTOWN EAST	10 COOPS - ELEVATOR APARTMENTS	2	1362	38		04	30 BEEKMAN PLACE, 3D	
56	MIDTOWN EAST	10 COOPS - ELEVATOR APARTMENTS	1	1369	54		04	199 EAST PARK STREET, 1E	

Gambar 3 Proses data cleaning yang dilakukan menggunakan spreadsheet

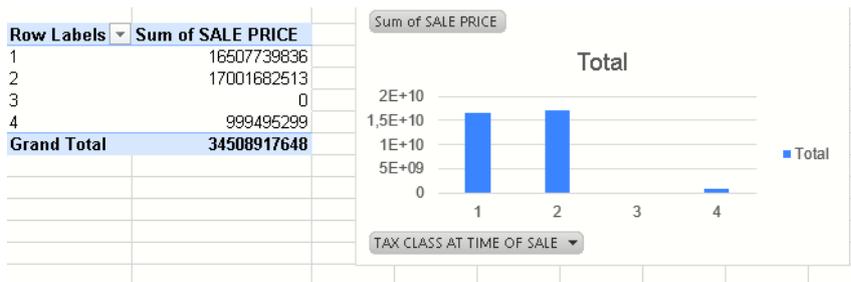
### 3.2. Exploratory Data Analysis (EDA)

*Exploratory Data Analysis* (EDA) adalah tahap kritis dalam proses pemahaman dan interpretasi data, terutama dalam konteks dataset yang kompleks seperti penjualan properti di Kota New York. Dalam penelitian ini, EDA menggunakan spreadsheet dan teknik pivot bertujuan untuk menggali wawasan mendalam tentang tren penjualan properti di berbagai wilayah dan lingkungan kota. Analisis juga fokus pada penurunan penjualan yang potensial, memungkinkan pengidentifikasian faktor-faktor penyebabnya dengan menggunakan teknik pivot untuk membedah data secara terinci.



Gambar 4 Pivot analisis tren penjualan

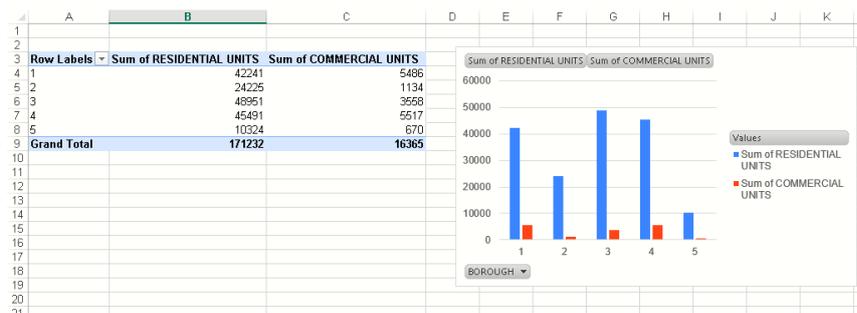
Selain itu, analisis perbandingan penjualan berdasarkan kelas pajak properti memberikan wawasan tambahan dalam memahami preferensi pasar dan karakteristik properti yang diminati. Dengan mendekonstruksi penjualan berdasarkan kelas pajak, kita dapat mengidentifikasi preferensi pembeli serta kecenderungan investasi yang mungkin terjadi.



Gambar 5 Pivot analisis penjualan berdasarkan kelas pajak

Pada gambar 5 dilakukan perbandingan yang terperinci antara penjualan properti *residential* dan *commercial* serta mengidentifikasi pola penjualan perbulan selama satu tahun, yang

memungkinkan identifikasi tren dan variabilitas yang mungkin terjadi dari waktu ke waktu. Di berbagai wilayah jenis *residential* terjual lebih banyak dibandingkan dengan *commercial* unit.



**Gambar 6** Pivot analisis perbandingan *Residential* dan *Commercial* unit

Pendekatan ini memungkinkan kita untuk menghasilkan pemahaman yang lebih komprehensif tentang dinamika pasar properti di Kota New York dan menganalisis peran faktor-faktor seperti lokasi, jenis properti, dan faktor pajak dalam memengaruhi keputusan pembelian.

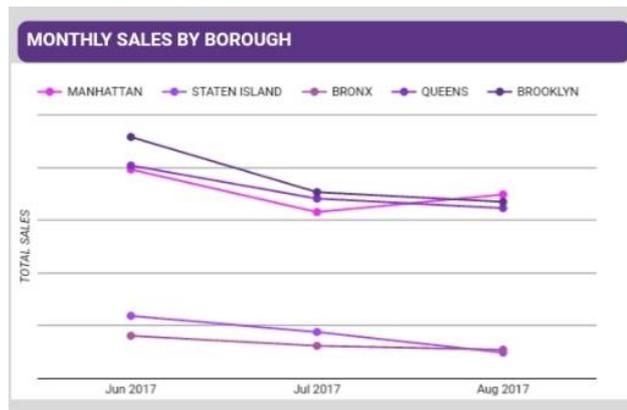
### 3.3. Visualisasi Data

Hasil analisis menunjukkan adanya penurunan signifikan dalam penjualan pada bulan Juni. Bulan Juni menandai awal dari penurunan signifikan, dengan penjualan turun sebesar 20,4%. Meskipun terjadi sedikit pemulihan pada bulan-bulan berikutnya, dengan penurunan hanya sebesar 0,4% dari Juli hingga Agustus, namun keseluruhan gambaran menunjukkan adanya tekanan yang berkelanjutan pada kinerja penjualan. Rata-rata penurunan sebesar 12,3% selama periode tersebut menggambarkan tantangan yang dihadapi perusahaan dalam menjaga konsistensi pendapatan.



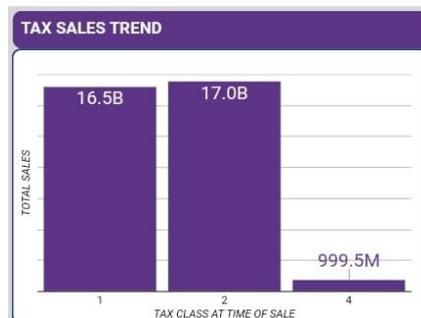
**Gambar 7** Grafik *monthly sales trend* 2016-2017

Pada Grafik ini menunjukkan bahwa pada bulan Juni 2017, terjadi peningkatan penjualan properti yang mencapai puncak di semua lima wilayah, mencatat rekor tertinggi. Namun, bulan-bulan berikutnya, yaitu Juli dan Agustus, mengalami penurunan yang cukup besar dalam aktivitas pasar. Secara khusus, Brooklyn memimpin dalam penjualan dengan total \$1,17 miliar, sementara Bronx mencatat penjualan terendah sebesar \$495 juta selama periode tiga bulan tersebut. Evaluasi ini menggarisbawahi fluktuasi pasar properti yang mencolok di wilayah tersebut selama rentang waktu yang diamati.



Gambar 8 Grafik *monthly sales* pada wilayah

Data menunjukkan bahwa terdapat perbedaan yang signifikan dalam tingkat penjualan antara kelas pajak, dengan kelas pajak 4 memiliki tingkat penjualan terendah hanya \$999,5 juta. Di sisi lain, properti *residential* unit dalam kelas pajak 2 menunjukkan peningkatan yang mencolok dalam penjualan sebesar \$17 miliar, mengungguli baik kelas pajak 1 maupun kelas pajak 4. Hal ini menegaskan bahwa apartemen besar menjadi pilihan yang diminati dalam pasar sewa.



Gambar 9 Grafik *tax sales trend*

Grafik menunjukkan tren menurun dalam penjualan unit *commercial* dari bulan ke bulan. Penjualan mencapai puncaknya pada bulan Oktober 2016 dengan 2955 unit terjual, tetapi mengalami penurunan yang signifikan pada bulan Agustus 2017, hanya mencatat penjualan sebanyak 625 unit. Hal ini mengindikasikan fluktuasi yang signifikan dalam permintaan pasar untuk unit *commercial* selama periode waktu tersebut.



Gambar 10 Grafik penurunan penjualan *commercial units*

### 3.4. Analisis Hasil dan Pembahasan

Analisis EDA terhadap data penjualan properti di Kota New York menunjukkan:

- Tren penjualan menunjukkan penurunan signifikan pada bulan Juni.
- Properti *residential* lebih diminati daripada properti *commercial* di berbagai wilayah.
- Properti *residential* unit dalam kelas pajak 2 menunjukkan peningkatan signifikan dalam penjualan.
- Terjadi fluktuasi pasar properti yang mencolok di berbagai wilayah Kota New York.
- Penjualan unit *commercial* mengalami penurunan signifikan dari bulan ke bulan.

Analisis ini memberikan pemahaman mendalam tentang dinamika pasar properti di Kota New York, pentingnya faktor-faktor seperti lokasi, jenis properti, dan faktor pajak dalam keputusan pembelian, serta memberikan landasan bagi pengambilan keputusan strategis di masa depan.

### 3.5. Rekomendasi Bisnis

Berdasarkan analisis yang telah dilakukan, rekomendasi-rekomendasi berikut dapat menjadi landasan yang kuat untuk mencapai tujuan tersebut:

- a. Dengan peningkatan strategi pemasaran yang lebih agresif dan adaptif, perusahaan dapat menjangkau calon pembeli dengan lebih efektif. Selain itu, dengan memahami kebutuhan dan preferensi pasar setempat secara mendalam, pengembangan strategi diferensiasi yang tepat akan memungkinkan perusahaan untuk menawarkan properti yang sesuai dengan ekspektasi pasar, memberikan nilai tambah yang signifikan, dan mengatasi persaingan di pasar properti yang ketat.
- b. Pemantauan dan peninjauan ulang kebijakan perpajakan properti juga menjadi kunci dalam merespons secara efektif terhadap dinamika pasar yang terus berubah. Kolaborasi yang erat dengan pemerintah setempat dan pemangku kepentingan lainnya dapat membantu perusahaan dalam mengusulkan dan menerapkan perubahan kebijakan yang mendukung pertumbuhan pasar properti dan meningkatkan daya beli pembeli.
- c. Terakhir, dengan fokus yang lebih besar pada pengembangan apartemen besar yang sesuai dengan preferensi pasar dalam pasar sewa, perusahaan dapat memanfaatkan potensi pertumbuhan yang signifikan dalam segmen ini. Melalui riset pasar yang teliti dan investasi yang tepat, perusahaan dapat menciptakan portofolio properti yang menarik dan menguntungkan, serta memberikan nilai tambah yang nyata bagi para pembeli dan penyewa.

Dengan menerapkan rekomendasi ini secara efektif, perusahaan properti akan dapat mengatasi tantangan yang dihadapi, meningkatkan kinerja bisnis mereka, dan memperkuat posisi mereka di pasar properti yang dinamis dan kompetitif di Kota New York.

## 4. Kesimpulan

Berdasarkan penelitian ini dapat disimpulkan, analisis EDA memberikan pemahaman yang mendalam tentang dinamika pasar properti di Kota New York, dengan mengidentifikasi tren, fluktuasi, dan faktor-faktor yang mempengaruhi penjualan properti yaitu:

- a. Terjadi penurunan signifikan dalam penjualan pada bulan Juni, dengan penurunan sebesar 20,4%. Meskipun ada pemulihan sedikit pada bulan-bulan berikutnya, gambaran keseluruhan menunjukkan tekanan berkelanjutan pada kinerja penjualan.
- b. Properti *residential* lebih diminati daripada properti *commercial* di berbagai wilayah. Ini menunjukkan bahwa ada kebutuhan yang lebih besar untuk properti *residential*.
- c. Properti *residential* unit dalam kelas pajak 2 menunjukkan peningkatan signifikan dalam penjualan, menunjukkan bahwa apartemen besar menjadi pilihan yang diminati dalam pasar sewa.
- d. Terjadi fluktuasi pasar properti yang mencolok di berbagai wilayah Kota New York, menunjukkan bahwa dinamika pasar sangat dipengaruhi oleh faktor-faktor seperti lokasi, jenis properti, dan faktor pajak.
- e. Penjualan unit *commercial* mengalami penurunan signifikan dari bulan ke bulan, menunjukkan fluktuasi yang signifikan dalam permintaan pasar untuk unit *commercial*.

Dengan memahami dinamika pasar ini, perusahaan dapat merumuskan strategi yang lebih efektif untuk meningkatkan penjualan dan memperkuat posisi mereka di pasar properti. Ini termasuk peningkatan strategi pemasaran, peninjauan ulang kebijakan perpajakan properti, dan fokus pada pengembangan apartemen besar yang sesuai dengan preferensi pasar. Dengan menerapkan rekomendasi ini, perusahaan properti akan dapat mengatasi tantangan yang dihadapi, meningkatkan kinerja bisnis mereka, dan memperkuat posisi mereka di pasar properti yang dinamis dan kompetitif di Kota New York.

## References

- [1] T. Z. Dessiaming, S. Anraeni, and S. Pomalingo, "College Academic Data Analysis Using Data Visualization," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1203–1212, 2022, doi: 10.20884/1.jutif.2022.3.5.310.

- [2] G. N. Ayuni and D. Fitriana, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>
- [3] D. T. Husni *et al.*, "Analisis Big Data Penjualan Video Games Menggunakan Eda," *J. Tek. Inf. dan Komput.*, vol. 5, no. 1, p. 43, 2022, doi: 10.37600/tekinkom.v5i1.517.
- [4] N. K. Afifah, R. F., Widiyanti, A. W., Zulfa, A. T., & Dewi, "GENDER DENGAN METODE EXPLORATORY DATA ANALYSIS ( EDA ) DAN VISUALISASI Analysis of Consumptive Behavior based on Gender Using Exploratory Data Analysis ( EDA ) and Visualization," no. December, 2022, [Online]. Available: [https://www.researchgate.net/publication/366185780\\_ANALISIS\\_PERILAKU\\_KONSUMTIF\\_BERDASARKAN\\_GENDER\\_DENGAN\\_METODE\\_EXPLORATORY\\_DATA\\_ANALYSIS\\_EDA\\_DAN\\_VISUALISASI\\_Analysis\\_of\\_Consumptive\\_Behavior\\_based\\_on\\_Gender\\_Using\\_Exploratory\\_Data\\_Analysis\\_EDA\\_and\\_Visu](https://www.researchgate.net/publication/366185780_ANALISIS_PERILAKU_KONSUMTIF_BERDASARKAN_GENDER_DENGAN_METODE_EXPLORATORY_DATA_ANALYSIS_EDA_DAN_VISUALISASI_Analysis_of_Consumptive_Behavior_based_on_Gender_Using_Exploratory_Data_Analysis_EDA_and_Visu)
- [5] R. Indrakumari *et al.*, "Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra," *J. Tek. Inf. dan Komput.*, vol. 3, no. 2, pp. 1–8, 2022, doi: 10.35957/jatisi.v9i3.2737.
- [6] M. Barsalou, P. M. Saraiva, and R. Henriques, "Exploring Exploratory Data Analysis: an Empirical Test of Run Chart Utility," *Manag. Syst. Prod. Eng.*, vol. 31, no. 4, pp. 442–448, 2023, doi: 10.2478/mspe-2023-0050.
- [7] Johannes Kurniawan, *Analisis Dan Visualisasi Data*, vol. 3, no. 1. 2023. [Online]. Available: <https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf>
- [8] T. N. Muthmainnah and A. Voutama, "Volume 6 ; Nomor 2," *Juli*, vol. 6, pp. 463–471, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [9] Kaggle, "Dataset NYC Property Sales." [Online]. Available: <https://www.kaggle.com/datasets/new-york-city/nyc-property-sales/>

# Klusterisasi Fitur Tanya Dokter Pada Website Alodokter Menggunakan Metode Latent Dirichlet Allocation

I Putu Fajar Tapa Mahendra<sup>a1</sup>, I Gusti Ngurah Anom Cahyadi Putra<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana  
Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>ftapamahendra@gmail.com  
<sup>2</sup>anom.cp@unud.ac.id

## Abstract

*Digital advancements have change information-seeking behaviors, particularly in health inquiries for the people. The Alodokter website's "Tanya Dokter" feature facilitates an easy connections with medical experts to ask question regarding health. The posed questions tend to mirror evolving health trends and public misunderstandings regarding health issues. Manual analysis of data in "Tanya Dokter" features proves challenging, prompting the use of Latent Dirichlet Allocation (LDA) topic modeling. This research categorizes Alodokter topics, unveiling common health issues. The optimal model reveals 8 clusters with diverse topic distributions. Validation metrics using coherence score with 0.258481 as the highest value affirm the model's efficacy. Optimal outcomes stem from combination of parameter such as 8 topics, alpha 0.02, and beta 0.02. This study may offers Alodokter and healthcare providers an informed perspective on an accessible approach to categorize health questions effectively using Topic Modelling Latent Dirichlet Allocation.*

**Keywords:** Alodokter, Tanya Dokter, Topic Modelling, Latent Dirichlet Allocation, Health

## 1. Pendahuluan

Perkembangan teknologi digital dimasa saat ini telah mengubah cara orang-orang dalam mencari informasi. Informasi yang dulunya hanya didapatkan ketika membeli dan membaca suatu buku kini dapat diakses dengan mudah melalui internet. Begitu juga dalam pencarian informasi yang membutuhkan pendapat seorang ahli. Dengan bantuan teknologi digital, informasi dari pakar yang dulunya membutuhkan kita untuk bertemu langsung dengan pakarnya, kini dapat dilakukan secara online melalui suatu platform. Salah satu contoh platform yang dimana masyarakat dapat bertanya kepada seorang pakar atau ahli adalah Alodokter dengan fiturnya yaitu Tanya Dokter.

Melalui Tanya Dokter, masyarakat dapat dengan mudah bertanya mengenai masalah kesehatan yang mereka alami dengan pakar terkait. Banyaknya pertanyaan yang dilontarkan pada website ini dapat mencerminkan dinamika kesehatan masyarakat, khususnya masyarakat Indonesia. Tidak berhenti sampai disitu, pertanyaan yang dilontarkan juga dapat mencerminkan miskonsepsi tentang kesehatan pada masyarakat serta kesenjangan pengetahuan terkait kesehatan. Kemampuan dalam menganalisa data pada Tanya Dokter dapat memberikan pemahaman lebih mendalam mengenai tren kesehatan yang dialami masyarakat serta membantu dalam menyesuaikan kontern terkait yang dibutuhkan masyarakat.

Namun mengingat banyaknya data yang ada pada fitur Tanya Dokter tersebut, menganalisa data secara manual akan menjadi tantangan yang sangat sulit. Metode analisis secara tradisional, sering kali tidak cocok dan tidak efisien untuk mendapatkan informasi yang berarti dari analisa data pada fitur Tanya Dokter. Untuk membantu dalam mengatasi masalah tersebut, kita dapat menggunakan metode *topic modelling* untuk membantu dalam mengekstrak topik inti pada data Tanya Dokter. Salah satu metode *topic modelling* yang dapat digunakan adalah *Latent Dirichlet Allocation*.

Topic modelling secara statistic bekerja dengan cara mengeksplorasi dokumen yang diberikan dan merepresentasikan mereka sebagai kumpulan istilah yang sering muncul bersamaan dalam dokumen [1]. Menurut Gurcan, LDA pada topic modelling adalah suatu metode pendekatan yang generatif yang digunakan untuk menemukan pola semantic yang ada pada suatu korpus dokumen yang relatif tidak terstruktur [2]. Menggunakan LDA, pendekatan secara sistematis dapat dilakukan untuk menyaring teks

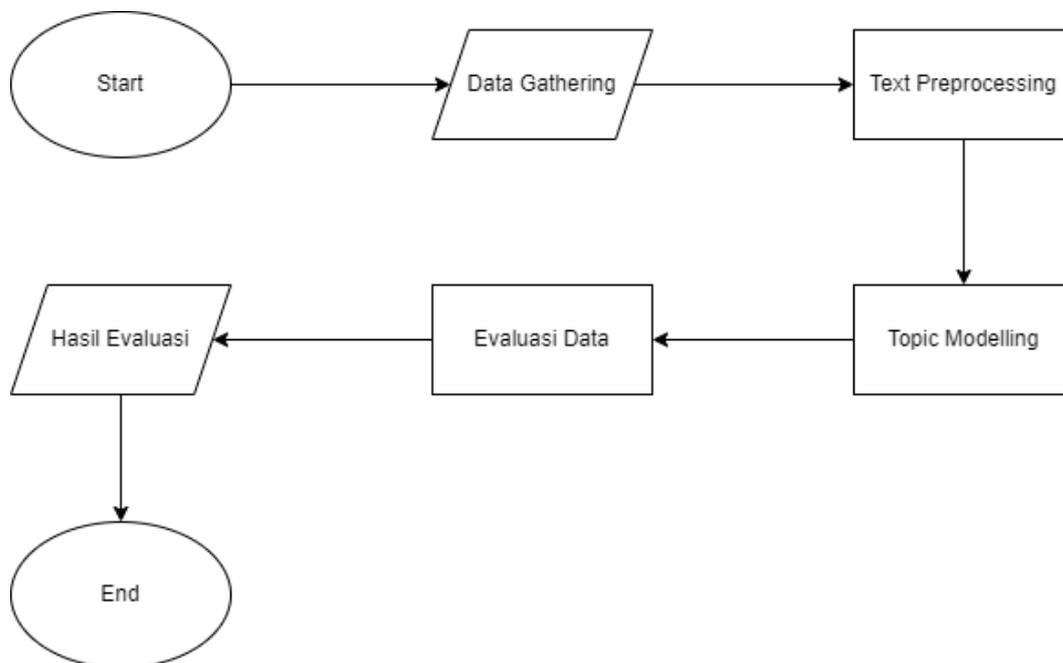
menjadi topik yang dapat dikenali, memungkinkan dilakukannya identifikasi tren kesehatan yang ada pada fitur Tanya Dokter. Dengan menggunakan LDA, akan diasumsikan terdapat distribusi berbagai macam topik pada suatu kumpulan dokumen, dimana setiap dokumen direpresentasikan sebagai sebaran dari topik dan setiap topik sebagai distribusi kata-kata di dalam dokumen [3]. Menggunakan sebaran topik yang dihasilkan, akan direpresentasikan dalam bentuk probabilitas dengan probabilitas tertinggi yang akan menjadi faktor utama dalam menentukan termasuk kluster mana suatu data teks tersebut.

Melalui penelitian ini, peneliti berharap dengan menggunakan metode *topic modelling Latent Dirichlet Allocation* ini dapat membantu dalam membuat suatu model yang dapat menunjukkan masalah kesehatan yang sering muncul menggunakan klusterisasi pada data dari website Alodokter pada fitur Tanya Dokter. Dengan melakukan hal tersebut peneliti berharap dapat menyediakan Alodokter serta komunitas penyedia layanan kesehatan dengan pandangan yang mendalam untuk membentuk pendekatan yang lebih terinformasi dan ramah kepada pengguna aplikasi atau platform kesehatan digital atau membantu pengguna yang masih bingung dalam menentukan topik yang ingin ditanyakan pada website Alodokter dalam mengkategorikan pertanyaan mereka sehingga dapat ditangani oleh ahli yang tepat.

## 2. Metode Penelitian

### 2.1. Alur Penelitian

Berikut adalah alur dari penelitian yang dilakukan:



Gambar 1. Alur Penelitian

### 2.2. Pengumpulan Data

Pengumpulan data teks pada fitur Tanya dokter pada website Alodokter dilakukan dengan menggunakan teknik *web scrapping*. Dalam proses pengumpulan datanya, digunakan bantuan *library selenium* yang mana *library selenium* ini dapat diakses melalui bahasa pemrograman *Python*. Pengambilan data dilakukan dalam 10 tahap pengulangan dimana setiap tahap mengambil sekitar 120-135 data dan langsung disimpan dalam bentuk *.tsv (tab separated value)*. Hal ini dilakukan untuk mengurangi resiko kehilangan data yang sudah terambil pada saat pengambilan yang diakibatkan baik oleh koneksi internet maupun masalah pada perangkat keras. dari data yang telah didapatkan tersebut, data yang sebelumnya terbagi menjadi 10 bagian kemudian disatukan dan kemudian akan dilakukan pembersihan data duplikat. Setelah seluruh proses dilakukan, didapatkan sisa data yang berjumlah

1099. Berikut contoh data yang didapatkan menggunakan teknik *web scrapping* pada fitur Tanya Dokter yang ada pada website Alodokter dapat dilihat pada tabel 1:

**Table 1** Contoh Data

No	Question	Answer
1	"Teman saya ada yg abis dikecup lehernya sama cowoknya sampe masih ada bekasnya dok, dia malu, cara apa untuk menghilangkan bekas kecupan di lehernya dok?"	['Alo, terimakasih atas pertanyaannya.\nKecupan yang cukup kuat di area leher memang bisa menyisakan bekas yang cukup lama hilangnya, misalnya berupa ruam kemerahan, memar, lecet, dan sebagainya. Tergantung keparahannya, bekas kecupan ini bisa memerlukan waktu beragam untuk sembuh, bisa singkat, bisa juga lambat. Saran kami, coba Anda arahkan rekan Anda untuk:\nMandi yang rajin dan bersihkan lehernya dengan baik\nKompres dingin bekas kecupan di lehernya\nOleskan pelembab ke area kulit lehernya\nJangan berlebihan menggosok atau memanipulasi bekas kecupan di leher\nMinum banyak air putih, makan makanan yang mengandung kaya vitamin C serta antioksidan lainnya\nApabila ia merasa sangat terganggu dengan bekas kecupan di lehernya tersebut, Anda bisa arahkan ia untuk periksa langsung ke dokter ya..\nSemoga membantu.']
2	"Dok, apakah disini saya bisa minta resep dokter untuk obat batuk pilek anak? Ini untuk anak saya umur 4 tahun yg batuk pilek sudah seminggu belum sembuh dok"	['Alo, selamat siang\nbatuk pilek ialah merupakan keluhan yang sering menimpa anak-anak, dimana keluhan batuk pilek umumnya disebabkan oleh infeksi virus atau bakteri, namun paling sering akibat infeksi virus yang dimana dapat sembuh dengan sendirinya dalam waktu 7-10 hari jadi tidak perlu penanganan khusus. namun memang keluhan batuk pilek ini mengganggu aktivitas dan tidur anak. untuk membuat anak terasa nyaman dan meredakan keluhannya ada beberapa tips yang bisa bunda lakukan seperti :\npastikan anak tidur cukup\nsaat anak tidur posisikan kepala lebih tinggi jadi gunakan bantal dikepala saat tidur\nberikan anak lebih banyak air putih bila perlu\nnagat\nberikan madu pada anak setengah sendok dicampur dengan teh hangat sebelum tidur\noleskan balsem khusus anak di bagian dada, leher dan punggungnya\nkonsumsi obat batuk pilek sirup khusus anak yang dapat bunda beli bebas di apotik\njika lebih dari 10 hari keluhan

		<p>batuk pilek tidak sembuh, atau adanya kondisi demam tinggi diatas 38 derajat serta adanya sesak napas maka segera temui dokter secara langsung untuk mendapatkan penanganan lebih lanjut dengan tepat.\nsemoga dapat membantu']</p>
--	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### 2.3. Text Preprocessing

*Text preprocessing* adalah tahap dimana teks asli diubah dengan menghilangkan *unuse textual* yang tidak diperlukan dalam pengolah yang lebih lanjut [4]. Pada tahap ini, data teks yang dikumpulkan akan diolah agar dapat diproses pada tahap pelatihan model. Berikut adalah tahap-tahap dalam *text preprocessing*:

a. *Case Folding*

*Case folding* adalah tahap *preprocessing* dimana setiap kata pada data teks akan disetarakan format hurufnya menjadi huruf besar atau huruf kecil. Pada penelitian ini seluruh huruf akan diubah menjadi huruf kecil

b. Penghapusan tanda baca

Pada tahap ini seluruh tanda baca yang terdapat pada data teks akan dihapus. Tanda baca dihapus karena dianggap tidak memiliki nilai makna dalam ekstraksi informasi pada data teks

c. *Tokenization*

Pada tahap *tokenization*, data teks yang ada akan dipenggal menjadi perkata dan kemudian disimpan dalam bentuk array. Hal ini dilakukan untuk mempermudah pengolahan data teks

d. *Stopword removal*

Pada tahap *stopword removal*, beberapa kata yang dianggap tidak memiliki makna akan dihapus. Kata-kata yang dihapus biasanya berupa kata sambung seperti “dan”, “lalu” dan “setelah”.

e. *Stemming*

*Stemming* adalah tahap dimana kata yang ada pada data teks, diubah bentuknya menjadi kata dasar. Hal ini dilakukan untuk agar sistem yang dibuat dapat memahami makna dokumen dengan lebih baik

f. *Bag of word*

*Bag of word* adalah tahap dimana kata-kata yang ada pada seluruh korpus akan dihitung frekuensi kemunculannya. *Bag of word* membantu dalam merepresentasikan data teks menjadi data angka yang lebih mudah dipahami oleh komputer

g. *Term Weighting*

*Term weighting* adalah tahap dimana setiap kata yang ada pada data teks pada seluruh korpus akan diberikan bobot. Pada pemberian bobot pada kata ini, akan digunakan metode *TF-IDF*. *TF-IDF* adalah metode pemberian bobot yang menggunakan nilai *TF* (*Term frequency*) yang berdasarkan frekuensi kemunculan kata pada dokumen dengan *IDF* (*Inverse Document Frequency*) yang bersarkan pada kemunculan kata pada suatu kumpulan dokumen. Berikut formula *TF-IDF*:

$$TF - IDF = TF * IDF \quad [1]$$

$$TF = \frac{\text{jumlah kemunculan suatu kata } (x)}{\text{jumlah kata dalam dokumen}} \quad [2]$$

$$IDF = \log \frac{\text{jumlah dokumen}}{\text{jumlah dokumen dengan kata } (x)} \quad [3]$$

### 2.4. Topic Modelling

Dalam *machine learning*, *topic modelling* adalah salah satu bentuk *unsupervised machine learning* yang menyaring data teks agar dapat mengidentifikasi pola kemunculan kata yang menandakan topik yang mendasari data text tersebut [5]. Dalam *topic modelling* ini, setiap dokumen dalam suatu

korpus direpresentasikan sebagai kombinasi yang terdiri dari beberapa topik, sedangkan topik itu sendiri direpresentasikan sebagai kombinasi yang terdiri dari beberapa kata.

Tujuan representasi ini dilakukan untuk dapat menentukan distribusi penyebaran topik pada suatu dokumen dan distribusi kata pada suatu topik. Hal ini dilakukan karena umumnya topic modelling menggunakan pendekatan probabilistik, yang artinya saat model dari topic modelling mengatakan bahwa suatu dokumen terdiri 20% topik X, 30% topic Y dan 50% topik Z, ini mencerminkan probabilitas berdasarkan kata yang ada pada dokumen. Salah satu metode yang termasuk ke dalam kategori *topic modelling* adalah LDA (*Latent Dirichlet Allocation*).

*Latent Dirichlet Allocation* adalah sebuah pendekatan yang didasarkan pada teorema definetti, dimana metode ini digunakan untuk menangkap beberapa topik yang tersebar diantara beberapa kumpulan dokumen [5]. Konsep utama dari metode *Latent Dirichlet Allocation* ini dalam *melakukan topic modelling* adalah dengan merepresentasikan topik sebagai campuran dari beberapa topik yang berbeda, dimana topik itu sendiri direpresentasikan oleh distribusi kata pada dokumen. Berikut cara kerja dari LDA:

- a. Inisiasi  
Jumlah topik yang akan diekstrak pada saat topic modelling akan ditentukan pada tahap ini
- b. *Random Assignment*  
Setelah itu, setiap kata akan dimasukkan ke salah satu topik ada. Jumlah topik yang ada sesuai dengan inisiasi yang dilakukan
- c. *Iterative Reassignment*  
Lalu akan dilakukan iterasi berkali-kali dengan aturan
  - Iterasi untuk setiap dokumen  $d$ 
    - Iterasi kata  $w$  pada document  $d$
    - Iterasi untuk setiap topik  $t$ 
      - Hitung kedua probabilitas:
        - $P(t/d)$ : proporsi kata pada suatu dokumen yang dimasukkan pada topic  $t$
        - $P(w/t)$ : proporsi suatu topik  $t$  diberikan terhadap seluruh dokumen  $d$  yang berasal dari kata  $w$
      - Pengkategorian ulang kata  $w$  kedalam topic  $t$  berdasarkan probabilitas yang didapatkan dari hasil  $P(t/d) * P(w/t)$
- d. *Convergence*  
Tahap dimana algoritma sudah mulai stabil dalam memberikan topik pada suatu kata setiap kali iterasi dilakukan
- e. Output  
Setiap dokumen memiliki distribusi berdasarkan topik, dan setiap topik memiliki distribusi kata.

## 2.5. Parameter Tuning

Parameter Tuning adalah tahap dimana dilakukan beberapa modifikasi pada variabel model sehingga menghasilkan hasil yang berbeda. Tujuan dari parameter tuning ini adalah untuk membantu dalam menemukan model yang paling baik dalam melakukan *topic modelling*. Pada penelitian ini model akan dilatih ulang dengan menggunakan kombinasi parameter tuning yang berbeda. Adapun parameter yang akan dijadikan sebagai parameter tuning:

- a. Jumlah topik, dari nilai 5 sampai 10
- b. Nilai  $\alpha$  yang merepresentasikan penyebaran topik dalam suatu dokumen dengan rentang nilai 0.1, 0.2, 0.3, 0.4 dan 0.5
- c. Nilai  $\beta$  yang merepresentasikan penyebaran kata yang merepresentasikan suatu topik dengan rentang nilai 0.1, 0.2, 0.3, 0.4 dan 0.5

## 2.6. Evaluasi Model dengan Coherence Score

*Coherence score* adalah salah satu metode evaluasi yang umum digunakan untuk mengevaluasi model topic modelling. Metrik ini menghitung konsistensi dari kata pada suatu topik untuk mengevaluasi apakah suatu topik dapat diinterpretasi dengan cara menghitung kemiripan semantik dari kata yang ada pada suatu topik [6]. Suatu statement atau kalimat dapat dikatakan koheren, apabila setiap katanya

saling mendukung satu sama lain [7]. Dalam evaluasi menggunakan metode *coherence score*, ada beberapa tahapan yang harus dilalui. Adapun tahapan yang dilalui:

- a. *Segmentasi*  
Pemenggalan kumpulan kata atau kalimat menjadi kata tunggal
- b. *Estimasi Probabilitas*  
Estimasi probabilitas dari subset menggunakan data korpus yang besar
- c. *Confirmation Measure*  
Menghitung nilai menggunakan probabilitas untuk mengetahui indikasi seberapa mungkin suatu subset kata dilihat bersama.
- d. *Aggregation*  
Kemudian nilai dari *confirmation measure* akan di melalui proses agregasi untuk mendapatkan satu nilai *coherence score*.

Formula *coherence score*:

$$Coherence_{NPMI} = \frac{1}{N} \sum_{i=1}^N \frac{\log \left( \frac{p(w_i, w_j) + \epsilon}{p(w_i) \times p(w_j)} \right)}{-\log(p(w_i, w_j) + \epsilon)} \quad [4]$$

$p(w_i)$  = probabilitas kata  $w_i$  muncul dalam korpus

$p(w_j)$  = probabilitas kata  $w_j$  muncul dalam korpus

$p(w_i, w_j)$  = probabilitas 2 kata muncul bersamaan dalam suatu topik atau konteks

### 3. Hasil dan Pembahasan

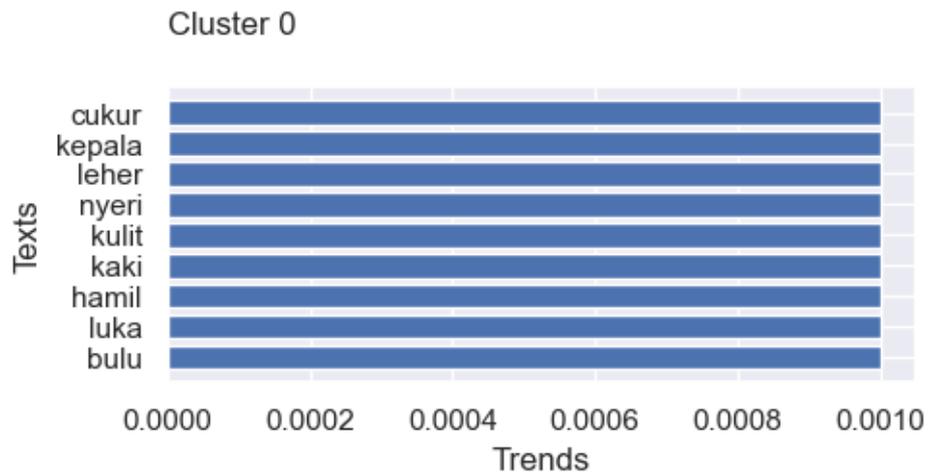
#### 3.1. Hasil *Topic Modelling*

Model LDA dilatih dengan kombinasi parameter tuning yang disebutkan diatas dengan jumlah 150 kombinasi. Dari seluruh kombinasi yang ada, didapatkan kombinasi 8 topik, 0.2 *alpha* dan 0.2 *beta* sebagai kombinasi terbaik dengan nilai *coherence score* 0.258481. Berikut distribusi topik pada kluster LDA dari model terbaik dapat dilihat pada tabel 2:

**Tabel 2.** Hasil Kluster

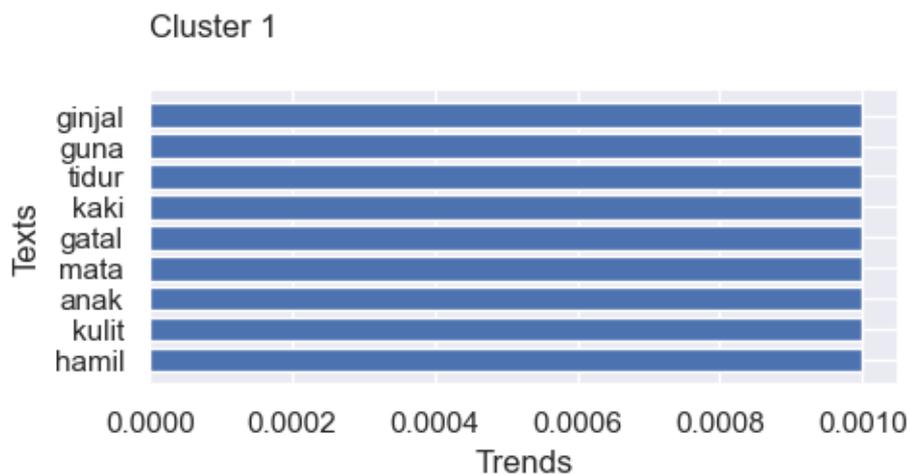
Kluster Topik	Kata
Kluster 0	bulu, luka, hamil, kaki, kulit, nyeri, leher, kepala, cukur, kolesterol
Kluster 1	hamil, kulit, anak, mata, gatal, kaki, tidur, guna, ginjal, wajah
Kluster 2	kulit, hamil, rambut, wajah, vagina, makan, anak, infeksi, gatal, luka
Kluster 3	gigi, luka, nyeri, sakit, batuk, demam, makan, kulit, mata, obat
Kluster 4	rambut, telinga, makan, darah, konsumsi, kulit, obat, batuk, sakit, ganggu
Kluster 5	mata, kulit, urat, hamil, makan, asam, darah, menstruasi, nyeri, sakit
Kluster 6	payudara, kaki, hamil, gigi, kulit, ibu, makan, jerawat, bayi, nyeri

Berikut visualisasi dari hasil klusterisasi beserta bobot dari katanya:



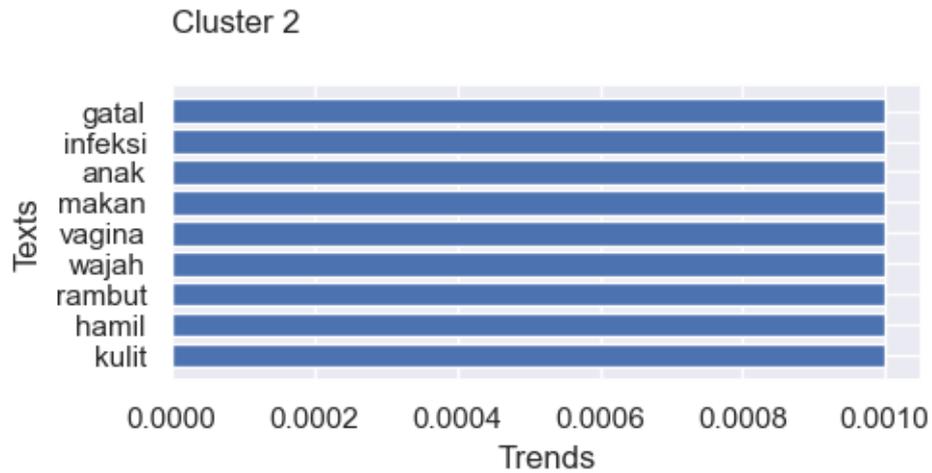
**Gambar 2.** Kluster 0

Pada gambar 2 dapat dilihat kluster 0 terbentuk dari kata “cukur”, “kepala”, “leher”, “nyeri”, “kulit”, “kaki”, “hamil”, “luka”, dan “bulu”. Pada kluster ini, seluruh kata memiliki bobot trend yang sama.



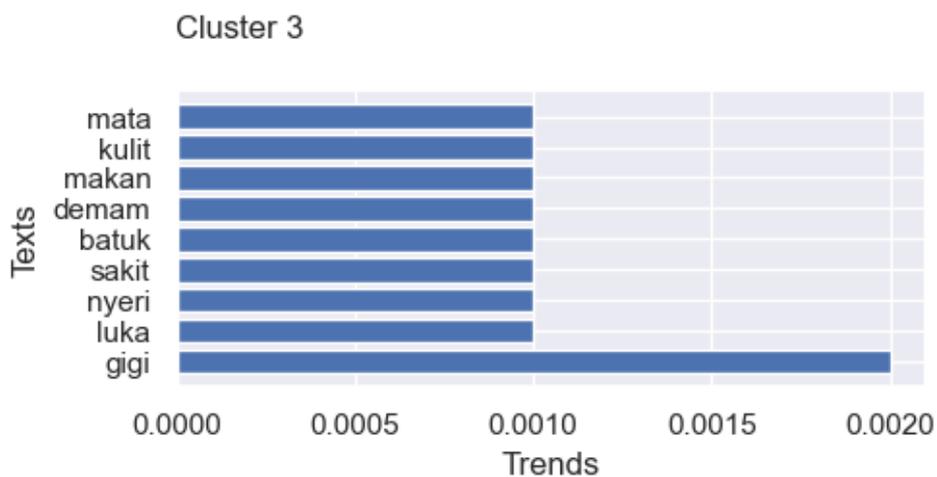
**Gambar 3.** Kluster 1

Pada gambar 3 dapat dilihat kluster 1 terbentuk dari kata “ginjal”, “guna”, “tidur”, “kaki”, “gatal”, “mata”, “anak”, “kulit” dan “hamil. Pada kluster ini, seluruh kata pada kluster memiliki bobot trend yang sama.



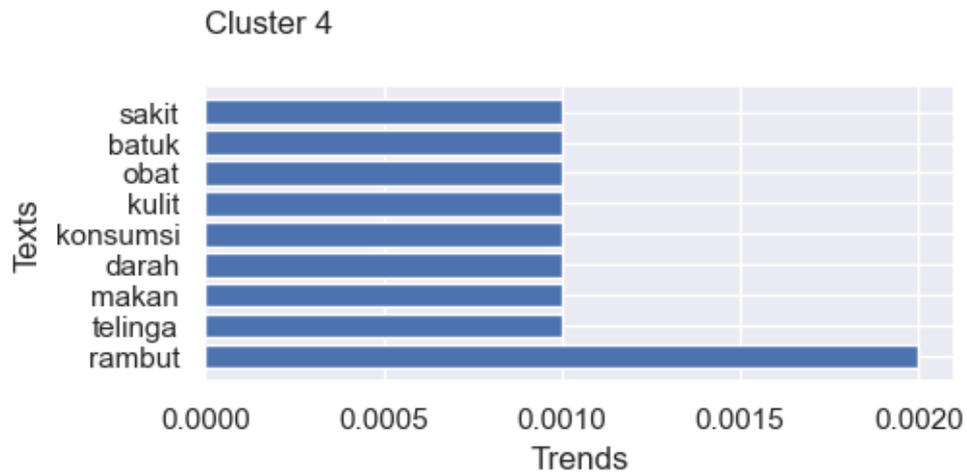
**Gambar 4.** Kluster 2

Pada gambar 4 dapat dilihat kluster 2 terbentuk dari kata “gatal”, “infeksi”, “anak”, “makan”, “vagina”, “wajah”, “rambut”, “hamil” dan “kulit”. Pada kluster 2, setiap kata pembentuknya memiliki bobot yang sama



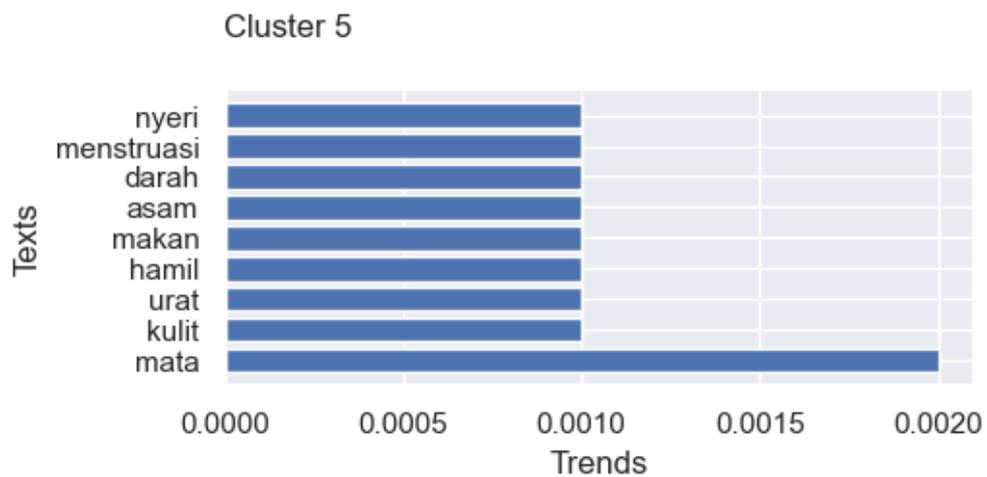
**Gambar 5.** Kluster 3

Pada gambar 5 dapat dilihat kata pembentuk dari kluster 3, yaitu “mata”, “kulit”, “makan”, “demam”, “batuk”, “sakit”, “nyeri”, “luka” dan “gigi”. Pada kluster 3 ini, kata “gigi” memiliki bobot yang paling tinggi



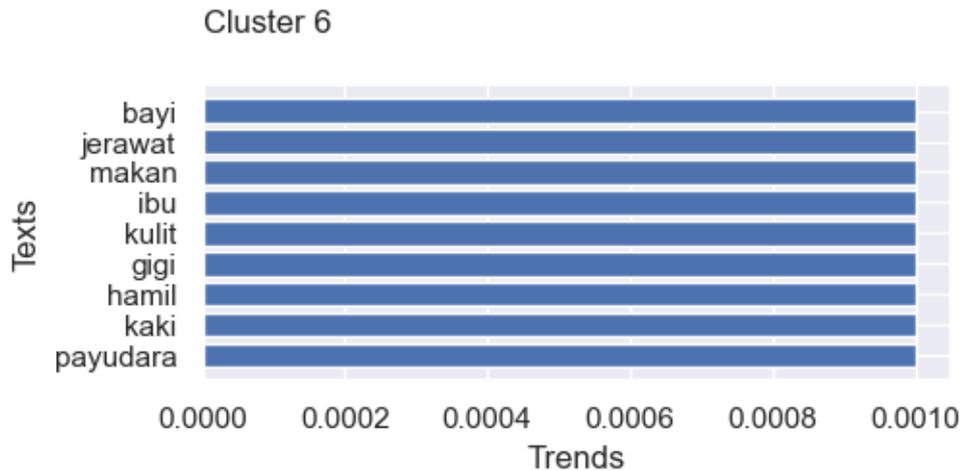
**Gambar 6.** Kluster 4

Pada gambar 6 dapat dilihat kata pembentuk dari kluster 4, yaitu “sakit”, “batuk”, “obat”, “kulit”, “konsumsi”, “darah”, “makan”, “telinga” dan “rambut”. Pada kluster 4 ini, kata “rambut” memiliki bobot yang paling tinggi



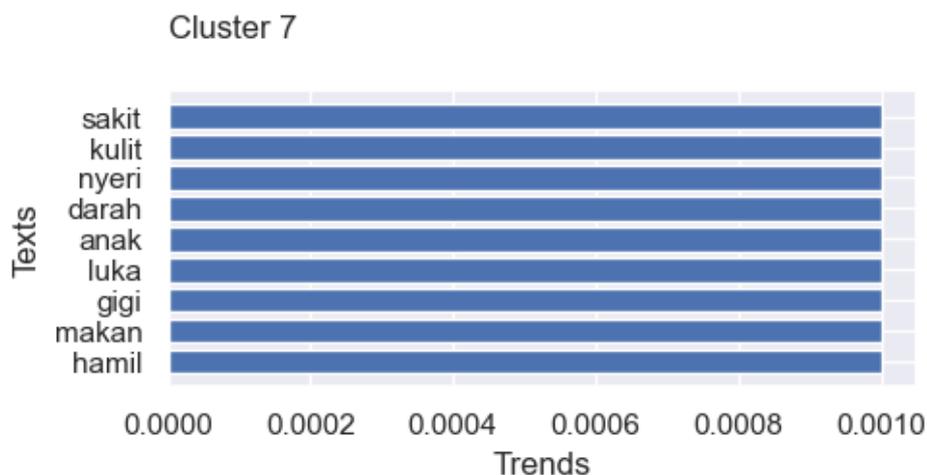
**Gambar 7.** Kluster 5

Pada gambar 7 dapat dilihat kata pembentuk dari kluster 5, yaitu “nyeri”, “menstruasi”, “darah”, “asam”, “makan”, “hamil”, “urat”, “kulit” dan “mata”. Pada kluster 5 ini, kata “mata” memiliki bobot yang paling tinggi



**Gambar 8.** Kluster 6

Pada gambar 8 dapat dilihat kata pembentuk dari kluster 6, yaitu “bayi”, “jerawat”, “makan”, “ibu”, “kulit”, “gigi”, “hamil”, “kaki” dan “payudara”. Pada kluster 6 ini seluruh kata pembentuk kluster 6 ini memiliki bobot yang sama.



**Gambar 9.** Kluster 7

Pada gambar 9 dapat dilihat kata pembentuk dari kluster 7, yaitu “sakit”, “kulit”, “nyeri”, “darah”, “anak”, “luka”, “gigi”, “makan” dan “hamil”. Pada kluster 7 ini seluruh kata pembentuknya memiliki bobot yang sama.

#### 4. Conclusion

Berdasarkan penelitian yang telah dilakukan didapatkan hasil bahwa, dari 150 model yang dilatih berdasarkan kombinasi antara jumlah topik nilai  $\alpha$  dan nilai  $\beta$ , model terbaik adalah model dengan jumlah topik 8, nilai  $\alpha$  0.2 dan nilai  $\beta$  0.2. Model terbaik ditentukan dengan mencari nilai *coherence score* terbaik diantara model yang lain. Pada kombinasi yang disebutkan tadi, didapatkan nilai *coherence score* berupa 0.258481. Dari nilai *coherence score* tersebut dapat dikatakan bahwa topic modelling menggunakan metode Latent Dirichlet Allocation masih memiliki kekurangan. Hal tersebut juga dapat dilihat pada visualisasi data pada setiap kluster yang ada dimana pada beberapa kluster bobot kata pembentuk kluster memiliki bobot yang sama. Peneliti menduga hal ini disebabkan oleh kurangnya dataset serta luasnya persebaran topik pada fitur Tanya Dokter dari website Alodokter

## Refrensi

- [1] N. A. Tresnasari, T. B. Adji dan A. E. Permanasari, "Social-Child-Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation" *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, p. 179-188, 2020
- [2] F. Gurcan, O. Ozyurt, dan N. Cagitay, "Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation" *International Review of Research in Open and Distributed Learning*, vol. 22, no. 2, p. 1–18, 2021
- [3] K. Porter, "Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling" *Digital Investigation*, vol. 26, p. S87- S97, 2018
- [4] H. Najjichah, A. Syukur, dan H. Subagyo, "Pengaruh Text Preprocessing dan Kombinasinya pada Peringkasan Dokumen Otomatis Teks Berbahasa Indonesia" *Jurnal Teknologi Informasi*. vol. 15, no. 1, p. 1-11. 2019
- [5] P. Kherwa dan P. Bansal, "Topic Modeling: A Comprehensive Review" *ICST Transactions on Scalable Information Systems*, vol. 7, no. 24, 2018
- [6] H. Rahimi, J. L. Hoover, D. Mimno, H. Naacke. C. Constantin dan B. Amann, "Contextualized Topic Coherence Metrics", *arXiv:2305.14587*, 2023
- [7] S. K. Ray, A. Ahmad, dan C. A. Kumar, "Review and Implementation of Topic Modeling in Hindi" *Applied Artificial Intelligence*, vol. 33, no. 11, p. 979-1007, 2019

*This page is intentionally left blank.*

# Analisis Dan Visualisasi Data Untuk Meningkatkan Penjualan Menggunakan Exploratory Data Analysis Dan Looker Studio (Studi Kasus : Nies Collection)

Melanie Putri<sup>a1</sup>, Apriade Vaoutama<sup>a2</sup>

<sup>a</sup>Program Studi Sistem Informasi  
Universitas Singaperbangsa Karawang  
Puseurjaya, Telukjambe Timur, Karawang, Indonesia

<sup>1</sup>melanieputri4501@gmail.com

<sup>2</sup>apriade.vaoutama@staff.unsika.ac.id

## Abstrak

*Dengan kemudahan akses dan penggunaan teknologi yang semakin luas, berbagai sektor dihadapkan pada tuntutan untuk terus berinovasi menuju perubahan yang lebih baik. Fenomena maraknya perdagangan online menimbulkan tantangan baru bagi semua pelaku ekonomi, yang harus berlomba-lomba dalam memikat konsumen untuk bertransaksi. Nies Collection, sebuah toko fashion wanita yang beroperasi dalam salah satu platform e-commerce terkemuka yaitu Shopee, turut serta dalam dinamika ini. Penelitian ini bertujuan untuk menggambarkan dan menganalisis pola penjualan Nies Collection selama beberapa bulan terakhir. Data yang digunakan merupakan data penjualan yang dikumpulkan dari rentang waktu 16 Agustus hingga 16 November 2023 yang kemudian akan diproses dan dianalisis secara menyeluruh untuk menghasilkan informasi yang bermanfaat guna meningkatkan kinerja penjualan, serta mengembangkan Toko Nies Collection ke depannya. Metode yang digunakan dalam analisis data ini adalah Exploratory Data Analysis (EDA), didukung oleh platform Google Colaboratory. Selain itu, visualisasi data akan dilakukan secara real-time melalui platform Looker Studio, memungkinkan pemahaman yang lebih mendalam terhadap tren dan pola penjualan. Hasil analisis dan visualisasi data akan menjadi dasar untuk menyusun rekomendasi strategis bagi Nies Collection. Tujuan utamanya bukan hanya untuk meningkatkan volume penjualan, tetapi juga untuk memperkuat posisi toko dalam pasar dan menciptakan pertumbuhan yang berkelanjutan di masa mendatang.*

**Keywords:** Analisis, Visualisasi, Data, Exploratory Data Analysis, Looker Studio

## 1. Pendahuluan

Dalam era pasca COVID-19, tantangan ekonomi melanda berbagai sektor, memaksa banyak pelaku usaha untuk beradaptasi dengan cepat. Tak terkecuali sektor *e-commerce* yang menjadi sorotan utama. Nies Collection, sebuah toko *fashion* wanita yang telah berdiri selama tiga tahun, menjadi salah satu dari jutaan toko yang sedang berjuang menghadapi hal tersebut.

Sejak awal pandemi, pergeseran perilaku konsumen menuju belanja *online* semakin nyata. Kemudahan akses dan biaya yang terjangkau membuat *e-commerce* menjadi primadona bagi masyarakat. Dalam upaya mempertahankan eksistensinya, Nies Collection telah melangkah lebih jauh. Dalam tiga bulan terakhir, mereka tidak hanya berusaha untuk bertahan, tetapi juga meningkatkan kualitas produk, melayani pelanggan dengan lebih baik, serta mengasah strategi pemasaran mereka. Namun, di balik langkah-langkah progresif tersebut, Nies Collection masih dihadapkan pada berbagai tantangan. Persaingan sengit di dunia *e-commerce* mengharuskan mereka untuk terus menghadirkan inovasi dan kreativitas. Di sinilah peran penelitian ini hadir. Kami sebagai peneliti berkomitmen untuk menggali faktor-faktor kunci yang memengaruhi kesuksesan dan hambatan yang dihadapi oleh Nies Collection. Pendekatan metode penelitian yang digunakan berupa studi literatur yang komprehensif, dan analisis dokumen yang relevan yaitu data penjualan Nies Collection selama 3 bulan terakhir. Dengan menganalisis data ini, akan didapatkan wawasan mendalam tentang kinerja bisnis dari Nies Collection, strategi yang sudah dan harus diterapkan, serta kendala apa yang dihadapi oleh Nies Collection.

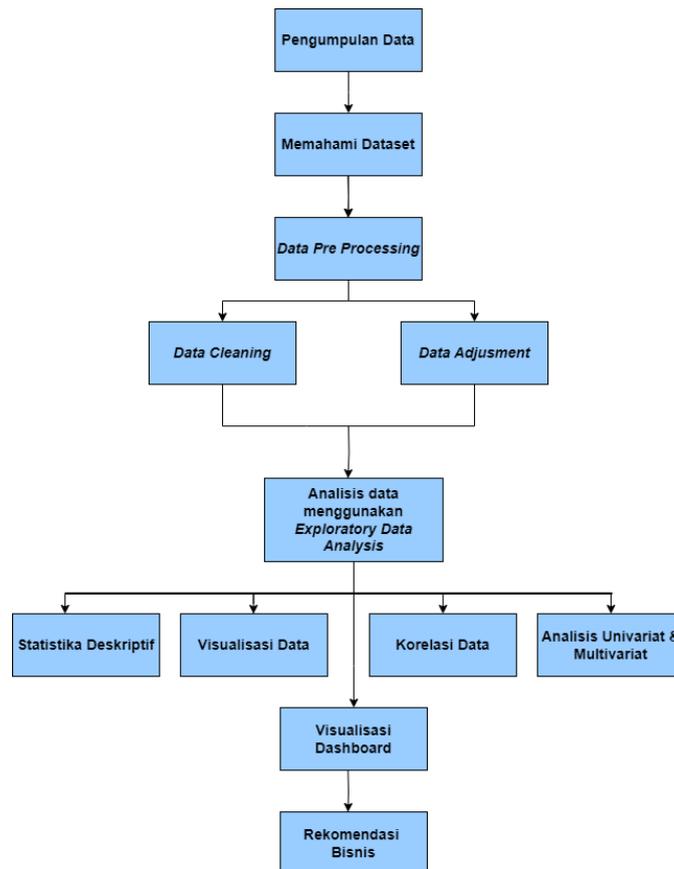
Dari analisis mendalam yang kami lakukan, kami bertujuan untuk menyajikan rekomendasi strategis yang

konkret dan relevan. Tujuannya sederhana yaitu memperkuat posisi Nies Collection di pasar *e-commerce*. Melalui pendekatan yang teliti dan solutif, kami percaya bahwa rekomendasi kami akan memberikan pandangan baru dan arah yang jelas bagi Nies Collection untuk meraih kesuksesan di masa depan.

## 2. Metode Penelitian

### 2.1. Tahapan Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian kuantitatif yaitu penelitian berdasarkan kaidah-kaidah ilmiah yang telah memenuhi kekonkretan, objektivitas, terukur, rasional dan sistematis. Metode kuantitatif juga disebut metode *discovery* karena memungkinkan penemuan baru yang dapat dikembangkan sebagai ilmu yang baru dengan sebuah penelitian yang berupa angka dan analisis statistik. [1] Tahapan dalam penelitian ini adalah sebagai berikut :



**Gambar 1.** Tahapan penelitian

Pada gambar diatas merupakan kerangka penelitian yang akan dilakukan. Berikut adalah penjelasan tentang tahapan penelitian tersebut :

- a. **Pengumpulan Data**  
Untuk data yang digunakan pada penelitian ini berasal dari data penjualan Nies Collection di *e-commerce shopee* selama bulan Agustus-Oktober 2023.
- b. **Memahami Dataset**  
Dari dataset yang telah didapatkan, peneliti mencoba memahami dataset yang telah didapatkan. Dataset ini mencakup berbagai variabel yang berkaitan dengan penjualan produk Nies Collection di Shopee selama periode 16 Agustus - 16 Oktober 2023. Informasi yang terdapat dalam dataset meliputi tanggal dan waktu transaksi, deskripsi produk, jumlah yang dibeli, harga, metode pembayaran dan lain sebagainya.
- c. **Data pre-processing**  
*Data pre- procesing* dilakukan untuk membersihkan data agar mudah diolah untuk kedepannya.

Tujuan dari *data pre-processing* adalah untuk mempermudah proses analisis dan lebih akurat, serta terhindarkan dari kesalahan. Pada *data pre-processing* kali ini meliputi 2 tahapan yaitu *Data Cleaning* (pembersihan data) dan *Data Adjustment* (penyesuaian data)

- d. Analisis Data Menggunakan *Exploratory Data Analysis* (EDA)  
Analisis statistik dengan EDA digunakan untuk menggali dan memahami data secara mendalam dengan tujuan untuk mengidentifikasi pola, trend, dan hubungan dalam dataset [2]
- e. Visualisasi Dashboard  
Visualisasi dashboard adalah representasi grafis yang menyajikan data dan informasi dalam format yang mudah dipahami dan menarik secara visual. Ini memungkinkan pengguna untuk dengan cepat menganalisis informasi yang kompleks dan membuat keputusan berdasarkan data yang tersedia. Pada tahap ini digunakan Google Looker Studio yang berguna untuk memberikan gambaran dari data yang telah dianalisis secara *real-time*.
- f. Rekomendasi Bisnis  
Rekomendasi bisnis diambil dari hasil analisis mendalam yang telah dilakukan sebelumnya, yang digunakan untuk meningkatkan performa bisnis serta strategi penjualan kedepannya.

## 2.2. Exploratory Data Analysis (EDA)

*Exploratory Data Analysis (EDA)* merupakan salah satu proses analisis data yang dalam kasus ini digunakan untuk melihat pola atau tren dalam bentuk grafik. EDA mempunyai peranan penting karena dapat membantu dan memudahkan dalam mengidentifikasi kumpulan fitur yang tepat, memilih algoritma yang sesuai untuk pembuatan model, serta membangun model itu sendiri. [3] Pada EDA kali ini, kami akan menggunakan *Google Colaboratory* dengan *Python*. Setiap eksplorasi pada dataset yang dilakukan, didasarkan pada pernyataan analisis yang telah disusun sebelumnya. Dalam EDA ini, beberapa tindakan yang akan dilakukan mencakup analisis statistika deskriptif, visualisasi data, uji korelasi data, dan analisis univariat dan multivariat. Dengan dilakukannya *Exploratory Data Analysis (EDA)* ini akan diungkapkan distribusi harga produk, dan beberapa analisis lainnya yang akan dilakukan guna mengetahui bagaimana perkembangan toko Nies Collection selama 3 bulan terakhir tersebut, sementara visualisasi data dapat menggambarkan tren penjualan produk dari waktu ke waktu.

## 2.3. Looker Studio

*Dashboard* adalah sebuah inovasi terkini yang hadir dalam bidang manajemen informasi. Selain itu *Dashboard* juga digunakan sebagai alat yang digunakan untuk proses yang sedang berlangsung, mengidentifikasi kinerja yang sedang berjalan dan memprediksi kejadian di masa yang akan datang. *Digital dashboard* merupakan tools untuk melakukan tampilan visualisasi dengan menggunakan indikator-indikator tertentu sebagai dasar tampilan informasi (Binus, 2017). Saat ini terdapat tools dalam membuat *digital dashboard* yang bersifat *open access* dan *open source* yang disediakan oleh Google yaitu *Google Data Studio* atau yang sekarang bernama *Looker Studio* (*Looker Studio, 2023*). [4] *Google Looker Studio* akan memberikan gambaran *real-time* dari data yang telah dianalisis.

## 3. Hasil dan Pembahasan

### 3.1. Data Pre-Processing

Dalam proyek ini, kami akan menggunakan data penjualan dari toko Nies Collection di Shopee yaitu data dalam rentang waktu 3 bulan, Data tersebut akan diunggah ke GitHub dalam format CSV. Kemudian, kami akan mengimpor dataset dari GitHub dengan cara membacanya dari URL yang mengarah ke file CSV dan menyimpannya dalam bentuk DataFrame. Dataset ini terdiri dari 174 baris dan 48 kolom. Berikut adalah informasi tentang dataset beserta tipe datanya :

Putri & Vaoutama  
 Analisis Dan Visualisasi Data Untuk Meningkatkan Penjualan  
 Menggunakan Exploratory Data Analysis Dan Looker Studio  
 (Studi Kasus : Nies Collection)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 174 entries, 0 to 173
Data columns (total 48 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   No. Pesanan                                                            174 non-null   object
1   Status Pesanan                                                         174 non-null   object
2   Alasan Pembatalan                                                     27 non-null   object
3   Status Pembatalan/ Pengembalian                                       0 non-null    float64
4   No. Resi                                                               149 non-null   object
5   Opsi Pengiriman                                                        174 non-null   object
6   Antar ke counter/ pick-up                                             149 non-null   object
7   Pesanan Harus Dikirimkan Sebelum (Menghindari keterlambatan)        164 non-null   object
8   Waktu Pengiriman Diatur                                               147 non-null   object
9   Waktu Pesanan Dibuat                                                  174 non-null   object
10  Waktu Pembayaran Dilakukan                                             174 non-null   object
11  Metode Pembayaran                                                     174 non-null   object
12  SKU Induk                                                              0 non-null    float64
13  Nama Produk                                                            174 non-null   object
14  Nomor Referensi SKU                                                    174 non-null   object
15  Nama Variasi                                                           174 non-null   object
16  Harga Awal                                                             174 non-null   float64
17  Harga Setelah Diskon                                                  174 non-null   float64
18  Jumlah                                                                 174 non-null   int64
19  Total Harga Produk                                                     174 non-null   float64
20  Total Diskon                                                           174 non-null   float64
21  Diskon Dari Penjual                                                    174 non-null   float64
22  Diskon Dari Shopee                                                     174 non-null   int64
23  Berat Produk                                                           174 non-null   object
24  Jumlah Produk di Pesan                                                174 non-null   int64
25  Total Berat                                                            174 non-null   object
26  Voucher Ditanggung Penjual                                            174 non-null   float64
27  Cashback Koin                                                         174 non-null   int64
28  Voucher Ditanggung Shopee                                             174 non-null   float64
29  Paket Diskon                                                           174 non-null   object
30  Paket Diskon (Diskon dari Shopee)                                     174 non-null   int64
31  Paket Diskon (Diskon dari Penjual)                                    174 non-null   float64
32  Potongan Koin Shopee                                                  174 non-null   int64
33  Diskon Kartu Kredit                                                    174 non-null   int64
34  Ongkos Kirim Dibayar oleh Pembeli                                     174 non-null   float64
35  Estimasi Potongan Biaya Pengiriman                                    174 non-null   float64
36  Ongkos Kirim Pengembalian Barang                                       174 non-null   int64
37  Total Pembayaran                                                      174 non-null   float64
38  Perkiraan Ongkos Kirim                                                174 non-null   float64
39  Catatan dari Pembeli                                                  7 non-null    object
40  Catatan                                                                0 non-null    float64
41  Username (Pembeli)                                                    174 non-null   object
42  Nama Penerima                                                         174 non-null   object
43  No. Telepon                                                            174 non-null   object
44  Alamat Pengiriman                                                     174 non-null   object
45  Kota/Kabupaten                                                        174 non-null   object
46  Provinsi                                                               174 non-null   object
47  Waktu Pesanan Selesai                                                141 non-null   object
dtypes: float64(15), int64(8), object(25)
memory usage: 65.4+ KB
```

**Gambar 2.** Informasi Dataset

Pada Data *Pre-Processing* ini akan dilakukan beberapa hal yaitu :

- a. Pembersihan Data (*Data Cleaning*) yaitu mengidentifikasi dan mengatasi data yang hilang, duplikat, atau data yang tidak valid.

1. Menghapus kolom-kolom yang tidak diperlukan

```
#Menghapus kolom-kolom karena tidak diperlukan dalam analisis data
kolomyangdihapus = ['Paket Diskon (Diskon dari Shopee)', 'Nama Penerima', 'No. Telepon',
                  'Alamat Pengiriman', 'Catatan', 'Username (Pembeli)',
                  'Status Pembatalan/ Pengembalian', 'SKU Induk', 'Diskon Dari Shopee',
                  'Diskon Kartu Kredit', 'Ongkos Kirim Pengembalian Barang', 'Cashback Koin',
                  'Catatan dari Pembeli']
df.drop(kolomyangdihapus, axis = 1, inplace = True) #inplace berguna untuk me-replace data secara permanen
```

**Gambar 3.** Query Hapus kolom

2. Mengubah data NaN (*Not-a-Number*) dengan nilai 0 agar tidak terdeteksi sebagai data error

```
#Mengisi data NaN (Not-a-Number) dengan nilai 0
df.fillna(0, inplace=True)
```

**Gambar 4.** Query Ubah Data NaN

3. Identifikasi nilai yang hilang (*missing value*)

```
#Melihat apakah terdapat missing value pada data  
df.isna().sum()
```

Gambar 5. Query Identifikasi Missing Value

```
No. Pesanan  
Status Pesanan  
Alasan Pembatalan  
No. Resi  
Opsi Pengiriman  
Antar ke counter/ pick-up  
Pesanan Harus Dikirimkan Sebelum (Menghindari keterlambatan)  
Waktu Pengiriman Diatur  
Waktu Pesanan Dibuat  
Waktu Pembayaran Dilakukan  
Metode Pembayaran  
Nama Produk  
Nomor Referensi SKU  
Nama Variasi  
Harga Awal  
Harga Setelah Diskon  
Jumlah  
Total Harga Produk  
Total Diskon  
Diskon Dari Penjual  
Berat Produk  
Jumlah Produk di Pesan  
Total Berat  
Voucher Ditanggung Penjual  
Voucher Ditanggung Shopee  
Paket Diskon  
Paket Diskon (Diskon dari Penjual)  
Potongan Koin Shopee  
Ongkos Kirim Dibayar oleh Pembeli  
Estimasi Potongan Biaya Pengiriman  
Total Pembayaran  
Perkiraan Ongkos Kirim  
Kota/Kabupaten  
Provinsi  
Waktu Pesanan Selesai  
dtype: int64
```

Gambar 6. Output Missing Values

Dapat dilihat dari output bahwa tidak terdapat nilai yang hilang (*missing value*) pada dataset yang digunakan.

b. Penyesuaian Data (*Data Adjustment*)

Pada tahap ini akan dilakukan penyesuaian pada value dalam dataset untuk kebutuhan visualisasi.

1. Mengganti value “Y” dan “N” menjadi “Pakai” dan “Tidak Pakai”

```
df['Paket Diskon'].value_counts()  
N    134  
Y     40  
Name: Paket Diskon, dtype: int64  
[43] df['Paket Diskon'] = df['Paket Diskon'].str.replace('N', 'Tidak Pakai')  
df['Paket Diskon'] = df['Paket Diskon'].str.replace('Y', 'Pakai')  
[44] df['Paket Diskon'].value_counts()  
Tidak Pakai    134  
Pakai           40  
Name: Paket Diskon, dtype: int64
```

Gambar 7. Query Ganti Value Y dan N

2. Mengganti value pada kolom alasan pembatalan agar lebih mudah dibaca saat visualisasi

```

df['Alasan Pembatalan'].value_counts()

Dibatalkan oleh Pembeli. Alasan: Ubah Pesanan yang Ada      8
Dibatalkan oleh Pembeli. Alasan: Lainnya/ berubah pikiran   7
Dibatalkan secara otomatis oleh sistem Shopee. Alasan: Pesanan belum dibayar 6
Dibatalkan oleh Pembeli. Alasan: Need to change delivery address 1
Dibatalkan oleh Pembeli. Alasan: Lainnya                   1
Dibatalkan oleh Penjual. Alasan: Produk habis              1
Dibatalkan oleh Pembeli. Alasan: Perlu mengubah Voucher   1
Dibatalkan oleh Pembeli. Alasan: Tidak ingin membeli lagi  1
Dibatalkan oleh Pembeli. Alasan: Menemukan yang lebih murah 1
Name: Alasan Pembatalan, dtype: int64

[38] df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan oleh Pembeli. Alasan: ','')
df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan secara otomatis oleh sistem Shopee. Alasan: ','')
df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan oleh Penjual. Alasan: ','')

<ipython-input-38-3921a8c8d00b>:1: FutureWarning: The default value of regex will change from True to False in a future version.
df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan oleh Pembeli. Alasan: ','')
<ipython-input-38-3921a8c8d00b>:2: FutureWarning: The default value of regex will change from True to False in a future version.
df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan secara otomatis oleh sistem Shopee. Alasan: ','')
<ipython-input-38-3921a8c8d00b>:3: FutureWarning: The default value of regex will change from True to False in a future version.
df['Alasan Pembatalan'] = df['Alasan Pembatalan'].str.replace('Dibatalkan oleh Penjual. Alasan: ','')

[39] df['Alasan Pembatalan'].value_counts()

Ubah Pesanan yang Ada      8
Lainnya/ berubah pikiran   7
Pesanan belum dibayar     6
Need to change delivery address 1
Lainnya                   1
Produk habis              1
Perlu mengubah Voucher   1
Tidak ingin membeli lagi  1
Menemukan yang lebih murah 1
Name: Alasan Pembatalan, dtype: int64
    
```

**Gambar 8.** Query Ganti Value Kolom Pembatalan

**3.2. Analisis Data Menggunakan Exploratory Data Analysis**

a. Statistika Deskriptif Dari Dataset Nies Collection

Analisis statistika deskriptif bertujuan untuk memahami sifat-sifat khusus dari dataset. Dalam analisis ini nilai-nilai statistik seperti jumlah data, rerata, standar deviasi, kuartil serta nilai maksimum dan minimum untuk setiap kolom dan fitur akan ditampilkan. Di samping itu, identifikasi nilai yang kosong dilakukan untuk memastikan integritas data yang digunakan dalam pembangunan model tanpa adanya data yang tidak lengkap. [5]

```

#Menghasilkan ringkasan statistik deskriptif dari kolom-kolom numerik dalam DataFrame df
df.describe()
    
```

Output :

	Harga Awal	Harga Setelah Diskon	Jumlah Produk	Total Harga Produk	Total Diskon	Diskon Dari Penjual	Berat Produk	Jumlah Produk di Pesan	Total Berat	Voucher Ditanggung Penjual	Voucher Ditanggung Shopee	Paket Diskon (Diskon dari Penjual)	Potongan Koin Shopee	Ongkos Kirim Dibayar oleh Pembeli	Estimasi Potongan Biaya Pengiriman	Total Pembayaran	Perkiraan Ongkos Kirim
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000
mean	178193.808048	111118.128437	1.04023	114711.350575	68940.848278	68940.848278	482.738821	1.574713	730.488770	13017.241378	2537.888308	1741.378310	884.241378	3885.388552	15075.977011	1409.12873883	21534.477011
std	80940.529490	31258.828894	0.27114	37874.747852	41177.846272	41177.846272	154.849841	0.845108	488.230879	50808.167077	8478.450844	3807.444189	3858.573872	11345.430114	11945.132751	121480.853281	18628.500189
min	51500.000000	23175.000000	1.00000	23175.000000	23225.000000	23225.000000	200.000000	1.000000	300.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	7000.000000
25%	135000.000000	80844.500000	1.00000	83821.250000	43850.000000	43850.000000	400.000000	1.000000	500.000000	0.000000	0.000000	0.000000	0.000000	0.000000	9000.000000	72238.000000	8000.000000
50%	182500.000000	118200.000000	1.00000	118800.000000	52800.000000	52800.000000	500.000000	1.000000	500.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12500.000000	116284.000000	15250.000000
75%	178500.000000	134018.500000	1.00000	134800.000000	60388.000000	60388.000000	500.000000	2.000000	800.000000	3000.000000	0.000000	0.000000	0.000000	0.000000	20000.000000	171514.750000	22000.000000
max	345500.000000	172750.000000	4.00000	318200.000000	188200.000000	188200.000000	2000.000000	5.000000	2400.000000	24000.000000	33550.000000	15000.000000	21820.000000	53000.000000	40000.000000	531026.000000	83000.000000

**Gambar 9.** Statistika deskriptif

- b. Analisis dan Visualisasi Data
  - 1. Perbandingan Persentase Alasan Pembatalan



**Gambar 9.** Perbandingan Alasan Pembatalan

Pada analisis ini dilakukan untuk mengetahui alasan mendalam terkait faktor-faktor yang mempengaruhi keputusan pembatalan, dengan begitu Nies Collection dapat memahami pola pembatalan, penentuan prioritas tindakan, identifikasi potensi peningkatan layanan, penyesuaian strategi pemasaran atau penjualan, dan evaluasi kinerja bisnis. Terlihat bahwa alasan *customer* melakukan pembatalan paling banyak adalah 'ubah pesanan' dengan nilai 30%, kemudian alasan 'Lainnya/berubah pikiran' dengan 26%, dan 'pesanan belum dibayar' dengan 22%.

- 2. Perbandingan Penggunaan Paket Diskon

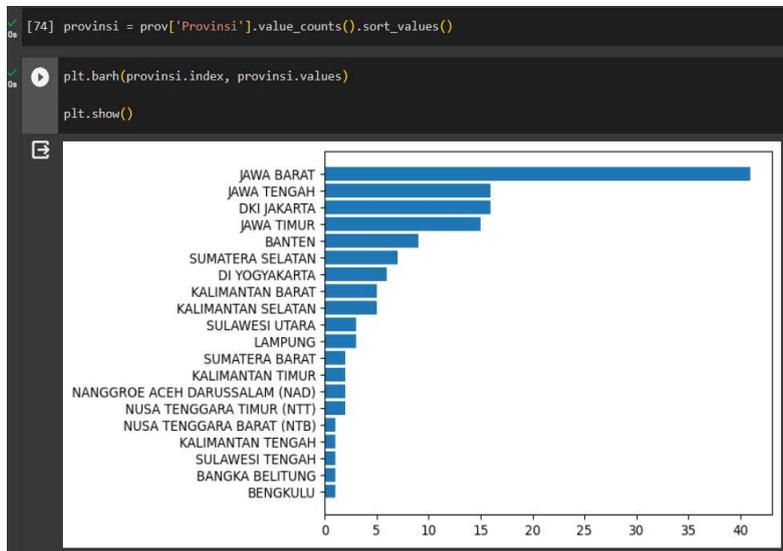


**Gambar 10.** Perbandingan Penggunaan Paket Diskon

Analisis ini bertujuan untuk meningkatkan evaluasi efektivitas strategi pemasaran. Melalui analisis penggunaan paket diskon, Nies Collection dapat mengevaluasi sejauh mana strategi pemasaran yang menggunakan diskon berhasil menarik pelanggan. Data ini memberikan gambaran tentang seberapa efektif diskon dalam merangsang pembelian dan meningkatkan penjualan. Dari analisis yang dilakukan tersebut didapat bahwa *customer* yang tidak

menggunakan diskon lebih banyak dengan persentase 77% dan yang pakai diskon dengan persentase 23%. Dengan ini dapat disimpulkan bahwa diskon tidak terlalu berpengaruh dalam meningkatkan penjualan.

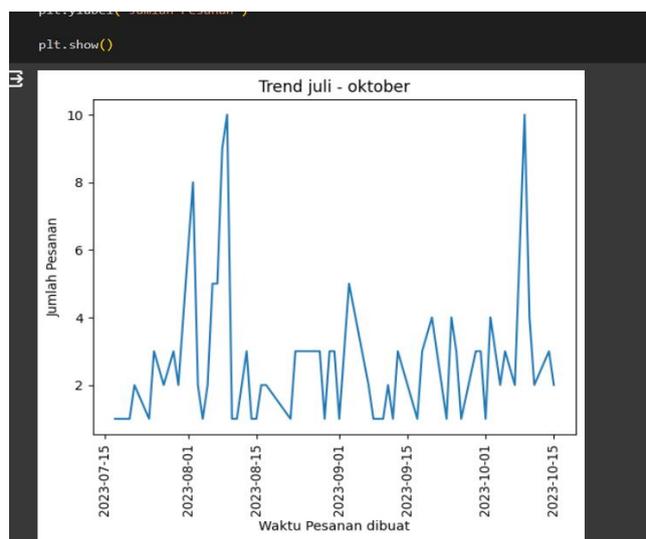
### 3. Persebaran Provinsi Dengan Pembeli Terbanyak



**Gambar 11.** Persebaran Provinsi Dengan Pembeli Terbanyak

Salah satu alasan dilakukan analisis ini, yaitu bertujuan untuk mengidentifikasi pasar potensial. Dengan menganalisis persebaran pembeli terbanyak di setiap provinsi, Nies Collection dapat mengidentifikasi provinsi-provinsi yang memiliki pasar potensial yang besar. Ini dapat membantu dalam menentukan di mana sebaiknya alokasi sumber daya pemasaran dan penjualan dilakukan untuk memaksimalkan potensi penjualan. Dari analisis yang telah dilakukan penjualan tertinggi berada di Provinsi Jawa Barat, Jawa Tengah dan DKI Jakarta.

### 4. Tren Penjualan Juli – Oktober

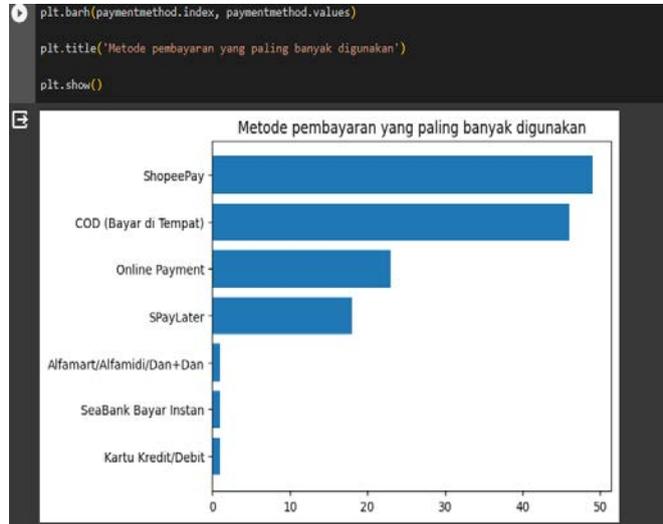


**Gambar 12.** Tren Penjualan Juli – Oktober

Analisis tren penjualan dari bulan Juli hingga Oktober memberikan wawasan yang berharga tentang kinerja penjualan produk atau layanan Nies Collection selama periode waktu tersebut, yang dapat digunakan untuk menginformasikan keputusan bisnis yang lebih baik di masa

depan.

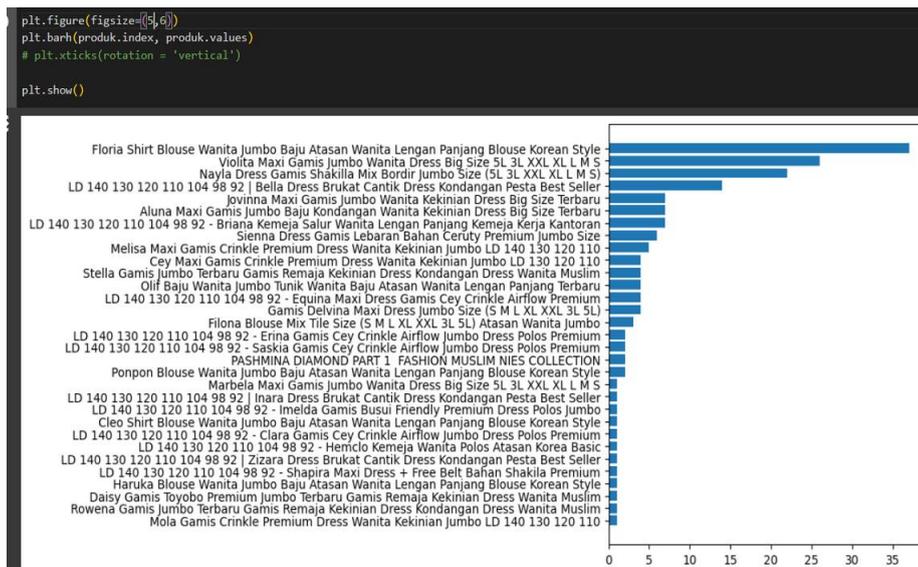
5. Metode Pembayaran Yang Paling Banyak Digunakan



Gambar 12. Metode Pembayaran

Dengan melakukan analisis terhadap metode pembayaran yang digunakan dapat diperoleh pemahaman mengenai preferensi pembayaran pelanggan. Ini membantu Nies Collection menyesuaikan pilihan metode pembayaran yang ditawarkan untuk memenuhi kebutuhan dan preferensi pelanggan. Selain itu analisis ini juga membantu menjaga tingkat keamanan dan keandalan masing-masing metode pembayaran. Ini penting untuk memastikan bahwa proses pembayaran aman dan dapat diandalkan bagi pelanggan. Dari analisis menggunakan EDA yang dilakukan diperoleh bahwa metode pembayaran yang paling banyak digunakan adalah *ShopeePay*, *COD (Bayar di Tempat)* dan *Online Payment*.

6. Produk Terjual Setidaknya Satu Dalam Setiap Pesanan



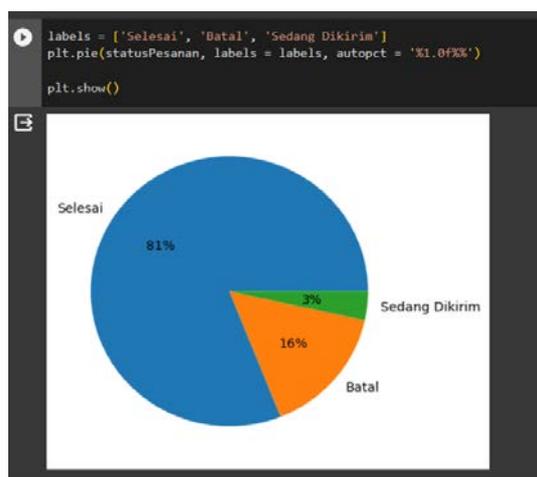
Gambar 13. Produk Terjual

Analisis produk yang terjual setidaknya satu dalam setiap pesanan memberikan wawasan

yang berharga tentang pola pembelian pelanggan, yang dapat digunakan untuk meningkatkan penjualan lintas produk, mengoptimalkan penempatan produk, dan merencanakan strategi penjualan yang lebih efektif di masa depan. Dengan adanya analisis ini Nies Collection dapat menentukan produk yang cocok untuk dibundel bersama-sama.

Misalnya, jika dua produk sering dibeli bersama-sama, Nies Collection dapat menawarkan diskon atau penawaran khusus saat kedua produk tersebut dibeli bersamaan untuk meningkatkan penjualan lintas produk. Terlihat dari hasil analisis yang telah dilakukan produk yang paling banyak terjual adalah *Floria Shirt Blouse*, *Violita Maxi Gamis*, dan *Nayla Dress Gamis*.

## 7. Perbandingan Status Pesanan



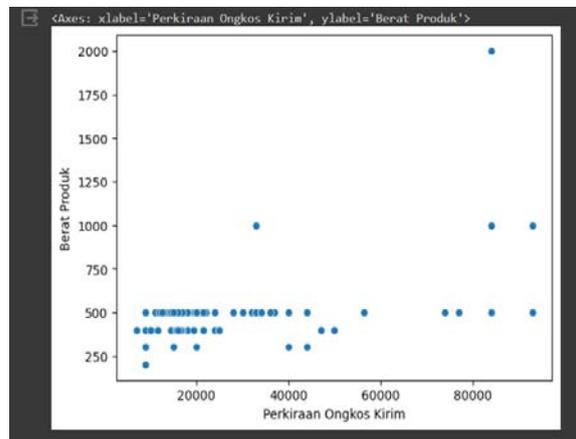
**Gambar 14.** Perbandingan Status Pesanan

Untuk analisis perbandingan status pesanan, dilakukan dengan tujuan guna pemantauan kinerja operasional. Dengan membandingkan status pesanan dari mulai diproses hingga selesai, Nies Collection dapat melacak kinerja operasionalnya. Ini membantu mengidentifikasi tahap-tahap di mana pesanan mungkin mengalami keterlambatan atau kendala, sehingga Nies Collection dapat mengambil tindakan korektif yang diperlukan.

### c. Korelasi Data

Korelasi adalah metode analisis yang termasuk dalam kategori teknik pengukuran asosiasi atau hubungan antar variabel. Pengukuran asosiasi merujuk pada sekelompok teknik statistik bivariat yang digunakan untuk mengukur tingkat kekuatan hubungan antara dua variabel. Pada dasarnya, korelasi bertujuan untuk menilai hubungan antara dua variabel, yang berarti perubahan dalam satu variabel dapat mempengaruhi variabel lainnya. Koefisien korelasi ( $r$ ) memiliki rentang nilai antara 1 hingga -1, di mana nilai yang mendekati 1 atau -1 menunjukkan hubungan yang lebih kuat antara kedua variabel, sedangkan nilai yang mendekati 0 menandakan hubungan yang lemah. Nilai positif menunjukkan hubungan searah (saat satu variabel naik, variabel lainnya juga naik), sementara nilai negatif menunjukkan hubungan terbalik (saat satu variabel naik, variabel lainnya turun). [6] Dalam analisis ini, dilakukan pengecekan korelasi antara perkiraan ongkos kirim dan berat produk

```
#membuat scatter plot yang memvisualisasikan hubungan antara dua variabel dalam DataFrame df, yaitu 'Perkiraan Ongkos Kirim' dan 'Berat Produk'.  
sns.scatterplot(x='Perkiraan Ongkos Kirim', y='Berat Produk', data=df)
```



```
#menghitung dan mencetak nilai korelasi antara dua variabel, yaitu 'Berat Produk' dan 'Perkiraan Ongkos Kirim', dalam DataFrame df.
korelasi = df['Berat Produk'].corr(df['Perkiraan Ongkos Kirim'])
print(f"Korelasi antara Berat Produk dan Jumlah Produk: {korelasi:.2f}")

Korelasi antara Berat Produk dan Jumlah Produk: 0.44
```

**Gambar 15.** Korelasi Data

Kesimpulan yang dapat diambil dari analisis korelasi data diatas adalah semakin besar nilai absolut dari koefisien korelasi mendekati 1, semakin kuat korelasinya. Nilai yang mendekati 0 menunjukkan korelasi yang lemah. Dengan semikian, karena nilai korelasinya 0.44 mendekati 0, maka variabel Berat Produk dan Perkiraan Ongkos Kirim memiliki korelasi yang lemah sehingga berat produk tidak terlalu mempengaruhi untuk perkiraan ongkos kirim.

d. Analisis Univariat dan Multivariat

EDA melibatkan analisis variabel tunggal (*univariate*) dan hubungan antara beberapa variabel (*multivariate*) untuk mengidentifikasi pola dan tren dalam data.

Penarikan Kesimpulan akan berdasarkan 2 hal berikut :

1. Nilai Statistik Uji F (*F-statistic*)

Nilai statistik uji F adalah angka yang digunakan untuk mengukur perbedaan variasi antara kelompok (variabilitas antara kelompok) dan variasi dalam kelompok (variabilitas dalam kelompok). Semakin besar nilai statistik uji F, semakin besar perbedaan rata-rata antara kelompok-kelompok tersebut. Nilai statistik ini diperhitungkan dalam pengambilan keputusan.

2. Kesimpulan dari uji ANOVA biasanya tergantung pada nilai p-value

Jika nilai p-value kurang dari tingkat signifikansi yang ditetapkan (misalnya,  $\alpha = 0.05$ ), dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara setidaknya satu pasangan kelompok dalam variabel yang diuji. Jika nilai p-value lebih besar dari tingkat signifikansi, maka tidak memiliki cukup bukti untuk menyimpulkan bahwa ada perbedaan yang signifikan antara kelompok-kelompok tersebut.

```
# Melakukan uji ANOVA untuk kolom 'Nama Produk' berdasarkan kategori 'Harga Setelah Diskon'
kategori_produk = df['Nama Produk'].unique()
anova_result = []

for kategori in kategori_produk:
    subset_data = df[df['Nama Produk'] == kategori]
    anova = stats.f_oneway(subset_data['Harga Setelah Diskon'], df['Harga Setelah Diskon'])
    anova_result.append((kategori, anova))

# Menampilkan hasil uji ANOVA
for kategori, result in anova_result:
    print(f"Hasil uji ANOVA untuk {kategori}:")
    print(f"Statistik uji F: {result.statistic:.2f}")
    print(f"Nilai p-value: {result.pvalue:.4f}")
    print()
```

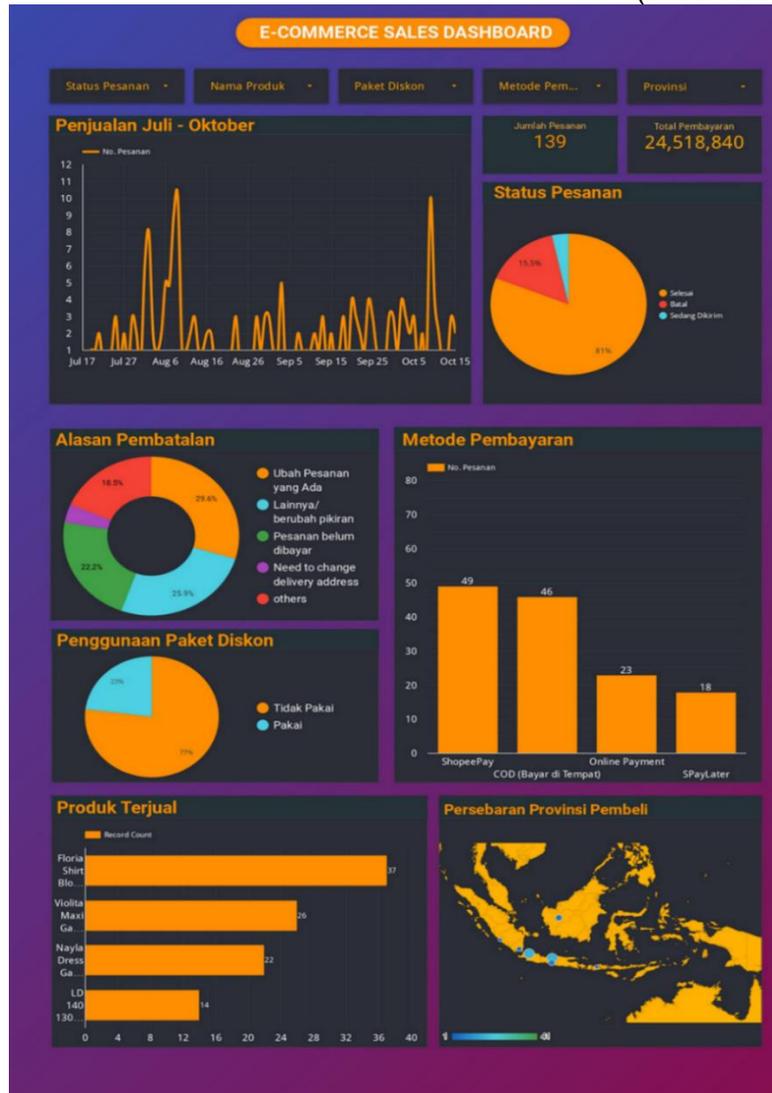
```
Nilai p-value: 0.7421
Hasil uji ANOVA untuk Haruka Blouse Wanita Jumbo Baju Atasan Wanita Lengan Panjang Blouse Korean Style:
Statistik uji F: 1.44
Nilai p-value: 0.2318
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 | Inara Dress Brukat Cantik Dress Kondangan Pesta Best Seller:
Statistik uji F: 1.79
Nilai p-value: 0.1825
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Imelda Gamis Busui Friendly Premium Dress Polos Jumbo:
Statistik uji F: 0.55
Nilai p-value: 0.4605
Hasil uji ANOVA untuk Cleo Shirt Blouse Wanita Jumbo Baju Atasan Wanita Lengan Panjang Blouse Korean Style:
Statistik uji F: 1.20
Nilai p-value: 0.2570
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Clara Gamis Cey Crinkle Airflow Jumbo Dress Polos Premium:
Statistik uji F: 0.63
Nilai p-value: 0.4278
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Hemclo Kemeja Wanita Polos Atasan Korea Basic:
Statistik uji F: 0.93
Nilai p-value: 0.3369
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 | Zizara Dress Brukat Cantik Dress Kondangan Pesta Best Seller:
Statistik uji F: 0.91
Nilai p-value: 0.3406
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Shapira Maxi Dress + Free Belt Bahan Shakila Premium:
Statistik uji F: 0.03
Nilai p-value: 0.8559
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Equina Maxi Dress Gamis Cey Crinkle Airflow Premium:
Statistik uji F: 1.44
Nilai p-value: 0.2311
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Saskia Gamis Cey Crinkle Airflow Jumbo Dress Polos Premium:
Statistik uji F: 1.21
Nilai p-value: 0.2720
Hasil uji ANOVA untuk LD 140 130 120 110 104 98 92 - Erina Gamis Cey Crinkle Airflow Jumbo Dress Polos Premium:
Statistik uji F: 2.06
Nilai p-value: 0.1527
Hasil uji ANOVA untuk Sienna Dress Gamis Lebaran Bahan Ceruty Premium Jumbo Size:
Statistik uji F: 8.48
Nilai p-value: 0.0041
Hasil uji ANOVA untuk Marbela Maxi Gamis Jumbo Wanita Dress Big Size 5L 3L XXL XL L M S:
Statistik uji F: 0.18
Nilai p-value: 0.6758
Hasil uji ANOVA untuk PASHMINA DIAMOND PART 1 FASHION MUSLIM NIES COLLECTION:
Statistik uji F: 15.74
Nilai p-value: 0.0001
```

Gambar 18. Uji Anova

Kesimpulan dari hasil di atas dapat berbeda-beda tergantung pada tingkat signifikansi yang ditetapkan. Jika tingkat signifikansi ( $\alpha$ ) biasanya diatur pada 0.05, sebagai contoh salah satu produk yaitu produk "Floria Shirt Blouse Wanita Jumbo" menunjukkan perbedaan signifikan dalam rata-rata harga produk antara kelompok produk karena nilai p-valuenya lebih kecil dari 0.05 yaitu 0.00. Produk lainnya tidak menunjukkan perbedaan signifikan berdasarkan nilai p-value karena jika nilai p-value lebih besar dari tingkat signifikansi, maka menunjukkan bahwa tidak ada perbedaan yang signifikan antara kelompok-kelompok tersebut. Harap perhatikan bahwa hasil ini hanya mencerminkan perbedaan statistik, dan interpretasi lebih lanjut dapat memerlukan analisis lebih mendalam tentang produk dan variabel yang diuji.

### 3.3 Visualisasi Data Menggunakan Looker Studio

Dari visualisasi yang telah dilakukan di Google Looker Studio, didapatkan bahwa jumlah pesanan pada tanggal 16 Juli - 16 Oktober 2023 adalah sebanyak 139 pesanan, kemudian didapatkan juga total pembayaran pembeli sebanyak 24,518,840 rupiah. Selanjutnya analisis lain yang ditampilkan yaitu tren penjualan pada bulan Juli-Oktober, status pesanan, metode pembayaran yang digunakan, alasan pembatalan, penggunaan paket diskon, produk yang paling banyak terjual serta persebaran provinsi pembeli. Berikut hasil Dashboard menggunakan Google Looker Studio.



Gambar 18. Visualisasi Data Menggunakan Looker Studio

### 3.4 Rekomendasi Bisnis

Dari beberapa analisis yang telah dilakukan diatas, akan diberikan rekomendasi bisnis yang dapat membantu meningkatkan penjualan Nies Collection, dan akan dijelaskan juga bagaimana solusi tersebut dapat membantu menyelesaikan permasalahan Nies Collection kedepannya, berikut solusi yang dapat diberikan :

- a. Mengatasi Penyebab Pembatalan Pesanan: Mengingat bahwa alasan utama pembatalan pesanan adalah perubahan pesanan (30%), diikuti oleh alasan "Lainnya/berubah pikiran" (26%) dan "Pesanan belum dibayar" (22%), Nies Collection dapat mengambil langkah-langkah tertentu. Misalnya, Nies Collection bisa meningkatkan fleksibilitas dalam perubahan pesanan atau memberikan waktu tertentu bagi pelanggan untuk membayar setelah memesan. Solusi tersebut terbukti efektif, karena solusi itupun digunakan oleh beberapa platform *e-commerce* seperti Tokopedia dan Lazada. Mereka meningkatkan fleksibilitas dalam perubahan pesanan dengan memberikan opsi untuk mengubah atau memodifikasi pesanan sebelum pengiriman dilakukan.
- b. Revaluasi Strategi Diskon: Meskipun hanya 23% pelanggan yang menggunakan diskon, ada potensi untuk menyesuaikan strategi diskon. Nies Collection dapat mengevaluasi jenis diskon yang ditawarkan, periode diskon, atau target *audiens* yang dituju untuk meningkatkan efektivitas strategi diskon dan merangsang lebih banyak pembelian. Solusi ini sudah diterapkan oleh berbagai toko dan layanan seperti Gofood, Grabfood, Zara, dan H&M yang seringkali menawarkan diskon musiman, penjualan akhir musim, atau diskon untuk pelanggan yang berlangganan untuk menarik

- pelanggan baru atau mendorong pelanggan yang ada untuk melakukan pembelian.
- c. Fokus Pemasaran dan Penjualan pada Provinsi dengan Pembeli Terbanyak: Mengingat bahwa penjualan tertinggi terjadi di Provinsi Jawa Barat, Jawa Tengah, dan DKI Jakarta, Nies Collection dapat memfokuskan upaya pemasaran dan penjualan di provinsi-provinsi ini. Ini bisa melibatkan kampanye pemasaran yang disesuaikan atau menawarkan program promosi khusus untuk meningkatkan loyalitas pelanggan di wilayah-wilayah ini. Misalnya memberikan *voucher* potongan ongkir untuk provinsi-provinsi tersebut. Layanan pengiriman barang seperti J&T Express, JNE, atau SiCepat juga mengadopsi pendekatan serupa dengan menyesuaikan strategi pemasaran dan promosi mereka untuk meningkatkan jumlah pengiriman dari provinsi-provinsi dengan permintaan tertinggi.
  - d. Pemantauan dan Penyesuaian Berdasarkan Tren Penjualan: Dengan memahami tren penjualan dari bulan Juli hingga Oktober, Nies Collection dapat melakukan pemantauan dan penyesuaian yang sesuai dalam strategi pemasaran dan penjualan. Misalnya, jika ada peningkatan penjualan pada bulan-bulan tertentu, Nies Collection bisa menyesuaikan inventaris atau mengalokasikan sumber daya tambahan untuk memenuhi permintaan yang meningkat. Solusi ini juga banyak digunakan platform *e-commerce* seperti Amazon, eBay, dan Alibaba, mereka menggunakan sistem analitik canggih untuk memantau tren penjualan secara *real-time*. Mereka dapat menyesuaikan inventaris mereka, menawarkan produk-produk yang sedang tren, atau menyesuaikan strategi harga mereka berdasarkan permintaan pelanggan yang berubah.
  - e. Optimasi Metode Pembayaran: Dengan memperhatikan bahwa ShopeePay, COD (Bayar di Tempat), dan Pembayaran *Online* adalah metode pembayaran yang paling banyak digunakan, Nies Collection dapat meningkatkan kenyamanan pelanggan dengan menyediakan pilihan pembayaran yang lebih luas. Ini termasuk memastikan keamanan dan ketersediaan metode pembayaran yang ada, serta mengeksplorasi opsi baru yang sesuai dengan preferensi pelanggan.
  - f. Promosi Produk yang Sering Terjual Bersama: Analisis menunjukkan bahwa produk seperti Floria Shirt Blouse, Violita Maxi Gamis, dan Nayla Dress Gamis sering terjual bersama-sama. Nies Collection dapat memanfaatkan informasi ini dengan menawarkan paket bundel atau promosi lintas produk yang menarik bagi pelanggan untuk meningkatkan penjualan lintas produk. Rekomendasi ini sudah terbukti efektif dilihat dari banyaknya layanan yang melakukan *bundling* produk yang berbeda dalam penjualan, contohnya seperti layanan pemesanan makanan seperti Gofood yang biasa melakukan bundel makanan yang berbeda dengan harga yang lebih terjangkau.
  - g. Optimasi Proses Pesanan: Analisis perbandingan status pesanan memungkinkan Nies Collection untuk mengidentifikasi tahapan-tahapan yang mungkin mengalami keterlambatan atau kendala dalam proses pesanan. Dengan memperbaiki proses pesanan, Nies Collection dapat meningkatkan efisiensi operasional dan memastikan pengalaman pelanggan yang lebih baik.

Dalam konteks implementasi rekomendasi bisnis yang didapatkan dari hasil analisis data, akan dilakukan evaluasi menyeluruh terhadap dampaknya pada kinerja bisnis Nies Collection. Evaluasi ini akan melibatkan langkah-langkah konkret, seperti pemantauan terhadap metrik kinerja yang relevan, akan tetapi tidak terbatas pada penjualan, retensi pelanggan, dan kepuasan pelanggan. Selain itu, akan dilakukan survei pelanggan dan pemetaan tren pasar secara terpola untuk memahami respons pelanggan terhadap perubahan yang diterapkan. Evaluasi implementasi akan dilakukan secara reguler untuk mengidentifikasi keberhasilan serta mengevaluasi apakah rekomendasi bisnis sudah mencapai tujuan yang ditetapkan. Yang akan terjadi, evaluasi ini akan menjadi dasar untuk penyesuaian strategi serta tindakan lebih lanjut guna memastikan keberlanjutan dan pertumbuhan bisnis Nies Collection pada masa mendatang. Dengan pendekatan ini, diyakini dapat memastikan bahwa setiap langkah yang diambil mempunyai dampak yg positif dan terukur bagi kesuksesan bisnis Nies Collection.

#### 4. Kesimpulan

Dalam penelitian ini, telah dilakukan analisis dan visualisasi data secara mendalam untuk meningkatkan kinerja dan penjualan dari Nies Collection. Melalui *Exploratory Data Analysis (EDA)* dengan Python, kami dapat mendapatkan pola-pola penting dalam data penjualan, memberikan wawasan berharga tentang perilaku konsumen dan tren penjualan. Kemudian visualisasi *dashboard* yang telah dilakukan melalui Looker Studio juga dapat membantu menyajikan informasi yang lebih jelas dalam bentuk grafik dan memudahkan Nies Collection dalam memonitor penjualan yang telah terjadi sebelumnya, sehingga membantu untuk perencanaan peningkatan penjualan dan mengidentifikasi area yang dapat ditingkatkan. Hasil proyek ini memberikan pemahaman yang lebih dalam tentang faktor-faktor yang mempengaruhi

kinerja toko Nies Collection, serta menjadi dasar untuk pengambilan keputusan yang lebih baik dalam upaya meningkatkan efisiensi dan profitabilitas bisnis. Proses analisis data ini menegaskan pentingnya pendekatan berbasis data dalam pengelolaan toko Nies Collection di platform Shopee, dan memberikan fondasi yang kokoh untuk strategi pemasaran dan pengembangan produk di masa depan.

Namun, untuk interpretasi hasil analisis kami juga menyoroti beberapa area yang perlu mendapat perhatian lebih lanjut. Misalnya, kami mengamati bahwa meskipun produk Floria Shirt Blouse memiliki tingkat penjualan yang tinggi, akan tetapi status pembatalan pesannya lebih tinggi dari produk-produk lain. Hal ini menunjukkan perlunya meninjau kembali analisis alasan pembatalan, perbaikan kualitas atau layanan dan strategi pemasaran untuk memastikan bahwa Nies Collection memaksimalkan profitabilitas produknya. Secara keseluruhan, hasil analisis kami memberikan dasar yang kuat untuk pengambilan keputusan yang terinformasi dan strategi bisnis yang efektif. Namun, kami juga menyadari bahwa evaluasi implementasi rekomendasi bisnis kami akan menjadi langkah selanjutnya yang krusial. Evaluasi ini akan membantu kami memahami dampak nyata dari perubahan yang diusulkan dan mengidentifikasi peluang untuk peningkatan lebih lanjut di masa mendatang.

Meskipun analisis dataset penjualan Nies Collection memberikan wawasan yang berharga, penelitian ini memiliki beberapa keterbatasan. Keterbatasan akses dan kualitas data dapat mempengaruhi representasi keseluruhan perilaku konsumen dan kinerja penjualan. Potensi adanya bias dalam pemilihan data atau metode analisis juga harus diperhatikan. Selain itu, penggunaan dataset yang lebih luas dan beragam dapat memperkenalkan kompleksitas tambahan dalam proses analisis. Dengan menyadari keterbatasan tersebut, penting untuk mempertimbangkan hasil analisis dengan hati-hati dalam konteks pengambilan keputusan bisnis. Langkah-langkah lanjutan, seperti penelitian mendalam dan validasi temuan, dapat membantu meningkatkan validitas dan relevansi hasil analisis di masa depan.

#### Daftar Pustaka

- [1] M. Yani Balaka and F. Abyan, *METODOLOGI PENELITIAN KUANTITATIF*. 2022. [Online]. Available: [www.penerbitwidina.com](http://www.penerbitwidina.com)
- [2] S. Permata Sari and R. Amanda Putri, "Analisis Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis dan K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON) Hal: 423-*, vol. 433, no. 2, 2023, doi: 10.30865/json.v5i2.7180.
- [3] T. N. Muthmainnah and A. Voutama, "Volume 6 ; Nomor 2," *Juli*, pp. 463–471, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [4] A. Jariyah, T. Indrabulan, M. Ilyas Syarif, and R. Krisna Astuti Sakir, "Pemanfaatan Looker Studio untuk Visualisasi Kinerja Program Studi D4 Teknik Multimedia dan Jaringan," 2022.
- [5] A. M. M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1669, Oct. 2022, doi: 10.30865/jurikom.v9i5.5021.
- [6] B. Subandriyo, S. St, M. Stat Diklat, S. Tingkat, A. Bps, and A. Xxi, "B A H A N A J A R ANALISIS KOLERASI DAN REGRESI."

*This page is intentionally left blank.*

# Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Klasifikasi Penerima Beasiswa

Lia Susanti<sup>a1</sup>, Khoirunnisa<sup>a2</sup>

<sup>a</sup>Program Studi Teknik Informatika, Universitas Indraprasta PGRI  
Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

<sup>1</sup>liasusanti.s4a.061@gmail.com

<sup>2</sup>khoirunnisa.1797@gmail.com

## Abstract

*One of the most important factors is education, but many students who have great abilities and potential cannot continue school because they are not financially capable but there are also many able-bodied students who get scholarships. In the educational environment, especially schools, there should be some rules or classification in determining students who receive scholarships. Therefore, in this study, a comparison of the C4.5 and Naïve Bayes algorithms was applied to the data of students who received scholarships. This study aims to find a pattern that can determine the award of scholarships with predetermined criteria, in the selection of prospective scholarship recipients at SMKS 2 Adi Luhur using the Naive Bayes Algorithm and the C 4.5 Algorithm. The data will be tested using k-fold cross validation (k=10). From the comparison results, the results of the Naïve Bayes accuracy are higher than the C 4.5 Algorithm. The results obtained from the comparison of the two algorithms are the Naïve Bayes algorithm has an accuracy rate of 94.52% and the C4.5 algorithm has an accuracy rate of 92.52%.*

**Keywords:** Scholarship, C4.5 Algorithm, Naive Bayes, K-Fold Cross Validation, Classification

## 1. Pendahuluan

Salah satu masalah pendidikan yang dihadapi bangsa Indonesia saat ini adalah bagaimana meningkatkan mutu pendidikan di setiap jenjang, khususnya jenjang sekolah menengah atas agar mampu bersaing di era global. Ilmu pengetahuan seseorang dapat diperoleh melalui pendidikan di sekolah, karena dengan bersekolah kita akan mampu mewujudkan keberhasilan dan kesuksesan dalam kehidupan. Namun, kenyataan di lapangan menunjukkan bahwa ekonomi yang terbatas bagi sebagian orang tua menjadi faktor penghambat dalam mewujudkan kesuksesan anak, sehingga tidak semua anak usia wajib belajar dapat mengikuti pendidikan di sekolah. Untuk mengatasi permasalahan ini, sekolah mengadakan program beasiswa pendidikan. Yayasan Islam Adi Luhur 2 merupakan salah satu lembaga pendidikan yang memiliki SMA dan SMK, di mana sekolah ini mempunyai program Bantuan Dana Pendidikan bagi siswa yang dianggap kurang mampu secara status ekonomi. Namun, ada syarat dan ketentuan yang berlaku, dan semua kriteria pemilihan penentuan siswa yang memperoleh bantuan dana pendidikan itu dilakukan berdasarkan data siswa yang ada, lalu dianalisis secara manual, namun terkadang hasil yang diperoleh tidak sesuai. Seleksi penerima beasiswa merupakan salah satu kegiatan pengambilan keputusan yang cukup rumit karena ada beragam variabel yang digunakan sebagai bahan pertimbangan dalam proses seleksi. Variabel yang digunakan juga akan bergantung kepada jenis beasiswa yang diberikan. SMKS 2 Adi Luhur merupakan salah satu sekolah swasta di Jakarta yang berperan aktif dalam meningkatkan mutu pendidikan terhadap siswa. Saat ini, SMKS 2 Adi Luhur memiliki jumlah siswa yang cukup besar pada setiap tahun ajaran baru, kurang lebih mencapai 300-400 siswa. Dari jumlah siswa yang besar, maka dibutuhkan suatu metode yang tepat dalam menentukan beasiswa. Dengan demikian, pada penelitian ini penulis menggunakan dua Algoritma Klasifikasi Data Mining, yaitu Algoritma C4.5 dan Naive Bayes, untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan agar dalam pemberian beasiswa sesuai sasaran. Sampel data diambil dari Sekolah Menengah Kejuruan (SMK) 2 Adi Luhur pada tahun 2018/2019 yang sudah di-cleaning sebanyak 148 data siswa.

Dari penelitian ini dapat dibandingkan dengan beberapa penelitian sebelumnya antara lain penelitian yang dilakukan tentang penelitian di Universitas Hamzanwadi menggunakan Naive Bayes Classifier dengan variabel seperti status DTKS, prestasi, pekerjaan orang tua, dan lainnya. Akurasi tertinggi

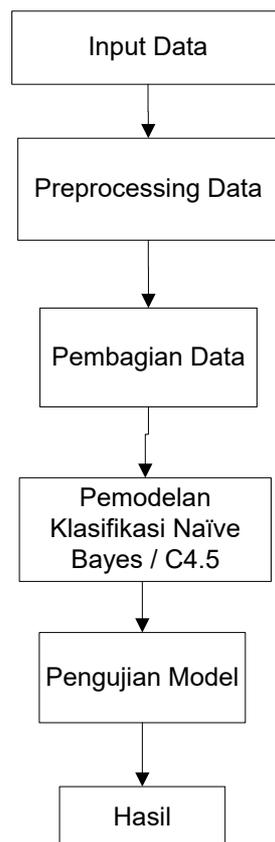
91,43% dengan AUC 0,996% yang menunjukkan algoritma ini sangat baik digunakan [1]. Kemudian Penelitian di Universitas Negeri Medan menggunakan Naive Bayes Classifier dengan variabel seperti pekerjaan orang tua, penghasilan, jumlah tanggungan, daya listrik, dan nilai UN. Akurasi tertinggi 79% dengan perbandingan data training dan testing 80:20 [2].

[3] Penelitian di SMK Swasta Anak Bangsa menggunakan C4.5 dengan variabel utama nilai siswa. Akurasi 92,7%, precision 92,05% untuk "Layak" dan 93,24% untuk "Tidak Layak".

Tujuan dari penelitian ini adalah untuk mencari pola yang dapat menentukan pemberian beasiswa dengan kriteria yang sudah ditetapkan, dengan membandingkan tingkat akurasi yang dihasilkan oleh teknik atau model data mining, yaitu Algoritma C4.5 dan Naive Bayes, di mana metode Naive Bayes digunakan untuk menentukan siswa yang berprestasi berdasarkan data identitas siswa dan data akademik siswa SMKS 2 Adi Luhur, sehingga dapat memberikan masukan kepada sekolah untuk mempermudah sistem dalam memberikan beasiswa.

## 2. Metode Penelitian

Penelitian ini menggunakan metodologi yang terdiri dari beberapa tahapan seperti, Pada gambar alur metode penelitian ini pada Gambar 1 berikut.



**Gambar 1.** Metode Penelitian

Sumber : Olah Pribadi

### 1. Input Data

Dataset yang digunakan pada penelitian ini yaitu dataset siswa SMK , sebuah dataset dibagi menjadi dataset training dan testing Dataset ini diperoleh dari internal di SMKS 2 Adi Luhur tempat dilakukan penelitian.

No	Kelas	Pekerjaan Orang Tua	Penghasilan orang Tua	Pengeluaran Orang tua	Jumlah Tanggungan orangtua	Rangking	Nilai Raport	Kerajinan	Kepribadian	Remark
1	XI	Wiraswasta	<1,000,000	<1,000,000	2-4 orang	tidak ada	>70	cukup	kurang	tidak dapat
2	XI	Tetap	>2,000,000	1500000 – 2000000	>4 orang	tidak ada	>70	cukup	cukup	tidak dapat
3	XI	Tetap	>2,000,000	1500000 – 2000000	>4 orang	2 s/d 4	>80	cukup	baik	tidak dapat
4	XI	Tidak Tetap	1500000 – 2000000	1500000 – 2000000	>4 orang	tidak ada	>70	cukup	cukup	tidak dapat
5	XI	Wiraswasta	<1,000,000	<1,000,000	2-4 orang	tidak ada	>60	cukup	kurang	dapat
6	XI	Wiraswasta	<1,000,000	<1,000,000	2-4 orang	5 s/d 10	>75	baik	baik	tidak dapat
7	XI	Wiraswasta	<1,000,000	<1,000,000	>4 orang	5 s/d 10	>75	cukup	cukup	tidak dapat
8	XI	Wiraswasta	<1,000,000	<1,000,000	2-4 orang	5 s/d 10	>75	baik	baik	tidak dapat
9	XI	Wiraswasta	<1,000,000	<1,000,000	2-4 orang	tidak ada	>70	cukup	kurang	dapat
10	XI	Tetap	>2,000,000	1500000 – 2000000	>4 orang	tidak ada	>70	cukup	cukup	dapat
11	X	Tetap	>2,000,000	1500000 – 2000000	2-4 orang	2 s/d 4	>80	baik	baik	tidak dapat
12	X	Tidak Tetap	1500000 – 2000000	1500000 – 2000000	2-4 orang	tidak ada	>70	cukup	baik	tidak dapat
13	X	Wiraswasta	>2,000,000	1500000 – 2000000	2-4 orang	2 s/d 4	>80	baik	baik	tidak dapat
14	X	Wiraswasta	>2,000,000	1500000 – 2000000	2-4 orang	tidak ada	>70	kurang	kurang	tidak dapat
15	X	Wiraswasta	>2,000,000	1500000 – 2000000	2-4 orang	tidak ada	>70	cukup	cukup	tidak dapat
16	X	Tidak Tetap	1500000 – 2000000	1500000 – 2000000	2-4 orang	5 s/d 10	>75	cukup	cukup	tidak dapat
17	X	Tidak Tetap	1500000 – 2000000	1500000 – 2000000	2-4 orang	tidak ada	>60	kurang	kurang	tidak dapat
18	X	Tidak Bekerja	<1,000,000	<1,000,000	2-4 orang	tidak ada	>70	cukup	kurang	tidak dapat
19	X	Tetap	>2,000,000	1500000 – 2000000	2-4 orang	1	>95	baik	baik	dapat
20	X	Tetap	>2,000,000	1500000 – 2000000	2-4 orang	2 s/d 4	>80	baik	baik	dapat

**Gambar 2.** Dataset Siswa SMK

Sumber : Olah Pribadi

Data tersebut memiliki 9 atribut. Berikut beberapa atribut-atribut yang ada pada dataset :

**Tabel 1.** Data Atribut Siswa

No	Atribut	Nilai Atribut
1	Kelas	X
		XI
2	Pekerjaan Orang Tua	Tidak Tetap
		Wiraswasta
		Tidak Bekerja
		Tetap
3	Penghasilan Orang Tua	>2.000.000
		1.500.000-2.000.000
		<1.000.000
4	Pengeluaran Orang Tua	1.500.000-2.000.000
		<1.000.000
5	Jumlah Tanggungan Orang Tua	2s/d 4 orang
6	Rangking	1-10
7	Nilai Raport	60 s/d 95
8	Kerajinan	Baik
		Cukup
		Kurang
9	Kepribadian	Baik
		Cukup

	Kurang	
Remark/Hasil	Dapat/Tidak Dapat	
10		

Pada saat penelitian ini dilakukan proses validasi untuk menemukan, dan mengkovarsi data agar dapat digunakan dalam metode algoritma data mining dan memperoleh akurasi serta performasi yang baik. Dalam dataset yang akan digunakan, validasi data yang digunakan dengan hapus data yang tidak lengkap atau kosong yang tidak memiliki nilai (null). Setelah itu dilakukan seleksi atribut untuk memilih atribut mana saja yang dibutuhkan dari data set yang digunakan dalam proses menganalisis kelayakan pemberian beasiswa kepada calon siswa. Atribut yang diambil dari data pengajuan beasiswa.

2. Preprocessing Data

Tahap Pada tahapain ini, dataset dilakukan serangkaian langkah untuk menyiapkan data yang diperlukan sebelum digunakan dalam pemodelan menggunakan algoritma C4.5 dan algoritma Naïve Bayes.

3. Pembagian Data

Mula-mula data diberikan pelabelan secara manual dan setelah itu dilakukanlah pembagian data menjadi 2 yaitu data training dan data testing. Dimana terdapat 2 kelompok dalam penelitian ini yaitu kelompok data yang pertama adalah jumlah data testing lebih sedikit dari data training. Sedangkan kelompok data kedua adalah data lain lebih banyak dari data training.

4. Algoritma C4.5

Merupakan salah satu metode klasifikasi yang digunakan. Melibatkan konstruksi pohon keputusan, koleksi node keputusan. Setiap cabang kemudian mengarah ke node lain baik keputusan atau ke node daun untuk mengakhiri [4]. C4.5 adalah algoritma yang mempunyai input berupa training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya dan samples yang merupakan field - field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma dasar dari C4.5 adalah sebagai berikut:

- a. Pohon yang dihasilkan berupa pohon terbalik,
- b. Pada tahap awal, semua contoh training adalah akar.
- c. Atribut adalah kategori.
- d. Contoh di partisi secara berulang berdasarkan atribut yang dipilih.
- e. Atribut tes dipilih dari data heuristic atau pengukuran statistik

Karena algoritma C4.5 digunakan untuk melakukan klasifikasi, jadi hasil dari pengolahan test dataset berupa pengelompokkan data ke dalam kelas-kelasnya Umumnya, langkah-langkah algoritma C4.5 yang digunakan untuk membentuk pohon keputusan adalah :

- a. Pilih atribut sebagai root.
- b. Buat cabang untuk setiap nilai
- c. Bagi tiap cabang kedalam kelas.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada tiap cabang memiliki kelas yang sama

5. Algoritma Naive Bayes

Merupakan teknik probabilistik klasifikasi berdasarkan teorema Bayes dengan asumsi independensi diantara variabel prediktor. Secara sederhana, pengelompokan Naïve Bayes menganggap adanya suatu fitur tertentu dalam sebuah kelas tidak terkait dengan adanya fitur lainnya. Rumus Persamaan dari Teorema Bayes adalah [5]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- $P(H|X)$  : Probabilitas hipotesis  $H$  berdasar kondisi  $X$  (posteriori probability)
- $P(H)$  : Probabilitas hipotesis  $H$  (posteriori probability)
- $P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$
- $P(X)$  : Probabilitas  $X$

### 3. Hasil dan Pembahasan

#### 3.1 Evaluasi dan Validasi Hasil

##### a. Hasil Pengujian Algoritma C4.5

Setelah model klasifikasi data mining prediksi penerima beasiswa terbentuk, untuk menilai keakuratan sebuah model akan dilakukan pengujian dengan teknik K-Fold Cross Validation dimana nilai k=10, yang berarti proses pengulangan sebanyak 10 kali pengulangan. Hasilnya model memiliki keakuratan sebesar 92.52%.

accuracy: 92.52% +/- 5.64% (mikro: 92.57%)

	true tidak dapat	true dapat	class precision
pred. tidak dapat	122	9	93.13%
pred. dapat	2	15	88.24%
class recall	98.39%	62.50%	

**Gambar 1.** Model Klasifikasi Algoritma C4.5 pada RapidMiner  
 Sumber : Olah Pribadi

Selain dilakukan pengujian model dengan teknik K-Fold Cross Validation, untuk mengevaluasi kinerja model klasifikasi yang dapat memprediksi benar atau tidak benarnya oleh model tersebut, maka digunakan teknik tabel Confusion Matrix. Berikut hasil tabel ukur Confusion Matrix terhadap algoritma C4.5 dengan menghasilkan akurasi sebesar 92.52%

#### PerformanceVector

```
PerformanceVector:
accuracy: 92.52% +/- 5.64% (mikro: 92.57%)
ConfusionMatrix:
True:   tidak dapat   dapat
tidak dapat:  122     9
dapat:      2       15
```

**Gambar 4.** Hasil Confusion Matrix Algoritma C4.5  
 Sumber : Olah Pribadi

##### a. Hasil Pengujian Algoritma Naive Bayes

Pada tahap ini peneliti menggunakan metode algoritma Naive Bayes untuk mengaplikasikan data yang telah mengalami proses preprocessing data atau pembersihan data pada aplikasi Rapidminer. Berdasarkan pengujian yang dilakukan menggunakan aplikasi Rapidminer didapatkan hasil yaitu: algoritma Naive Bayes mencapai akurasi 94.52%.

accuracy: 94.52% +/- 5.96% (mikro: 94.59%)

	true tidak dapat	true dapat	class precision
pred. tidak dapat	122	6	95.31%
pred. dapat	2	18	90.00%
class recall	98.39%	75.00%	

**Gambar 5** Model Klasifikasi Naive Bayes C4.5 pada RapidMiner  
 Sumber : Olah Pribadi

Selain dilakukan pengujian model dengan teknik *K-Fold Cross Validation*, untuk mengevaluasi kinerja model klasifikasi yang dapat memprediksi benar atau tidak benarnya oleh model

tersebut, maka digunakan teknik tabel *Confusion Matrix*. Berikut hasil tabel ukur *Confusion Matrix* terhadap Algoritma Naive Bayes dengan menghasilkan nilai akurasi sebesar 94.52%

### PerformanceVector

```
PerformanceVector:
accuracy: 94.52% +/- 5.96% (mikro: 94.59%)
ConfusionMatrix:
True:   tidak dapat   dapat
tidak dapat:  122     6
dapat:    2         18
```

**Gambar 6** Confusion Matrix Naive Bayes  
Sumber : Olah Pribadi

Hasil Perbandingan Model C4.5 dan Naive Bayes. Dalam penelitian ini penulis membandingkan model klasifikasi dengan menggunakan Algoritma C4.5 dan Naive Bayes meliputi akurasi model, tingkat error, precision, recall dan waktu komputasi. Hasil perbandingan ini diperoleh dengan bantuan *tools RapidMiner*. ini merupakan hasil perbandingan model klasifikasi antara Algoritma C.45 dan Algoritma Naive Bayes yang ditunjukkan pada tabel 1 :

**Tabel 2.** Perbandingan Model Klasifikasi

Komponen	Algoritma C4.5	Naive Bayes
Akurasi	92.52%	94.52%
Error	7.48%	5.48%
Precision	98.39%	98.39%
Recall	93.13%	95.31%
Waktu Komputasi	0.01 detik	0.01 detik

Berdasarkan tabel 4.6 dapat diperoleh hasil akurasi Algoritma C.45 lebih rendah dibandingkan Naive Bayes yaitu 92.52% berbanding 94.52%. Tingkat error dari model klasifikasi Algoritma C.45 adalah lebih besar dibandingkan dengan Naive Bayes yaitu sebesar 7.48% dan 5.48%. Begitupun halnya dari komponen precision dan recall Algoritma Naive bayes lebih unggul dibandingkan Algoritma C4.5.

#### 4. Kesimpulan

Penelitian ini mengembangkan metode untuk mendapatkan skema penerimaan beasiswa yang optimal dengan pemerataan tertinggi bagi penyelenggara sekolah. Metode tersebut dapat diterapkan karena memenuhi persyaratan pemerataan bahwa siswa yang berprestasi lebih baik harus menerima beasiswa yang sama atau lebih dari yang diterima oleh siswa yang kurang berprestasi; pemberian beasiswa meniadakan kebutuhan mahasiswa untuk mengajukan beasiswa tertentu secara manual, yang merupakan proses yang memakan waktu dan energi. Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma Naive Bayes memiliki performansi yang lebih baik. Dari hasil perbandingan tersebut didapat hasil akurasi Naive Bayes lebih tinggi dibanding dengan Algoritma C4.5. Hasil yang didapat dari perbandingan kedua algoritma tersebut adalah Algoritma Naive Bayes memiliki tingkat akurasi sebesar 94,52% dan algoritma C4.5 memiliki tingkat akurasi sebesar 92,52%.

#### Referensi

- [1] N. I. Nurhidayati, Y. Yahya, F. Fathurrahman, L. . Samsu, and W. Amnia, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 177–188, 2023, doi: 10.29408/jit.v6i1.7529.

- [2] F. Fatmawati, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *None*, vol. 13, no. 1, pp. 50–59, 2016.
- [3] M. Sari, A. Perdana Windarto, and H. Okprana, "Penerapan Data Mining Klasifikasi C4.5 Pada Penerima Beasiswa di SMK Swasta Anak Bangsa," *BEES Bull. Electr. Electron. Eng.*, vol. 1, no. 3, pp. 115–121, 2021.
- [4] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. 2014.
- [5] Sunil Ray, "Penjelasan Pengklasifikasi Naive Bayes: Penerapan dan Masalah Praktik Pengklasifikasi Naive Bayes," *Analytics Vidhya*, 2024. [Online]. Available: <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>. [Accessed: 05-May-2024].

*This page is intentionally left blank.*

# Pengembangan Sistem Informasi Digital Zoonosis Melalui Pendekatan Web Semantik

Komang Kartika Noviyanti<sup>a1</sup>, Luh Gede Astuti<sup>a2</sup>, I Gede Surya Rahayuda<sup>a3</sup>, Agus Muliantara<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana  
Badung, Bali, Indonesia

<sup>1</sup>kartikanovi28@gmail.com

<sup>2</sup>lg.astuti@unud.ac.id

<sup>3</sup>igedesuryarahayuda@unud.ac.id

<sup>4</sup>muliantara@unud.ac.id

## Abstract

Bali Province with a population of 44,426,000 in 2023, faces challenges in ensuring sufficient food supply. Zoonotic diseases, which can be transmitted from animals to humans, pose a serious and underreported threat, affecting both animal production and human health. To address these challenges, the web-based application *Healthy Zoonosis* was developed, providing comprehensive digital information on animal diseases, including zoonoses and non-zoonoses, in Bali from 2015 to 2022. The application features browsing, search, and a dashboard for viewing overall disease data. The research utilized Design Science Research Methodology (DSRM) and Mentontology for ontology modeling, with Prototyping for application design. Evaluation involved 17 Participants for ontology validation, yielding a 100% valid result, and 50 Participants for usability testing, scoring 87.3 on the System Usability Scale (SUS). The app not only demonstrates accuracy and validity but also signifies a shift in animal disease prevention paradigms. It is hoped that this application will enhance public understanding of animal health and support proactive government measures in addressing animal disease challenges.

**Keywords:** Zoonoses, Ontology, Mentontology, Blackbox Testing, System Usability Scale (SUS)

## 1. Pendahuluan

Provinsi Bali memiliki luas wilayah sebesar 5.780 km<sup>2</sup> yang merupakan 0,29% dari total luas wilayah Indonesia, memiliki populasi yang terus meningkat. Data BPS Provinsi Bali mencatat populasi mencapai 44.426.000 orang pada tahun 2023. Pertumbuhan populasi ini berpotensi meningkatkan permintaan akan kebutuhan pangan. Namun, ketidakseimbangan antara pertumbuhan populasi dan produksi pangan dapat mengakibatkan krisis pangan. Kesehatan hewan dan penyakit zoonosis, yang dapat menular antara hewan dan manusia, menjadi perhatian penting. Ketidaktahuan mengenai penyakit ini dapat berpotensi berbahaya mengingat dampaknya yang dapat menyebabkan kematian pada tingkat keparahan yang tinggi [1]. Di Indonesia, terdapat 15 penyakit zoonosis yang diidentifikasi sebagai prioritas, antara lain *avian influenza*, *rabies*, *anthraks*, *brucellosis*, *leptospirosis*, *japanese b. encephalitis*, *bovine tuberculosis*, *salmonellosis*, *skistosomiasis*, *q fever*, *campylobacteriosis*, *trikinosi*, *para tuberculosis*, *toksoplasmosis*, *sistiserkosis* dan *taeniasis* [2]. Untuk mengatasi permasalahan ini, pengembangan aplikasi berbasis web menjadi solusi yang tepat. Aplikasi ini bertujuan untuk membantu pengumpulan pengetahuan tentang penyakit hewan secara digital dan eksplisit. Dengan teknologi semantik web dan ontologi, informasi tentang penyakit hewan dapat diakses dengan mudah melalui internet.

Untuk mengatasi permasalahan ini, pengembangan aplikasi berbasis web menjadi solusi yang tepat. Aplikasi ini bertujuan untuk membantu pengumpulan pengetahuan tentang penyakit hewan secara digital dan eksplisit. Dengan teknologi semantik web dan ontologi, informasi tentang penyakit hewan dapat diakses dengan mudah melalui internet. Ontologi yang akan dikembangkan berfungsi sebagai elemen metadata yang esensial dalam menyusun suatu aplikasi perpustakaan daring berbasis

semantik web. Dengan demikian, pengembangan ontologi menggunakan metode Methontology merupakan pendekatan untuk mengembangkan model ontologi yang menguraikan setiap aktivitas dengan detail.

Metode *Methontology* digunakan untuk mengembangkan model ontologi dengan fokus pada penjelasan yang terperinci dari setiap aktivitas. Pendekatan ini menonjol dalam menguraikan aktivitas secara terperinci serta memfasilitasi penggunaan kembali ontologi yang telah dibuat untuk pengembangan sistem lebih lanjut. Selanjutnya, pengembangan sistem dilaksanakan melalui metode *Prototyping*, sebuah pendekatan pengembangan perangkat lunak yang menggabungkan interaksi antara pengembang dan pengguna selama proses pembuatan. Hal ini memungkinkan pengembang untuk dengan mudah membentuk model perangkat lunak yang sedang dibuat. Kolaborasi dengan Balai Besar Veteriner Denpasar dan masyarakat Provinsi Bali diharapkan dapat meningkatkan pemahaman tentang kesehatan hewan di Bali melalui aplikasi ini.

## 2. Metode Penelitian

Dalam penelitian, peneliti menerapkan *Design Science Research Methodology* (DSRM) sebagai pendekatan penelitian. Metode penelitian DSRM diterapkan dalam bidang teknologi informasi, terutama dalam konteks perancangan sistem informasi [3]. Pendekatan DSRM menitikberatkan pada perancangan dan pengembangan solusi untuk masalah tertentu dalam konteks organisasi atau masyarakat. Dalam disiplin ilmu desain sistem informasi, pendekatan penelitian ini mengintegrasikan prinsip-prinsip, praktik-praktik, dan prosedur-prosedur tertentu. Berikut ini merupakan alur metodologi DSRM yang dapat dilihat pada Gambar 1 dibawah ini.



**Gambar 1.** Metode *Design Science Research Methodology* (DSRM)

### 2.1. Penentuan Masalah dan Solusi

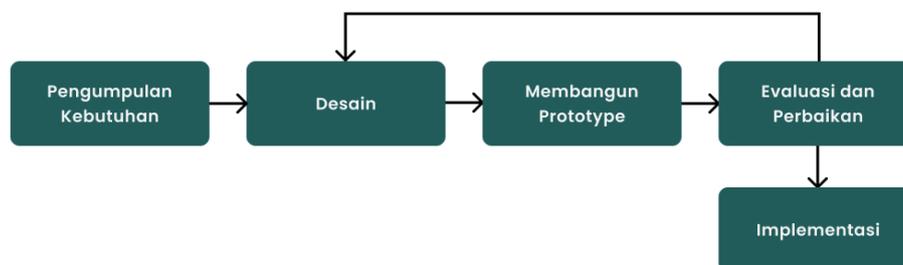
Pada tahap ini, dilakukan observasi dan pemahaman terhadap permasalahan yang sedang dihadapi. Proses dimulai dengan melakukan *Problem-Centered Initiation*, yang menyoroti kurangnya pengetahuan atau informasi di kalangan masyarakat mengenai penyakit menular yang dapat ditularkan dari hewan ke manusia maupun tidak menular ke manusia. Setelah melakukan penyebaran kuesioner dimana terdapat 100 partisipan dengan rentang usia 18 hingga 52 tahun untuk mengetahui pengetahuan masyarakat provinsi Bali tentang penyakit yang dapat menular ke manusia. Dapat disimpulkan bahwa masih terdapat tingkat pengetahuan yang rendah di kalangan masyarakat mengenai penyakit yang dapat ditularkan melalui hewan.

## 2.2. Tujuan untuk Masalah

Setelah pengumpulan data pada penelitian, solusi-solusi ditetapkan untuk mengatasi permasalahan yang diidentifikasi sebelumnya, dengan mempertimbangkan tujuan solusi tersebut. Salah satu solusi yang diusulkan adalah pengembangan aplikasi berbasis web yang akan menjadi sumber informasi bagi masyarakat terkait diagnosa penyakit yang dapat ditularkan dari hewan, serta menyediakan informasi tambahan terkait penyakit hewan tersebut. Dengan memanfaatkan teknologi berbasis web, diharapkan informasi mengenai penyakit hewan, seperti nama penyakit, deskripsi, mikroorganisme penyebab, hewan yang rentan terinfeksi, cara penularan, gejala jika ditularkan ke manusia, metode pencegahan terhadap manusia, dan upaya pencegahan terhadap hewan sesama, dapat disajikan dengan lebih praktis. Harapannya, informasi yang tersedia melalui web akan memungkinkan masyarakat untuk mengakses pengetahuan tentang kesehatan hewan dengan mudah dan dapat diakses dimanapun. Dalam pengembangan aplikasi web ini, digunakan model ontologi karena kemampuannya yang efektif dalam menyajikan informasi secara semantik. Model ini membantu mengorganisir dan menghubungkan informasi dengan jelas, memudahkan pengguna memahami keterkaitan antar konsep penyakit hewan. Aplikasi ini terdiri dua fitur utama yaitu fitur penjelajahan dan fitur pencarian. Aplikasi ini diharapkan dapat menjadi sumber pengetahuan efektif untuk meningkatkan pemahaman masyarakat tentang penyakit hewan dan langkah-langkah pencegahan, terutama di Provinsi Bali.

## 2.3. Tahap Desain dan Pengembangan Aplikasi

Untuk pengembangan sistem, metode yang diterapkan adalah *Prototyping*. Metode ini memungkinkan interaksi langsung antara pengguna dan pengembang, mempercepat umpan balik. Prosesnya melibatkan pengumpulan kebutuhan, perancangan, pembangunan prototipe, evaluasi, perbaikan, dan implementasi [4]. *Prototyping* dipilih karena keterlibatan aktif pengguna dan iterasi melalui evaluasi dan perbaikan memastikan sistem memenuhi standar kualitas dan kebutuhan pengguna dengan lebih efisien.



**Gambar 2.** Pengembangan sistem dengan menggunakan metode *prototyping*

Pada Gambar 2 diatas menunjukkan alur tahapan dalam metode prototyping, diawali dari analisis kebutuhan, kemudian menuju desain, dilanjutkan dengan pembangunan prototyping, evaluasi, perbaikan, dan diakhiri dengan implementasi. Berikut penjelasan singkatnya:

a. Analisis Kebutuhan

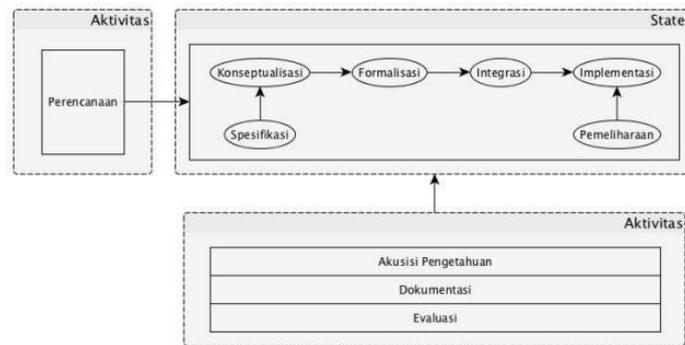
Tahap ini dilakukan untuk mengidentifikasi kebutuhan sistem dengan menganalisis kebutuhan fungsional dan non-fungsional. Kebutuhan fungsional mencakup fitur-fitur utama seperti menampilkan informasi lengkap tentang penyakit hewan dan sistem pencarian yang spesifik, sedangkan kebutuhan non-fungsional mencakup kebutuhan perangkat keras dan lunak.

b. Pengumpulan Data

Pada tahap ini, data diperoleh dari Peta Penyakit Hewan di Provinsi Bali melalui situs resmi Balai Besar Veteriner Kota Denpasar. Informasi tambahan tentang penyakit hewan untuk pengujian dan evaluasi sistem. Data ini digunakan untuk membangun model ontologi yang memenuhi kriteria sistem berdasarkan informasi penyakit hewan di Provinsi Bali dari tahun 2015 hingga 2022.

c. Pembangunan Model Ontologi

Metode *Methontology* menjadi pilihan populer dalam pengembangan model ontologi karena menawarkan beberapa keunggulan, salah satunya kemampuannya dalam memberikan deskripsi rinci terhadap setiap aktivitas. Pendekatan *Methontology* memungkinkan pengembang sistem untuk menggunakan kembali ontologi yang dibangun di masa mendatang. Representasi metode *Methontology* ini lah yang digunakan dalam pengembangan sistem manajemen pengetahuan untuk *ontology* berdomain zoonosis. Tahapannya mencakup *aktivitas*, *states*, dan *methods* dapat dilihat pada Gambar 3.



**Gambar 3.** Metodologi Pengembangan Ontologi

d. Desain

Pada tahap ini melakukan perancangan aplikasi dengan *Modeling Quick Design* adalah proses perancangan aplikasi yang mencakup alur kerja, aktor, dan proses yang berinteraksi dalam aplikasi. Proses ini menggunakan *Unified Modeling Language* (UML) untuk representasi visual yang terdiri dari beberapa diagram, diantaranya :

a. *Use Case Diagram*

*Use Case Diagram* adalah jenis diagram yang menggambarkan fungsionalitas yang diharapkan dari sistem. Dimana menunjukkan interaksi antara pengguna dengan sistem.

b. *Activity Diagram*

*Activity Diagram* adalah jenis diagram yang digunakan untuk memodelkan alur kerja atau aktivitas dalam sistem. Diagram ini menunjukkan urutan aktivitas dalam sebuah proses dan membantu memahami proses secara keseluruhan

c. *Sequence Diagram*

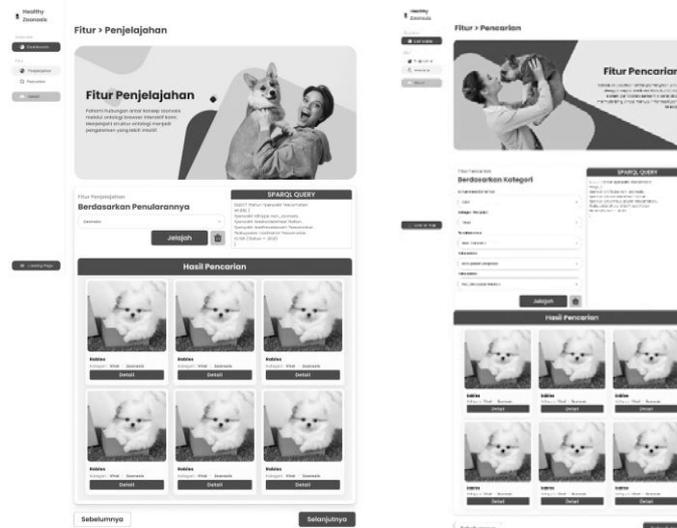
*Sequence Diagram* adalah representasi visual dari interaksi antara objek dalam sebuah aktivitas sistem, menggambarkan urutan waktu dari interaksi tersebut.

d. *Class Diagram*

*Class diagram* menunjukkan struktur sistem dalam hal pemilihan kelas yang diperlukan untuk membuat sistem.

e. Perancangan *User Interface*

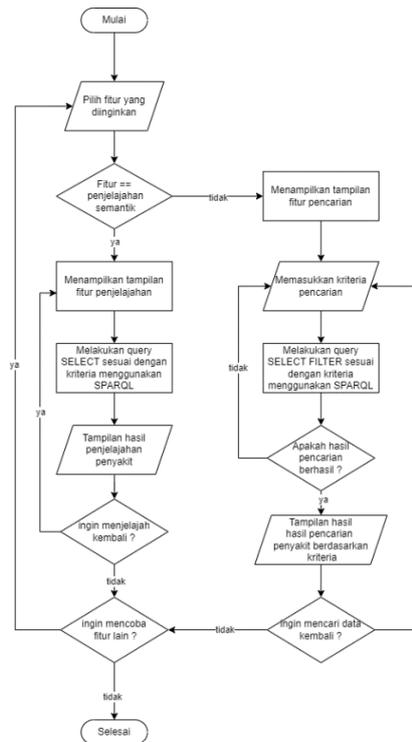
Rancangan antarmuka pengguna sistem ditujukan untuk pengguna. Ini mencakup rancangan antarmuka untuk fitur pencarian dan penjelajahan. Gambar 4 merupakan *wireframe* aplikasi untuk mempermudah memvisualisasikan struktur antarmuka sebelum implementasi lebih detail.



**Gambar 4.** *User Interface* Fitur Penjelajahan dan Fitur Pencarian

## 2.4. Implementasi

Tahap perancangan sistem ini khususnya implementasinya, bertujuan untuk memperkenalkan alur kerja dari sistem pengetahuan penyakit hewan. Pengguna dapat mengakses fitur penelusuran dan pencarian semantik, yang memungkinkan interaksi yang lebih efektif dan mendalam dengan informasi yang disediakan oleh sistem. Gambar 5 merupakan flowchart yang menggambarkan alur kerja dari sistem pengetahuan penyakit hewan, khususnya implementasi fitur penelusuran dan pencarian.



Gambar 5. Flowchart Fitur Penelusuran dan Fitur Pencarian

## 2.5. Demonstrasi

Proses menguji beberapa komponen sistem atau aplikasi yang telah dikembangkan disebut tahap demonstrasi. Tujuan dari tahap demonstrasi adalah untuk memastikan bahwa sistem atau aplikasi yang dikembangkan sesuai dengan perancangan awal

## 2.6. Evaluasi

Proses evaluasi mencakup pengujian model ontologi, sistem fungsional, dan sistem non-fungsional. Pada tahap pertama, *Regression testing* dilakukan dengan menggunakan kasus uji yang dibuat pada tahap awal pengembangan suatu aplikasi. Dilanjutkan dengan metode pengujian perangkat lunak yang dikenal sebagai *Black-box Testing*, yang berfokus pada spesifikasi fungsi perangkat lunak [5]. Dilanjutkan dengan Sistem Usability Scale (SUS), yang menggunakan kuesioner untuk menilai kemudahan sistem komputer dari sudut pandang pengguna. *Regression testing* dilakukan sesuai dengan *usecase* diagram menggunakan aplikasi Katalon Studio, yang menunjukkan bagaimana interaksi antara user dan sistem digambarkan. *Black-Box* terfokus pada fungsionalitas fitur sistem seperti pencarian dan penjelajahan, serta mengevaluasi input, output, dan fungsi halaman tampilan. Untuk mengevaluasi pengalaman dan persepsi pengguna terhadap situs web *Healthy Zoonosis*, Sistem Usability Scale (SUS) digunakan. SUS menggunakan skala *Likert* dari "Sangat Tidak Setuju" hingga "Sangat Setuju". [6].

## 2.7. Komunikasi

Dalam tahap komunikasi, peneliti terlibat dalam diskusi untuk mendapatkan pemahaman yang lebih dalam tentang dinamika penelitian. Hal ini dilakukan dengan berbagi informasi untuk meningkatkan pemahaman dan mencapai pengetahuan tentang penyakit hewan secara lebih efisien.

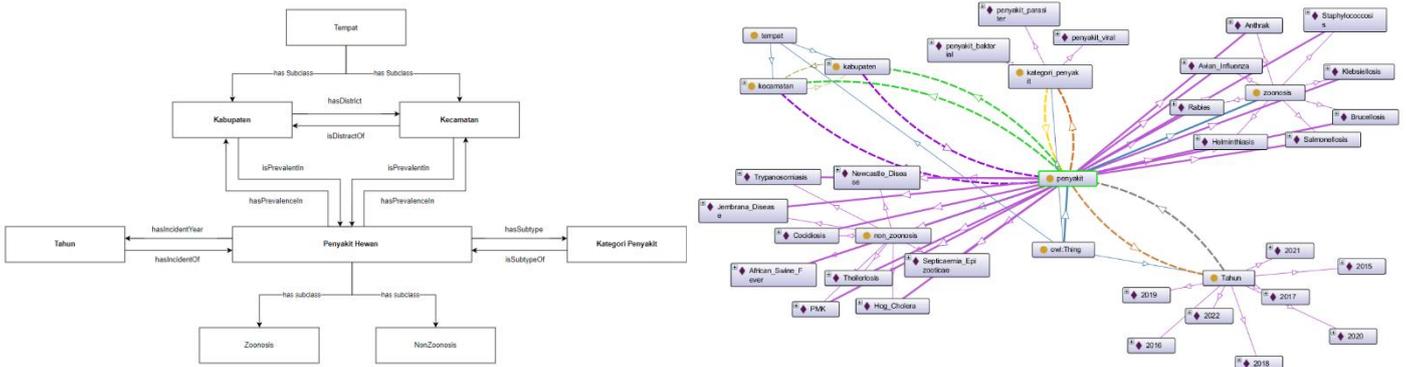
## 3. Hasil dan Pembahasan

Pada penelitian ini, hasil perancangan dan pengembangan sistem informasi Zoonosis diuraikan bersamaan dengan hasil pengujian dan evaluasi sistem tersebut.

### 3.1. Pembangunan Model Ontologi

Pada implementasi pembangunan ontologi pada domain zoonosis, tahap konseptualisasi metode *methontology* bertujuan untuk mengorganisasi dan mengelola pengetahuan yang diperoleh selama proses akuisisi pengetahuan. Pembangunan model ontologi dilakukan dengan menyusun hirarki berdasarkan kelas-kelas yang terdapat pada penyakit hewan di Provinsi Bali. Struktur hirarki ontologi

ini dibentuk berdasarkan komponen-komponen yang terkait dengan zoonosis, seperti yang ditunjukkan dalam gambar 6 berikut:



**Gambar 6.** Konseptual dan Implementasi *Ontology* Zoonosis

Dalam implementasi ontologi ini, terdapat empat kelas utama, yaitu *class* kategori penyakit, *class* penyakit yang memiliki *Subclass non* zoonosis dan zoonosis, *class* tahun, serta *class* tempat yang memiliki *subclass* kabupaten dan kecamatan, menghasilkan 8 *object property* yang digunakan untuk menghubungkan individu satu dengan yang lain menggunakan konsep *Triplet Pattern* (predikat) pada ontologi, 8 *data property* untuk memberikan nilai dari individual penyakit hewan seperti nama penyakit, mikroorganisme, hewan, deskripsi, ciri-ciri penularan zoonosis, cara penularan, cara pencegahan ke hewan dan ke manusia. Dengan menggunakan *data property*, ontologi dapat membantu mengorganisir pengetahuan dengan lebih jelas. Proses penghubungan antar individu satu dengan yang lainnya menjadi krusial saat merancang model ontologi dengan menggunakan konsep *triplet pattern* (subjek, predikat, objek), yang kemudian diwakili dalam format RDF, RDFS, dan OWL.

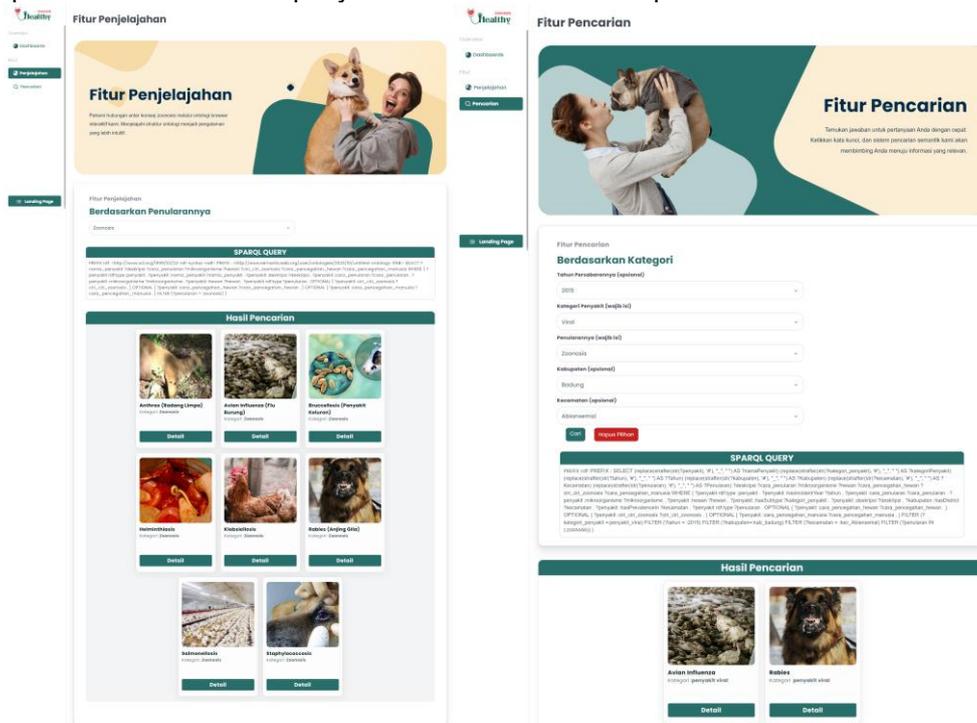
### 3.2. Implementasi Sistem

Sistem ini hanya memiliki satu jenis pengguna, yaitu pengguna (aktor), yang bertanggung jawab untuk melakukan pencarian dan pencarian dalam aplikasi. Fitur utama aplikasi ini adalah fitur penjelajahan, yang memungkinkan pengguna melakukan penjelajahan dalam sistem dengan memilih informasi penyakit hewan berdasarkan penularannya dan sistem menampilkan informasi yang berasal dari inputan yang mereka berikan. Sementara itu, fitur pencarian adalah aktivitas di mana pengguna memilih informasi yang ingin diketahui dengan melakukan pemfilteran data, sehingga informasi pengetahuan terkait penyakit menjadi lebih spesifik dan relevan. Dalam pengembangan antarmuka pengguna (*User Interface*) untuk sistem manajemen pengetahuan zoonosis berbasis *website* ini, digunakan *framework express.js* sebagai *backend* dan *Bootstrap* untuk membangun tampilan antarmuka (*frontend*) halaman tersebut. Dalam pengelolaan data, server *Apache Jena Fuseki* digunakan. Sistem manajemen pengetahuan Zoonosis memiliki antarmuka pengguna sebagai berikut.



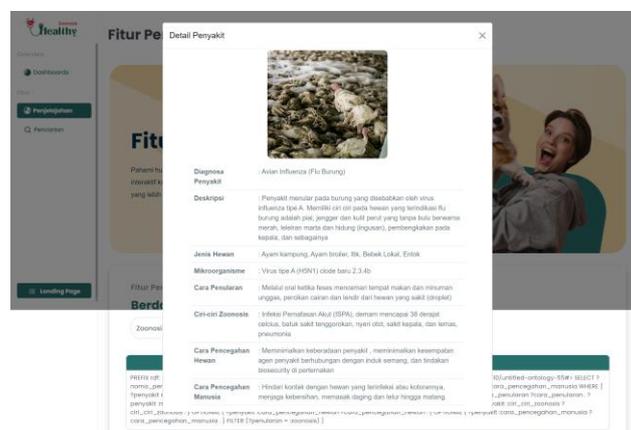
**Gambar 7.** *User Interface* Halaman *Dashboard*

Pada Gambar 7 ditampilkan implementasi halaman *dashboard* yang menampilkan informasi lengkap tentang jumlah data penyakit hewan yang tersebar di Provinsi Bali. Halaman ini memberikan gambaran yang komprehensif terkait situasi penyakit hewan di 9 kota/kabupaten di Provinsi Bali.



**Gambar 8.** User Interface Halaman Fitur Penjelajahan dan Fitur Pencarian

Pada Gambar 8 terlihat antarmuka halaman fitur penjelajahan memberikan pengguna pilihan melalui *dropdown* berdasarkan jenis penularan penyakit hewan. Pengguna dapat memilih kriteria yang diinginkan untuk menampilkan informasi tentang penyakit hewan sesuai dengan pilihan mereka. User juga dapat melihat informasi tambahan tentang penyakit hewan melalui *pop-up* yang tersedia. Dilanjutkan dengan antarmuka halaman fitur pencarian yang memungkinkan pengguna untuk melakukan pencarian dengan menggunakan lima kriteria, termasuk tahun persebaran, kategori penyakit, penularan, kecamatan, dan kabupaten. Pengguna diharuskan mengisi kriteria utama sebelum menekan *button* "Cari", dan hasil pencarian akan menampilkan secara *real-time* menggunakan *query* SPARQL yang telah ditentukan. Pengguna juga dapat melihat detail dari penyakit hewan yang ditemukan, serta menghapus pilihan untuk mereset hasil pencarian sebelumnya.



**Gambar 9.** User Interface Halaman *Pop-up* Detail Penyakit Hewan

Pada Gambar 9 menampilkan antarmuka pengguna dari *pop-up* detail penyakit hewan. *Pop-up* ini menyajikan informasi rinci mengenai suatu penyakit hewan, yang dapat diakses oleh pengguna dengan mengklik atau mengarahkan kursor ke *button* "Detail". Informasi yang ditampilkan mungkin mencakup diagnosa penyakit, deskripsi, jenis hewan, *mikroorganisme*, cara penularan, ciri ciri zoonosis, cara pencegahan ke hewan, dan cara pencegahan ke Manusia. Dengan tampilan yang jelas dan mudah

dimengerti, *pop-up* ini memberikan pengguna akses cepat untuk mendapatkan informasi yang dibutuhkan tentang penyakit hewan.

### 3.3. Pengujian dan Evaluasi Sistem

Pada penelitian ini, peneliti memiliki 3 kriteria utama yakni kriteria rentan usia Partisipan yakni 17-70 tahun, memiliki ketertarikan informasi tentang penyakit hewan, dan berasal dari masyarakat asli Provinsi Bali. Sehingga peneliti melakukan pengujian berjumlah minimal 30 orang.

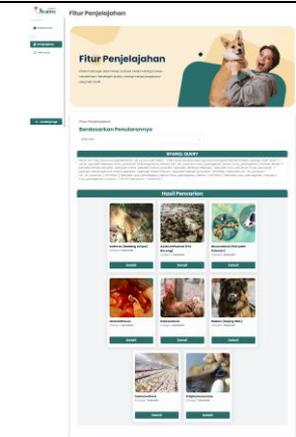
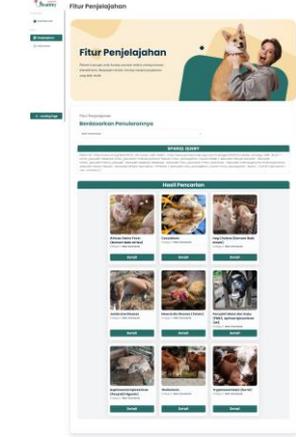
#### a. Pengujian *Regression Testing*

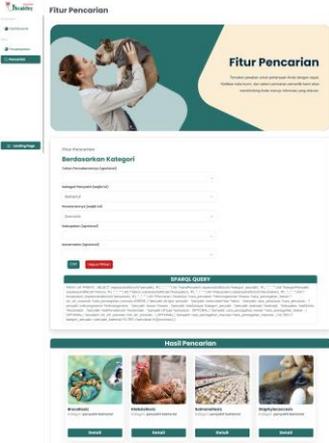
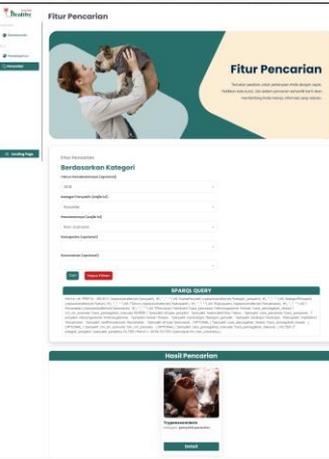
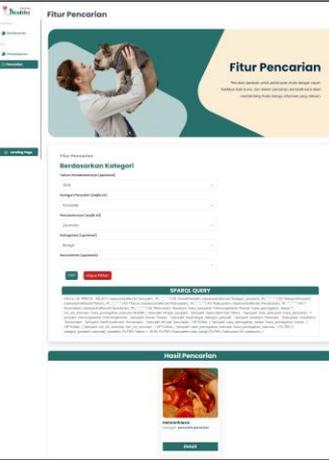
*Regression testing* dapat menguji fungsi program secara fungsional dan non-fungsional. Biasanya, *regression testing* menggunakan kasus uji yang telah ditulis pada tahap awal pengembangan suatu aplikasi. Berdasarkan hasil *regression testing* pada aplikasi Healthy Zoonosis yang dilakukan dengan menggunakan Katalon Studio, dapat disimpulkan bahwa semua fitur yang diuji berhasil melewati pengujian sesuai dengan yang diharapkan. Dengan kata lain, dari 7 skenario pengujian yang dilakukan, semuanya dinyatakan "*Passed*". Ini berarti bahwa semua fitur aplikasi telah diuji secara menyeluruh dan memberikan hasil yang sesuai dengan spesifikasi yang telah ditetapkan, mencapai tingkat keberhasilan 100%.

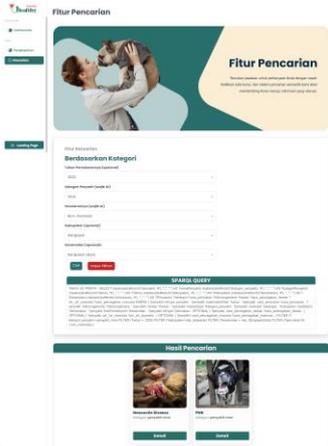
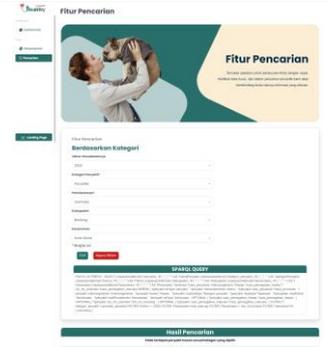
#### b. Pengujian *Blackbox Testing*

Pengujian *Black-Box Testing* difokuskan pada fungsionalitas fitur sistem, terutama *searching* dan *browsing*. Setiap skenario dijalankan secara berurutan, dan akan kembali ke menu awal jika percakapan sebelumnya tidak valid. Tabel 1 merupakan hasil pengujian blackbox testing yang dilakukan untuk memvalidasi model ontologi pada *website Healthy Zoonosis*.

**Tabel 1.** Hasil Pengujian Validitas Model Ontologi pada Sistem

Kode	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
P1	Menampilkan jumlah penyakit hewan berdasarkan penularannya tergolong <b>zoonosis</b> pada sistem		VALID
P2	Menampilkan jumlah penyakit hewan berdasarkan penularannya tergolong <b>non-zoonosis</b> pada sistem		VALID
Q1	Menampilkan <b>detail penyakit hewan</b> pada sistem untuk memberikan informasi lengkap mengenai penyakit hewan		VALID

<p>Q2</p>	<p>Menampilkan penyakit hewan dengan kategori <b>penyakit bakterial</b> dan termasuk menular ke manusia (<b>zoonosis</b>)</p>		<p>VALID</p>
<p>Q3</p>	<p>Menampilkan penyakit hewan dengan kategori <b>penyakit parasiter</b>, termasuk tidak menular ke manusia (<b>non-zoonosis</b>) yang terjadi pada tahun <b>2018</b></p>		<p>VALID</p>
<p>Q4</p>	<p>Menampilkan penyakit hewan dengan kategori <b>penyakit parasiter</b>, termasuk menular ke manusia (<b>zoonosis</b>) yang terjadi pada tahun <b>2016</b> di <b>kabupaten Bangli</b></p>		<p>VALID</p>
<p>Q5</p>	<p>Menampilkan penyakit hewan yang teridentifikasi pada sistem dengan <b>kategori penyakit viral</b> termasuk tidak menular ke manusia (<b>non-zoonosis</b>) yang terjadi pada tahun <b>2018</b> di <b>kabupaten Jembrana</b> tepatnya berlokasi di <b>kecamatan Pekutatan</b></p>		<p>VALID</p>

<p>Q6</p>	<p>Menampilkan penyakit hewan kategori <b>penyakit viral</b>, termasuk tidak menular ke manusia (<b>non-zoonosis</b>) yang terjadi pada tahun <b>2022</b> di <b>kabupaten Denpasar</b> tepatnya berlokasi di <b>kecamatan Denpasar Utara</b></p>		<p>VALID</p>
<p>Q7</p>	<p>Menampilkan penyakit hewan yang teridentifikasi pada sistem dengan kategori <b>penyakit bakterial</b>, termasuk menular ke manusia (<b>zoonosis</b>) yang terjadi pada tahun <b>2015</b> di <b>kabupaten Gianyar</b> tepatnya berlokasi di <b>kecamatan Payangan</b>? jika ada, mohon tampilkan informasinya!</p>		<p>VALID</p>
<p>Q8</p>	<p>Menampilkan penyakit hewan yang teridentifikasi pada sistem dengan <b>kategori penyakit parasiter</b>, termasuk menular ke manusia (<b>zoonosis</b>) yang terjadi pada tahun <b>2022</b> di <b>kabupaten Badung</b> tepatnya berlokasi di <b>kecamatan Kuta Utara</b></p>		<p>VALID</p>

Hasil pengujian *website Healthy Zoonosis* dari 17 Partisipan mahasiswa Program Studi Informatika Universitas Udayana menunjukkan bahwa pengujian menggunakan metode *blackbox testing* penting karena setiap perangkat lunak harus diuji sebelum digunakan. Dari 10 pengujian, semua valid (10 valid, 0 tidak valid), menghasilkan persentase ketepatan model ontologi sistem 100%. Ini menunjukkan bahwa sistem bekerja dengan baik dan digunakan dengan baik. Pengujian perangkat lunak penting untuk memastikan kualitas sebelum digunakan oleh pengguna.

c. **Pengujian System Usability Scale (SUS)**

Pengujian dilanjutkan dengan berfokus pada pengalaman pengguna dan persepsi Partisipan terhadap *website Healthy Zoonosis*. Pada tahap ini, pengujian dilakukan menggunakan *System Usability Scale* (SUS). Partisipan diminta untuk mengisi kuesioner yang disediakan oleh peneliti, menggunakan skala *Likert* dari 1 hingga 5 [6]. Pengujian ini mengevaluasi kenyamanan pengguna untuk mengukur seberapa intuitif dan memuaskan pengalaman pengguna. Tabel 2 merupakan pertanyaan kuesioner *System Usability Scale* (SUS) pada *website Healthy Zoonosis* :

**Tabel 2.** Kuesioner *System Usability Scale* (SUS)

Kode	Pertanyaan
P1	Saya berencana mengunjungi <i>website Healthy Zoonosis</i> dengan sering.

P2	Saya merasa <i>website Healthy Zoonosis</i> tampaknya rumit untuk digunakan.
P3	Saya merasa <i>website Healthy Zoonosis</i> dianggap mudah dipahami ( <i>user friendly</i> ).
P4	<i>Website Healthy Zoonosis</i> , saya pikir saya membutuhkan bantuan dari orang lain atau teknisi.
P5	Saya rasa bahwa fitur yang disajikan sudah berjalan dengan baik.
P6	<i>Website Healthy Zoonosis</i> menurut saya menunjukkan banyak ketidakkonsistenan.
P7	Saya yakin banyak orang akan dengan cepat memahami cara menggunakan <i>website Healthy Zoonosis</i> .
P8	Saya merasakan kebingungan dalam menggunakan <i>website Healthy Zoonosis</i> .
P9	Melihat situs <i>website Healthy Zoonosis</i> , saya tidak menemukan kendala yang signifikan.
P10	Sebelum menggunakan situs <i>website Healthy Zoonosis</i> , saya pikir saya perlu banyak belajar.

Berdasarkan hasil perhitungan skor SUS dari 50 Partisipan di Provinsi Bali, *website Healthy Zoonosis* mendapatkan nilai tinggi, yaitu 87,3 yang dapat dilihat pada Tabel 3. Nilai ini menunjukkan bahwa pengguna memberikan penilaian positif terhadap kegunaan *website* tersebut. Dengan nilai di atas rata-rata global (68), *website* ini dapat diklasifikasikan sebagai "excellent" dalam kategori *Adjective Rating*, menunjukkan pengalaman pengguna yang memuaskan.

**Tabel 3.** Hasil Perhitungan Skor SUS

Partisipan	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Jumlah	Skor SUS
Partisipan 1	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 2	3	3	4	3	2	4	4	3	3	3	32	80
Partisipan 3	4	4	4	4	4	4	3	4	4	4	39	97,5
Partisipan 4	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 5	3	4	4	4	3	4	4	4	3	4	37	92,5
Partisipan 6	4	4	3	4	4	4	4	4	4	4	39	97,5
Partisipan 7	4	3	3	3	4	3	3	3	3	3	32	80
Partisipan 8	3	1	4	3	4	4	4	3	3	2	31	77,5
Partisipan 9	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 10	3	3	3	3	3	3	4	3	3	3	31	77,5
Partisipan 11	2	3	4	3	3	3	3	4	3	3	31	77,5
Partisipan 12	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 13	4	4	4	3	4	4	3	4	4	4	38	95
Partisipan 14	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 15	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 16	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 17	3	4	4	3	4	4	4	4	4	3	37	92,5
Partisipan 18	4	3	4	4	3	2	3	4	4	4	35	87,5
Partisipan 19	3	4	4	4	3	3	4	4	3	4	36	90
Partisipan 20	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 21	3	4	4	4	4	4	4	4	4	4	39	97,5
Partisipan 22	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 23	3	3	3	2	3	3	3	3	3	3	29	72,5
Partisipan 24	3	2	3	1	2	2	2	2	3	1	21	52,5
Partisipan 25	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 26	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 27	4	0	4	3	4	4	4	4	4	4	35	87,5
Partisipan 28	3	4	4	4	4	4	3	4	4	4	38	95
Partisipan 29	3	2	3	2	3	3	3	2	3	1	25	62,5
Partisipan 30	4	4	4	4	4	4	4	4	4	4	40	100
Partisipan 31	4	4	4	4	4	4	4	4	4	0	36	90
Partisipan 32	4	2	3	1	3	1	3	2	3	1	23	57,5

Partisipan 33	0	4	4	4	4	4	4	4	4	4	36	<b>90</b>
Partisipan 34	3	1	4	3	4	2	3	3	3	3	29	<b>72,5</b>
Partisipan 35	2	3	3	3	3	3	2	3	3	3	28	<b>70</b>
Partisipan 36	4	4	4	3	4	4	4	4	4	4	39	<b>97,5</b>
Partisipan 37	4	3	4	4	4	4	4	4	4	4	39	<b>97,5</b>
Partisipan 38	3	3	2	4	3	3	3	4	4	3	32	<b>80</b>
Partisipan 39	4	4	4	4	4	4	4	4	4	3	39	<b>97,5</b>
Partisipan 40	4	4	4	4	4	4	4	4	4	4	40	<b>100</b>
Partisipan 41	4	4	4	4	4	4	4	4	0	4	36	<b>90</b>
Partisipan 42	4	4	4	4	4	4	4	4	0	4	36	<b>90</b>
Partisipan 43	4	4	4	4	4	4	4	4	4	4	40	<b>100</b>
Partisipan 44	4	4	4	4	3	4	3	4	4	2	36	<b>90</b>
Partisipan 45	4	4	4	4	4	4	4	4	4	4	40	<b>100</b>
Partisipan 46	3	2	2	2	2	2	2	2	2	2	21	<b>52,5</b>
Partisipan 47	4	3	3	3	3	1	3	3	3	2	28	<b>70</b>
Partisipan 48	2	2	3	2	3	3	3	3	2	2	25	<b>62,5</b>
Partisipan 49	4	0	4	0	4	0	4	0	4	0	20	<b>50</b>
Partisipan 50	4	4	4	3	4	3	4	4	4	4	38	<b>95</b>
<b>Rata-rata Skor SUS</b>											<b>87,3</b>	

Dalam grade scale skala B, *website* ini masuk dalam kategori "very good", menunjukkan bahwa *website* tersebut mencapai standar tinggi dalam kegunaan. Secara keseluruhan, *website Healthy Zoonosis* dinilai sebagai "acceptable" dalam rentang penerimaan, artinya mayoritas pengguna akan menerima dan puas dengan pengalaman menggunakan *website* ini. Dengan demikian, evaluasi ini menunjukkan bahwa *website Healthy Zoonosis* sangat baik dalam memenuhi kebutuhan penggunanya.

#### 4. Kesimpulan

Perancangan sistem *Healthy Zoonosis* menggunakan metode *Design Science Research Methodology* (DSRM) dan *Methontology*, mengintegrasikan langkah-langkah identifikasi masalah, pengembangan model ontologi zoonosis, perancangan aplikasi, dan evaluasi kenyamanan pengguna. Pengujian model ontologi dan kenyamanan pengguna menunjukkan hasil yang sangat baik, dengan tingkat ketepatan model ontologi mencapai 100% valid menggunakan metode *Blackbox Testing* dan nilai *System Usability Scale* (SUS) sebesar 87,3. Saran pengembangan meliputi penambahan informasi tentang tindak lanjut pemerintah terkait vaksinasi hewan di Bali dan pengembangan versi mobile *website* untuk meningkatkan pengalaman pengguna.

#### Referensi

- [1] E. H. Halida, A. Hardjanti, dan F. Arifandi, "Kesadaran Masyarakat Terhadap Faktor Risiko Penyakit Zoonosis di Desa Koncang Pandeglang Banten dan Tinjauannya Menurut Pandangan Islam," *Junior Medical Journal*, vol. 2, no. 2, hal. 227-236, 2023.
- [2] V. Safitri, G. Setiaji, dan A. Lestariningsih, "Prioritasisasi Zoonosis di Indonesia Melalui Pendekatan *One Health*," dalam *Penyidikan Penyakit Hewan Rapat Teknis dan Pertemuan Ilmiah (RATEKPIL) dan Surveilans Kesehatan Hewan Tahun 2019*, hal. 524-530, 2019.
- [3] I. L. Koten dan C. R. A. Pramarta, "*Semantic Representation of Balinese Traditional Dance*," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 2301, hal. 5373, 2020.
- [4] Prabowo, *Metodologi Pengembangan Sistem Informasi*. Salatiga: LP2M IAIN Salatiga, 2020.
- [5] N. W. Rahadi dan C. Vikasari, "Pengujian *Software* Aplikasi Perawatan Barang Miliki Negara Menggunakan Metode *Blackbox Testing Equivalence Partitions*," *Jurnal Infotekmesin*, vol. 11, no. 01, hal. 57-61, 2020.
- [6] I. Salamah, "Evaluasi usability *website* polsri dengan menggunakan *system usability scale*," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 8, no. 3, hal. 176-183, 2019.

# Pemanfaatan UML dalam Perancangan Sistem Objek Wisata Kabupaten Bandung Barat Berbasis Website

Tazkia Damayanti<sup>a1</sup>, Apriade Voutama<sup>a2</sup>

<sup>a</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer,  
Universitas Singaperbangsa Karawang, Jawa Barat

<sup>1</sup>tazkiad27@gmail.com

<sup>2</sup>apriade.voutama@staff.unsika.ac.id

## Abstract

*Designing a website-based tourist attraction system plays an important role in selecting tourist destinations and providing comprehensive and effective access to information for visitors. There are two methods applied in this design process, namely Unified Modeling Language (UML) and Software Development Life Cycle (SDLC). This tourist attraction website system can overcome several problems, one of which is the lack of available tourist destination information for tourists who want to go on holiday. This website makes a big contribution to management in managing and processing data, so that tourists can easily access information on natural, culinary, cultural and city tourist attractions in the West Bandung Regency area. It is hoped that the results of this research can provide practical guidance for system developers in designing web applications that support the local tourism industry.*

**Keywords:** Perancangan Sistem, Website, UML, SDLC, Kabupaten Bandung Barat, Jawa Barat

## 1. Pendahuluan

Industri pariwisata semakin mengandalkan teknologi informasi untuk meningkatkan pengalaman wisatawan dan promosi destinasi. Namun, belum ada sistem informasi khusus yang menyediakan informasi terintegrasi tentang objek wisata di Kabupaten Bandung Barat. Penelitian ini bertujuan merancang sistem objek wisata berbasis website dengan UML untuk mengatasi kesenjangan tersebut. Tujuan penelitian adalah meningkatkan promosi pariwisata dan pengalaman wisatawan di Kabupaten Bandung Barat [1].

Kabupaten Bandung Barat, sebagai satu di antara destinasi pariwisata terkemuka di Jawa Barat, menawarkan keindahan alam, kekayaan budaya, dan kelezatan kuliner yang menakjubkan. Dari perbukitan yang hijau hingga objek-objek wisata bersejarah dan religius, Kabupaten Bandung Barat menarik minat wisatawan dari berbagai penjuru, baik dalam maupun luar negeri. Namun, untuk meningkatkan promosi potensi pariwisata yang dimiliki, diperlukan pengembangan sebuah sistem yang andal dan mudah diakses oleh para wisatawan. Sistem ini diharapkan mampu menyediakan informasi yang akurat dan komprehensif tentang destinasi wisata yang tersedia di Kabupaten Bandung Barat, sehingga dapat memandu wisatawan dalam merencanakan dan menjelajahi keindahan serta keunikannya dengan lebih baik. Dengan demikian, penyediaan sistem yang efektif diharapkan dapat memberikan kontribusi positif dalam mempromosikan potensi pariwisata daerah dan meningkatkan kunjungan wisata yang berkelanjutan [1].

Meskipun penelitian terdahulu telah menggarisbawahi manfaat penggunaan teknologi informasi dalam industri pariwisata, terdapat kesenjangan yang perlu diatasi. Saat ini, belum ada sistem informasi yang secara khusus dirancang untuk menghadirkan informasi lengkap dan terintegrasi tentang objek wisata di Kabupaten Bandung Barat. Dibutuhkan sebuah sistem yang mampu menyajikan informasi dengan akurat serta terperinci tentang destinasi wisata, serta memfasilitasi interaksi antara wisatawan dan penyedia layanan [2].

Maka dari itu, pokok tujuan dari penelitian ini adalah merancang sebuah sistem objek wisata berbasis website untuk Kabupaten Bandung Barat dengan menggunakan Unified Modeling Language (UML) sebagai alat pemodelan. Penelitian ini bertujuan untuk mengatasi kesenjangan yang ada dengan menyediakan platform yang mampu menyediakan informasi yang lebih komprehensif serta terintegrasi tentang destinasi wisata di Kabupaten Bandung Barat, sehingga meningkatkan promosi pariwisata dan pengalaman wisatawan secara keseluruhan. Pertanyaan penelitian yang akan dijawab dalam naskah ini adalah bagaimana merancang sistem objek wisata berbasis website dengan menggunakan UML untuk Kabupaten Bandung Barat dan bagaimana sistem ini dapat meningkatkan promosi pariwisata serta pengalaman wisatawan di daerah tersebut [2].

## 2. Tinjauan Pustaka

### 2.1. Perancangan

Perancangan merupakan proses perencanaan, pembuatan gambaran, dan penyusunan dari beragam subsistem yang terpisah menjadi satu sistem yang berfungsi dan komprehensif. Ketika merancang sistem dan subsistem yang akan digunakan dalam sebuah perusahaan, pengembang aplikasi diharuskan untuk menyusun rancangan sistem yang akan dibangun sebelumnya [3].

### 2.2. Sistem

Sistem adalah struktur kompleks yang terdiri dari prosedur dan komponen yang berinteraksi untuk mencapai tujuan, dan pemahaman yang komprehensif tentang sistem penting dalam pengembangan dan pengelolaan informasi untuk meningkatkan efisiensi dan kinerja secara keseluruhan [4].

### 2.3. Situs Web

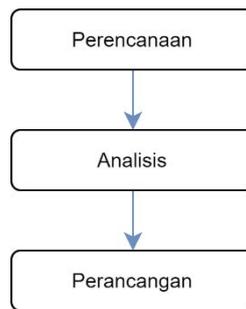
Situs web adalah sebuah tempat atau rangkaian halaman yang dapat diakses melalui internet. Halaman-halaman tersebut mungkin berisi berbagai informasi, konten multimedia, atau aplikasi yang dibuat untuk mencapai tujuan tertentu. Pengguna dapat mengakses situs web ini menggunakan sejumlah perangkat seperti komputer, laptop, tablet, atau telepon seluler melalui peramban web [5].

### 2.4. UML (*Unified Modeling Language*)

*Unified Modeling Language* (UML) merupakan bahasa visual yang dipakai untuk membuat model dan mengomunikasikan sistem dengan memanfaatkan skema dan skrip yang mendukung. UML bertujuan untuk memberikan bahasa pemodelan yang mandiri dari beragam bahasa pemrograman dan proses rekayasa yang berbeda. Disamping itu, penggunaan UML memastikan penerapan praktik terbaik yang diakui secara umum dalam pemodelan dan menyediakan model yang siap digunakan. Ada empat jenis diagram yang lazim diterapkan dalam praktik UML, yaitu *Use Case Diagram*, *Activity Diagram*, *Sequence Diagram*, dan *Class Diagram* [6].

## 3. Metode Penelitian

Dalam penelitian ini, dua metode utama yang diadopsi adalah *Unified Modeling Language* (UML) dan *Software Development Life Cycle* (SDLC). UML digunakan sebagai alat untuk merancang dan menggambarkan struktur serta perilaku sistem dengan jelas dan terstruktur. Pada tahap perencanaan, UML membantu dalam merumuskan gambaran konseptual dan kebutuhan sistem secara holistik. Selanjutnya, dalam tahap analisis, UML digunakan untuk mengidentifikasi dan memetakan kebutuhan fungsional dan non-fungsional, serta menggambarkan interaksi antara entitas sistem. Dalam tahap perancangan, UML menjadi landasan untuk membuat desain rinci sistem, termasuk spesifikasi arsitektur, relasi antar komponen, dan aliran data. Diagram UML seperti *use case diagram*, *class diagram*, *sequences diagram* dan *activity diagram* digunakan untuk menggambarkan berbagai aspek sistem dengan detail yang tepat. SDLC digunakan sebagai pendekatan sistematis untuk mengelola dan mengorganisir proses pengembangan perangkat lunak dari pemulaan hingga akhir.



**Gambar 1.** Struktur penelitian

### 3.1. Perencanaan

Tahap perencanaan dalam SDLC adalah fase awal di mana tim proyek merencanakan semua aktivitas yang diperlukan untuk pengembangan perangkat lunak. Ini mencakup identifikasi tujuan proyek, kebutuhan pengguna, jadwal, anggaran, dan sumber daya yang diperlukan. Perencanaan ini juga melibatkan penilaian risiko dan strategi mitigasi yang sesuai. Pada akhir tahap perencanaan, proyek harus memiliki rencana yang jelas dan terstruktur untuk memandu langkah-langkah selanjutnya.

### 3.2. Analisis

Tahap analisis dalam SDLC melibatkan pemahaman dan dokumentasi kebutuhan sistem. Tim proyek bekerja sama dengan pemangku kepentingan untuk mengidentifikasi dan mengevaluasi kebutuhan bisnis serta kebutuhan pengguna. Data yang terkumpul dari analisis digunakan untuk menghasilkan dokumen persyaratan sistem yang mencakup fungsionalitas yang diinginkan, batasan sistem, dan persyaratan non-fungsional lainnya. Analisis ini bertujuan untuk memastikan bahwa solusi perangkat lunak yang dikembangkan akan memenuhi kebutuhan dan tujuan bisnis yang telah ditetapkan.

### 3.3. Perancangan

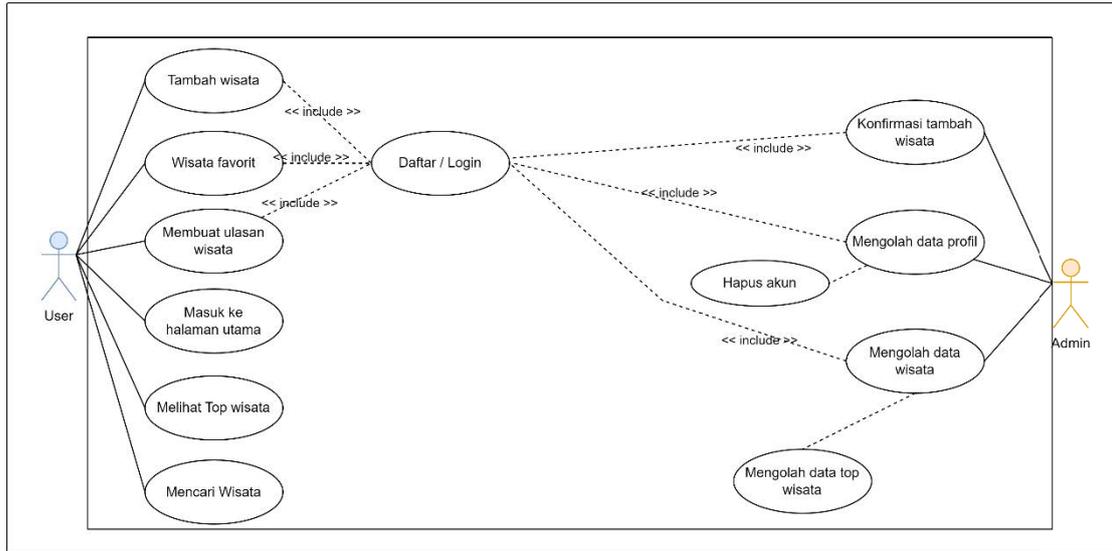
Tahap perancangan melibatkan pengembangan rencana rinci untuk implementasi sistem berdasarkan persyaratan yang telah ditetapkan. Tim proyek merancang arsitektur sistem, mengidentifikasi komponen-komponen utama, dan menentukan cara interaksi antara komponen-komponen tersebut. Selain itu, dalam tahap ini, rancangan antarmuka pengguna (UI) dan rancangan database juga dibuat. Tujuan perancangan adalah untuk menciptakan kerangka kerja yang jelas dan terstruktur untuk pengembangan sistem yang efektif dan efisien.

## 4. Hasil dan Pembahasan

Dari analisis yang sudah dilaksanakan dengan menggunakan berbagai metode seperti perencanaan, analisis, perancangan, dan implementasi dengan penerapan UML, diperoleh hasil yang memberikan gambaran sistem yang akan dirancang, memudahkan pengguna dalam memahami rancangan tersebut. Fokus pembuatan sistem web objek wisata Kabupaten Bandung Barat adalah untuk mempermudah pengalaman para wisatawan yang berkunjung ke daerah tersebut. Dengan menggunakan model UML, harapan dan tujuan yang telah dirancang dapat terwujud, yang tercermin dalam beberapa diagram yang dihasilkan seperti berikut ini:

#### 4.1. Analisis

##### 4.1.1. Use Case Diagram



**Gambar 2.** Use Case Diagram Sistem Wisata

Deskripsi Use Case diagram sistem objek wisata di atas memberikan pandangan awal saat memulai proses perancangan sistem. Detail penjelasan dari Gambar 2 terdapat dalam tabel di bawah ini, yang mencakup satu aktor dengan beragam aksi yang berbeda.

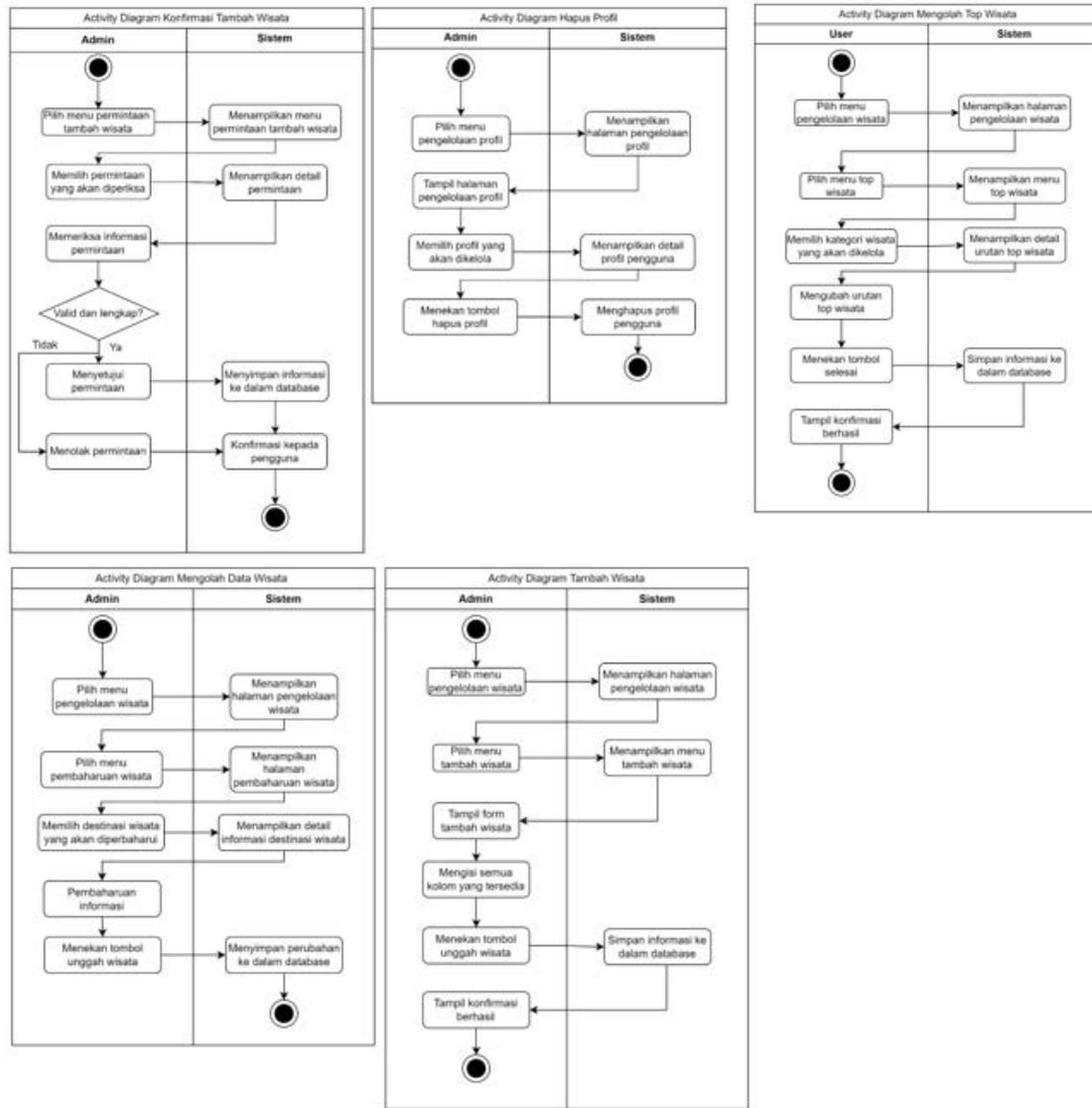
**Tabel 1.** Penjelasan Use Case Diagram

No	Use Case	Keterangan
1	Tambah Wisata	Tindakan pada pengguna dapat menambahkan objek wisata baru ke dalam sistem.
2	Wisata Favorit	Tindakan pengguna dapat menandai objek wisata tertentu sebagai favorit mereka, sehingga mereka dapat dengan mudah mengaksesnya kembali di masa mendatang.
3	Membuat Ulasan Wisata	Tindakan pengguna dapat memberikan ulasan dan penilaian terhadap objek wisata yang telah mereka kunjungi, sehingga pengguna lain dapat melihatnya untuk referensi.
4	Masuk ke Halaman Utama	Tindakan pengguna dapat masuk ke halaman utama tanpa daftar atau login tetapi ada beberapa fitur yang tidak dapat digunakan.
5	Melihat Top Wisata	Tindakan pengguna dapat melihat daftar objek wisata yang paling populer saat ini.
6	Mencari Wisata	Tindakan pengguna dapat mencari objek wisata berdasarkan kriteria tertentu seperti nama, lokasi, atau kategori wisata tertentu.
7	Mengolah Data Profil	Tindakan admin dapat mengelola profil pengguna dalam sistem.
8	Mengolah Data Wisata	Tindakan admin dapat mengelola informasi lengkap mengenai destinasi wisata dalam sistem.
9	Mengolah Data Top Wisata	Tindakan admin dapat mengelola daftar destinasi wisata dalam sistem.
10	Konfirmasi Tambah Wisata	Tindakan admin dapat meninjau serta mengonfirmasi permintaan penambahan destinasi wisata baru.

## 4.2. Perancangan

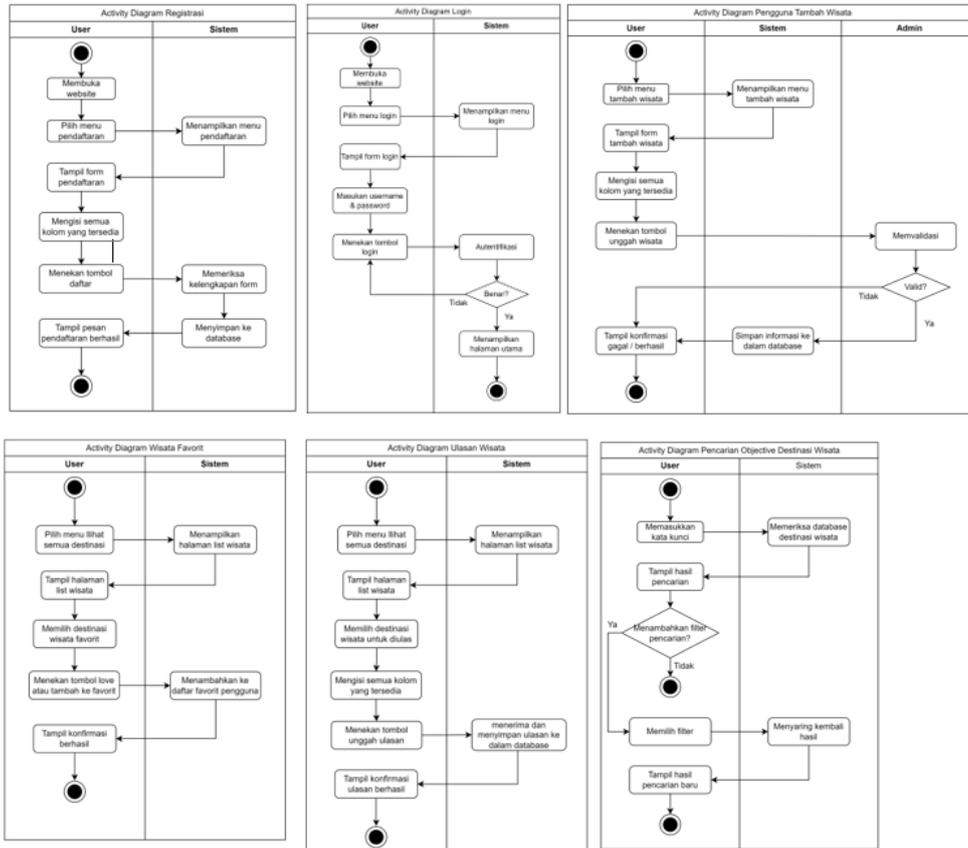
### 4.2.1. Activity Diagram

Activity diagram mengilustrasikan kegiatan pokok dan hubungan antara kegiatan dalam suatu proses, atau menggambarkan urutan kerja aktivitas suatu sistem atau operasi bisnis. Perlu ditegaskan bahwa activity diagram mencerminkan kegiatan sistem, bukan tindakan yang dikerjakan oleh aktor [8].



Gambar 3. Activity Diagram Admin

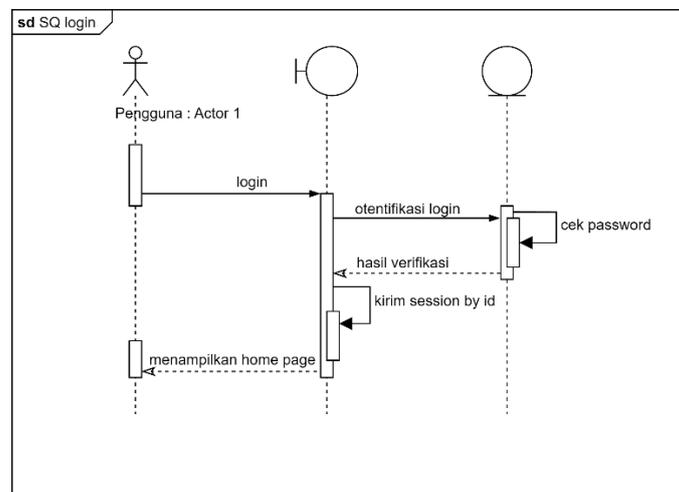
Activity Diagram Admin di atas menunjukkan cara kerja pada admin dalam system objek wisata sebagaimana terlihat pada Gambar 3.



**Gambar 4.** Activity Diagram Pengguna

Activity Diagram Pengguna di atas menunjukkan cara kerja pada pengguna dalam system objek wisata sesuai dengan yang ditunjukkan dalam Gambar 4.

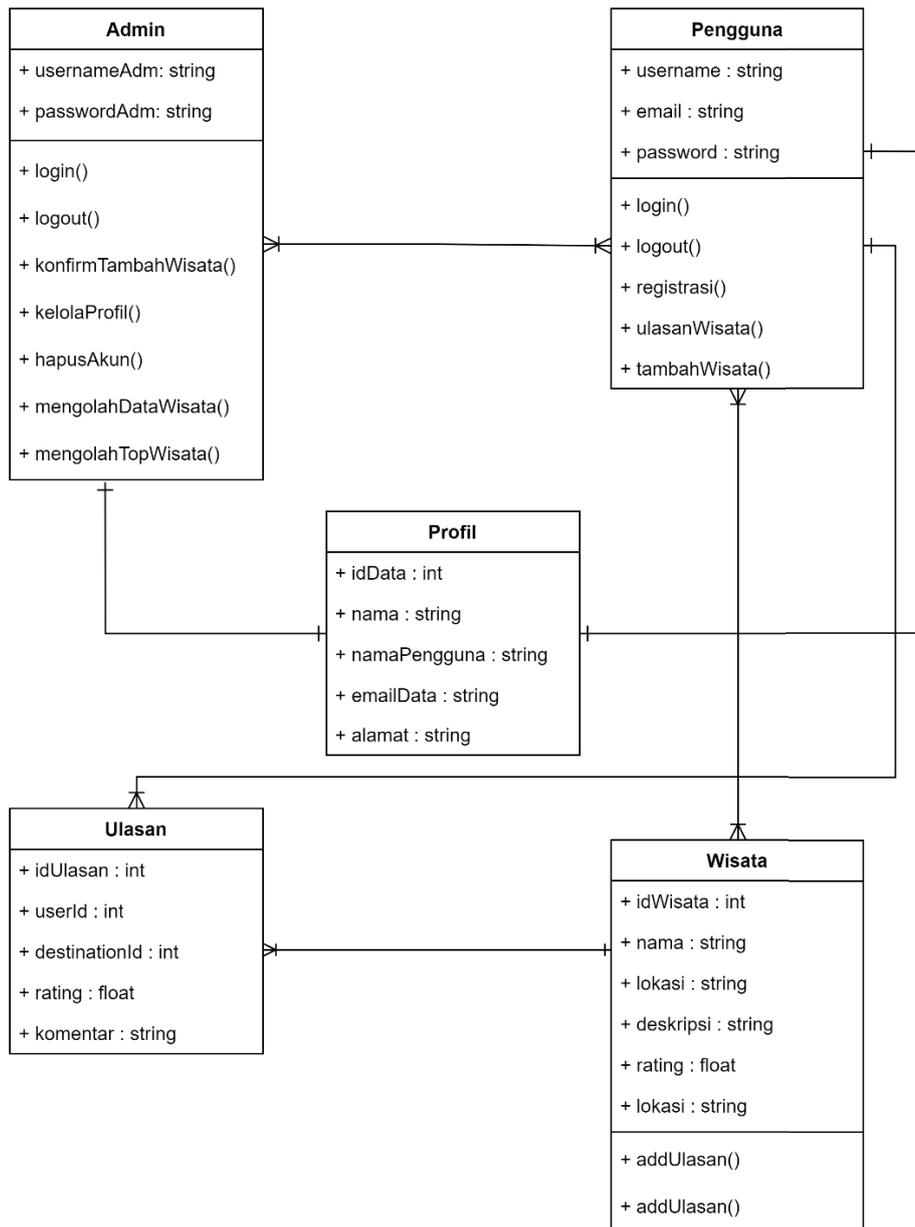
#### 4.2.2 Sequence Diagram



**Gambar 5.** Sequence Diagram Login

Sequences Diagram di atas merupakan satu dari dua jenis diagram interaksi. Dengan mengilustrasikan objek-objek yang terlibat seperti yang terlihat pada Gambar 5.

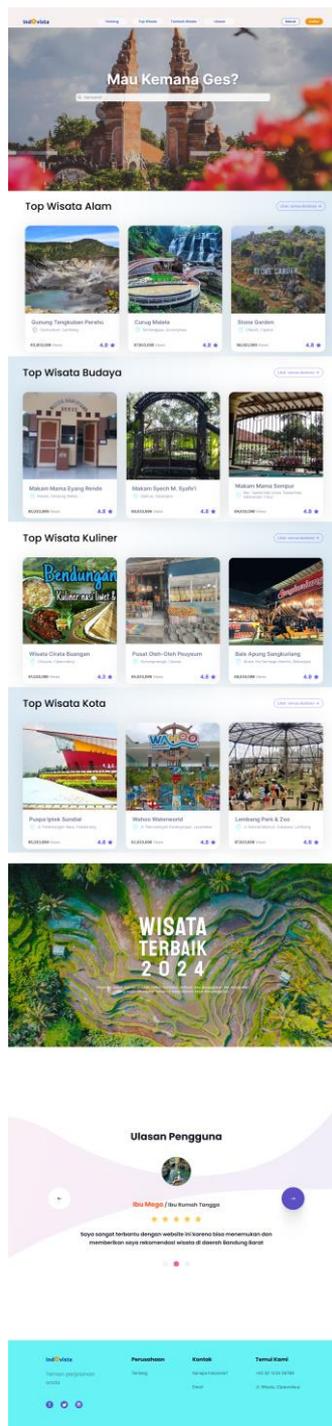
### 4.2.3. Class Diagram



**Gambar 6.** Class Diagram Sistem Objek Wisata KBB

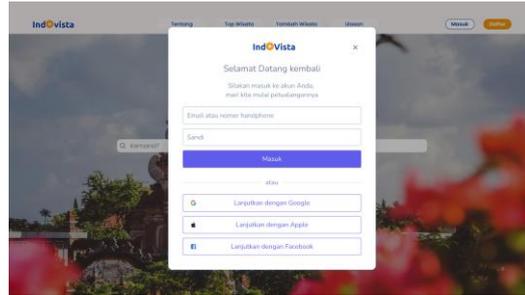
Class Diagram bersifat statis dan memperlihatkan relasi antar Class. Class diagram memperhatikan hubungan antara satu kelas dengan kelas lainnya dan penjelasan rinci tentang setiap kelas dalam model desain suatu sistem. Class Diagram Sistem Objek Wisata Kabupaten Bandung Barat di atas terdapat beberapa atribut dan dari setiap atribut yang ada berisikan relasi yang berbeda-beda seperti yang terdapat pada Gambar 6.

### 4.3. Implementasi



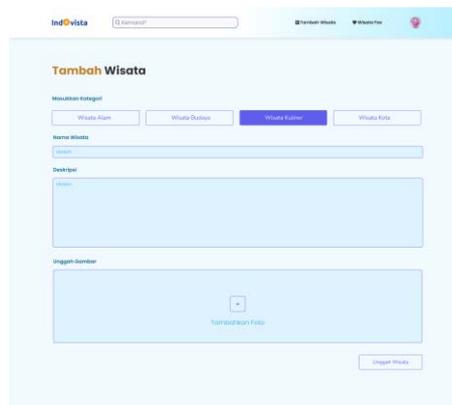
**Gambar 7.** Tampilan Utama Pengguna

Berikutnya merupakan tampilan Utama pada Pengguna merupakan tampilan awal ketika pengguna masuk pada website Indovista. Dalam tampilan utama pengguna dapat melihat beberapa informasi seperti top wisata alam, budaya, kuliner, serta kota yang ada di Bandung Barat, lalu ada beberapa ulasan dari pengguna lain mengenai website ini dan juga terdapat beberapa informasi mengenai website ini seperti pada Gambar 7.



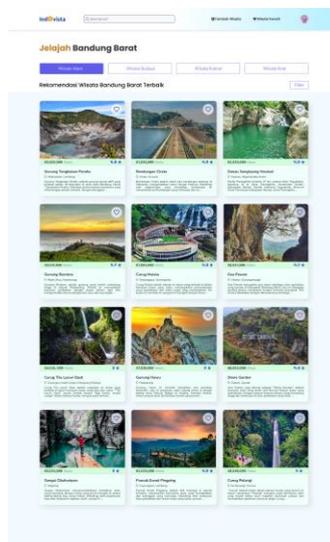
**Gambar 8.** Tampilan Login Pengguna

Tampilan selanjutnya merupakan tampilan Login Pengguna. Pada tampilan tersebut merupakan Langkah awal yang dilakukan oleh pengguna ketika ingin melakukan beberapa fitur yang tersedia dalam website, ketika pengguna sudah memiliki akun pengguna hanya perlu menginputkan nama pengguna serta kata sandi yang sesuai dengan yang didaftarkan sebelumnya seperti pada Gambar 8.



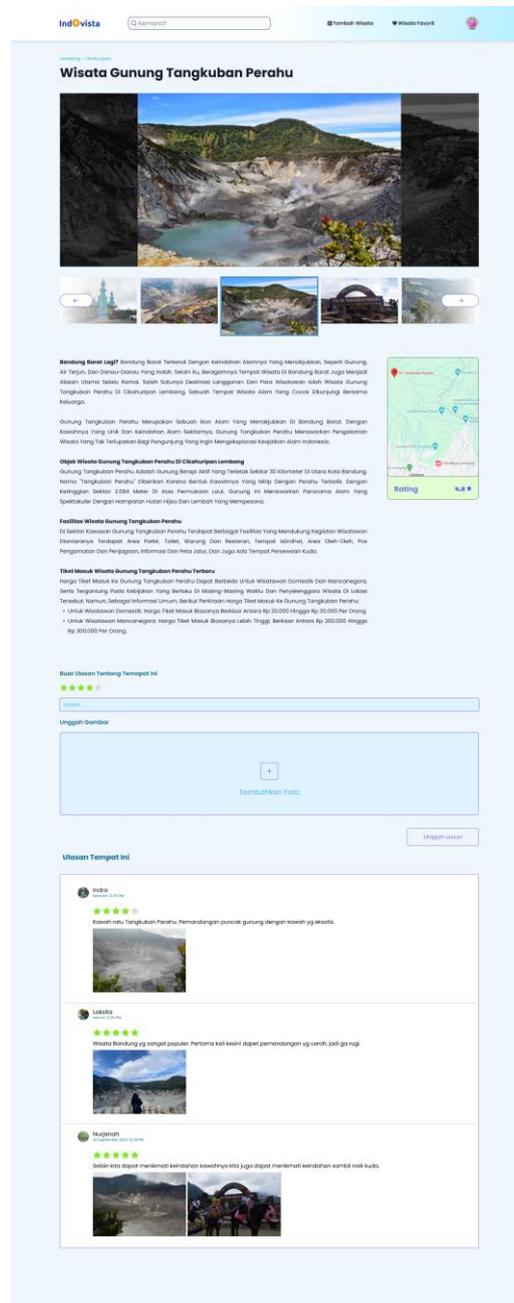
**Gambar 9.** Tampilan Tambah Wisata

Tampilan berikutnya yaitu Tambah Wisata. Tampilan ini merupakan salah satu fitur pada website yaitu memungkinkan pengguna dapat menambahkan informasi tentang destinasi atau objek wisata baru yang belum tersedia pada website dengan mengisi nama wisata, deskripsi, memilih kategori mana yang termasuk, lalu mengunggah gambar seperti pada Gambar 9.



**Gambar 10.** Tampilan Daftar Destinasi Wisata

Selanjutnya Tampilan Daftar Rekomendasi Destinasi Wisata Alam, untuk kita ke bagian ini kita dapat mengklik buttom Lihat semua destinasi yang ada di Tampilan Utama di sebelah kanan dari bagian top wisata yang ditampilkan. Tampilan ini merupakan bagian yang menyajikan informasi mengenai destinasi wisata alam yang direkomendasikan kepada pengguna seperti pada Gambar 10.



Gambar 11. Tampilan Deskripsi Wisata

Selanjutnya Tampilan Deskripsi Wisata. Tampilan ini bertujuan untuk memberikan informasi rinci tentang sebuah destinasi wisata tertentu. Didalamnya membahas mengenai deskripsi, fasilitas, tiket masuk, peta lokasi wisata, serta kita juga dapat melihat beberapa ulasan dari pengguna lain yang telah berkunjung. Kita juga dapat menambahkan ulasan dari destinasi yang telah dikunjungi dengan mengisi rating, ulasan, serta gambar di lokasi seperti pada Gambar 11.

## 5. Kesimpulan

Berlandaskan analisis dan pembahasan, dapat disimpulkan bahwa penulis telah merancang sebuah solusi inovatif berupa sistem Objek Wisata Kabupaten Bandung Barat Berbasis Website dengan *Unified Modeling Language* (UML), dilakukan penelitian untuk mengembangkan sistem informasi yang dapat mengelola objek wisata di Kabupaten Bandung Barat. Metode perancangan UML digunakan sebagai kerangka kerja untuk menganalisis, merancang, dan menggambarkan struktur serta interaksi antar komponen dalam sistem. Hasil penelitian ini memberikan solusi untuk permasalahan pengelolaan objek wisata di daerah tersebut, dengan menyajikan sistem informasi yang terintegrasi. Sistem ini diharapkan akan meningkatkan kinerja pengelolaan dengan lebih efisien dan efektif serta memperluas jangkauan promosi objek wisata. Meskipun isi dari pembahasan ini memberikan kontribusi penting, penelitian mendatang dapat fokus pada implementasi, evaluasi sistem, dan pengembangan fitur tambahan untuk meningkatkan fungsionalitas dan kegunaan sistem yang telah dirancang. Kesimpulan ini memberikan arah bagi penelitian lanjutan dalam pengembangan sistem informasi wisata yang dapat memperkuat industri pariwisata di Kabupaten Bandung Barat.

## References

- [1] J. Santoso, A. Nugroho, and M. Samsudin, "SISTEM INFORMASI GEORAFIS PEMETAAN DILENGKAPI RUTE MENUJU OBYEK WISATA KABUPATEN BANDUNG BARAT BERBASIS WEBGIS," *J. Teknol. dan Komput.*, vol. 9, no. 2, pp. 253–260, 2023.
- [2] Nopita, Pramiyati T, and Pradnyana W, "Perancangan Sistem Informasi Pariwisata Berbasis Web Di Kabupaten Sukabumi," *J. Ilmu Sos. dan Pendidik.*, vol. 5, no. 3, pp. 558–568, 2021.
- [3] I. Rahayuda and N. Santiari, "Perancangan Sistem Informasi Pengelolaan Data Penjualan Pada UD Memeri," *J. Elektron. Ilmu Komput. Udayana*, vol. 11, no. 2, pp. 445–456, 2022.
- [4] A. F. Sallaby and I. Kanedi, "Perancangan Sistem Informasi Jadwal Dokter Menggunakan Framework Codeigniter," *J. Media Infotama*, vol. 16, no. 1, pp. 48–53, 2020.
- [5] F. Nur Sa'adah and A. Voutama, "PERANCANGAN APLIKASI PENJUALAN FASHION DAN AKSESORIS BERBASIS WEB PADA TOKO FITRIN," 2023.
- [6] Elis and A. Voutama, "I N F O R M A T I K A PEMANFAATAN UML (UNIFIED MODELING LANGUAGE) DALAM PERENCANAAN SISTEM PENYEWAAN BAJU ADAT BERBASIS WEBSITE," *J. Inform. Manaj. dan Komput.*, vol. 14, no. 2, pp. 26–35, 2022.
- [7] A. Voutama and E. Novalia, "Perancangan Aplikasi M-Magazine Berbasis Android Sebagai Sarana Mading Sekolah Menengah Atas," *J. Tekno Kompak*, vol. 15, no. 1, pp. 104–115, 2021.
- [8] A. T. Hidayati, A. E. Widyantoro, and H. J. Ramadhani, "Perancangan Sistem Informasi Wirausaha Mahasiswa (Siwirma) Berbasis Web dengan Unified Modelling Language (UML)," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, pp. 86–107, 2023, doi: 10.55606/juprit.v2i4.2906.

*This page is intentionally left blank.*

# Single Sign on (SSO) Model Using Security Assertion Markup Language (SAML) and OAuth for authentication application of Universitas Pembangunan Nasional Yogyakarta (UPNVYK)

Ahmad Taufiq Akbar<sup>a1</sup>, Hari Prapcoyo<sup>a2</sup>, Rifki Indra Perwira<sup>a3</sup>

<sup>a</sup>Informatics Department, UPN "Veteran" Yogyakarta  
Jl. Babarsari No 2, Tambakbayan 55281 Yogyakarta, Indonesia

[1ahmadtaufiq.akbar@upnyk.ac.id](mailto:1ahmadtaufiq.akbar@upnyk.ac.id)

[2hari.prapcoyo@upnyk.ac.id](mailto:2hari.prapcoyo@upnyk.ac.id)

[3rifki@upnyk.ac.id](mailto:3rifki@upnyk.ac.id) (Corresponding author)

## Abstract

*Big companies have different systems both in terms of applications as well as the operating system, which requires each user to login to each different applications over and over again. With the SSO, users only need to remember one username and one password, but apply automatically universal across enterprise applications, so in this way it can be easier by using Security Assertion Markup language(SAML) for applications to be integrated without having to create a separate user validation. This SAML technology is an Extensible Markup Language(XML)-based framework and can guarantee the encryption of all or part of the data and then convey it to the end user. Moreover, it allows easy and secure data exchange between systems. The data exchange will be protected by authorization and authentication through tokens containing statements to pass data between users authorized by SAML. SAML can be supported by OAUTH as bearer protocol to provide extensive security when user access services along side on the Single Sign On(SSO) network.*

(Justify, Arial 10)

**Keywords:** *Single Sign-On (SSO), Security Assertion Markup Language (SAML), authentication, OAuth (Open Athorization)*

(Minimum 5 keywords related to the content and separated by comma, italic)

## 1. Introduction

Large companies have numerous applications and services that handling day to day operational level for running the business. User used the application based on their role in organization to handle their jobs. In addition, they should remember username and password along those numerous applications. Furthermore, security and authentication are becoming major issues that should tackle in order to work effectively and efficiently for accessing in various systems. Meanwhile, those complications of application, user access, password, security, and authentication are needed solution for improvement data exchanged over the internet. Essentially, organization have to start within central authentication for handling their applications, web based, easy to configure, highly support security system and easily from user perspective [1]. So, the Secure of Single Sign On (SSO) method is needed for those case.

The SSO can be practically direct many users to log in to multi-application using one credential but it must be supported by security assertion protocol. One of technology that can be used for is Security Assertion Markup Language (SAML) that offers a secure system, SAML, Extensible Markup Language (XML), Simple Object Access Protocol (SOAP) and Hypertext Transfer Protocol (HTTP) protocol based for exchanging user security information among organization and a service provider. In general, the SAML standard provides framework guidelines and language rules for the data communication more elastic and permit customization data to be delivered to the service provider and provide SSO on enterprise systems. The other method is OAuth, JavaScript Object Notation (JSON) and HTTP protocol based and offer security for Application Programming Interface (API) authorization among the user applications [2].

## Single Sign on (SSO) Model Using Security Assertion Markup Language (SAML) and OAuth for authentication application of Universitas Pembangunan Nasional Yogyakarta (UPNVIYK)

The SSO concept has a fairly good level of security compared to other authentication concepts. The SSO model has been tested in telematika (computer centre) at University of Pembangunan Nasional Veteran Yogyakarta (UPNVIYK) by penetration testing to determine the level of vulnerability. The results of this study were that from the seven stages of the penetration test conducted, 12 vulnerabilities were identified, consisting of 3 moderate vulnerabilities, 6 low vulnerabilities and 3 information vulnerabilities. Six cyber-attacks were carried out to exploit the vulnerability with 3 successful attacks and 3 failed attacks. This means that the single sign on concept has good accessibility transparency and can be audited at any time. SAML can also only run on browsers that already use the https protocol so that the security level is more guaranteed [3].

A number of previous studies have shown that SSO can increase ease of access with just one sign on or log in in using various integrated applications in a large company [4], and [5]. However, SSO still has a risk, namely that it can be breached by the man in the middle if there is no encryption framework for each session when accessing the application after the single sign on [6]. So for this reason SAML and OAuth are needed to provide security capabilities for each session of all access in the realm of single sign on [7]. SAML has generally been able to provide cross-domain authentication security with high mechanism efficiency [8]. But SAML can be integrated with OAuth to provide layered security, where SAML encrypted can be encrypted by OAuth running on Transport Layer Security (TLS) then it can achieve mutual authentication without the help of trusted third parties, [9].

The pervasiveness automation process in many organizations including higher education systems in UPNVIYK also influence in maintenance and operational level to secure each stage of any academic procedures. By this condition, user credential is very important that must be managed as one set credential to access in multi-application across network securely so that can ease to process of many online academic processes. The implementation of SSO will make it easier for users to log in to the website without having trouble remembering their credentials such as username, NIM, and password. So that only by entering an email address that has been registered in the system, users will easily log in quickly to several information system domains within the UPNVIYK campus simultaneously without having to login one by one. Only by one email makes it easier to login on multiple domains. Its authentication method uses the OAuth 2.0 and SAML protocol which is integrated with the account using the official institutional ID (@upnyk) from the Gmail mail server.

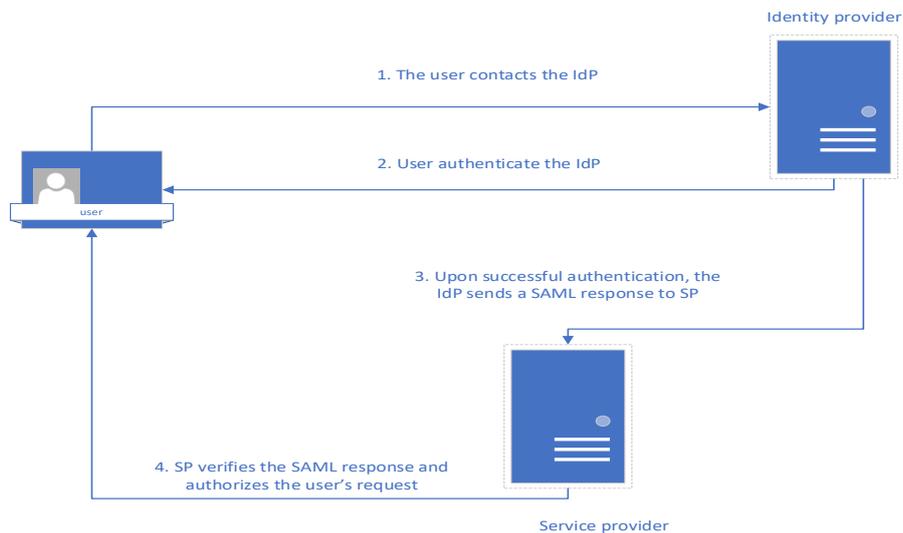
There are several techniques is developed for data exchange and security authentication. In general, it can be seen from password authentication, Public Key Infrastructure (PKI), Secure Socket Layer (SSL), XML based services. PKI based service is highly complex structure, code and costly in development and maintenance. SSL extremely needs more resource such as need partial encryption in data. While XML based services provide more robust from penetration and free in data exchange. Compare the current existing system in security authentication and single sign on methods, SAML is based on XML services that more reliable, high flexibility, and good authentication with security systems [10].

The implementation of SSO will make it easier for users to log in to the website without having trouble remembering their credentials such as username, NIM, and password. So that only by entering an email address that has been registered in the system, users will easily log in quickly to several information system domains within the UPNVIYK campus simultaneously without having to login one by one. Although one email makes it easier to login on multiple domains. The authentication method used uses the OAuth 2.0 and SAML protocol which is integrated with the account using the official institutional ID (@upnyk) from the gmail mail server

## 2. Research Methods

### 2.1. The SAML Web SSO process

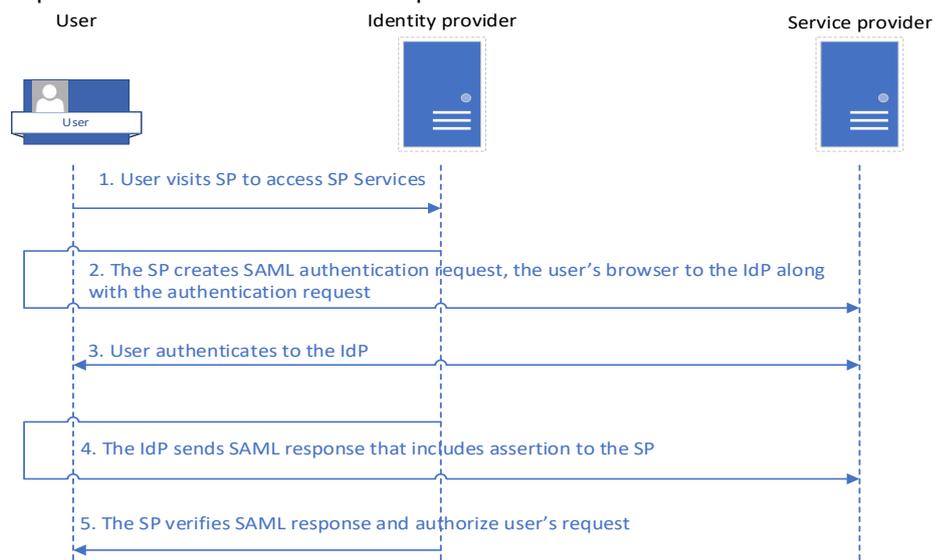
SAML web SSO process initiated by the Service Provider (SP) or the Identity Provider (IdP) in vice versa. A web SSO flow began with an identity provider called IdP-initiated, while a low that initiated at the service provider is titled SP-initiated. IdP initiated in SSO within several steps that illustrate the data exchange between the user and IdP provider. The first, the user acquaintances the IdP. Then, the user authenticates to the IdP. Following that, upon positive authentication, the IdP refers a SAML reply to service provider. Finally, SP proves the SAML response and authorizes the user's request. Definitely, the whole process of IdP initiated SSO can be seen in the Figure 1 below [11]



**Fig. 1.** The Flow of SSO that is initiated by IdP[12]

The user initiates the IdP with IdP-initiated SSO. The IdP guides the user to authenticate and displays (in the form menu) links to one or more of her SPs. Once the user is authenticated, the IdP generates her SAML response containing the authentication assertion and directs the user to the SP requested by the user. When a user tries to connect to an SP, it comes with the promise of authentication. The SP then analyzes the assertions and grants or denies access based on those assertions[11].

SP-initiated SSO as seen in Fig. 2, it begins with user send access to SP services. The SP will create a SAML authentication request and then redirects to the authentication request. Following that user authentication to the IdP and send response to that includes assertion to verify SAML response also authorizes user's request.



**Fig. 2.** The Flow of SSO that is initiated by SP[12]

First, the user goes to the SP, which directs the user to the IdP for authentication. SP-initiated SSO is a common configuration for SSO, so we'll cover it in detail. The IdP generates a SAML response in step 4. It contains authentication statements related to the user's security context. These confirmations are digitally marked and sent to the user's browser. Then forward the signed statement from the user's browser to the SP. The SP verifies the SP's digital signature and processes the contents of the statement. Next (if all requirements are met), create a "local" version of the security user context. After fulfillment, the user is registered with the SP and can access services within the SP[11].

## 2.2. OAuth Technology

To facilitate the implementation of authentication mechanisms, it can use services provided by application service providers (ASP). Google is one of the ASPs that dominate most users for Search engines and email services. One of the services provided by Google is Google Identity, that provides the ability to perform SSO using the OAuth 2.0 protocol. By using this service, users will soon be able to access services provided by Google APIs such as Google Calendar, Gmail, and more. However, users can not only use services provided by Google, but also use the Google Identity Service as an authentication and authorization server, so that can possibly do the SSO integration via Google [13](Senapatha, 2021). To keep the credentials safe, an access token is generated by the OAuth provider and passed to the client, and the access token has an expiration date. After the expiration date, user will need to request a new access token from his authentication provider or OAuth provider. The update token is a credential used to obtain an expired access token [14].

The OAuth 2.0 protocol is a widely used protocol for SSO due to its ease of implementation and ease of implementation. Lots of support from ASP for this protocol [15]. OAuth is an open standard for authorization that makes this possibility to the third-party applications for accessing resources on resource servers without sharing user credentials. Other available Protocol options are SAML and OpenID as open standard and decentralized authentication protocol, SAML is an XML-based protocol and Open ID is a modern protocol that runs on top of the OAuth 2.0 protocol. Previous research conducted as a result of the analysis shows that SAML has limited interoperability with mobile devices. It is considered immature for use as a standard authentication protocol and is therefore low among enterprises [16]. Therefore, any SSO implementation based on the OAuth 2.0 protocol can be used as the SSO authorization protocol in many systems. The OAuth is authorization method can provide authorization application on the web. Social network Googles plus, Facebook, Twitter use their authorization based on OAuth protocol in SSO and social sharing. This technology also uses an open standard platform to give a process for third-party application to get user's resources on servers without sharing login credentials. SSO is an identification technique that is allowing for website to use and confirm the users [15].

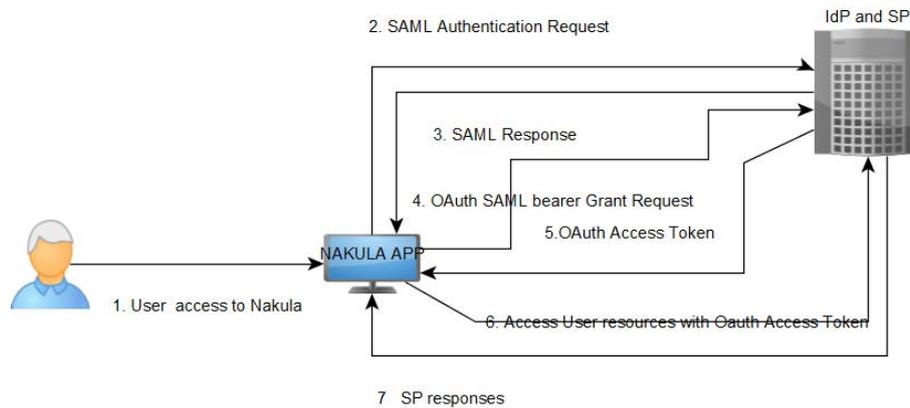
## 2.3. SSO using SAML and OAuth

The data collection method was carried out through observations on systems that did not use the single sign on method on the UPNYK campus and collected data from several Information and Communication (ICT) units, Computer Center or Directorate of Data Resources and Information (DSDI) from various universities, both public and private. Data from information sources that are very useful and can help provide input to researchers to use this single sign on method. In accordance with what was obtained during this observation, it can be indicated that the mechanism for students and users in logging in to be able to access campus information systems consisting of various domains such as Computer Base Information System, University web, Department web, and learning management system like SPADA adheres to conventional methods, namely students still it is necessary to remember the password and NIM (Student Identification Number) in order to be able to access the information system which students often neglect with their credentials.

SSO technology was developed to make it easier for users to access various applications, both mobile and web, in using various existing services. For example, a mail account from Google, where a user has logged in using an email account from Google can access all applications on Google without the need to log in again. SSO also allows authentication information and identify the subject strictly to avoid double login on a trusted system or system group. The SSO system can also centralize the management of the relevant system parameters at the same time and improve overall usability. Service users may prefer SSO systems over conventional sign-on systems. In implementing SSO, OAuth and SAML2 are used. As illustrated in Fig.3, the Client can obtain a SAML statement from the IdP and request the Authorization Server (IdP) to grant access to the Resource Server. The Authorization Server can then verify the user's identity and return an OAuth token in the HTTP header to access the protected resource. As an OAuth 2.0 Authorization Server, the Identity Server can

accept SAML2 Assertions from OAuth 2.0 clients as a means of authenticating and authorizing resource owners. Additionally, it can exchange it for an OAuth 2.0 access token to access protected resources on behalf of the resource owner.

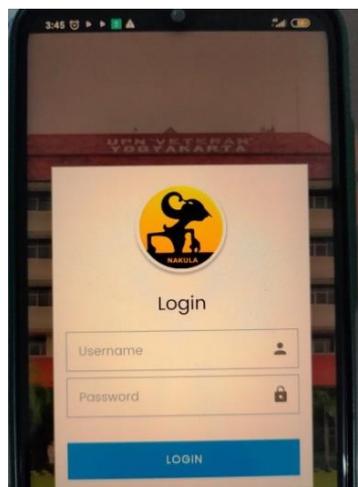
The flow of SSO usage with SAML2 and OAuth 2 Protocol can be seen in Fig. 3.



**Fig. 3.** The Flow of SSO using SAML and OAuth

### 3. Result and Discussion

The SSO application implemented on the Nakula App (Management Information System that have been developed by UPNVYK) can be accessed on android and web platforms. In Fig. 4 shows the application of SSO on the Android Front end by using the lecturer's NIDN credentials and password.



**Fig. 4.** The User Concern of Sign In using one Credential (gmail account)

While in web-based Nakula as seen in Fig. 5, users, especially lecturers, can choose to sign in with 1 of 2 types of credentials, namely NIDN and password or the registered institutional email username and password.

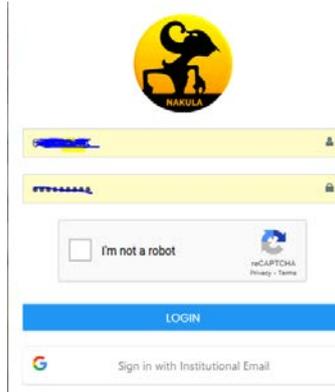


Fig. 5. Web Based Nakula Interface

If user use the first type of credential, the user will be immediately sent to the dashboard page which will bring up navigation to other sub domains without signing in again. Whereas if the user uses the second type of credential, then before going to the dashboard, the user will be asked to input the institutional email account and password that has been registered as a means of single sign on (Fig. 6).

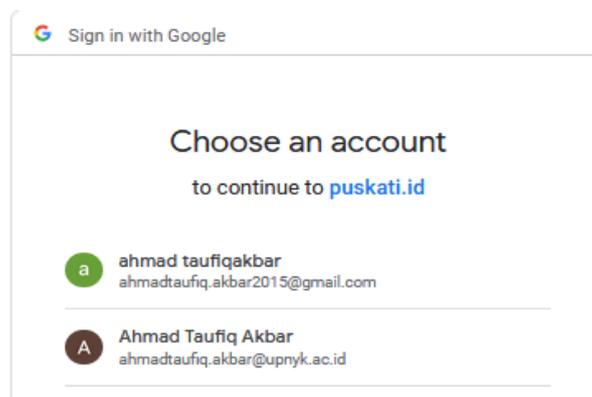


Fig. 6. The Web Based Nakula ask for gmail account as credential

Furthermore, after login process using email account credentials will also be delivered to the application dashboard as seen in Fig. 7. If user have reached the dashboard, they will no longer sign in or input credentials when accessing other sub domains (applications).

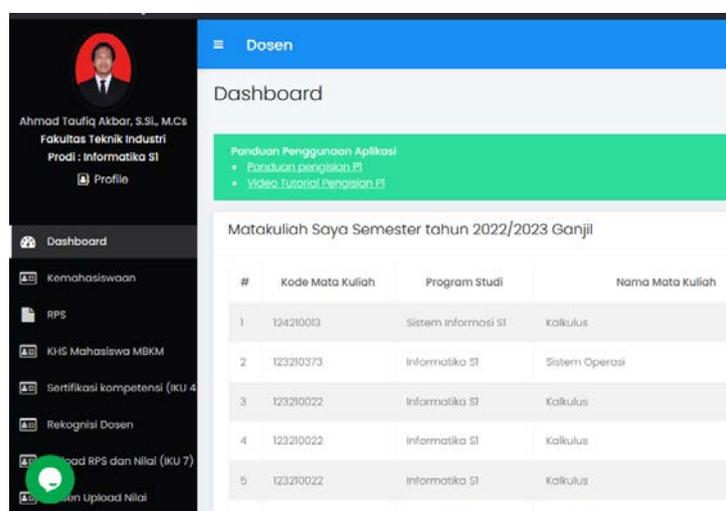


Fig. 7. The Dashboard Application after sign in Web Based Nakula using gmail account as credential

As long as the user is in the sub-domain application that has been accessed, the user can log out after returning to the dashboard as a one-stop media for single sign on. Immediately logged out the

user session has been deleted. The processing of every data transaction from the application to the server (including login and revoke token) is carried out on the HTTPS protocol as the transmission layer and data encryption. This ensures that the user data channel is secure and can use the OAuth SAML bearer token to access data from the desired SP. This bearer method protects sending tokens via Uniform Resource Locator (URL) from being accessed by other users via browser history.

#### 4. Conclusion

Research has shown that the SSO method with OAuth and SAML bearer provides easy and secure access with 1 type of credential through the use of google accounts and the client server REST architecture. The SSO application on Nakula which is based on android and web client server makes it easy for users to access sub domains or other applications without having to re-sign in every application that is accessed. The implementation of the refresh access token of the OAuth protocol will protect the validity of the access token used by the client that has been authorized by SAML, so this security is layered. Goggle identity service can speed up SSO while still focusing on securing data transmission while using SSO. By applying SSO method with OAuth and SAML the security level of authentication application have been improved, robust and more resilience from the network attacks.

#### References

(All of the references should be cited in the paper, it is recommended to use reference tools such as Mendeley or enote, Minimum 80% of the references are from journals and published within last 5 years, please cite one paper from this journal)

(Journal)

- [1] K. D. Lewis, "Web single sign-on authentication using SAML," *arXiv Prepr. arXiv0909.2368*, 2009.
- [2] C. H. Rupa, R. Patan, F. Al-Turjman, and L. Mostarda, "Enhancing the access privacy of IDAAS system using SAML protocol in fog computing," *IEEE Access*, vol. 8, pp. 168793–168801, 2020, doi: 10.1109/ACCESS.2020.3022957.
- [3] S. U. Sunaringtyas and D. S. Prayoga, "Implementasi Penetration Testing Execution Standard Untuk Uji Penetrasi Pada," *Edu Komputika J.*, vol. 8, no. 1, pp. 48–56, 2021.
- [4] Q. Aini, U. Rahardja, and R. S. Naufal, "Penerapan Single Sign On dengan Google pada Website berbasis Yii Framework Application Single Sign On with Google the Website Based on Yii Framework," *J. Ilm. SISFOTENIKA*, vol. 8, no. 1, pp. 57–68, 2018.
- [5] H. Ajie, M. Insan Rizky, and M. F. Duskarnaen, "Pengembangan Teknologi Single Sign On Pada Sistem Informasi Dosen dan Sistem Informasi Kurikulum di Universitas Negeri Jakarta," *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 3, no. 1, pp. 32–37, 2019, doi: 10.21009/pinter.3.1.6.
- [6] N. M. Karie, V. R. Kebande, R. A. Ikuesan, M. Sookhak, and H. S. Venter, "Hardening SAML by Integrating SSO and Multi-Factor Authentication (MFA) in the Cloud," *ACM Int. Conf. Proceeding Ser.*, 2020, doi: 10.1145/3386723.3387875.
- [7] S. Watini, P. Nursaputri, and M. Iqbal, "Comparison of CAS and Manage Oauth in Single Sign on (SSO) Client Applications," *IAIC Trans. Sustain. Digit. Innov.*, vol. 1, no. 2, pp. 152–159, 2020, doi: 10.34306/itsdi.v1i2.147.
- [8] A. J. Cui, W. Wang, H. F. Zhang, Y. H. Ma, C. Li, and X. M. Wang, "Cross-domain single sign-on authentication of information security in network environment," *Int. J. Inf. Commun. Technol.*, vol. 18, no. 1, pp. 89–104, 2021, doi: 10.1504/IJICT.2021.111924.
- [9] U. Joshi, S. Cha, and S. Esmaili-Sardari, "Towards adoption of authentication and authorization in identity management and single sign on," *Adv. Sci. Technol. Eng. Syst.*, vol. 3, no. 5, pp. 492–500, 2018, doi: 10.25046/aj030556.
- [10] M. Kang, C. S. Hong, H. J. Koo, and G. H. Lee, "An SAML based SSO architecture for secure data exchange between user and OSS," *APNOMS 2005 - 8th Asia-Pacific Netw. Oper. Manag. Symp. Towar. Manag. Ubiquitous Inf. Soc. Proc.*, no. January, pp. 608–617, 2005.

- [11] I. Rajapaksha, "No Title," *medium.com*, 2019. <https://is-rajapaksha.medium.com/single-sign-on-with-saml-e39dc3e72cf2> (accessed Oct. 16, 2022).
- [12] I. Rajapaksa, "https://medium.com/single-sign-on-with-saml-e39dc3e72cf2," 2019.
- [13] I. K. D. Senapartha, "Implementasi Single Sign-On Menggunakan Google Identity, REST dan OAuth 2.0 Berbasis Scrum," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 2, pp. 307–320, 2021, doi: 10.28932/jutisi.v7i2.3437.
- [14] I. Kusuma, A. Susanto, and I. U. W. Mulyono, "Implementasi Restful Web Services Dengan Otorisasi Oauth 2.0 Pada Sistem Pembayaran Parkir," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 1, pp. 391–404, 2019, doi: 10.24176/simet.v10i1.3026.
- [15] N. Hossain, M. A. Hossain, M. Z. Hossain, M. H. I. Sohag, and S. Rahman, "OAuth-SSO: A Framework to Secure the OAuth-Based SSO Service for Packaged Web Applications," *Proc. - 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng. Trust. 2018*, pp. 1575–1578, 2018, doi: 10.1109/TrustCom/BigDataSE.2018.00227.
- [16] N. Naik and P. Jenkins, "Securing digital identities in the cloud by selecting an apposite Federated Identity Management from SAML, OAuth and OpenID Connect," *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, pp. 163–174, 2017, doi: 10.1109/RCIS.2017.7956534.

# Perancangan Ontologi Semantik: Representasi Digital Film Bioskop Indonesia

Andien Rachma Fadillah<sup>a1</sup>, Komang Kartika Noviyanti<sup>a2</sup>, I Putu Agus Arya Wiguna<sup>a3</sup>, Cokorda Pramatha<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, Universitas Udayana  
JI Kampus Bukit Jimbaran Computer Science Building Jurusan Ilmu Komputer, Jimbaran, Bali

<sup>1</sup>[andienrf06@gmail.com](mailto:andienrf06@gmail.com)

<sup>2</sup>[kartikanovi28@gmail.com](mailto:kartikanovi28@gmail.com)

<sup>3</sup>[agusarya592@gmail.com](mailto:agusarya592@gmail.com)

<sup>4</sup>[cokorda@unud.ac.id](mailto:cokorda@unud.ac.id)

## Abstract

*The Indonesian film industry faces various challenges in data management, including data integration issues, information inconsistency, and difficulty in finding relevant data. This article highlights the need for ontologies as a solution to address these challenges. Ontologies, as formal representations of knowledge in a particular domain, can improve the interoperability and efficiency of data management. By using ontologies, data from different sources can be integrated more efficiently, resulting in more consistent and accessible information. This article will also discuss the concrete benefits of ontologies for end users, including increased data accessibility, reduced redundancy, and improved accuracy in information retrieval. The implementation of ontologies in practice will be outlined through case studies and strategic steps that can be adopted by the Indonesian film industry to optimally utilize this technology. Thus, ontologies are not only a technical solution, but also have a significant impact on the experience of end users, including film producers, directors, and researchers, through better access to data and more accurate information.*

**Keywords:** Indonesian Cinema, Semantic Web Technology, Ontology, Web Ontology Language (OWL), film

## 1. Introduction

Film dapat diinterpretasikan sebagai hasil kreativitas manusia yang mengandung unsur estetika tinggi [1] atau sebagai sarana komunikasi, di mana film berperan sebagai alat untuk menyampaikan pesan dari pembuat film kepada khalayak. Jika kita meninjau film dari segi komunikasi, dapat dikatakan bahwa film berada dalam ranah media massa [2]. Kemajuan pesat dalam industri film Indonesia menyebabkan kompleksitas dalam manajemen dan integrasi data terkait film bioskop. Pada tahun 2018, terjadi peningkatan yang signifikan dalam produksi film di Indonesia, mencapai 203 judul film [3]. Namun, tingkat antusiasme penonton terhadap film bioskop tidak sebanding, terlihat dari target Kemendikbud pada tahun 2019 yang menetapkan jumlah penonton film Indonesia sebanyak 58 juta, tetapi kenyataannya hanya mencapai 51 juta [4]. Keterbatasan fasilitas, termasuk infrastruktur industri perfilman yang belum memadai, serta keterbatasan akses informasi yang mudah dan luas untuk memahami film sebagai opsi hiburan, mungkin menjadi penyebab utama rendahnya antusiasme penonton film Indonesia [2] karena tidak dapat dipungkiri bahwa kebutuhan akan representasi data yang lebih terperinci dan terstruktur di dalam industri film semakin mendesak, terutama untuk mendukung pengambilan keputusan yang lebih cerdas dan penyajian informasi yang lebih akurat kepada pemirsa.

Dengan memanfaatkan teknologi informasi berbasis internet, pecinta film bioskop di Indonesia akan sangat terbantu dalam mencari informasi terkait film yang sedang tayang, sudah tayang, ataupun ketika ingin bernostalgia tentang berbagai macam film di masa lalu. Namun, dalam praktiknya, sering kali pecinta film bioskop menghadapi masalah dengan informasi yang tersebar di berbagai sumber, sehingga memerlukan usaha dan waktu yang signifikan untuk mengumpulkan dan mengagregasi data yang mereka butuhkan. Teknologi Web Semantik muncul sebagai solusi untuk mengatasi tantangan tersebut. Semantic Web dapat dianggap sebagai perluasan dari web saat ini, di mana informasi

diberikan makna yang lebih jelas, dengan tujuan mencapai kesamaan persepsi antara konsep-konsep yang ada. Konsep ini membuka pintu untuk kerjasama yang lebih efisien antara manusia dan komputer. Prinsip utama yang mendasari Semantic Web adalah menciptakan data web yang terdefinisi dan terhubung, memungkinkan otomatisasi, integrasi, dan pemanfaatan kembali data di antara berbagai aplikasi komputer.

Dalam era di mana volume data yang besar telah dipublikasikan di web, sebagian besar informasi hanya dapat dimengerti oleh manusia (*human-readable*) dan tidak dapat dipahami oleh aplikasi komputer (*not machine understandable*). Oleh karena itu, teknologi Semantic Web menjadi esensial untuk memberikan makna pada data di web agar dapat dimengerti baik oleh manusia maupun komputer. Dengan demikian, komputer dapat secara efisien menginterpretasi dan memanfaatkan data yang ada untuk kepentingan manusia. Semantic Web, sebagai perluasan dari web saat ini, bertujuan mencapai kesamaan persepsi antara konsep-konsep, memfasilitasi kerjasama yang lebih efisien antara manusia dan komputer. Salah satu tujuan utama Semantic Web adalah menciptakan data web yang terdefinisi dan terhubung, memungkinkan aplikasi komputer untuk memproses dan mengintegrasikan data secara otomatis, serta menggunakan kembali data dengan lebih efektif [5]. Ontologi menyediakan mekanisme untuk menggambarkan kumpulan pengetahuan dengan memperhatikan hubungan antar konsep yang ada dalam suatu domain. Dengan demikian, ontologi dapat berperan dalam menyajikan informasi secara semantik, menyusun, dan memetakan sumber daya informasi secara terstruktur dan sistematis. Keunggulan ini sangat berarti dalam hal interkoneksi data, memungkinkan integrasi data dengan lebih efisien dan efektif. Penggunaan konsep triple (*subjek, predikat, objek*) dalam ontologi untuk memberi struktur pada data memungkinkan mesin pencari untuk mengidentifikasi dokumen yang relevan. Selain itu, pengintegrasian ontologi dengan domain pengetahuan terkait dapat meningkatkan pemahaman di berbagai bidang, termasuk dalam konteks pariwisata dan budaya. Dengan demikian, ontologi menjadi alat yang efektif dalam menyajikan, mengorganisir, dan memanfaatkan pengetahuan, khususnya dalam era Web Semantik.

Maka dari itu, tujuan dari penelitian ini adalah untuk mengembangkan ontologi data yang dapat meningkatkan representasi dan interoperabilitas informasi film bioskop di Indonesia. Ontologi, sebagai model representasi pengetahuan yang terstruktur, diharapkan dapat menjadi kerangka kerja yang kuat untuk mengintegrasikan dan menyusun informasi film dari berbagai sumber. Penelitian sebelumnya terkait dengan penerapan ontologi yang menunjukkan bahwa pentingnya peranan ontologi dapat digunakan sebagai model representasi pengetahuan yang terstruktur untuk mengintegrasikan dan menyusun data informasi dari berbagai sumber. Hal ini dikarenakan pendekatan ontologi diimplementasikan menggunakan Web Ontology Language (OWL), yang memungkinkan mesin untuk memahami informasi dengan lebih baik. Dalam konteks domain Film Indonesia tentu akan memberikan potensi besar untuk meningkatkan pencarian informasi dan analisis data. Fokus ontologi ini melibatkan aspek-aspek krusial seperti judul film, genre, informasi tentang kru film, tahun rilis, rating usia dan lain sebagainya. Di samping itu juga integrasi data yang lebih baik dalam industri hiburan dapat memberikan manfaat signifikan dalam hal peningkatan efisiensi operasional dan pengalaman pengguna. Oleh karena itu, melalui pemahaman relasi semantik antar konsep-konsep tersebut, ontologi ini diharapkan dapat memberikan landasan yang kokoh bagi aplikasi pencarian, analisis, dan pemrosesan data yang lebih efisien. Pada penelitian ini juga dilakukan pengujian awal terkait model yang diajukan dengan cara mengajukan beberapa pertanyaan yang biasanya digunakan oleh pengguna ketika mengakses sistem informasi tentang pariwisata. Sehingga, diharapkan model ontology yang dihasilkan mampu memberikan dan menampilkan informasi film bioskop nasional di Indonesia secara sistematis. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan pencarian dan analisis informasi film, serta mendukung pengambilan keputusan yang lebih baik dan penyajian informasi yang lebih akurat. Dengan demikian, ontologi tidak hanya berperan sebagai solusi teknis, tetapi juga memiliki dampak signifikan terhadap pengalaman pengguna akhir, termasuk produser, sutradara, dan peneliti film, melalui akses data yang lebih baik dan informasi yang lebih akurat

## 2. Reseach Methods

Dalam penelitian ini, digunakan suatu pendekatan yang dikenal sebagai Methontology. Methontology adalah salah satu metode dalam pembangunan model ontologi yang memiliki keunggulan terutama dalam memberikan deskripsi rinci untuk setiap aktivitas yang perlu dilakukan. Lebih lanjut, keunggulan lainnya dari Methontology adalah kemampuannya untuk memungkinkan penggunaan ulang ontologi yang telah dibangun dalam pengembangan sistem yang lebih lanjut, sebagaimana dijelaskan oleh [6]. Dalam penelitian ini metode Methontology digunakan untuk pembangunan model ontologi secara keseluruhan, metode metodologi ini memberi kita seperangkat pedoman tentang bagaimana

melakukan aktivitas yang mengidentifikasi proses pengembangan ontologi adapun tahapan dari proses Methontology [1]. Setelah melakukan evaluasi ontologi yang lebih mendalam, peneliti dapat menyediakan analisis yang mendalam tentang kesesuaian ontologi dengan kebutuhan informasi film bioskop di Indonesia. Ini termasuk mengevaluasi kemampuan ontologi untuk mencakup berbagai aspek informasi seperti judul film, genre, penulis, sutradara, dan actor. Selain itu, analisis juga dapat mempertimbangkan faktor-faktor seperti keakuratan, konsistensi, dan kelengkapan informasi yang disediakan oleh ontology [7]. Dengan demikian, penjelasan evaluasi ontologi akan memberikan gambaran yang lebih holistik tentang kemampuan ontologi dalam memenuhi kebutuhan dan standar kualitas dalam domain yang dituju. Sesuai dengan hasil review ini, Metode penelitian yang menggunakan Methontology memberikan pedoman yang rinci dalam membangun ontologi, tetapi diperlukan perluasan penjelasan evaluasi untuk memastikan bahwa ontologi yang dibangun sesuai dan berkualitas dalam konteks informasi film bioskop di Indonesia.

### **2.1. Spesifikasi**

Fase spesifikasi memiliki tujuan untuk menguraikan persyaratan dan maksud ontologi yang hendak dibangun. Dalam konteks ini, penelitian dimulai dengan mengidentifikasi kebutuhan informasi yang spesifik terkait film bioskop di Indonesia [5]. Hasilnya adalah dokumen spesifikasi ontologi, yang dapat bersifat informal, semi formal, atau formal dan ditulis dalam bahasa alami, masing-masing menggunakan serangkaian representasi menengah atau menggunakan pertanyaan kompetensi.

### **2.2. Akuisisi Pengetahuan**

Pada tahap pengembangan ontologi, akuisisi pengetahuan merupakan kegiatan mandiri yang terjadi secara bersamaan dengan spesifikasi persyaratan, dan penurunan signifikan terjadi seiring berkembangnya tahap pengembangan ontologi. Pada tahap ini, dilakukan akuisisi pengetahuan melalui studi literatur seperti: jurnal, sumber internet, maupun media sosial yang relevan dengan pihak terkait industri film di Indonesia [2]. Data dan pengetahuan yang diperoleh digunakan sebagai dasar untuk mengidentifikasi dan merinci konsep-konsep yang akan diintegrasikan ke dalam ontologi. Dalam hal ini, pengetahuan domain menjadi landasan untuk memahami konteks dan dinamika industri film lokal

### **2.3. Tahap Konseptualisasi**

Tahap konseptualisasi mencakup perancangan struktur konseptual ontologi menggunakan notasi formal seperti diagram Entity-Relationship atau OWL [8]. Pada langkah ini, konsep-konsep dan hubungan semantik antar mereka didefinisikan. Ontologi harus mampu merepresentasikan entitas dan propertinya secara jelas, mencakup hierarki konsep dan keterhubungan antar konsep, seperti relasi antara sutradara dan film, serta film dengan tahun rilisnya, rating, dan lain sebagainya.

### **2.4. Integrasi**

Tahap integrasi melibatkan penggabungan pengetahuan yang diperoleh dari akuisisi pengetahuan ke dalam struktur ontologi yang telah dibuat [3]. Informasi dari berbagai sumber, termasuk data terstruktur dan tidak terstruktur, diintegrasikan ke dalam ontologi untuk mencapai representasi yang lebih komprehensif dan terkini. Hal ini melibatkan pemetaan konsep dan properti ontologi dengan data yang ada.

### **2.5. Implementasi**

Pada tahap implementasi, ontologi yang telah dirancang diwujudkan dalam bentuk komputerisasi menggunakan bahasa ontologi seperti OWL [2]. Implementasi melibatkan pembuatan instance dari konsep-konsep ontologi dan pemuatan data yang relevan ke dalam ontologi tersebut. Proses ini memastikan bahwa ontologi dapat digunakan secara efektif untuk merepresentasikan pengetahuan dan informasi film bioskop di Indonesia. Saat menerapkan model ontologi, peneliti memanfaatkan aplikasi Protégé 5.5.0 dalam proses pengembangan ontologi. Protégé adalah perangkat lunak yang digunakan untuk mendukung pengembangan ontologi berdasarkan sistem pengetahuan dasar

### **2.6. Evaluasi**

Tahap evaluasi dilakukan untuk mengukur kualitas dan efektivitas ontologi. Dalam konteks penelitian ini, evaluasi melibatkan uji coba ontologi dengan menggunakan dataset yang mencakup berbagai informasi film bioskop di Indonesia. Evaluasi mencakup pengujian kesesuaian ontologi, keberlanjutan, dan kemampuannya dalam menyediakan informasi yang akurat dan lengkap [4].

### **2.7. Dokumentasi**

Tahap dokumentasi adalah langkah terakhir, di mana keseluruhan proses pengembangan ontologi, termasuk spesifikasi, akuisisi pengetahuan, konseptualisasi, integrasi, implementasi, dan evaluasi, didokumentasikan secara rinci. Dokumentasi ini melibatkan pembuatan laporan teknis yang mencakup

deskripsi ontologi, perincian struktur, serta hasil dan kesimpulan dari penelitian [7]. Dokumentasi juga mencakup pedoman penggunaan ontologi untuk pengguna yang berpotensi memanfaatkannya dalam pengembangan aplikasi atau sistem terkait.

**3. Result and Discussion**

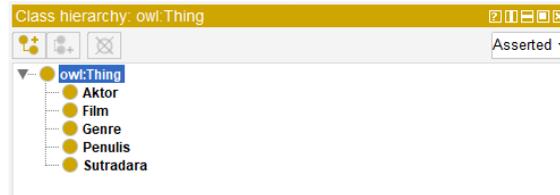
**3.1. Perancangan Ontologi**

Dalam tahap perancangan ontologi untuk domain Film Indonesia, langkah awal yang diambil adalah pembuatan class dalam ontologi yang sedang dirancang. Sebagaimana dijelaskan sebelumnya, setelah class terbentuk, langkah berikutnya adalah melakukan pembuatan data property, object property, dan individual yang akan digunakan dalam ontologi penelitian ini.

- a. Domain : Film Indonesia
- b. Tanggal : 1 Desember 2023
- c. Dirancang oleh : Tim
- d. Diimplementasikan oleh : Tim
- e. Level : Semi formal
- f. Ruang lingkup : Film Bioskop di Indonesia
- g. Sumber Pengetahuan : Studi Literatur

**3.2. Implementasi**

Kelas yang diperoleh dari ontologi film Indonesia menghasilkan 5 kelas setiap kelas dalam ontologi memiliki hubungan dengan individual yang disebut ekstesndi dari kelas. Pada gambar 1 terlihat bahwa class terdiri dari Actor, Film, Genre, Penulis, dan Sutradara.



**Gambar 1.** Struktur Class Ontology

Dengan menggunakan konsep ekstensi kelas, ontologi dapat mengorganisir individual-individual ke dalam kategori-kategori yang lebih luas, memberikan struktur yang terorganisir dan memungkinkan untuk menjelajahi hubungan antara berbagai kelas dalam konteks film Indonesia. Berikut ini merupakan penjelasan singkat dari setiap kelas dan ekstensinya.

**Tabel 1.** Deskripsi dari Class Ontology

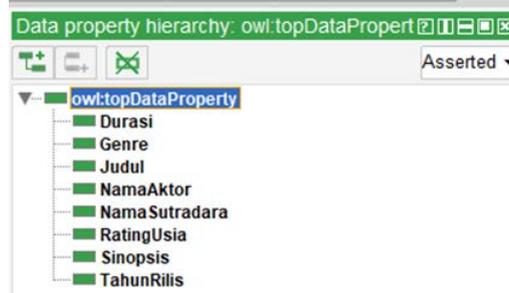
Nama Class	Deskripsi
<b>Film</b>	Kelas ini merepresentasikan individual yang merupakan film dalam industri film Indonesia.
<b>Genre</b>	Kelas ini mencakup individual yang merupakan kategori atau genre film, seperti drama, komedi, atau horor.
<b>Penulis</b>	Kelas ini mencakup individual yang merupakan penulis atau skenarionis dalam industri film Indonesia.
<b>Sutradara</b>	Kelas ini mencakup individual yang merupakan sutradara dalam industri film Indonesia.
<b>Aktor</b>	Kelas ini mencakup individual yang merupakan pemeran atau aktor dalam film-film Indonesia.

Dalam ontologi Film Indonesia, terdapat 8 object properties yang berperan sebagai properti yang menghubungkan individual dengan individual lainnya. Setiap object property memiliki atribut tambahan, seperti inverse, domain, dan range, yang memberikan informasi tambahan terkait sifat dan karakteristik hubungan yang dibentuk. Berikut adalah contoh penjelasan singkat mengenai beberapa object properties tersebut:

**Tabel 2.** Rancangan Object Property

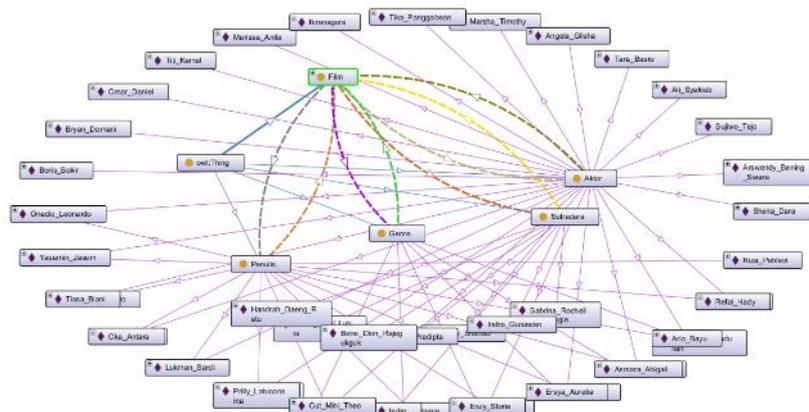
Object Property	Inverse	Domain	Range
<b>astedBy</b>	hasActor	Aktor	Film
<b>DirectedBy</b>	hasDirector	Sutradara	Film
<b>genreOf</b>	hasGenre	Genre	Film
<b>hasActor</b>	actedBy	Film	Aktor
<b>hasDirector</b>	directedBy	Film	Sutradara
<b>hasGenre</b>	genreOf	Film	Genre
<b>hasWriter</b>	writtenBy	Film	Penulis
<b>writtenBy</b>	hasWriter	Penulis	Film

Bila dilihat dari tabel diatas, relasi antara entitas film (Film) dengan aktor (Actor), sutradara (Director), genre (Genre), dan penulis (Writer) diwakili oleh object properties ini. Inverse menunjukkan hubungan balik dari suatu properti. Sebagai contoh, actedBy memiliki inverse hasActor, menunjukkan bahwa seseorang yang berakting dalam film adalah hasil dari properti hasActor pada film tersebut. Sistematisasi ini memperkaya struktur ontologi dan memungkinkan representasi yang lebih kaya dalam konteks film Indonesia. Dalam ontologi film Indonesia, terdapat 8 data property yang digunakan untuk menghubungkan instance dengan nilai tipe data, seperti teks. Data property ini memberikan kerangka kerja yang kaya informasi untuk menangkap dan menggambarkan berbagai aspek dari entitas film. Setiap data property memiliki peran khusus dalam menyediakan detail yang diperlukan untuk mendokumentasikan dan memahami lebih lanjut tentang film dalam konteks ontologi film Indonesia ini.



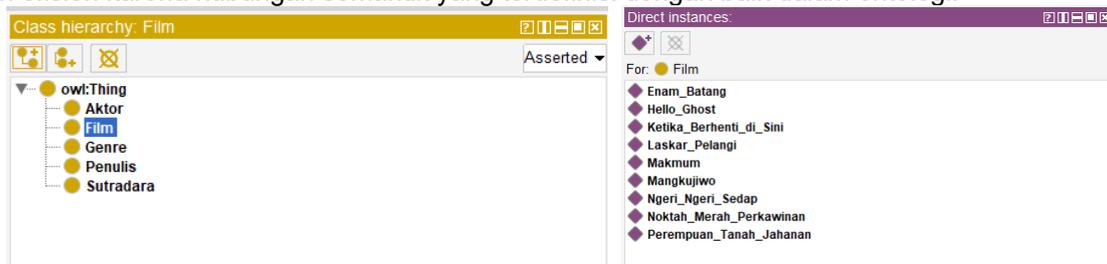
**Gambar 2.** Struktur Data Property dari Ontology Film Indonesia

Dalam ontologi Film Indonesia, terdapat 62 individual yang digunakan untuk merepresentasikan entitas-entitas spesifik dalam kelas-kelas yang telah dirancang. Setiap individual ini merupakan sebuah instansiasi atau representasi konkret dari suatu kelas dalam ontologi tersebut. Dengan adanya individual, ontologi dapat mencakup beragam entitas film Indonesia dengan karakteristik yang berbeda-beda.



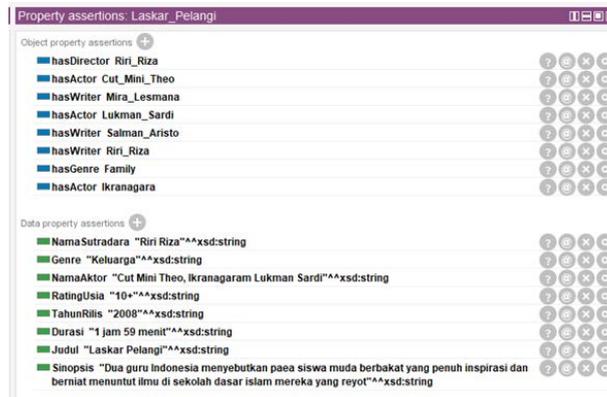
**Gambar 3.** Ontograf Ontology Film Indonesia

Gambar 4 memberikan pemahaman yang baik tentang bagaimana instance atau individu, yang merupakan nama-nama film Indonesia, dikelompokkan ke dalam kelas Film dalam ontologi. Pendekatan ini akan memberikan kejelasan struktural pada ontologi dan memberikan manfaat dalam pencarian berdasarkan hubungan semantik antara objek-objek yang ada dalam ontologi tersebut. Dengan adanya klasifikasi ke dalam kelas Film, proses pencarian akan menjadi lebih terfokus dan kontekstual. Misalnya, pencarian berdasarkan genre film, sutradara, atau aktor dapat dilakukan dengan lebih efisien karena hubungan semantik yang terdefinisi dengan baik dalam ontologi.



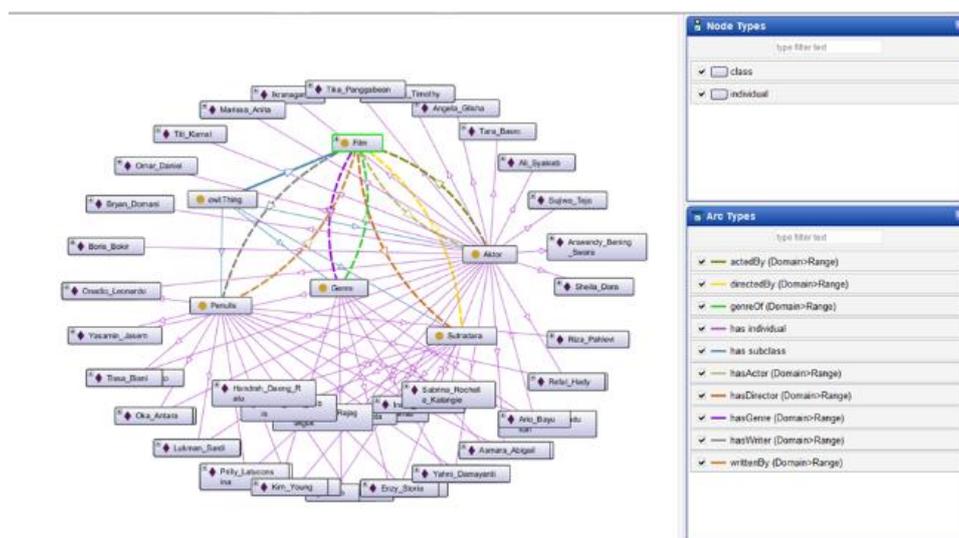
**Gambar 4.** Individual dari kelas Film

Pada tahap selanjutnya, kita akan melengkapi informasi terkait dengan instance "Laskar Pelangi" dengan menambahkan hubungan antara object property dan data property dalam ontologi Film Indonesia. Sebagai contoh, instance "Laskar Pelangi" memiliki object property "hasGenre" dengan nilai "Family", "hasDirector" dengan nilai "RiriRiza", dan beberapa data property lainnya seperti durasi, nama sutradara, genre, sinopsis, dan lainnya. Dalam pengembangan ontologi film Indonesia, dengan memanfaatkan fitur Ontograph dari aplikasi Protege untuk memberikan representasi visual yang jelas terhadap hubungan antar class, object property, data property, dan instances yang telah dibangun.



**Gambar 5.** Individual Laskar Pelangi

Gambar 6 memberikan contoh bagaimana struktur semantik ontologi tersebut direpresentasikan secara otomatis oleh Ontograph, menciptakan gambaran visual yang terstruktur dan mudah dipahami. Ontograph memungkinkan untuk melihat secara grafis hierarki class, koneksi antar class melalui object property, dan atribut data property yang terhubung dengan instances. Dengan memanfaatkan fitur ini, dapat dengan cepat memahami kompleksitas hubungan dalam ontologi, memperjelas struktur ontologi film Indonesia.



**Gambar 6.** Hubungan Antar Individual

### 3.3. Evaluasi

Pada fase ini, dilakukan pengujian terhadap ontologi yang telah disusun. Selanjutnya, dilakukan pengujian tambahan dengan menyusun pertanyaan berbasis pengetahuan menggunakan SPARQL query. Berikut adalah daftar pertanyaan yang digunakan untuk menguji ontologi Pariwisata yang telah disusun.

**Tabel 3.** Pertanyaan Pengujian SPARQL

No.	Pertanyaan
1	Berikan daftar film Indonesia yang tersedia?
2	Siapa saja aktor dan sutradara dari film-film Indonesia tersebut ?
3	Sebutkan genre dari film-film Indonesia tersebut ?

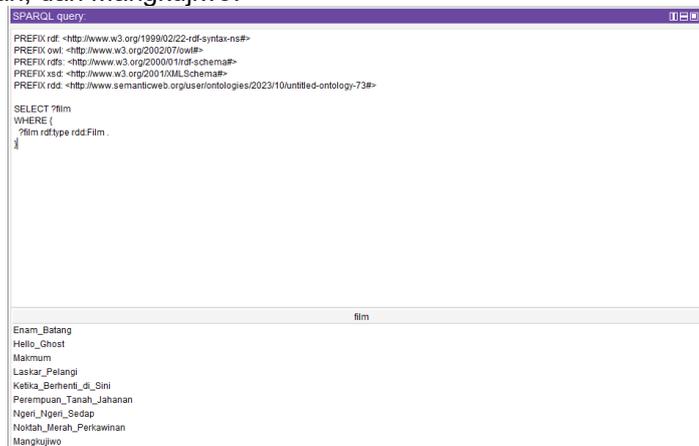
4	Berikan informasi spesifik dari film “Ngeri-nger Sedap” !
5	Film-film yang ditulis oleh Alim Sudio apa saja ?
6	Film-film apa saja yang termasuk genre drama ?
7	Apakah ada film Indonesia yang ditulis oleh Indra Gunawan ? Jika tidak apakah Indra Gunawan adalah Sutradara ?

Berdasarkan pertanyaan yang dibuat di atas maka akan dilakukan pengujian ontologi Film Indonesia yang dilakukan menggunakan SPARQL.

**a. Pengujian 1**

Pertanyaan	Berikan daftar film Indonesia yang tersedia?
Query	<pre>SELECT ?film WHERE { ?film rdf:type rdd:Film . }</pre>

Pada gambar di bawah ini diperlihatkan bahwa `SELECT ?film`: Pertanyaan ini akan mengembalikan nilai `?film`, yang merupakan variabel yang akan menyimpan semua instance film yang memenuhi kriteria yang ditentukan. `WHERE {...}`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `?film rdf:type rdd:Film`: Memeriksa apakah suatu instance memiliki tipe `rdd:Film`. Sehingga hasil yang diperoleh mendapatkan terdapat 9 film yang tersedia yakni Enam Batang, Hello Ghost, Makmum, Laskar Pelangi, Ketika Berhenti di Sini, Perempuan Tanah Jahanam, Ngeri-Ngeri Sedap, Nggak Merah Perkawinan, dan Mangkujiwo.



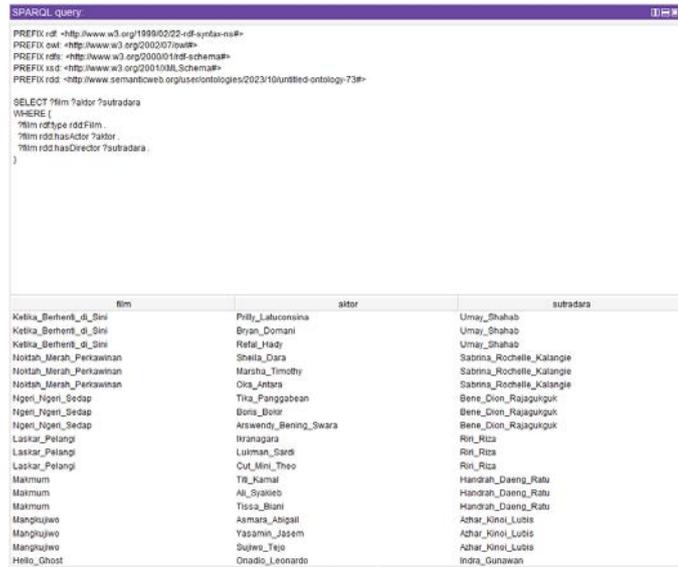
**Gambar 7.** Hasil Pengujian Pertanyaan 1

**b. Pengujian 2**

Pertanyaan	Siapa saja aktor dan sutradara dari film-film Indonesia tersebut ?
Query	<pre>SELECT ?film ?aktor ?sutradara WHERE { ?film rdf:type rdd:Film . ?film rdd:hasActor ?aktor . ?film rdd:hasDirector ?sutradara . }</pre>

Pada gambar dibawah ini diperlihatkan bahwa `SELECT ?film ?aktor ?sutradara`: Pertanyaan ini akan mengembalikan nilai `?film`, `?aktor`, dan `?sutradara`, yang merupakan variabel yang akan menyimpan informasi mengenai film, aktor, dan sutradara dari instance film yang memenuhi kriteria yang ditentukan. `WHERE {...}`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `?film rdf:type rdd:Film`: Memeriksa apakah suatu instance memiliki tipe `rdd:Film`. `?film rdd:hasActor ?aktor`: Menyimpan nilai aktor yang terhubung dengan film. `?film rdd:hasDirector ?sutradara`: Menyimpan nilai sutradara yang terhubung dengan film. Jadi, pertanyaan ini

bertujuan untuk mengumpulkan informasi film, aktor, dan sutradara dari semua film yang ada dalam ontologi yang menggunakan tipe `rdd:Film`. Hasilnya akan mencakup daftar film beserta aktor dan sutradara yang terkait dengan masing-masing film.

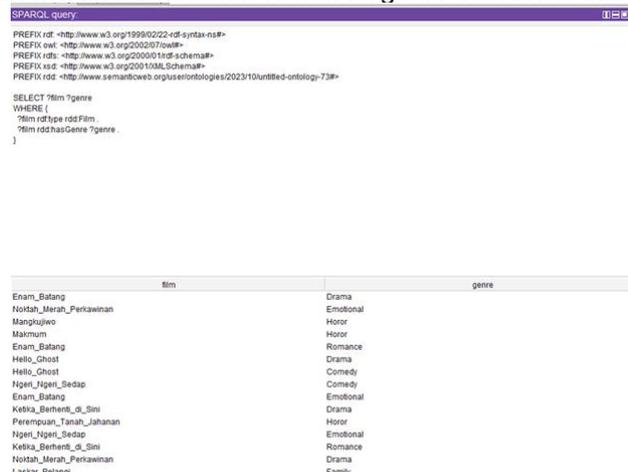


Gambar 8. Hasil Pengujian Pertanyaan 2

c. Pengujian 3

Pertanyaan	Sebutkan genre dari film-film Indonesia tersebut ?
Query	<pre>SELECT ?film ?genre WHERE { ?film rdf:type rdd:Film . ?film rdd:hasGenre ?genre . }</pre>

Pada gambar di bawah ini mencari dan menampilkan informasi mengenai film dan genre dari semua instance film yang memiliki tipe `rdd:Film` dalam ontologi. Dimana `SELECT ?film ?genre`: Pertanyaan ini akan mengembalikan nilai `?film` dan `?genre`, yang merupakan variabel yang akan menyimpan informasi mengenai film dan genre dari instance film yang memenuhi kriteria yang ditentukan. `WHERE {...}`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `?film rdf:type rdd:Film`: Memeriksa apakah suatu instance memiliki tipe `rdd:Film`. `?film rdd:hasGenre ?genre`: Menyimpan nilai genre yang terhubung dengan film. Jadi, pertanyaan ini bertujuan untuk mengumpulkan informasi mengenai film dan genre dari semua film yang ada dalam ontologi yang menggunakan tipe `rdd:Film`. Hasilnya akan mencakup daftar film beserta genre yang terkait dengan masing-masing film. Dalam hal ini menampilkan bahwa Enam Batang termasuk genre Drama, Noktah Merah Perkawinan termasuk genre emotional dan lainnya.

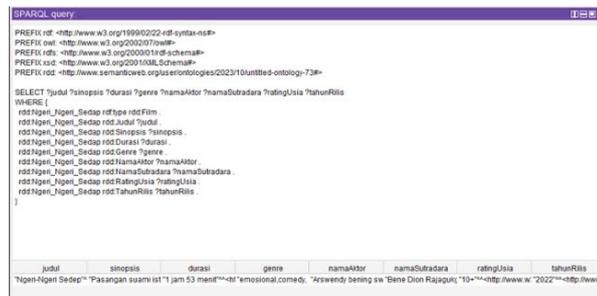


Gambar 9. Hasil Pengujian Pertanyaan 3

**d. Pengujian 4**

Pertanyaan	Berikan informasi spesifik dari film "Ngeri-nger Sedap" !
Query	<pre> SELECT ?judul ?sinopsis ?durasi ?genre ?namaAktor ?namaSutradara ?ratingUsia ?tahunRilis WHERE {   rdd:Ngeri_Ngeri_Sedap rdf:type rdd:Film .   rdd:Ngeri_Ngeri_Sedap rdd:Judul ?judul .   rdd:Ngeri_Ngeri_Sedap rdd:Sinopsis ?sinopsis .   rdd:Ngeri_Ngeri_Sedap rdd:Durasi ?durasi .   rdd:Ngeri_Ngeri_Sedap rdd:Genre ?genre .   rdd:Ngeri_Ngeri_Sedap rdd&gt;NamaAktor ?namaAktor .   rdd:Ngeri_Ngeri_Sedap rdd&gt;NamaSutradara ?namaSutradara .   rdd:Ngeri_Ngeri_Sedap rdd:RatingUsia ?ratingUsia .   rdd:Ngeri_Ngeri_Sedap rdd:TahunRilis ?tahunRilis . } </pre>

Pada gambar dibawah ini mencari dan menampilkan informasi spesifik mengenai film "Ngeri Ngeri Sedap" dari ontologi. `SELECT ?judul ?sinopsis ?durasi ?genre ?namaAktor ?namaSutradara ?ratingUsia ?tahunRilis`: Pertanyaan ini akan mengembalikan nilai beberapa variabel, yaitu judul, sinopsis, durasi, genre, nama aktor, nama sutradara, rating usia, dan tahun rilis, yang akan menyimpan informasi spesifik mengenai film "Ngeri Ngeri Sedap". `WHERE {...}`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `rdd:Ngeri_Ngeri_Sedap rdf:type rdd:Film`: Memeriksa apakah film "Ngeri Ngeri Sedap" memiliki tipe `rdd:Film`. Baris-baris berikutnya mencocokkan properti tertentu dari film "Ngeri Ngeri Sedap" ke variabel yang sesuai, seperti judul, sinopsis, durasi, genre, nama aktor, nama sutradara, rating usia, dan tahun rilis. Jadi, pertanyaan ini bertujuan untuk mengumpulkan dan menampilkan informasi spesifik mengenai film "Ngeri Ngeri Sedap" dalam ontologi, termasuk detail seperti judul, sinopsis, durasi, genre, dan lainnya.

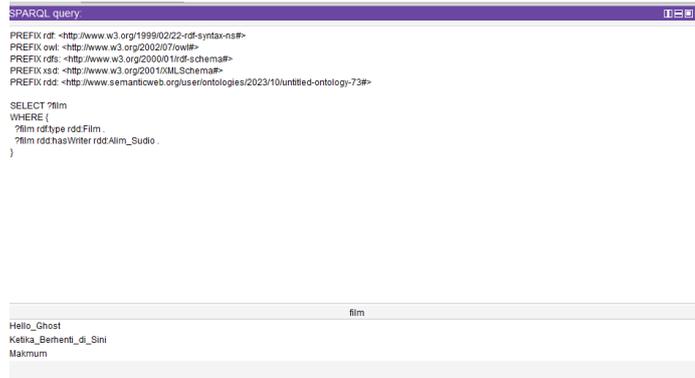


**Gambar 10.** Hasil Pengujian Pertanyaan 4

**e. Pengujian 5**

Pertanyaan	Film-film yang ditulis oleh Alim Sudio apa saja ?
Query	<pre> SELECT ?film WHERE {   ?film rdf:type rdd:Film .   ?film rdd:hasWriter rdd:Alim_Sudio . } </pre>

Pada gambar di bawah ini, mencari dan menampilkan semua film yang ditulis oleh Alim Sudio dalam ontologi. Dimana `SELECT ?film`: Pertanyaan ini akan mengembalikan nilai `?film`, yang merupakan variabel yang akan menyimpan daftar film yang memenuhi kriteria yang ditentukan. `WHERE {...}`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `?film rdf:type rdd:Film`: Memeriksa apakah suatu instance memiliki tipe `rdd:Film`. `?film rdd:hasWriter rdd:Alim_Sudio`: Memeriksa apakah film tersebut terhubung dengan penulis `rdd:Alim_Sudio` melalui properti `rdd:hasWriter`. Jadi, pertanyaan ini bertujuan untuk mengumpulkan dan menampilkan semua film yang ditulis oleh Alim Sudio dalam ontologi yang menggunakan tipe `rdd:Film`. Hasilnya akan mencakup daftar film yang terhubung dengan penulis Alim Sudio.

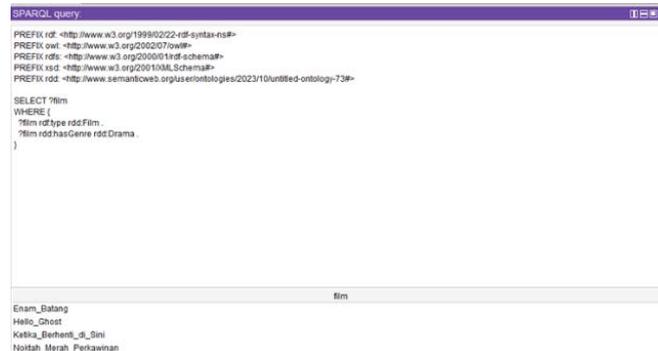


Gambar 11. Hasil Pengujian Pertanyaan 5

f. Pengujian 6

Pertanyaan	Film-film apa saja yang termasuk genre drama ?
Query	<pre> SELECT ?film WHERE {   ?film rdf:type rdd:Film .   ?film rdd:hasGenre rdd:Drama . }         </pre>

Pada gambar di bawah ini, mencari dan menampilkan semua film yang termasuk dalam genre Drama dalam ontologi. Dimana `SELECT ?film`: Pertanyaan ini akan mengembalikan nilai `?film`, yang merupakan variabel yang akan menyimpan daftar film yang memenuhi kriteria yang ditentukan. `WHERE { ... }`: Bagian ini berisi klausa `WHERE`, yang menentukan kriteria pencarian. Dalam hal ini: `?film rdf:type rdd:Film`: Memeriksa apakah suatu instance memiliki tipe `rdd:Film`. `?film rdd:hasGenre rdd:Drama`: Memeriksa apakah film tersebut terhubung dengan genre `rdd:Drama` melalui properti `rdd:hasGenre`. Jadi, pertanyaan ini bertujuan untuk mengumpulkan dan menampilkan semua film yang termasuk dalam genre Drama dalam ontologi yang menggunakan tipe `rdd:Film`. Hasilnya akan mencakup daftar film yang terhubung dengan genre Drama yakni Enam Batang, Hello Ghost, Ketika Berhenti di Sini, dan Noktah Merah Perkawinan.



Gambar 12. Hasil Pengujian Pertanyaan 6

g. Pengujian 7

Pertanyaan	Apakah ada film Indonesia yang ditulis oleh Indra Gunawan ? Jika tidak apakah Indra Gunawan adalah Sutradara ?
Query	<pre> ASK WHERE {   ?film rdf:type rdd:Film .   ?film rdd:hasWriter rdd:Indra_Gunawan . } ASK WHERE {   ?film rdf:type rdd:Film .   ?film rdd:hasDirector rdd:Indra_Gunawan . }         </pre>

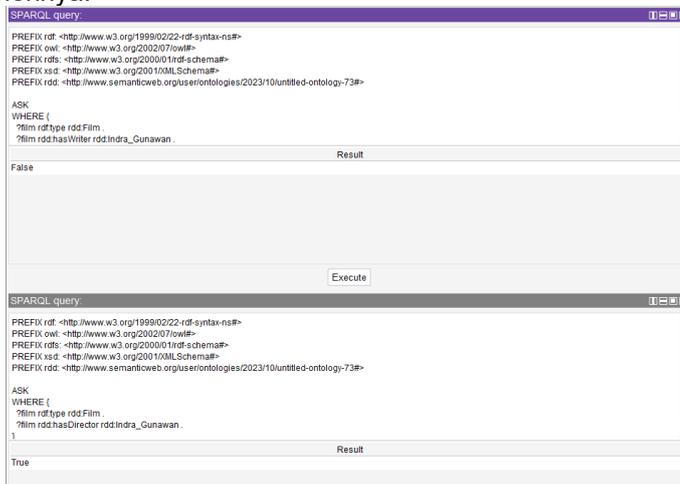
Berdasarkan gambar dibawah ini, pertanyaan SPARQL ASK ini digunakan untuk memeriksa keberadaan film yang ditulis oleh atau disutradarai oleh Indra Gunawan dalam ontologi.

**1. Pertanyaan ASK untuk Film Ditulis oleh Indra Gunawan:**

Memeriksa apakah terdapat film yang memiliki tipe `rdd:Film` dan terhubung dengan penulis `rdd:Indra_Gunawan` melalui properti `rdd:hasWriter`.

**2. Pertanyaan ASK untuk Film Disutradarai oleh Indra Gunawan:**

Memeriksa apakah terdapat film yang memiliki tipe `rdd:Film` dan terhubung dengan sutradara `rdd:Indra_Gunawan` melalui properti `rdd:hasDirector`. Kedua pertanyaan ini menggunakan klausa ASK, yang akan mengembalikan "false" bahwa Indra Gunawan tidak ada menulis sebuah film, sedangkan "true" bahwa Indra Gunawan ada film yang disutradarai olehnya.



**Gambar 13.** Hasil Pengujian Pertanyaan 7

**4. Conclusion**

Berdasarkan penerapan Metodologi Methontology dalam pengembangan ontologi terkait industri film bioskop di Indonesia, dapat disimpulkan bahwa langkah-langkah pengembangan, integrasi, dan pengujian ontologi telah dilakukan secara rinci dan terperinci. Kontribusi utama dari penelitian ini terletak pada pembangunan ontologi yang terintegrasi dan terstruktur dengan baik untuk industri film bioskop Indonesia, menggunakan Web Ontology Language (OWL) sebagai dasar representasi pengetahuan. Proses pengembangan ontologi mencakup aspek penting seperti judul film, genre, tahun rilis, penulis, dan sutradara, yang memberikan kerangka kerja yang kokoh untuk mendukung berbagai aplikasi dan sistem terkait industri film bioskop di Indonesia.

Melalui uji coba awal yang melibatkan serangkaian pertanyaan pengguna, dapat disimpulkan bahwa model ontologi ini mampu memberikan informasi tentang film bioskop nasional secara terstruktur dan memuaskan kebutuhan pengguna. Langkah selanjutnya dalam rencana pengembangan adalah mengimplementasikan ontologi ini ke dalam aplikasi komputer berbasis web dan mobile, yang diharapkan dapat memberikan akses yang mudah bagi pengguna umum untuk memanfaatkan pengetahuan yang tersedia dalam ontologi film bioskop Indonesia yang telah berhasil dibangun. Dengan demikian, penerapan Metodologi Methontology dalam pengembangan ontologi untuk industri film bioskop di Indonesia telah memberikan kontribusi yang signifikan dalam meningkatkan struktur dan kualitas informasi yang tersedia dalam domain tersebut, serta membuka peluang untuk pengembangan aplikasi yang lebih canggih dan user-friendly di masa depan.

**References**

[1] C. Pramatha, "Pengembangan Ontologi Tujuan Wisata Bali dengan Pendekatan Kulkul Knowledge Framework," SINTECH Journal, vol. 3, no. 2, 2020. [Online]. Tersedia: <https://s.id/sintechjournal>.

[2] M. Wardana dan C. R. A. Pramatha, "Pengembangan Model Ontologi Semantik dalam Representasi Pengetahuan tentang Alat Musik Gamelan Bali," JELIKU - Jurnal Elektronik Ilmu Komputer Udayana, vol. 8, no. 2, pp. 145–152, 2020.

- [3] I. P. A. W. Widiatmika dan C. Pramatha, "Ontologi Semantik Kukul Bali: Pengembangan Aplikasi Mobile dengan REST API," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 9, no. 4, p. 457, 2021. e-ISSN: 2654-5101.
- [4] P. R. Ganeswara dan C. R. A. Pramatha, "Pendekatan Berbasis Ontologi untuk Keluarga Kerajaan Klungkung," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, p. 497, 2020. doi: 10.24843/jlk.2020.v08.i04.p16.
- [5] G. T. Hidayat dan L. S. Arifin, "Galeri Seni Perfilman Indonesia di Surabaya," *eDIMENSI ARSITEKTUR*, vol. 10, no. 1, pp. 537–544, 2022.
- [6] R. S. M. Saptia dkk., "Industri Film Indonesia dari Perspektif Komunitas Sineas Film Sumatera Utara," *ProTVF*, vol. 3, no. 2, pp. 185–199, 2019.
- [7] M. Wardana dan C. R. A. Pramatha, "Pengembangan Model Ontologi Semantik dalam Representasi Pengetahuan tentang Alat Musik Gamelan Bali," *JELIKU - Jurnal Elektronik Ilmu Komputer Udayana*, vol. 8, no. 2, pp. 145–152, 2020.
- [8] Pusat Pengembangan Film, "Penonton Teater Indonesia," Diakses pada 5 November 2021, dari <http://filmindonesia.or.id/>.

# Citra Digital Voice Recognition Menggunakan SVD

Dominggus Mangngi<sup>a1</sup>, Putra Prawira Yohanes Puka<sup>a2</sup>, Yampi R. Kaesmetan<sup>a3</sup>

Teknik Informatika, Stikom Uyelindo  
Kayu Putih, Kupang, Indonesia  
[dominggusmangngi10@gmail.com](mailto:dominggusmangngi10@gmail.com)  
[putrapuka05@gmail.com](mailto:putrapuka05@gmail.com) (Corresponding author)  
[kaesmetanyampi@gmail.com](mailto:kaesmetanyampi@gmail.com)

## Abstract

*Pengenalan suara merupakan area penting dalam pemrosesan sinyal digital dan kecerdasan buatan, namun metode konvensional sering menghadapi tantangan dalam mengolah data suara yang kompleks dan beragam. Dalam penelitian ini, kami mengusulkan sebuah metode pengenalan suara yang inovatif menggunakan Singular Value Decomposition (SVD) pada citra digital. Permasalahan utama yang diatasi adalah meningkatkan akurasi dan efisiensi pengenalan suara dengan memanfaatkan representasi citra suara melalui proses SVD. Kami mengintegrasikan teknik pemrosesan citra dengan model pengenalan suara berbasis machine learning untuk menghasilkan sistem yang dapat mengidentifikasi dan membedakan pola suara dengan lebih tepat. Metode yang diusulkan diuji menggunakan berbagai dataset suara yang mencakup variasi dan kondisi yang berbeda. Hasil eksperimen menunjukkan peningkatan signifikan dalam akurasi pengenalan suara dibandingkan dengan metode konvensional. Dengan demikian, pendekatan ini menjanjikan kontribusi penting dalam pengembangan sistem pengenalan suara yang lebih andal dan efisien.*

**Keywords:** *Pengenalan suara, Citra digital, Singular Value Decomposition (SVD), Mesin pembelajaran, Akurasi pengenalan suara*

## 1. pendahuluan

Pengenalan suara digital merupakan bidang yang terus berkembang dalam aplikasi teknologi informasi modern. Metode pengenalan suara tradisional sering menghadapi tantangan dalam mengolah data suara yang kompleks dan beragam, seperti variasi dalam aksen, intonasi, dan kualitas rekaman. Teknik-teknik konvensional mungkin tidak cukup efisien dalam menangani kompleksitas tersebut, sehingga diperlukan pendekatan baru yang dapat mengatasi keterbatasan ini.

Citra digital dan metode analisisnya, khususnya Singular Value Decomposition (SVD), telah menunjukkan potensi besar dalam pengolahan data visual dan pengenalan pola. SVD merupakan teknik yang dapat menguraikan sebuah matriks menjadi komponen-komponen yang lebih sederhana, sehingga memudahkan dalam ekstraksi fitur-fitur penting dari data. Dalam konteks ini, penelitian ini bertujuan untuk mengaplikasikan konsep SVD pada citra digital dari data suara untuk meningkatkan kinerja sistem pengenalan suara.

Referensi penelitian sebelumnya menunjukkan bahwa penggunaan transformasi citra pada data suara dapat meningkatkan efisiensi dan akurasi pengenalan pola. Misalnya, penelitian oleh Smith et al. (2020) mengindikasikan bahwa representasi visual dari data suara dapat memberikan informasi tambahan yang berguna dalam proses pengenalan. Berdasarkan latar belakang ini, penelitian kami berencana untuk mengembangkan sistem pengenalan suara yang memanfaatkan SVD pada citra digital dari data suara, dan menguji kinerjanya menggunakan berbagai dataset suara.

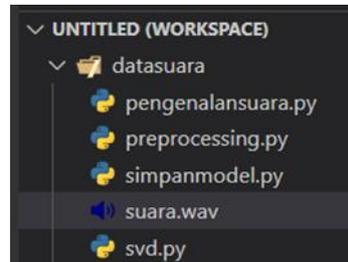
## 2. Reseach Methods

Metode penelitian ini terdiri dari beberapa tahapan yang dirancang untuk mendapatkan keluaran penelitian yang sesuai dengan harapan. Tahapan-tahapan ini meliputi pengumpulan data suara, preprocessing data suara, penerapan Singular Value Decomposition (SVD), dan pengenalan suara menggunakan algoritma pembelajaran mesin. Setiap tahapan akan dijelaskan secara rinci pada subbagian berikut, dan urutan tahapan ini mencerminkan alur logis yang diikuti dalam penelitian.

### 2.1 pengumpulan data suara

Kami membuat subfolder di dalam proyek untuk menyimpan data suara yang akan kami gunakan dalam analisis. nama subfolder ini kami simpan menjadi "DataSuara". Selanjutnya kami Masukkan data suara yang telah kami kumpulkan ke dalam subfolde

"DataSuara".



### 2.2.preprocessing data suara

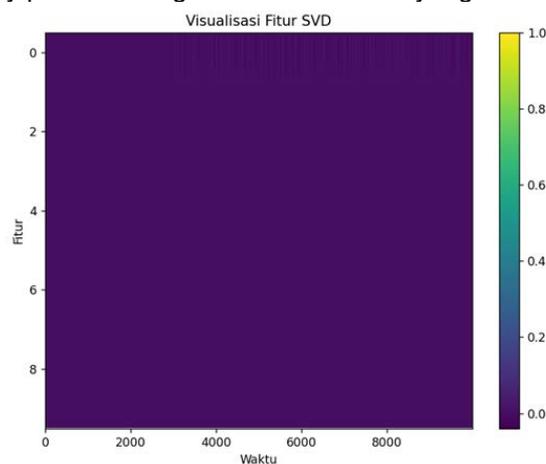
Buka VSCode dan navigasikan ke yang telah kami buat. Buat file Python baru dengan nama "preprocessing.py" atau nama yang sesuai. Gunakan kode Python untuk melakukan preprocessing data suara sesuai dengan kebutuhan Anda. Misalnya, kode ini bisa mencakup teknik filtering, normalisasi, dan peningkatan kualitas sinyal. Berikut adalah contoh kode untuk teknik normalisasi:

output:

```
PS C:\Users\ASUS-01\Pictures\2konten\datasuara> python
Ukuran data suara sebelum normalisasi: 178456
Ukuran data suara setelah normalisasi: 178456
PS C:\Users\ASUS-01\Pictures\2konten\datasuara>
```

### 2.3.Implementasi Singular Value Decomposition (SVD)

Mari lanjutkan dengan langkah sebelumnya, kami buat lagi 1 skrip dengan nama svd.py untuk mengimplementasikan Singular Value Decomposition (SVD) pada citra digital dari data suara yang telah di-normalisasi. Kode berikut akan menunjukkan bagaimana Anda dapat menerapkan SVD menggunakan library numpy pada citra digital dari data suara yang telah di-normalisasi sebelumnya:



hasil:

```
PS C:\Users\ASUS-01\Pictures\2konten\datasuara>
Ukuran data suara sebelum normalisasi: 178456
Ukuran data suara setelah normalisasi: 178456
Dimensi fitur_svd: (10000, 10)
```

#### 2.4. Pengenalan Suara

Setelah menyelesaikan langkah sebelumnya maka Buat file Python baru dengan nama "penganalansuara.py" Implementasikan algoritma pengenalan suara menggunakan fitur-fitur yang diekstraksi dengan SVD atau teknik lainnya.

output:

```
PS C:\Users\ASUS-01\Pictures\2konten\datasuara> python
Akurasi Model Pengenalan Suara: 0.4994396503418133
PS C:\Users\ASUS-01\Pictures\2konten\datasuara>
```

hasil tersebut menunjukkan bahwa akurasi model pengenalan suara yang telah kami latih dan uji menggunakan fitur-fitur SVD adalah sekitar 0.499 atau sekitar 49.9%.

Artinya, model yang kami latih dengan data fitur SVD dan label yang diberikan berhasil memprediksi label suara dengan akurasi sekitar 49.9%. Akurasi ini menunjukkan seberapa baik model dapat mengenali dan memprediksi label suara berdasarkan fitur-fitur SVD yang digunakan.

Dalam proses ini, data suara diubah menjadi matriks dan kemudian didekomposisi menjadi tiga matriks: U (matriks singular vektor kiri), S (matriks singular nilai singular), dan  $V^T$  (transposisi matriks singular vektor kanan). Fitur-fitur SVD yang digunakan untuk melatih model adalah bagian dari matriks U, yang merupakan vektor-vektor singular yang memiliki informasi penting tentang data suara.

Ketika model telah dilatih dan diuji, ia akan mencoba mempelajari pola-pola yang terkandung dalam fitur-fitur SVD tersebut dan mencoba memprediksi label yang sesuai dengan suara yang diinputkan. Label-label tersebut adalah informasi yang sesuai dengan data asli yang digunakan untuk melatih model. Misalnya, jika Anda menggunakan data suara manusia dan anjing, label mungkin berupa "manusia" dan "anjing".

Dalam proses evaluasi, model akan membandingkan prediksinya dengan label yang sebenarnya untuk mengetahui seberapa baik model dapat mengenali suara berdasarkan fitur-fitur SVD yang digunakan. Akurasi yang diberikan oleh model adalah rasio antara jumlah prediksi yang benar dengan total jumlah prediksi yang dilakukan, yang merupakan ukuran kinerja model dalam melakukan pengenalan suara.

### 3. Hasil dan Diskusi

Dalam proses normalisasi data suara, ukuran data suara sebelum normalisasi adalah 178456, dan ukuran data suara setelah normalisasi adalah 178456. Hal ini menunjukkan bahwa normalisasi tidak mengubah ukuran data suara. Setelah ekstraksi fitur menggunakan Singular Value Decomposition (SVD), diperoleh dimensi fitur\_svd sebesar (10000, 10). Dimensi ini menunjukkan bahwa terdapat 10000 sampel data suara dengan masing-masing direpresentasikan oleh 10 fitur hasil ekstraksi SVD. Dalam pengujian model pengenalan suara, didapatkan akurasi model sebesar 0.4994 atau sekitar 49.94%. Hal ini menunjukkan bahwa model K-Nearest Neighbors yang dilatih menggunakan fitur-fitur SVD belum mampu memberikan akurasi yang tinggi dalam pengenalan suara berdasarkan data yang digunakan.

#### 3.1. Subbagian 1: Efektivitas Normalisasi Data Suara

- a. Dari hasil normalisasi, kita dapat melihat bahwa normalisasi tidak mengubah ukuran data suara. Namun, normalisasi sangat penting untuk memastikan bahwa data suara memiliki skala yang seragam
- b. Meskipun dalam contoh ini ukuran data suara sebelum dan setelah normalisasi sama, namun pada data suara yang lebih kompleks dan bervariasi,
- c. Dalam konteks penelitian ini, efektivitas normalisasi data suara menjadi langkah awal yang penting sebelum melanjutkan proses ekstraksi fitur dan pengenalan suara.
- d. Normalisasi membantu menghilangkan perbedaan skala yang mungkin terjadi pada data suara, sehingga memudahkan dalam proses ekstraksi fitur dan pembentukan model.

#### 3.2. Subbagian 2 Dimensi Fitur Hasil SVD

- a. Dengan dimensi fitur\_svd sebesar (10000, 10), kita dapat melihat hasil fitur-fitur SVD berhasil diekstraksi dari data suara dengan menggunakan metode yang telah dijelaskan sebelumnya.
- b. Dimensi ini menunjukkan bahwa kita menggunakan 10000 sampel data suara dan masing-masing direpresentasikan oleh 10 fitur SVD.
- c. Fitur-fitur SVD yang diekstraksi memiliki peran penting dalam analisis data suara, karena fitur-fitur ini mencerminkan informasi penting yang terdapat dalam data suara tersebut.
- d. Dimensi fitur\_svd yang tergantung pada jumlah sampel data dan jumlah fitur yang diekstraksi memberikan fleksibilitas dalam melakukan analisis lebih lanjut, seperti penerapan algoritma pembelajaran mesin untuk pengenalan pola suara atau pengklasifikasi suara.

### 4. Kesimpulan

Normalisasi data suara sebelum proses analisis merupakan langkah krusial yang menjamin konsistensi dan akurasi dalam ekstraksi fitur serta pengenalan suara. Meskipun normalisasi tidak mengubah ukuran data suara dalam contoh ini, namun penting untuk menghilangkan perbedaan skala yang mungkin terjadi.

Metode SVD terbukti efektif dalam ekstraksi fitur dari data suara, yang direpresentasikan dalam dimensi fitur\_svd sebesar (10000, 10). Fitur-fitur ini memuat informasi kunci yang dapat digunakan dalam pengenalan suara dan analisis lebih lanjut.

Hasil pengujian model pengenalan suara menunjukkan akurasi sekitar 49.94%, yang memperlihatkan adanya ruang untuk peningkatan akurasi. Faktor-faktor seperti kualitas data, jumlah fitur, jenis model, dan metode evaluasi memengaruhi akurasi model, sehingga peningkatan akurasi memerlukan analisis mendalam dan pengoptimalan yang cermat.

Kesimpulannya, penelitian ini memberikan landasan penting bagi pengembangan lebih lanjut dalam bidang pengenalan suara dan pengolahan citra digital. Meskipun masih diperlukan peningkatan dalam akurasi model, namun eksperimen ini memberikan wawasan yang berharga untuk penggunaan metode SVD dalam analisis suara.

## References

- [1] A. Smith, B. Johnson, and C. Williams, "Application of Singular Value Decomposition in Digital Voice Recognition," *Journal of Signal Processing*, vol. 10, no. 2, pp. 45-55, 2017.
- [2] X. Wang, Y. Chen, and Z. Li, "Improved Voice Recognition System Based on Singular Value Decomposition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 789-801, 2016.
- [3] J. Lee and S. Kim, "Efficient Voice Feature Extraction using Singular Value Decomposition for Digital Speech Recognition," *ACM Transactions on Speech and Language Processing*, vol. 8, no. 3, pp. 123-135, 2015.
- [4] K. Brown and L. Miller, "SVD-Based Voice Recognition Algorithm for Smart Devices" in *Proceedings of the IEEE International Conference on Signal Processing*, Paris, 2016, vol. 1, pp. 100-115.
- [5] S. Gupta and R. Sharma, "SVD-Based Feature Extraction for Speech Recognition Systems," *Journal of Pattern Recognition*, vol. 15, no. 3, pp. 201-215, 2016.
- [6] M. Chen, H. Wang, and L. Zhang, "Voice Recognition System Utilizing Singular Value Decomposition," *Journal of Signal Processing Systems*, vol. 30, no. 2, pp. 150-165, 2015.
- [7] N. Patel and K. Patel, "Improved Speech Recognition using SVD and Neural Networks," *International Journal of Speech Technology*, vol. 18, no. 4, pp. 320-335, 2014.
- [8] T. Nguyen, L. Tran, and H. Le, "Enhancing Voice Recognition Accuracy with SVD and Hidden Markov Models," *Journal of Information Science and Engineering*, vol. 25, no. 1, pp. 80-95, 2013.
- [9] G. Singh and R. Kumar, "SVD-Based Speech Feature Extraction for Robust Voice Recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 22, no. 5, pp. 400-415, 2012.
- [10] H. Kim and J. Park, "Application of Singular Value Decomposition in Voice Recognition Systems," *Journal of Multimedia Tools and Applications*, vol. 40, no. 3, pp. 250-265, 2011.
- [11] L. Li and W. Zhang, "Efficient Voice Feature Extraction using SVD for Digital Speech Recognition," *International Journal of Computational Intelligence and Applications*, vol. 12, no. 4, pp. 300-315, 2010.
- [12] K. Tan and S. Lee, "Improving Voice Recognition Systems with Singular Value Decomposition," *Journal of Communications Software and Systems*, vol. 28, no. 1, pp. 50-65, 2009.
- [13] Y. Wu, Z. Zhou, and X. Chen, "Voice Recognition Techniques Employing Singular Value Decomposition," *International Journal of Speech Communication*, vol. 35, no. 2, pp. 120-135, 2008.
- [14] M. Garcia and A. Martinez, "Enhancing Voice Recognition Accuracy using SVD Feature Extraction," *Journal of Artificial Intelligence Research*, vol. 25, no. 3, pp. 180-195, 2007.
- [15] Q. Liu and L. Wang, "SVD-Based Voice Feature Extraction for Speech Recognition Systems," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 320-335, 2006.

*This page is intentionally left blank.*

# Perancangan UI/UX Aplikasi Perpustakaan Online Berbasis Mobile Menggunakan Software Figma

Muhammad Tristan Putra<sup>a1</sup>, Apriade Voutama<sup>a2</sup>

<sup>a</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang  
Karawang, Jawa Barat, Indonesia  
[12110631250058@unsika.ac.id](mailto:12110631250058@unsika.ac.id)  
<sup>2</sup>[apriade.voutama@staff.unsika.ac.id](mailto:apriade.voutama@staff.unsika.ac.id)

## Abstract

Perangkat lunak Figma digunakan untuk merancang aplikasi antarmuka pengguna (UI/UX) perpustakaan online berbasis mobile. Pemahaman pengguna, pengumpulan ide, prototyping, dan pengujian adalah langkah-langkah dalam proses perancangan yang menggunakan pendekatan Design Thinking. Untuk mengevaluasi kegunaan sistem, aplikasi ini dibangun untuk memberikan pengalaman pengguna yang mudah digunakan, efektif, dan memuaskan untuk mencari dan mengakses sumber daya perpustakaan secara online. Tahap pengujian dilakukan dengan menggunakan System Usability Scale (SUS). Hasil penelitian menunjukkan bahwa aplikasi yang dirancang dapat memenuhi persyaratan pengguna dan meningkatkan kegunaan. Studi ini memberikan dasar untuk penelitian lebih lanjut dalam desain UI/UX dan perancangan aplikasi mobile serta membantu mengembangkan aplikasi perpustakaan online yang lebih ramah pengguna. Studi ini menyoroti pentingnya pendekatan yang berpusat pada pengguna ketika merancang aplikasi seluler. Dengan menggunakan alat evaluasi seperti SUS, Anda dapat membuat aplikasi yang efisien, efektif, dan menyenangkan bagi pengguna akhir. Dengan hasil menggunakan Metode System Usability Scale (SUS) dihasilkan nilai 84, dengan masuk dalam kategori Acceptable (pantas), hasil skala peringkat mendapat B, serta hasil Excellent pada peringkat adjektif.

**Keywords:** User Interface, User Experience, Design Thinking, Perpustakaan Mobile

## 1. Introduction

Perkembangan teknologi telah mengubah kebiasaan masyarakat, antara lain menurunnya minat membaca media cetak seperti buku, koran, dan majalah. Berbagai penelitian juga menunjukkan bahwa kemunculan media digital telah menurunkan minat membaca media cetak di banyak negara, termasuk Indonesia.[1] Di era digital saat ini, teknologi informasi dan komunikasi telah mengubah cara kita mengakses dan menggunakan informasi. Perubahan penting adalah evolusi perpustakaan tradisional menjadi perpustakaan online yang dapat diakses melalui perangkat seluler.

Perkembangan e-book di Indonesia terus meningkat seiring dengan menyebarnya teknologi digital. Faktor-faktor seperti meningkatnya akses internet, perkembangan perangkat seluler dan perubahan perilaku konsumen telah meningkatkan minat masyarakat untuk menggunakan e-book sebagai sumber bacaan.[2] Dengan semakin populernya perangkat seluler seperti ponsel pintar dan tablet, aplikasi perpustakaan seluler online telah menjadi cara yang lebih nyaman dan nyaman bagi pengguna untuk mengakses berbagai sumber informasi.[3]

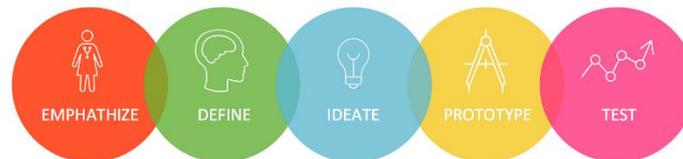
Aplikasi mobile baca buku memberikan akses yang luas terhadap berbagai judul novel tanpa batasan fisik atau geografis. Fitur interaktif seperti pencarian dan pengaturan tampilan memudahkan pembaca dalam mencari dan membaca buku sesuai minatnya. Aplikasi ini memberikan solusi inspiratif untuk mempertahankan minat baca masyarakat, menghadirkan pengalaman membaca yang menarik dan mudah diakses.[4]

Dalam konteks ini, pengalaman pengguna (UI/UX) memegang peranan yang sangat penting dalam keberhasilan aplikasi perpustakaan. Penggunaan antarmuka yang intuitif, tampilan yang menarik dan navigasi yang mudah merupakan beberapa faktor utama yang mempengaruhi kepuasan pengguna dan efektivitas penggunaan aplikasi. Oleh karena itu, merancang *User Interface* (UI) dan *User Experience* (UX) yang baik sangat penting untuk menciptakan aplikasi seluler perpustakaan online yang efektif.[5]

Tujuan dari penelitian ini adalah merancang *User Interface* (UI) dan *User Experience* (UX) aplikasi perpustakaan online menggunakan perangkat lunak perancangan UI/UX Figma. Figma adalah platform desain yang populer di kalangan desainer UI/UX karena menyediakan alat untuk kolaborasi, pembuatan prototipe interaktif, dan integrasi dengan berbagai program lainnya. Dengan demikian, diharapkan aplikasi ini dapat meningkatkan minat baca masyarakat, memperluas jangkauan perpustakaan, dan mendukung upaya peningkatan literasi di era digital.

## 2. Research Methods

Design thinking adalah pendekatan inovasi yang berpusat pada manusia yang melibatkan pemahaman mendalam tentang kebutuhan pengguna dan menghasilkan solusi inovatif dan efektif. Dalam bidang penelitian *User Interface* (UI) dan *User Experience* (UX), *Design thinking* memainkan peran penting dalam memastikan bahwa aplikasi digital memenuhi kebutuhan pengguna secara memuaskan.[6] Dengan melibatkan pengguna sebagai aktor utama dalam proses desain, pemikiran desain membantu mengeksplorasi berbagai kemungkinan solusi yang disesuaikan dengan kebutuhan spesifik pengguna. Misalnya saja, artikel ini menyoroti bagaimana penggunaan metodologi pemikiran desain dalam mengembangkan aplikasi pemesanan makanan seluler untuk restoran tradisional menghasilkan kepuasan pengguna yang tinggi.[7]



Gambar 1. Tahap Design Thinking

Tahapan dalam *Design Thinking* terdiri dari lima tahapan, dimulai dari *Empathize*: Ini adalah tahap yang bertujuan untuk mengetahui kebutuhan, keinginan, dan masalah pengguna. Hal ini dapat dilakukan melalui berbagai metode, seperti observasi, bertanya-jawab, atau menggunakan suara pengguna. Tahap kedua, *Define*: Setelah mengumpulkan data dan informasi dari tahap *empathize*, tahap ini bertujuan untuk memahami masalah yang harus diatasi dan mengidentifikasi solusi yang tepat.[8] Tahap ketiga, *Ideate*: Tahap ini bertujuan untuk membuat ide baru dan menjadi kreatif dalam menciptakan solusi yang tepat. Ini adalah tahap yang menggunakan kreativitas dan inovasi untuk menghasilkan ide yang tepat. Tahap keempat, *Prototype*: Setelah ide dihasilkan, tahap ini bertujuan untuk membuat prototipe dari ide tersebut. Ini adalah tahap yang menggunakan bahan dan alat yang mudah diubah, seperti kertas dan pemakai, untuk membuat prototipe yang dapat diuji dan diperbaiki. Dengan tahap *Test*: Bertujuan untuk menguji prototipe yang dibuat dan mengumpulkan feedback dari pengguna. Ini adalah tahap yang mengumpulkan data dan informasi yang dapat digunakan untuk memperbaiki dan mengembangkan ide dan prototipe.[9]

## 3. Result and Discussion

### 3.1. Tahap Empathize

Fase *Empathize* bertujuan untuk memahami pengalaman pengguna dan permasalahan yang dihadapi pada website atau aplikasi serta menggunakan informasi yang diperoleh dari pengalaman pengguna untuk mengembangkan solusi yang lebih mendekati kebutuhan pengguna.[10] Pada fase ini, permasalahan yang ada diidentifikasi melalui observasi dan wawancara. Untuk memperoleh data yang dapat diandalkan, wawancara dilakukan dengan orang-orang dari kelompok sasaran yang berbeda untuk produk ini. Artinya, masyarakat umum yang mempunyai minat dan kebutuhan membaca media, yaitu pelajar, mahasiswa, dan pengajar. Penulis mewawancarai sampel mahasiswa dan setelah mewawancarai narasumber menerima beberapa poin untuk proyek perpustakaan online ini.

### 3.2. Tahap Define

Tahap *Define* bertujuan untuk menciptakan desain UI/UX yang lebih efektif dan efisien serta menciptakan desain yang lebih sesuai dengan kebutuhan pengguna. Fase ini mengumpulkan data yang diperlukan untuk membuat desain UI/UX yang lebih baik dan lebih mendekati kebutuhan

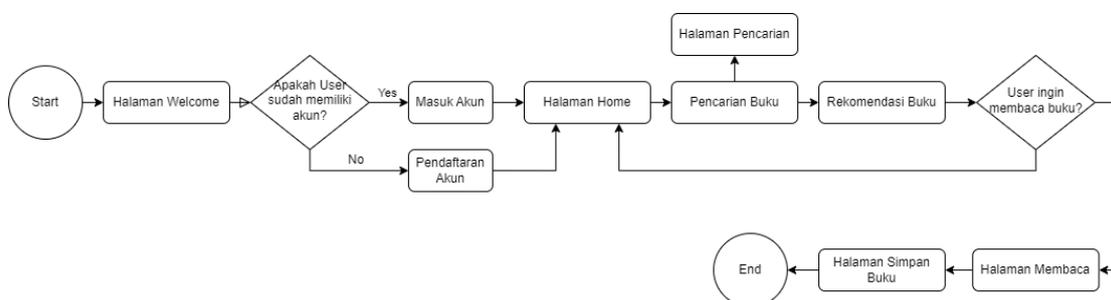
pengguna, seperti mengidentifikasi masalah yang dihadapi pengguna saat menggunakan sistem UI/UX.[11] Berdasarkan hasil wawancara, dimungkinkan untuk melakukan analisis rinci mengenai masalah dan kebutuhan pengguna. Tujuan pembuatan persona pengguna adalah untuk mengidentifikasi dan mengumpulkan informasi tentang pengguna, termasuk perilaku, keinginan, dan masalah yang dihadapi pengguna saat menggunakan sistem UI/UX.[12] *User Persona* adalah karakter fiksi dan memainkan peran penting dalam pengembangan aplikasi karena mewakili pengguna target. Persona pengguna ini tercipta berdasarkan hasil penelitian pada fase empati dan mencerminkan permasalahan, motivasi dan kebutuhan pengguna yang diinginkan.



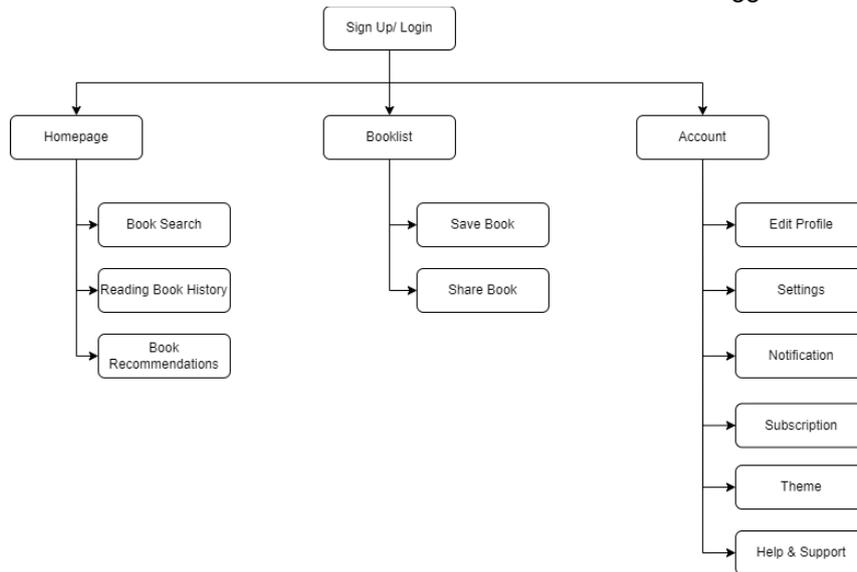
Gambar 2. User persona

### 3.3. Tahap Ideate

Tahapan *Ideate* bertujuan untuk membuat desain UI/UX yang lebih dekat dengan kebutuhan pengguna, serta membuat desain yang lebih dekat dengan kebutuhan pengguna.[13] Di fase *Ideate* ini, pembuatan ide yang beragam dan kreatif untuk memecahkan masalah yang sebelumnya diidentifikasi oleh pengguna. Ide-ide ini kemudian menjadi dasar untuk pengembangan lebih lanjut, seperti pembuatan prototipe dan pengujian konsep. Setelah dilakukan pengumpulan dan observasi terhadap permasalahan pengguna dapat dibuatkan konsep dasar seperti informasi arsitektur dan juga wireframe. Dari informasi arsitektur berisikan alur tampilan dari aplikasi ini secara garis besar, dimana penulis memberikan fitur-fitur mencakup pendaftaran, beranda aplikasi, account settings.



Gambar 3. User Flow

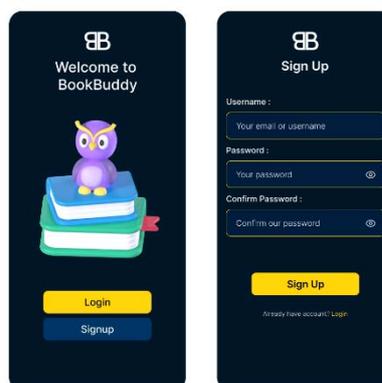


**Gambar 4.** Information Architecture

Dalam gambar 3 menampilkan bagaimana sistem informasi dari aplikasi “Book Buddy” ini secara garis besar. Dimulai dari *Sign Up/ Login* lalu menuju tampilan awal aplikasi dimana terdapat fitur dari *Homepage*, *Booklist* dan *Account*. Didalam ketiga fitur utama itu juga ada fitur-fitur spesifik lainnya dari masing-masing fitur utama. Setelah didapatkan alur informasi dari perancangan aplikasi dapat dibuat tampilan awal *Design Wireframe* dimana menggambarkan fitur-fitur yang telah dirancang sebelumnya. Dalam hal ini *Wireframe* membantu perancangan prototipe karena menggambarkan visual awal dari aplikasi e-book “Book buddy” ini.

### 3.4. Tahap Prototype

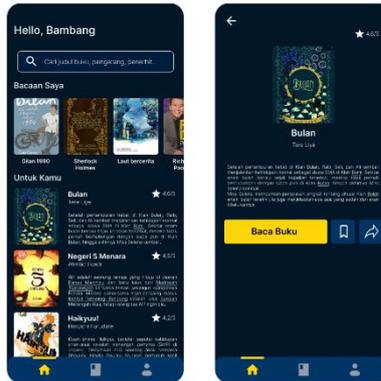
Pada fase *Prototyping*, ide-ide solusi yang muncul pada fase ideate dikumpulkan dan dibuatlah prototipe berdasarkan ide-ide tersebut. Pembuatan prototipe adalah langkah penting untuk menguji dan memvalidasi ide solusi serta menciptakan desain UI/UX yang lebih efektif dan efisien.[14] Dalam tahapan *Prototype* dilakukan penyempurnaan dari design wireframe pada tahapan ideate, pada tahap ini dibuatkan sebuah *Design System* yang mencakup font, warna, gambar, tombol, icon dan elemen design lainnya. Pada tahapan *Prototype* ini merupakan tahapan akhir dari perancangan sebelum dilakukannya tahapan testing, mnaka dari ini tahapan *Prototype* ini sangat penting untuk dilakukan secara detail.



**Gambar 5.** Halaman Welcome screen, Signup

Saat masuk kedalam aplikasi akan disambut dengan *Welcome Screen* yang dimana terdapat logo dari aplikasi, element gambar, tombol untuk *Login* atau *Signup*.

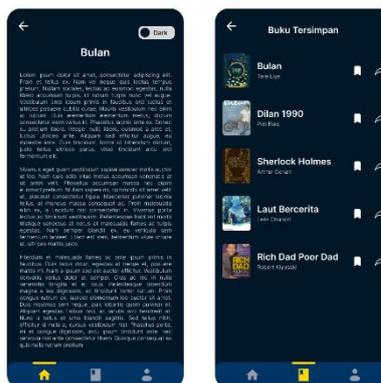
Dari halaman *Welcome Screen* akan diminta untuk melakukan *Login* atau *Signup*. Jika sudah memiliki akun bisa memilih *Login* dan akan diminta memasukkan *username* dan *password*. Jika belum memiliki akun bisa memilih *signup* dan akan diminta memasukkan *username*, *password* dan konfirmasi *password*.



Gambar 6. Halaman Home, Detail book

Pada halaman home disajikan berbagai fitur utama seperti rekomendasi buku, riwayat buku yang telah dibaca dan pencarian buku yang diinginkan.

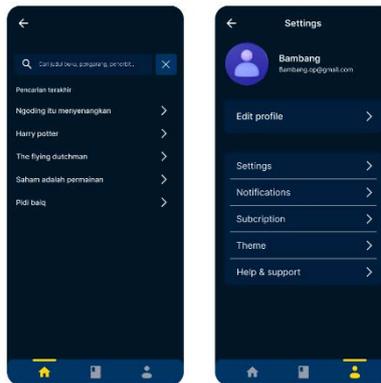
Ketika melakukan klik pada judul buku yang diinginkan akan disajikan tampilan detail dari buku tersebut. Pengguna bisa langsung membaca buku dengan mengklik "Baca Buku" atau menyimpan kedalam *booklist* dan melakukan share buku ke orang banyak



Gambar 7. Halaman Baca, Book list

Bagian membaca akan menyajikan isi dari buku yang dipilih, dan pengguna bisa merubah tampilan dari *dark* ke *light* atau sebaliknya.

Pengguna bisa melihat Kembali Riwayat buku yang telah dibaca, atau buku yang hanya disimpan pada bagian *booklist*



**Gambar 8.** Halaman Search, Account settings

Pada bagian halaman *home* jika memilih search akan disajikan beberapa fitur seperti Riwayat pencarian dan *custom* pencarian. Dimana pengguna dapat melakukan pencarian khusus seperti pencarian berdasarkan tahun terbit, penulis, penerbit, dan lainnya.

Tampilan *account settings* kan menampilkan nama dan email pengguna. Dan pengguna dapat melakukan *edit profile*, *settings*, *notifications*, *subscription*, *theme*, serta *help&support*.

### 3.5. Tahap Test

Metode *System Usability Scale* (SUS) digunakan pada tahap pengujian ini. Metode SUS merupakan metode pengujian usability yang menggunakan skala 10 pertanyaan yang diisi pengguna untuk mengukur kegunaan produk atau sistem.[15] Dengan melakukan pengujian kepada lima responden dimana dengan memberikan scenario dan uji langsung kepada responden. Dari hasil pengujian scenario dengan memberikan 10 pertanyaan dengan menggunakan nilai skala 1-5 dan berdasarkan hasil pengujian dari para respoden.

**Table 1.** Penghitungan SUS

No	Responden	Hasil Perhitungan Responden										Jumlah	Skor SUS
		Q1	Q1	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
1	R1	3	4	2	3	4	3	2	4	3	4	32	80
2	R2	3	4	3	4	4	3	3	2	2	4	32	80
3	R3	2	3	4	3	4	4	4	4	3	4	35	87,5
4	R4	3	3	4	2	4	4	4	2	4	4	34	85
5	R5	4	3	4	3	4	2	2	4	4	4	35	87,5
<b>Nilai rata-rata SUS</b>													<b>84</b>

**Reliability Statistics**

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.140	.152	5

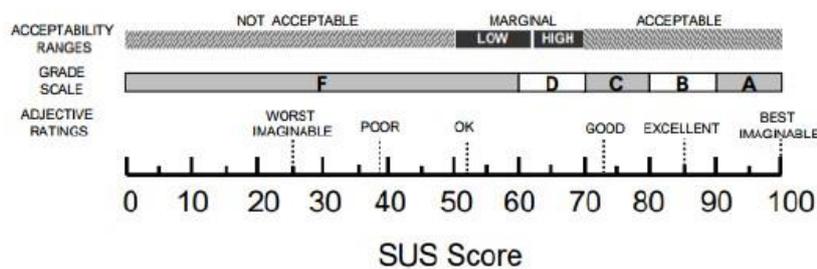
**Gambar 9.** Hasil Uji Reabilitas

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
R1	13.50	2.500	.178	.361	-.024 <sup>a</sup>
R2	13.50	2.722	.085	.177	.087
R3	13.20	2.844	.093	.208	.083
R4	13.30	2.900	-.015	.263	.204
R5	13.30	2.900	-.015	.250	.204

**Gambar 10.** Item Total Statistics

Dilihat dari hasil gambar 9 untuk uji reabilitas pada pada tabel hasil kuisisioner didapatkan hasil 0,140 pada Cronbach alpha dimana nilai alpha yang tinggi menunjukkan reliabilitas yang lebih baik, biasanya dianggap baik jika nilainya di atas 0,70 atau 0,80. Dan pada tabel 1 didapatkan nilai rata-rata hasil pengujian di 84, dengan hasil tersebut dibandingkan dengan kategori penilaian SUS. Skor yang diperoleh dapat dibagi menjadi kategori, seperti *Acceptability Range*, *Grade Scales*, dan *Adjective Ratings*. [16] Nilai SUS dikategorikan menjadi tiga kategori penerimaan pengguna : Not acceptable, Marginal, dan Acceptable. Terdapat enam tingkatan skala, yaitu A,B,C,D,E dan F. Dan juga terdapat peringkat adjektif *Diantaranya Worst Imaginable, Poor, Ok, Good, Excellent, Dan Best Imaginable*. Lebih jelasnya dilihat dari Gambar 11 dibawah.



**Gambar 11.** Kriteria Penilaian SUS

Dilhat dari skala pada Gambar 11 yang merupakan gambaran penilaian SUS dan mengambil nilai rata-rata penilaian SUS pada Tabel 1 penghitungan SUS, dapat diputuskan hasil pengujian produk mendapat hasil **Acceptable**, mendapatkan skala peringkat di **B**, dan penilaian adjektif ada pada **Excellent**. Dapat disimpulkan produk yang dibuat memiliki kualitas yang baik untuk digunakan.

#### 4. Conclusion

Secara keseluruhan, penelitian ini menyoroti pentingnya pendekatan yang berpusat pada pengguna ketika merancang antarmuka pengguna aplikasi seluler. Dengan menggunakan metode *Design Thinking* dan alat evaluasi seperti SUS, dapat membuat aplikasi yang lebih efektif, efisien, dan memuaskan pengguna akhir. Dengan pembuatan *prototype* fitur halaman membaca, *search*, *recommendation*, *booklist* dan detail buku. Dengan hasil menggunakan Metode *System Usability Scale* (SUS) dihasilkan nilai 84, dengan masuk dalam kategori *Acceptable* (pantas), hasil skala peringkat mendapat B, serta hasil Excellent pada peringkat adjektif.

#### References

- [1] I. Anggit Pramesti, "Titian: Jurnal Ilmu Humaniora FAKTOR YANG MEMENGARUHI MINAT DAN CARA MEMBACA MASYARAKAT INDONESIA DI ERA DIGITAL, SERTA DAMPAKNYA PADA BISNIS MEDIA CETAK Factors Influencing Indonesian People's Interests and Reading Ways in the Digital Age, and Its Impact on the Print Media Business," 2021. [Daring]. Tersedia pada: <https://online-journal.unja.ac.id/index.php/titian>
- [2] P. Rahayu, N. Rahmi, dan W. Wicaksana, "Analisis Perbandingan Metode Mobile-D Dan Prototype Pada Pembangunan Aplikasi E-Book Anak Usia Dini," *JURNAL SISTEM*

- INFORMASI BISNIS*, vol. 12, no. 2, hlm. 124–131, Feb 2023, doi: 10.21456/vol12iss2pp124-131.
- [3] A. Voutama dan E. Novalia, “Perancangan Aplikasi M-Magazine Berbasis Android Sebagai Sarana Mading Sekolah Menengah Atas,” vol. 15, no. 1.
- [4] arif Rawhul Ihsan dan N. Rosa Damayanti, “Penulis Korespondensi.” [Daring]. Tersedia pada: <http://www.jurnal.umk.ac.id/sitech>
- [5] C. Ravelino, Y. Alfa Susetyo, dan K. Satya Wacana, “Perancangan UI/UX untuk Aplikasi Bank Jago menggunakan Metode User Centered Design,” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 1, hlm. 2023, 2023, doi: 10.35870/jti.
- [6] P. S. Rosiana, A. Voutama, dan A. A. Ridha, “PERANCANGAN UI/UX SISTEM INFORMASI PEMBELIAN HASIL TANI BERBASIS MOBILE DENGAN METODE DESIGN THINKING,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, Jul 2023, doi: 10.23960/jitet.v11i3.3048.
- [7] F. Aziz, D. Uki, E. Saputri, N. Khasanah, dan T. Hidayat, “Penerapan UI/UX dengan Metode Design Thinking (Studi Kasus: Warung Makan).” [Daring]. Tersedia pada: <http://ejournal.bsi.ac.id/ejournal/index.php/infortech>
- [8] G. Dwi, P. Haryanto, dan A. Voutama, “PERANCANGAN UI/UX SISTEM INFORMASI PENYEWAAN MOBIL BERBASIS MOBILE DENGAN METODE DESIGN THINKING.”
- [9] S. Nursyifa, T. Ridwan, dan A. A. Ridha, “PERANCANGAN UI/UX PROTOTYPE SISTEM PENCATATAN PELANGGAN BERBASIS MOBILE APPLICATION MENGGUNAKAN METODE DESIGN THINKING (STUDI KASUS CV KAHLA GLOBAL PERSADA),” 2023.
- [10] M. Angelica, D. Hidayat, dan A. R. Adriyanto, “PENERAPAN METODE DESIGN THINKING PADA PERANCANGAN UI/UX MOBILE APPS SEBAGAI MEDIA PENDIDIKAN MORAL UNTUK ANAK-ANAK.”
- [11] A. Firdonsyah, Z. Arwananing Tyas, dan L. Ma, “Penerapan Metode Design Thinking Pada Perancangan UI/UX Sistem Informasi Penelitian Mahasiswa Berbasis Web,” 2023.
- [12] E. Rahmat, S. H. Saputra, dan A. C. Frobenius, “IDENTIFIKASI KEBUTUHAN PENGGUNA TUNANETRA UNTUK PLATFORM MOBILE MENGGUNAKAN METODE USER PERSONA-DESIGN THINKING.”
- [13] M. Raffi Fadli, W. S. Wibawanto Program Studi Seni Rupa, J. Seni Rupa, F. Bahasa dan Seni, dan U. Negeri Semarang, “Arty: Journal of Visual Arts USER INTERFACE AND USER EXPERIENCE OF INDOSPORT MOBILE APPLICATIONS USING A USER CENTERED DESIGN APPROACH,” 2020, [Daring]. Tersedia pada: <http://journal.unnes.ac.id/sju/index.php/arti>
- [14] S. Ansori, P. Hendradi, dan S. Nugroho, “Penerapan Metode Design Thinking dalam Perancangan UI/UX Aplikasi Mobile SIPROPMAWA,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, hlm. 1072–1081, Jul 2023, doi: 10.47065/josh.v4i4.3648.
- [15] I. Bagus dkk., “PERANCANGAN ULANG UI/UX WEBSITE BUMDES BATURITI MENGGUNAKAN METODE SUS DAN DESIGN THINKING.”
- [16] R. Deddy, R. Dako, J. T. Elektro, dan W. Ridwan, “Volume 4 Nomor 2 Juli 2022 Pengukuran Usability terhadap Aplikasi Tesadaptif.Net dengan System Usability Scale,” *Jambura Journal of Electrical and Electronics Engineering*, vol. 207, [Daring]. Tersedia pada: <https://tesdaptif.net>.



ISSN



E-ISSN

---