Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Penerapan Algoritma K-Nearest Neighbor untuk Memprediksi Pengunduran Diri Karyawan

Ni Luh Komang Indira Pramesti[a1], Made Agung Raharja, S.Si., M.Cs.[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]indiprames@gmail.com
[2]made.agung@unud.ac.id

Abstrak

Pengunduran diri karyawan adalah situasi di mana seorang karyawan meninggalkan perusahaan atas kehendak karyawan itu sendiri. Jika terlalu sering terjadi, hal ini merupakan salah satu masalah besar bagi perusahaan yang ingin terus berkembang. Oleh karena itu, penting bagi perusahaan untuk mengantisipasi hal tersebut dengan memprediksi kemungkinan karyawan akan berhenti atau tidak.
Prediksi pengunduran diri karyawan dapat dilakukan menggunakan algoritma *K-Nearest Neighbor*. Algoritma ini bekerja dengan cara menghitung jarak dari data uji ke data latih untuk menentukan tetangga terdekat. Evaluasi akhir dilakukan menggunakan *confusion matrix*. Hasil klasifikasi menggunakan *K-Nearest Neighbor* dengan nilai K = 5 menghasilkan nilai *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%.

Kata Kunci: *K-Nearest Neighbor, Pengunduran Diri, Prediksi, Klasifikasi, Karyawan*

1. Pendahuluan

Pengunduran diri karyawan adalah situasi di mana seorang karyawan meninggalkan perusahaan atas kehendak karyawan itu sendiri. Jika terlalu sering terjadi, hal ini merupakan salah satu masalah besar bagi perusahaan yang ingin terus berkembang karena biaya dan waktu yang dibutuhkan untuk mencari kandidat baru yang sesuai dengan standar perusahaan tidaklah sedikit [1]. Dalam beberapa kasus, terdapat karyawan yang mengundurkan diri tanpa memberitahu perusahaan terlebih dahulu. Hal ini dapat berujung pada kerugian bagi perusahaan dalam bentuk penurunan produktivitas dan efisiensi kerja [2]. Oleh karena itu, penting bagi perusahaan untuk mengantisipasi hal tersebut dengan memprediksi kemungkinan karyawan akan berhenti atau tidak. Dengan mengetahui kemungkinan tersebut, perusahaan dapat mulai mencari kandidat baru lebih awal tanpa khawatir terhadap penurunan produktivitas.

Prediksi tersebut dapat dilakukan dengan menerapkan algoritma *machine learning*, salah satunya adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan salah satu algoritma yang banyak digunakan dalam kasus memprediksi sesuatu. Algoritma ini termasuk ke dalam algoritma *supervised learning*. Prinsip kerjanya adalah data yang diuji akan termasuk ke dalam label terbanyak yang terdekat dari data uji [3].

Penelitian serupa sebelumnya pernah dilakukan dengan judul Prediksi Pengunduran Diri Karyawan Perusahaan "Y" Menggunakan *Random Forest* [4]. Penelitian tersebut menghasilkan nilai akurasi sebesar 87% dan *error* sebesar 13%. Dalam penelitian ini, akan dilakukan prediksi pengunduran diri karyawan menggunakan algoritma *K-Nearest Neighbor*.

2. Metode Penelitian

Proses penelitian dimulai dari pengumpulan data, *preprocessing* data, pembagian data, implementasi algoritma *K-Nearest Neighbor*, dan diakhiri dengan evaluasi menggunakan *confusion matrix*.

ISSN: 2301-5373
E-ISSN: 2654-5101
Volume 12 • Number 3 • February 2024

JELIKU

Jurnal Elektronik Ilmu Komputer Udayana

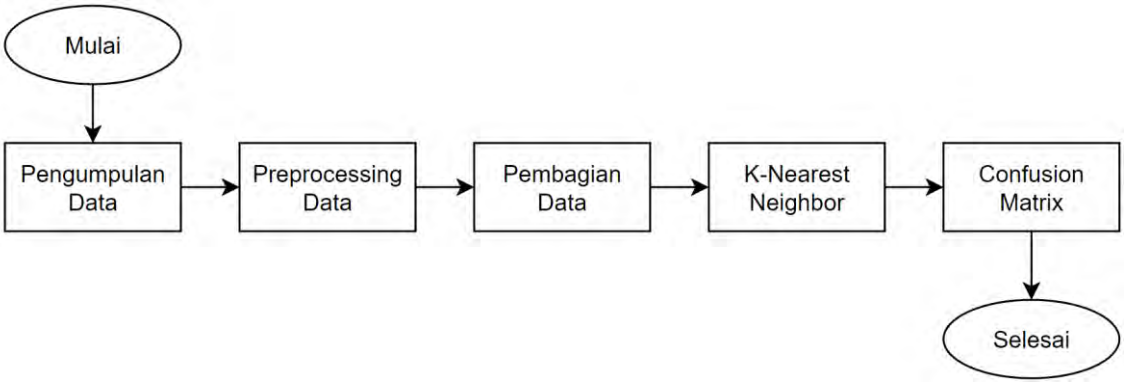
Informatics Study Program
Faculty of Mathematics and Natural Sciences
Udayana University

Table of Contents

This page is intentionally left blank.

SUSUNAN DEWAN REDAKSI
JURNAL ELEKTRONIK ILMU KOMPUTER
UDAYANA (JELIKU)

This page is intentionally left blank.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data pengunduran diri karyawan yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle. Data ini terdiri dari 8 atribut (Gender, Satisfaction, Business Travel, Department, EducationField, Salary, Home-Office, Attrition) dari 100 karyawan di mana sebanyak 18 karyawan meninggalkan perusahaan dan 82 karyawan memutuskan untuk tetap berada di perusahaan. Deskripsi dari masing-masing atribut dapat dilihat di tabel 1.

Tabel 1. Atribut Data

Atribut	Deskripsi	Nilai kategori
Gender	Gender pegawai	- Male - Female
Satisfaction	Tingkat kepuasan pegawai terhadap lingkungan kerja dan perusahaan	- High - Medium - Low
Business Travel	Seberapa sering karyawan melakukan perjalanan bisnis	- Rare - Frequent - No
Department	Departemen pegawai	- R&D - Sales
EducationField	Kualifikasi bidang pendidikan	- Engineering - Medical - Marketing - Other - Technical Degree
Salary	Golongan gaji karyawan	- High - Medium - Low
Home-Office	Jarak dari rumah ke kantor	- Near - Far
Attrition	Pegawai meninggalkan perusahaan atau tidak	- Yes - No

2.2 *Preprocessing* Data

Sebelum dapat digunakan, data sebaiknya dibersihkan terlebih dahulu. Pembersihan data pengunduran diri karyawan meliputi menghilangkan data yang memiliki nilai kosong dan mengubah data kategoris menjadi numerik agar lebih mudah diinterpretasikan oleh model *machine learning*.

2.3 Pembagian Data

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Setelah data dibersihkan, langkah selanjutnya yang dilakukan adalah memisahkan data menjadi 2, yaitu data latih dan data uji. Data latih terdapat sebanyak 80% dari seluruh data dan data uji merupakan 20% sisanya.

2.4 K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma klasifikasi yang termasuk ke dalam *supervised learning*. Cara kerjanya adalah dengan menghitung jarak dari data uji ke data latih untuk menentukan tetangga terdekat. Setelah itu, kategori mayoritas tetangga terdekat merupakan prediksi dari data uji tersebut. Jarak antar tetangga dapat dihitung menggunakan rumus *Euclidean* [5].

$$D(p,q) = \sqrt{\sum_{i=1}^{n} (pi - qi)^2} \quad (1)$$

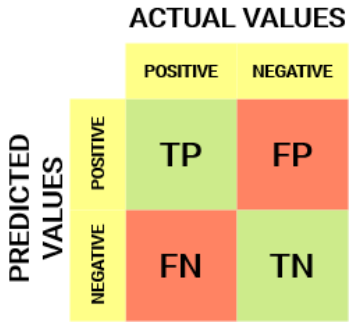
Keterangan:
D = jarak antara titik
p = data latih
q = data uji
i = nilai atribut
n = dimensi atribut

Selengkapnya, berikut langkah-langkah menggunakan algoritma KNN [5]:

a. Menentukan parameter K (jumlah tetangga paling dekat).
b. Menghitung jarak antara data uji dan data latih menggunakan rumus *Euclidean*.
c. Kemudian mengurutkan jarak yang terbentuk.
d. Menentukan jarak terdekat sampai urutan K.
e. Memasangkan kelas yang bersesuaian.
f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

2.5 *Confusion Matrix*

Confusion matrix merupakan metode evaluasi yang banyak digunakan untuk mengukur kinerja suatu algoritma klasifikasi. *Confusion matrix* mengandung informasi perbandingan hasil klasifikasi yang diprediksi sistem dan hasil yang seharusnya. Pada pengukuran ini, terdapat empat komponen yang dimiliki, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) [6]. Seperti yang dapat dilihat pada gambar 2 [7], kolom menunjukkan hasil klasifikasi yang sebenarnya dan baris menunjukkan hasil yang diprediksi oleh sistem.



Gambar 2. Confusion Matrix

Dari *confusion matrix* tersebut, dapat dihitung nilai *accuracy*, *precision*, dan *recall* menggunakan rumus berikut [8]:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$recall = \frac{TP}{TP+FN} \quad (4)$$

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data pengunduran diri karyawan yang terdiri dari 100 baris data dan 8 atribut. Karena tidak ada data dengan nilai yang kosong, maka semua data tersebut bisa digunakan. Kemudian, data yang aslinya memiliki nilai kategoris pada semua kolom diubah menjadi nilai integer agar lebih mudah diproses. Perbedaan data sebelum dan sesudah melalui *preprocessing* dapat dilihat pada tabel 2 dan 3.

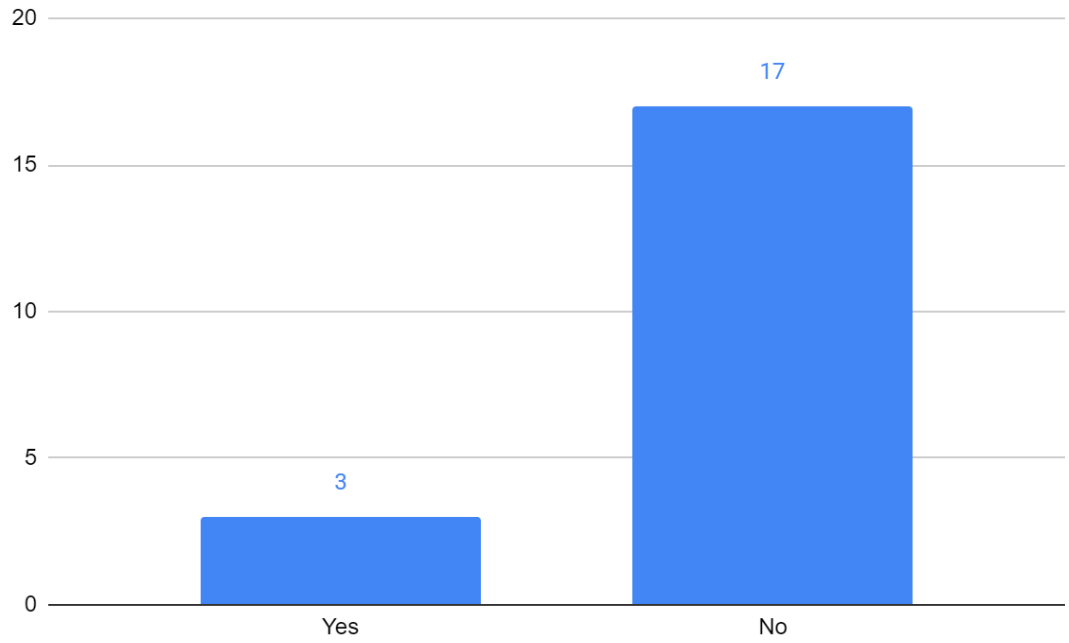
Tabel 2. Data Sebelum *Preprocessing*

Gender	Satisfaction	Business Travel	Department	EducationField	Salary	Home - Office	Attrition
Female	High	Rare	Sales	Engineering	Medium	Near	Yes
Male	Low	Frequent	Sales	Engineering	Low	Near	Yes
Male	Medium	Rare	R&D	Other	Medium	Far	No
Female	Low	Frequent	R&D	Engineering	Medium	Far	Yes
Male	Medium	Rare	R&D	Medical	Medium	Far	No

Tabel 3. Data Setelah *Preprocessing*

Gender	Satisfaction	Business Travel	Department	EducationField	Salary	Home-Office	Attrition
0	3	1	1	0	2	1	1
1	1	2	1	0	1	1	1
1	2	1	0	3	2	2	0
0	1	2	0	0	2	2	1
1	2	1	0	1	2	2	0

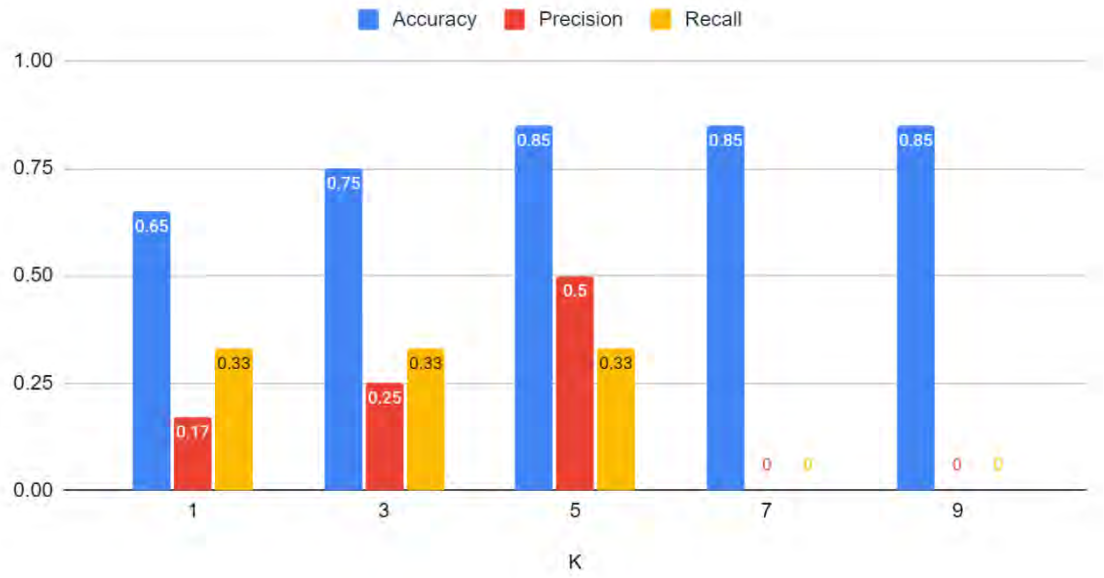
Sebelum mengimplementasikan algoritma KNN, data dibagi menjadi dua, yaitu data latih sebanyak 80% dan data uji sebanyak 20%. Data uji terdiri dari 3 data karyawan yang meninggalkan perusahaan dan 17 data karyawan yang tidak meninggalkan perusahaan seperti yang ditunjukkan pada gambar 3.



Gambar 3. Perbandingan Pengunduran Diri Karyawan pada Data Uji

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pemodelan *K-Nearest Neighbor* dilakukan sebanyak 5 kali dengan nilai K yang berbeda. Nilai K ditentukan dengan menguji bilangan ganjil dari 1 sampai 9 kemudian membandingkan hasil metriknya yaitu *accuracy*, *precision*, dan *recall*. Bilangan genap tidak digunakan untuk menghindari hasil jumlah tetangga yang seri. Perbandingan hasil metrik untuk masing-masing nilai K dapat dilihat pada gambar 4.



Gambar 4. Perbandingan Metrik Nilai K

Berdasarkan grafik tersebut, pemodelan terbaik diperoleh dengan menggunakan K = 5 karena memiliki nilai *accuracy*, *precision*, dan *recall* tertinggi dibandingkan nilai K lainnya. Dari 20 data yang diuji, diperoleh *confusion matrix* yang terdiri dari nilai *True Positive* (TP) sebanyak 1, *False Positive* (FP) sebanyak 1, *False Negative* (FN) sebanyak 2, dan *True Negative* (TN) sebanyak 16. Berdasarkan *confusion matrix* tersebut, diperoleh hasil *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%.

4. Kesimpulan

Dalam penelitian ini, prediksi pengunduran diri karyawan dilakukan menggunakan algoritma *K-Nearest Neighbor* dengan nilai K = 5. Kinerja algoritma dinilai menggunakan *confusion matrix* untuk menghitung nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Dari *confusion matrix* tersebut dapat dihitung nilai *accuracy*, *precision*, dan *recall*. Diperoleh hasil akhir yaitu nilai *accuracy* sebesar 85%, *precision* sebesar 50%, dan *recall* sebesar 33,33%. Nilai *accuracy* yang lumayan tinggi tersebut menandakan bahwa model yang telah dibuat sudah cukup baik dalam memprediksi data secara keseluruhan. Namun, nilai *precision* dan *recall* yang rendah menunjukkan bahwa model tidak bisa memprediksi kelas positif, dalam hal ini karyawan yang mengundurkan diri, dengan baik.

Diharapkan penelitian selanjutnya dapat mengembangkan algoritma yang lebih baik untuk memprediksi pengunduran diri karyawan agar dapat memprediksinya dengan lebih akurat dan tepat.

Daftar Pustaka

[1] N. N. Rahma, “Pengunduran Diri Karyawan Massal Pasca-Lebaran Diperkirakan Meningkat! Ini Saran EngageRocket,” 2022. https://wartaekonomi.co.id/read407407/pengunduran-diri-karyawan-massal-pasca-lebaran-diperkirakan-meningkat-ini-saran-engagerocket (accessed Oct. 02, 2022).

[2] E. I. Larasati, "KERUGIAN PERUSAHAAN AKIBAT PENGUNDURAN DIRI PEKERJA WAKTU TERTENTU TANPA ADANYA PEMBERITAHUAN KEPADA PERUSAHAAN," *Jurist-Diction*, vol. 2, no. 1, pp. 112–130, 2019.

[3] N. K. S. P. Rahayu and I. K. A. Mogi, "Implementation of K-Nearest Neighbor Algorithm in Heart Disease Classification," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 10, no. 1, pp. 39–44, 2021.

[4] D. D. E. Manurung, F. Sandi, F. Akbardipura, H. Ashfahan, and D. S. Prasvita, "Prediksi Pengunduran Diri Karyawan Perusahaan 'Y' Menggunakan Random Forest," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Sep. 2021, pp. 202–213.

[5] D. Cahyantia, A. Rahmayania, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, Jul. 2020.

[6] Karsito and S. Susanti, "PENGAJUAN KREDIT RUMAH DENGAN ALGORITMA NAÏVE BAYES DI PERUMAHAN AZZURA RESIDENCIA," *SIGMA – Jurnal Teknologi Pelita Bangsa*, vol. 9, no. 3, pp. 43–48, 2019.

[7] A. Bhandari, "Everything you Should Know about Confusion Matrix for Machine Learning," 2020. https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/ (accessed Oct. 03, 2022).

[8] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Metode AHP Pada Sistem Pendukung Keputusan Gaji Bonus Karyawan di PT Sadhana Adiwidya Bhuana

I Made Rian Wijaya[a1], I Gusti Ngurah Anom Cahyadi Putra[a2]

[a]Program Studi Informatika
Universitas Udayana
Kuta Selatan, Badung, Bali, Indonesia
[1]rianwijaya1352@gmail.com
[2]anom.cp@unud.ac.id

Abstract

PT Sadhana Adiwidya Bhuana is a company engaged in business and management consulting services which still applies manual methods in giving employee salary bonuses. Writer develop a Decision Support System (DSS) to assist in the provision of employee bonus salaries, using the Analytical Hierarchy Process (AHP) method to determine the weight criteria. There are eight criteria used in employee assessment, including Key Performance Indicators (KPI), Data Analysis, Critical Thinking, Creative and Innovative, Communication, Cooperation, Attitude, and Attendance. Then the award is given in four forms, namely less, sufficient, good, and very good, with scores of 1, 2, 3, and 4. The results of the weighting of the criteria with the AHP method obtained key performance indicators (KPI) scores 0.330, data analysis score 0.183, critical thinking score 0.125, creative and innovative score 0.119, communication score 0.063, cooperation score 0.063, attitude score 0.034, and attendance score 0.031. The weight will be a reference in calculating the performance score of the employee who will determine the bonus salary. This system has been tested using blackbox testing, with the test results being appropriate.

Keywords: *decision support system, analytical hierarchy process, bonus salary, weighting, performance score*

1. Pendahuluan

Perusahaan adalah organisasi atau kelompok yang didirikan oleh orang perseorangan atau kelompok orang untuk melakukan kegiatan produktif yang melibatkan unsur manusia, alam, dan modal. Dalam suatu perusahaan terdapat karyawan yang menjadi penggerak perusahaan dan pemain utama sistem produksi perusahaan. Karyawan adalah salah satu aset perusahaan yang paling penting dan kekuatan pendorong di belakang produktivitas perusahaan, pengembangan perusahaan, persaingan, dan keberlanjutan laba. [1]. Dalam perusahaan karyawan berhak untuk mendapatkan gaji dari hasil kerja seorang pegawai atau karyawan. Perhitungan gaji pada suatu perusahaan biasanya berbeda-beda sesuai denga ketentuan perusahaan. Secara umum perusahaan memiliki dua jenis gaji yaitu gaji pokok dan gaji bonus. Gaji pokok merupakan gaji yang akan diterima oleh seorang karyawan sesaui kontrak kerja yang sudah disepakati. Kemudian gaji bonus merupakan gaji yang diberikan kepada karyawan sebagai bentuk apresiasi terhadap hasil kerjanya yang biasanya dirancang berdasarkan performa karyawan dalam suatu periode tertentu. Untuk menentukan bonus gaji karyawan dilakukan penilaian berdasarkan kriteria-kriteria tertentu kemudian dari hasil penilaian tersebut akan dilakukan perhitungan untuk mendapatkan bonus gaji karyawan [2]. Dalam penentuan bonus gaji dapat dilakukan secara manual akan tetapi jumlah karyawan yang banyak tentu diperlukan otomasi dalam perhitungannya sehingga dapat menggunakan Sistem Pendukung Keputusan (SPK).

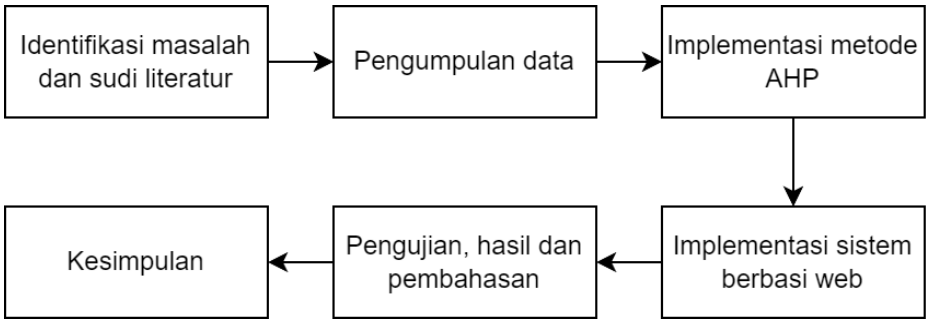
PT Sadhana Adiwidya Bhuana merupakan perusahaan yang berkonsentrasi pada bidang jasa konsultansi bisnis dan management yang berkomitmen penuh dalam memberikan dukungan ilmu pengetahuan dan metodologi, melalui konsultasi manajemen, *assessment & recruitment, business coaching, training & development*, jasa akuntansi dan audit, serta business development. Perusahaan ini memiliki jumlah karyawan yang lumayan banyak yaitu lebih dari 30 karyawan. PT Sadhana Adiwidya Bhuana menerapkan juga sistem gaji bonus yang diberikan kepada karyawannya dalam periode tahunan. Gaji bonus diberikan kepada karyawan setiap tahunnya jika laba yang diperoleh sudah melebihi target perusahaan. Dalam penentuan gaji bonus untuk setiap karyawan PT Sadhana Adiwidya

Bhuana masih memberlakukan sistem secara manual dalam menentukan karyawan yang memiliki potensi untuk mendapatkan gaji bonus tersebut.

Solusi yang penulis berikan disini adalah sebuah sistem pendukung keputusan yang mampu mendukung dalam menentukan gaji bonus karyawan PT Sadhana Adiwidya Bhuana yang dimana sistem ini tidak serta merta menggantikan peran HRD dalam pengambilan keputusan, tetapi sistem ini dapat membantu mendukung seorang HRD dalam pengambilan keputusan terhadap pemberian hak gaji bonus karyawan. Metode yang diimplementasikan dalam percobaan ini adalah metode Analytical Hierarchy Process (AHP) dalam menentukan bobot dari setiap kriteria. Kemudian dari penilaian yang diberikan berdasarkan kriteria dilakukan perhitungan untuk setiap karyawan untuk mendapatkan performance score dari setiap karyawan yang akan menjadi acuan dalam pemberian gaji bonus.

2. Metode Penelitian

Metode penelitian yang digunakan yaitu melakukan simulasi penerapan metode AHP dan implementasi sistem. Flow dari penelitian dapat dilihat pada gambar 1.



Gambar 1. Diagram alur penelitian

2.1. Studi Literatur

Sistem Pendukung Keputusan (SPK) adalah sebuah konsep atau metode yang dibuat untuk menghasilkan hasil atau keputusan dalam kondisi tertentu. SPK juga dapat digunakan sebagai alat bagi pengambil keputusan untuk merekomendasikan keputusan yang tepat berdasarkan studi metode tertentu, sehingga mengungkapkan keputusan alternatif yang lebih beragam dan relevan [3]. Salah satu metode yang dapat diimplementasikan dalam SPK metode Analytical Hierarchy Process (AHP). Metode Analytical Hierarchy Process (AHP) merupakan metode yang bertujuan untuk membantu dalam pengambil keputusan dengan menerapkan konsep perbandingan berpasang yang menggabungkan faktor kualitatif dan kuantitatif dalam masalah yang kompleks. Metode AHP dapat memberikan solusi atau hasil dari berbagai kondisi yang saling berlawanan, karena hal tersebut penggunaan AHP di berbagai bidang meningkat pesat dan menjadi populer. Dalam prosesnya AHP menggunakan analisis hierarki dalam tahapnya untuk menyelesaikan masalah meliputi dekomposisi, evaluasi komparatif, sintesis prioritas, dan konsistensi logis [4].

2.2. Data dan Pengumpulan Data

Dalam penelitian ini penulis menggunakan data latih dan data uji dalam pengimplementasian dan pengujian metode AHP dalam SPK. Pengumpulan data dilakukan dengan metode wawancara dengan narasumbernya yaitu HRD dari PT Sadhana Adiwidya Bhuana.

2.3. Implementasi Analytical Hierarchy Process (AHP) dan Performance Score

Nilai perbandingan berpasangan dari matriks yang dibuat dari hasil wawancara akan diubah sesuai dengan Skala Saaty. Skala Saaty ditunjukkan pada Tabel 1.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 1. Skala Saaty

Skala	Keterangan
1	Sama penting
3	Cukup penting
5	Lebih penting
7	Sangat penting
9	Mutlak penting
2, 4, 6, 8	Nilai-nilai antar dua nilai yang berdekatan

Pada bagian ini termasuk dalam analisa data yang akan melakukan perhitungan data-data yang kemudian hasilnya akan digunakan dalam metode *AHP* untuk menentukan bobot kriteria dari masing-masing alternatif. Berikut merupakan langkah-langkahnya:

1. Menentukan rencana atau tujuan dari analisa kemudian mendefinisikannya menjadi penyelesaian masalah.
2. Menentukan klasifikasi elemen perbandingan pasangan dengan membandingkan elemen sesuai dengan kriteria yang diberikan. Kemudian menentukan matriks perbandingan berpasangan dengan memberikan elemen berdasarkan skala Saaty.
3. Menormalisasi matriks perbandingan berpasangan dengan rumus (1).

$$\underline{a}_{lk}=\frac{a_k}{\sum_{l=1}^{m} a_{lk}} \quad (1)$$

Keterangan:
$\underline{a}_{lk}$ = Nilai matriks normalisasi
a_k = Nilai elemen
a_{lk} = Nilai masing-masing elemen dalam satu kriteria
m = Banyak kriteria

4. Nilai-nilai pada setiap baris dijumlahkan dan dibagi dengan berapa banyak kriteria yang kemudian mendapatkan nilai bobot kriteria, nilai tersebut dapat diperoleh dengan menggunakan rumus (2).

$$w_l = \frac{\sum_{k=1}^{m} \underline{a}_{lk}}{m} \quad (2)$$

Keterangan:
w_l = Nilai bobot prioritas
m = Jumlah kriteria
$\underline{a}_{lk}$ = Nilai matriks normalisasi
m = Banyak kriteria

5. Pengujian tingkat konsistensi terhadap nilai-nilai yang diberikan dalam matriks perbandingan berpasangan menggunakan perhitungan berikut:

$$CI = (\lambda_{max} - n) / (n-1) \quad (3)$$

Keterangan:
CI = Indeks konsistensi
n = Ukuran matriks
λ_{max} = Nilai eigen maksimum

Nilai eigen dari tiap kriteria dapat ditentukan dengan rumus (4)

$$\lambda = \frac{\sum_{l=1}^{m} a_{lk}}{w_k} \quad (4)$$

Selanjutnya menentukan rasio konsistensi dengan menggunakan rumus (5).

$$CR = \frac{CI}{RI} \quad (5)$$

Keterangan:

CR = Rasio konsistensi

CI = Indeks konsistensi

RI = Random indeks

Random Indeks yang digunakan penulis adalah Alonso-Lamata RI Values dapat dilihat pada tabel 2.

Tabel 2. Alonso-Lamata RI Values

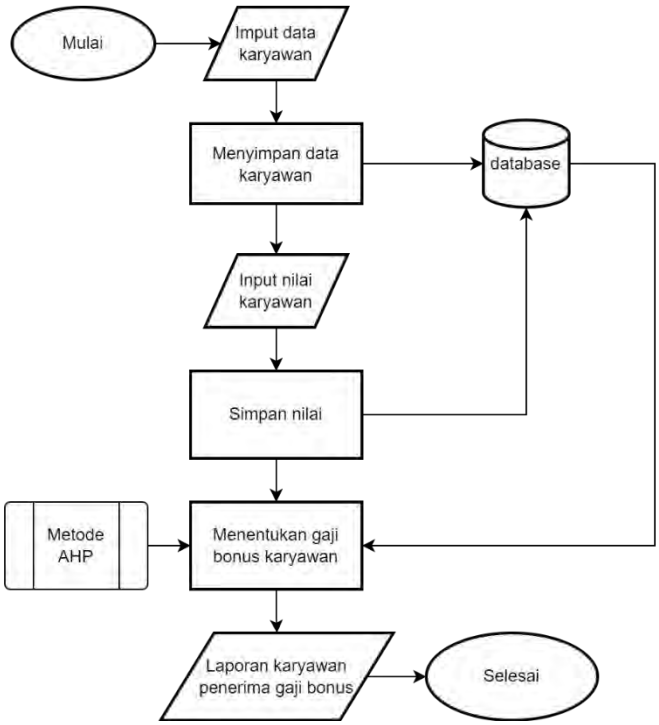
Nilai Matriks	Alonso-Lamata RI Values
3	0,5245
4	0,8815
5	1,1086
6	1,2479
7	1,3417
8	1,4056

6. Nilai matriks akan dinilai berdasarkan konsistensi nya jika nilai CR kurang dari 10% atau kurang dari 0,1. Jika lebih dari 0,1 maka dinyatakan tidak konsisten dan diperlukan pembuatan ulang matriks.

Setelah mendapatkan hasil bobot dari tiap-tiap kriteria selanjutnya menentukan *performance score* dari setiap karyawan. Terlebih dahulu adalah menentukan range penilaian tiap kriteria dengan range nilai yang disepakati bisa disebut (N_i). Kemudian tiap nilai kriteria harus di normalisasi dengan membagi nilai dengan nilai maksimal dalam range tersebut atau ($max\ (N_i)$) sehingga nilai yang sudah dinormalisasi bisa dituliskan dengan $\underline{N_i} = \frac{N_i}{max\ (N_i)}$. Selanjutnya $\underline{N_i}$ dari setiap kriteria dikalikan dengan bobot dari kriterianya, yang selanjutnya hasilnya akan dijumlahkan semuanya dari tiap kriteria yang hasil penjumlahannya merupakan nilai dari performance score satu karyawan.

2.4. Desain Sistem dan Database

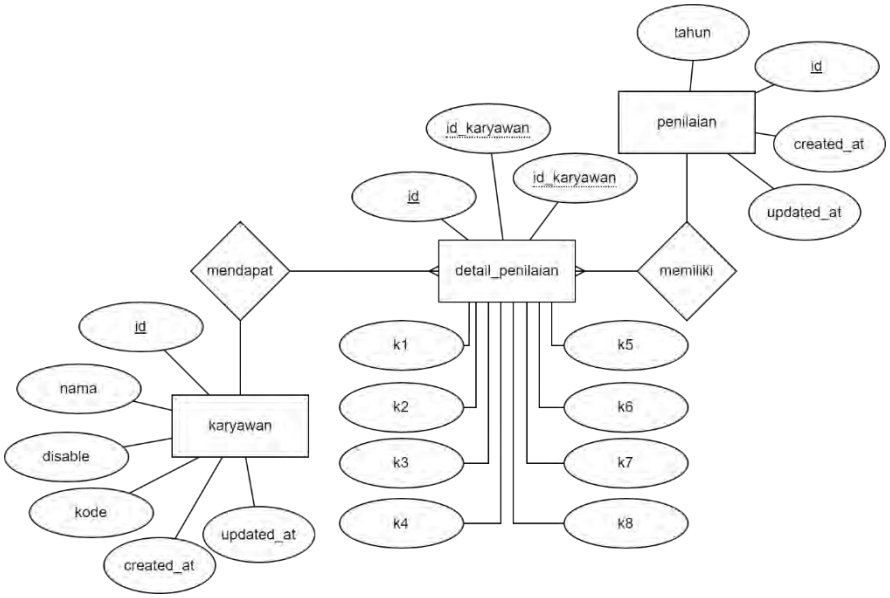
Desain sistem diilustrasikan dalam bentuk flowchart pada gambar 2.



Gambar 2. Flowchart alur sistem

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Dalam sistem ini admin atau HRD diperlukan untuk mengisi data karyawan yang akan dinilai. Kemudian admin memasukkan penilaian terhadap tiap karyawan berdasarkan kriteria-kriteria penilaian. Hasil tersebut disimpan ke dalam sebuah database dengan rancangan database atau ERD dapat dilihat pada gambar 3. Untuk memperoleh karyawan yang memiliki potensi mendapatkan gaji bonus disini menggunakan metode AHP dengan datanya diambil dari database.



Gambar 3. Desain database (ERD)

2.5. Pengujian Sistem

Pengujian sistem dilakukan dengan metode blackbox testing. Blackbox testing atau pengujian kotak hitam adalah pengujian yang dilakukan untuk mengamati kesesuaian fungsional dari masukan dan luaran suatu sistem. Dengan blackbox testing nantinya akan mendeteksi masalah sistem seperti kesalahan fungsional, kesalahan antarmuka, dan kesalahan deklarasi [5]. Blackbox testing yang dilakukan adalah dengan membuat pertanyaan atau pernyataan mengenai fungsional atau ekspektasi terhadap suatu proses dalam sistem yang hasil penilaiannya dibedakan menjadi dua yaitu sesuai dan tidak sesuai.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang diperlukan disini adalah data mengenai apa saja kriteria yang digunakan PT Sadhana Adiwidya Bhuana dalam melakukan penilaian terhadap karyawannya, yang akan menjadi tolak ukur dalam mendapatkan referensi untuk menentukan keputusan. Hasilnya terlihat pada tabel 3 merupakan data kriteria yang sudah disepakati dalam penelitian ini.

Tabel 3. Kriteria

Kode	Nama Kriteria
K1	Key Performance Indicator (KPI)
K2	Analisis data
K3	Berpikir kritis
K4	Kreatif dan inovatif
K5	Komunikasi
K6	Kerjasama
K7	Sikap
K8	Absensi

Selanjutnya adalah data nilai prioritas kriteria yang sudah disepakati dapat dilihat di tabel 4 berikut.

Tabel 4. Nilai prioritas perbandingan berpasang

Kriteria	K1	K2	K3	K4	K5	K6	K7	K8
K1	1	3	3	3	5	5	6	6
K2	1/3	1	2	2	3	3	4	5
K3	1/3	1/2	1	1	1/2	4	4	5
K4	1/3	1/2	1	1	1	3	4	4
K5	1/5	1/3	2	1	1	2	3	4
K6	1/5	1/3	1/4	1/3	1/2	1	3	3
K7	1/6	1/4	1/4	1/4	1/3	1/3	1	1
K8	1/6	1/5	1/5	1/4	1/4	1/3	1	1

3.2. Pembobotan Kriteria Dengan AHP

Dari data yang sudah didapatkan selanjutnya adalah melakukan langkah-langkah dalam pengimplementasian metode AHP dapat dilihat sebagai berikut.

a. Menormalisasi matriks perbandingan berpasangan.

Tabel 5. Matriks Perbandingan Berpasangan

Kriteria	K1	K2	K3	K4	K5	K6	K7	K8
K1	1	3	3	3	5	5	6	6
K2	0,333	1	2	2	3	3	4	5
K3	0,33	0,5	1	1	0,5	4	4	5
K4	3	0,5	1	1	1	3	4	4
K5	0,33	0,333	2	1	1	2	3	4
K6	3	0,333	0,2	0,33	0,5	1	3	3
K7	0,2	0,25	5	3	0,333	0,333	1	1
	0,2		0,2	0,25				
	0,166		5					
	6							
K8	0,166	0,2	0,2	0,25	0,25	0,333	1	1
	6							
sum	2,733	6,116	9,7	8,833	11,583	18,666	26	29

Untuk mendapatkan nilai matriks normalisasi menggunakan rumus (1).

$\sum_{l=1}^{m} a_{lk}$= sum – tabel 5

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$\underline{a}_{11} = \frac{1}{2,733} = 0,365$
$\underline{a}_{12} = \frac{3}{6,116} = 0,490$
$\underline{a}_{13} = \frac{3}{9,7} = 0,309$
$\underline{a}_{14} = \frac{3}{8,833} = 0,339$
$\underline{a}_{15} = \frac{5}{11,583} = 0,431$
$\underline{a}_{16} = \frac{5}{18,666} = 0,267$
$\underline{a}_{17} = \frac{1}{26} = 0,230$
$\underline{a}_{18} = \frac{8}{29} = 0,206$

Dari perhitungan diatas diperoleh nilai matriks normalisasi pada baris 1. Kemudian dilakukan perulangan pada baris selanjutnya sehingga menghasilkan matriks normalisasi perbandingan berpasangan yang hasilnya pada tabel 6.

Tabel 6. Hasil Normalisasi Matriks Perbandingan Berpasangan

Kriteria	K1	K2	K3	K4	K5	K6	K7	K8	sum
K1	0,365	0,490	0,309	0,339	0,431	0,267	0,230	0,206	2,642
K2	0,121	0,163	0,206	0,226	0,258	0,160	0,153	0,172	1,464
K3	0,121	0,081	0,103	0,113	0,043	0,214	0,153	0,172	1,003
K4	0,121	0,081	0,103	0,113	0,086	0,160	0,153	0,137	0,958
K5	0,121	0,054	0,103	0,113	0,086	0,107	0,153	0,137	0,893
K6	0,121	0,054	0,103	0,113	0,043	0,053	0,153	0,103	0,506
K7	0,073	0,040	0,206	0,113	0,028	0,017	0,115	0,034	0,275
	0,073		0,025	0,037			0,115		
	0,060		0,025	0,028			0,038		
K8	0,060	0,032	0,206	0,028	0,021	0,017	0,038	0,034	0,254

b. Menentukan Bobot Kriteria
Setelah menentukan matriks normalisasi dari tabel perbandingan berpasang selanjutnya menentukan bobot tiap kriteria dengan menggunakan rumus (2)

$\sum_{k=1}^{m} \underline{a}_{lk}$ = sum - tabel 6

$w_1 = \frac{2,642}{12} = 0,330$
$w_2 = \frac{1,464}{12} = 0,183$
$w_3 = \frac{1,003}{12} = 0,125$

$$w_4 = \frac{0{,}958}{12} = 0{,}119$$
$$w_5 = \frac{0{,}893}{12} = 0{,}111$$
$$w_6 = \frac{0{,}506}{12} = 0{,}063$$
$$w_7 = \frac{0{,}275}{12} = 0{,}034$$
$$w_8 = \frac{0{,}254}{12} = 0{,}031$$

Tabel 7. Bobot Kriteria

Kode Kriteria	Bobot (w_k)	Nama Kriteria
Q1	0,330	Key Performance Indicator (KPI)
Q2	0,183	Analisis data
Q3	0,125	Berpikir kritis
Q4	0,119	Kreatif dan inovatif
Q5	0,111	Komunikasi
Q6	0,063	Kerjasama
Q7	0,034	Sikap
Q8	0,031	Absensi

s

c. Melakukan Pengujian Konsistensi Matriks Perbandingan Berpasangan
Untuk melakukan pengujian konsistensi terlebih dahulu harus menentukan indeks konsistensi. Untuk menentukan indeks konsistensi terlebih dahulu harus menentukan nilai eigen dari setiap kriteria dengan menggunakan rumus (4).

$$\sum_{l=1}^{m} \quad a_{lk} = sum_k - tabel\ 5$$

$$\lambda_1 = \frac{2{,}733}{0{,}330} = 0{,}902$$
$$\lambda_2 = \frac{6{,}116}{0{,}183} = 1{,}119$$
$$\lambda_3 = \frac{9{,}7}{0{,}125} = 1{,}216$$
$$\lambda_4 = \frac{8{,}833}{0{,}119} = 1{,}058$$
$$\lambda_5 = \frac{11{,}583}{0{,}111} = 1{,}294$$
$$\lambda_6 = \frac{18{,}666}{0{,}063} = 1{,}182$$
$$\lambda_7 = \frac{26}{0{,}034} = 0{,}895$$
$$\lambda_8 = \frac{29}{0{,}031} = 0{,}924$$

$$\lambda_{max} = \sum_{l=1}^{m} \quad \lambda_l = 8{,}495$$

Kemudian masukkan nilai eigen maksimum dalam rumus (3) untuk mendapatkan indeks konsistensi.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$$CI = \frac{(8{,}495 - 8)}{(8 - 1)} = 0{,}084$$

Kemudian selanjutnya kita dapat menentukan rasio konsistensi dengan rumus (5)
RI = nilai dari tabel 2 dengan nilai pada nilai matriks 8 (banyak kriteria)

$$CR = \frac{0{,}084}{1{,}4056} = 0{,}060$$

Diperoleh rasio konsistensi (CR) dari matriks perbandingan berpasangan adalah 0,06. Dari hasil tersebut dapat menunjukkan bahwa perhitungan metode AHP dianggap konsisten karena sudah memenuhi syarat pengujian yaitu CR < 0,1.

d. Menentukan Performance Score Karyawan
Setelah mendapatkan nilai bobot setiap kriteria selanjutnya adalah menentukan nilai performance score dari setiap karyawan berdasarkan penilaian yang diberikan dalam setiap kriteria. Penilaian dilakukan dengan menggunakan data uji 5 karyawan dengan rentang nilai 1 (kurang), 2 (cukup), 3 (baik), dan 4 (sangat baik). Hasil penilaian dapat dilihat dalam tabel 8.

Tabel 8. Penilaian Karyawan

Nama Karyawan	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Nyoman Ardiani	4	4	3	4	3	4	4	3
Komang Merta	3	3	2	3	3	2	4	3
Putu Tarmika	3	4	3	2	3	3	3	4
Made Suryanto	4	2	3	3	4	2	3	4
Gede Wisnaya	2	3	3	4	3	4	1	4

Karena hanya menggunakan satu range nilai yang sapa pada tiap-tiap kriteria maka normalisasi nilai dengan membagi setiap nilai dengan maksimum rentang nilai tersebut. Hasilnya terlihat pada tabel 9.

Tabel 9. Hasil Normalisasi Data

Inisial	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
C1	1	1	0,75	1	0,75	1	1	0,75
C2	0,75	0,75	0,5	0,75	0,75	0,5	1	0,75
C3	0,75	1	0,75	0,5	0,75	0,75	0,75	1
C4	0,5	0,5	0,75	0,75	1	0,5	0,75	1
C5	1	0,75	0,75	1	0,75	1	0,25	1
	0,5							

Selanjutnya adalah menentukan penilaian berdasarkan pembobotan tiap kriteria dari tabel 6. Hasilnya terlihat pada tabel 10.

Tabel 10. Hasil Nilai Berdasarkan Bobot Kriteria

Inisial	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
C1	0,330	0,183	0,093	0,119	0,083	0,063	0,034	0,023

C2	0,24	0,13	0,06	0,08	0,083	0,03	0,034	0,02
	7	7	2	9		1		3
C3	0,24	0,18	0,09	0,05	0,083	0,04	0,025	0,03
C4	7	3	3	9	0,111	7	0,025	1
C5	0,33	0,09	0,09	0,08	0,083	0,03	0,008	0,03
	0	1	3	9		1		1
	0,16	0,13	0,09	0,11		0,06		0,03
	5	7	3	9		3		1

Terakhir performance score dapat ditentukan dengan $PS_n = \sum_{l=1}^{m} \quad Q_n$ hasilnya dapat dilihat pada tabel 11.

Tabel 11. Performance Score

Inisial	PS
C1	0,928
C2	0,706
C3	0,768
C4	0,801
C5	0,691

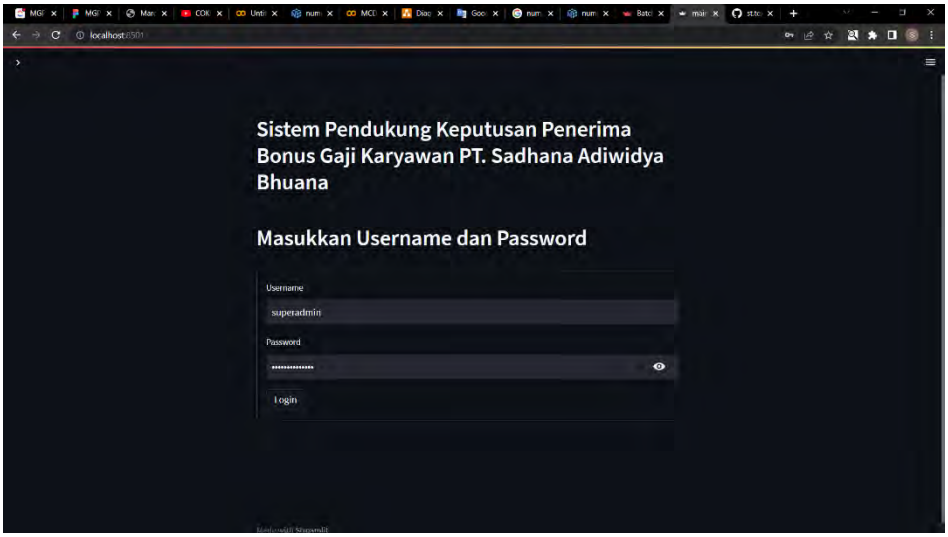
Dari performance score tersebut dapat menjadi nilai sebagai tolak ukur dalam memberikan gaji bonus terhadap seorang karyawan. Berdasarkan kesepakatan dengan PT Sadhana Adiwidya Bhuana minimal performance score yang dibutuhkan agar dapat menerima gaji bonus adalah > 0,7 dengan total gaji bonus yang akan diterima adalah ($Anggaran\ gaji\ bonus\ perkaryawan \times Performance\ score$). Sehingga karyawan dengan inisial C5 tidak mendapatkan gaji bonus karena performance scorenya yang kurang dari 0,7.

3.3. Implementasi Sistem

Dalam pengimplementasiannya sistem ini saya rancang dalam bentuk web dengan menggunakan framework python yaitu *streamlit.* Kemudian untuk menyimpan data saya menggunakan MySQL sebagai database managementnya, dan masih dalam scope localhost. Berikut merupakan hasil implementasi dan hasil testing dengan menggunakan balckbox testing.

a. Login

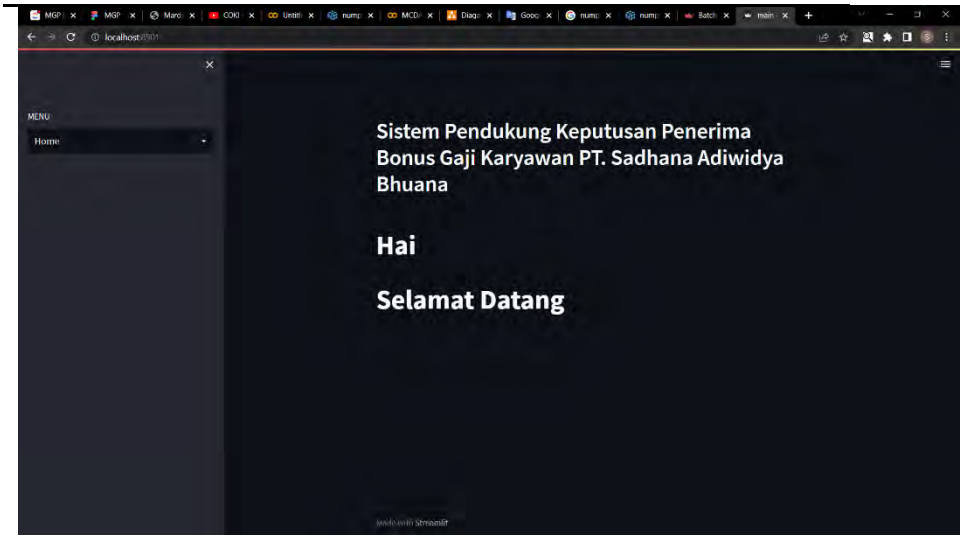
p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Login

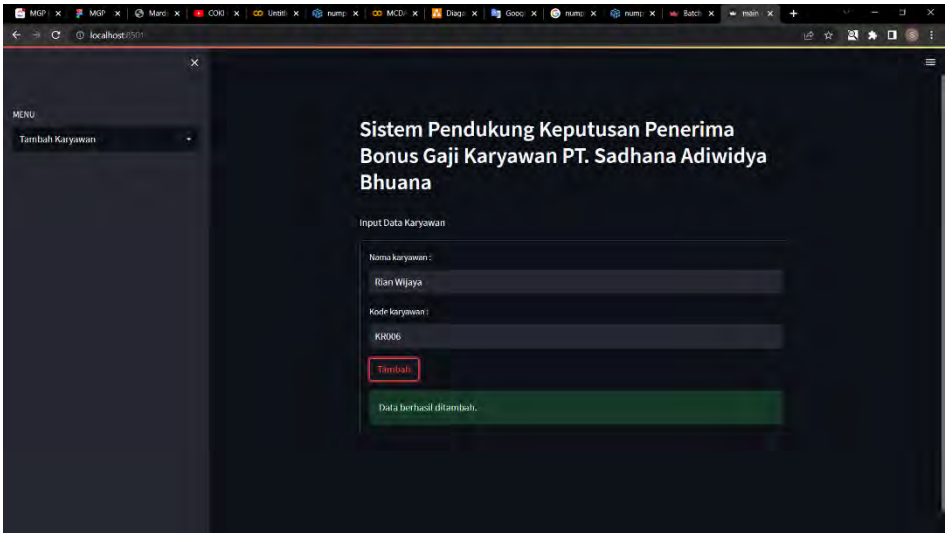
Tabel 12. Pengujian Login

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memasukkan username dan password.	Jika nama pengguna dan kata sandi valid maka dapat masuk ke dalam sistem dan dibawa ke home page, jika tidak sesuai user diminta mengulang.	sesuai



Gambar 5. Home Page

b. Input Data Karyawan

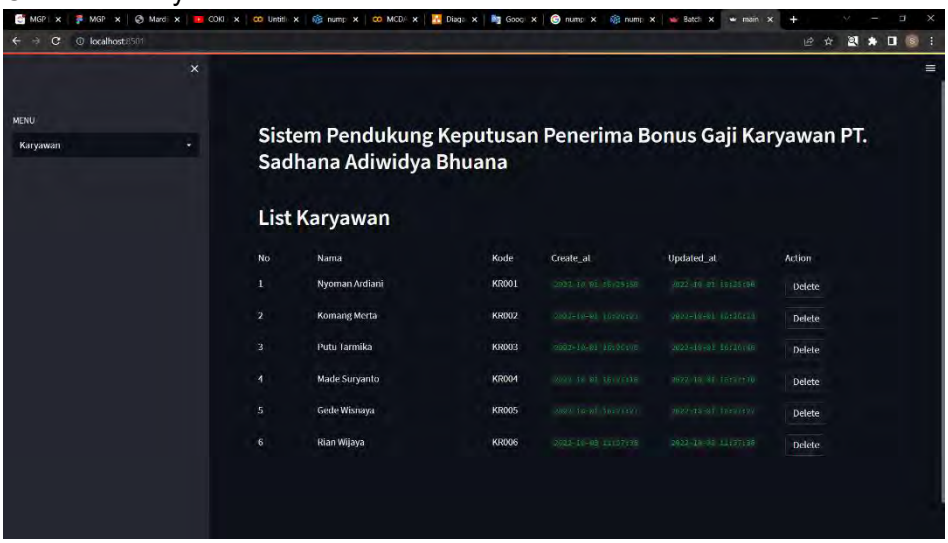


Gambar 6. Input Data Karyawan

Tabel 13. Pengujian Input Karyawan

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memasukkan nama karyawan dan kode karyawan	Jika sudah mengisi nama dan kode dan menekan tombol tambah maka data akan disimpan. Jika nama dan kode kosong maka data tidak disimpan dan muncul alert.	sesuai

c. Cek Data Karyawan



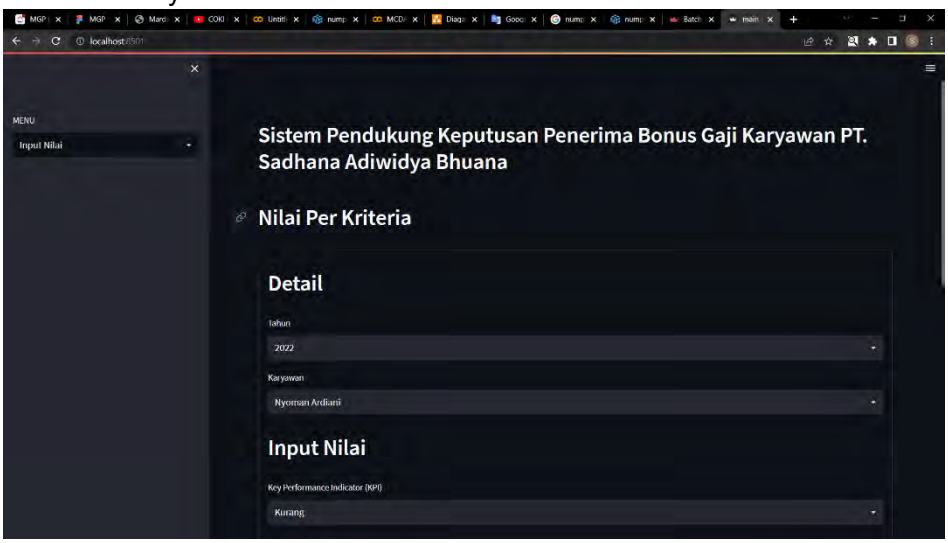
Gambar 7. Data Karyawan

Tabel 14. Pengujian Tampilan Data Karyawan

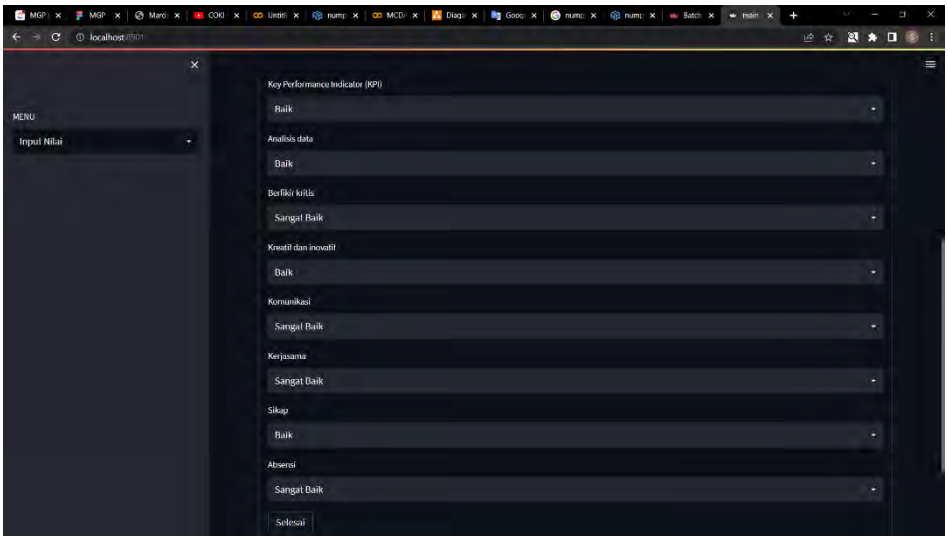
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memilih menu karyawan	Menampilkan data karyawan yang sesuai. Tombol delete dapat berfungsi dengan baik.	sesuai

d. Penilaian Karyawan



Gambar 8. Menu Penilaian

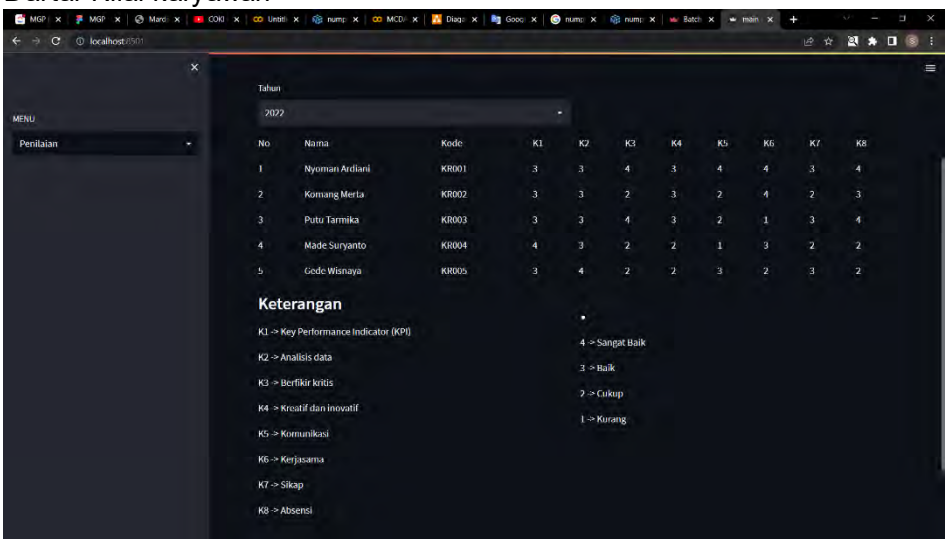


Gambar 9. Kriteria Penilaian

Tabel 15. Pengujian Penilaian Karyawan

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memasukkan tahun penilaian dan karyawan yang dinilai kemudian memberikan nilai	Jika karyawan pada tahun yang dipilih belum dinilai maka data dapat disimpan jika sudah dinilai maka muncul alert.	sesuai

e. Daftar Nilai karyawan

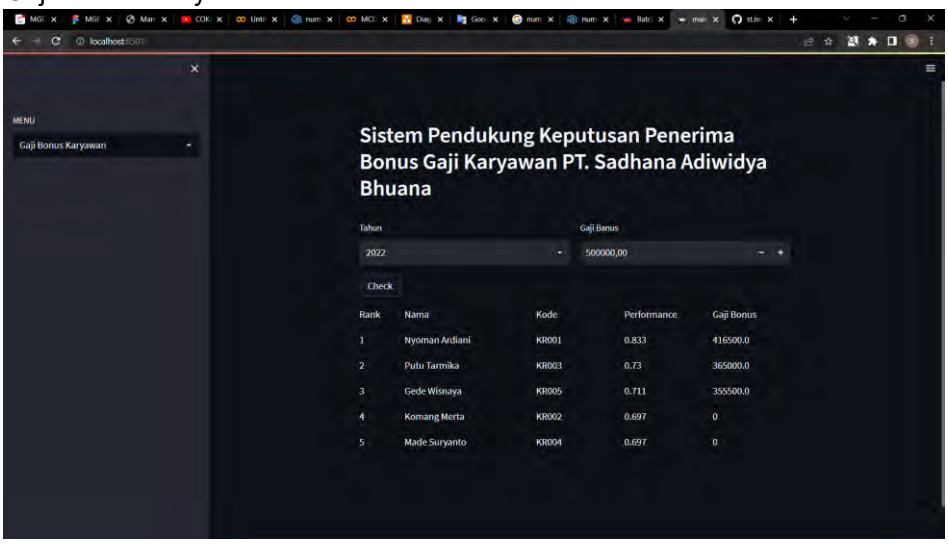


Gambar 10. Hasil Penilaian Karyawan

Tabel 16. Pengujian Menu Hasil Penilaian

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memilih menu penilaian dan memasukkan tahun penilaian.	Menampilkan data yang sesuai. Jika tahun yang dipilih penilaian karyawan belum lengkap atau belum ada penilaian maka muncul alert	sesuai

f. Gaji Bonus Karyawan



Gambar 11 Hasil Penentuan Gaji Bonus

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 17. Pengujian Penentuan Gaji Bonus

Skenario Pengujian	Hasil Yang diharapkan	Keterangan
User memasukkan tahun dan gaji bonus yang sudah dianggarkan di tahun tersebut.	Menampilkan nilai gaji yang relevan sesuai dengan penilaian karyawan. Jika performance kurang dari 0.7 maka karyawan tidak mendapatkan gaji bonus. Jika nilai gaji kosong atau belum ada karyawan di tahun yang dipilih maka muncul alert.	sesuai

4. Kesimpulan

Implementasi Analytical Hierarchy Process (AHP) dalam penentuan bobot kriteria pendukung keputusan mendapatkan hasil yang baik. Terdapat delapan kriteria yang telah disepakati dengan masing-masing bobot yang diperoleh dari metode AHP adalah key performance indicator (KPI) dengan bobot 0.330, analisis data dengan bobot 0.183, berpikir kritis dengan bobot 0.125, kreatif dan inovatif dengan bobot 0.119, komunikasi dengan bobot 0.111, kerjasama dengan bobot 0.063, sikap dengan bobot 0.034, dan absensi dengan bobot 0.031. Dari hasil pembobotan tersebut kemudian diimplementasikan dalam bentuk aplikasi berbasis website. Hasil pengujian menunjukkan fungsional aplikasi dapat berjalan dengan baik dan sesuai ekspektasi. Sistem dapat memberikan daftar karyawan yang berhak mendapatkan gaji dengan nominal yang berdasarkan dengan performance score yang diperoleh dari penilaian karyawan.

Daftar Pustaka

[1] D. Witasari and Y. Jumaryadi, "Aplikasi Pemilihan Karyawan Terbaik Dengan Metode Simple Additive Weighting (Saw) (Studi Kasus Citra Widya Teknik)," *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 10, no. 2, p. 115, 2020, doi: 10.24853/justit.10.2.115-122.

[2] D. Nababan and R. Rahim, "Sistem Pendukung Keputusan Reward Bonus Karyawan Dengan Metode Topsis," *Nababan, Darsono Rahim , Robbi*, vol. 3, no. 1, pp. 2528–5114, 2018, [Online]. Available: https://ejournal.medan.uph.edu/index.php/isd/article/view/185

[3] M. Yanto, "Sistem Penunjang Keputusan Dengan Menggunakan Metode Ahp Dalam Seleksi Produk," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 3, no. 1, pp. 167–174, 2021, doi: 10.47233/jteksis.v3i1.161.

[4] N. Sari, "Implementation of the AHP-SAW Method in the Decision Support System for Selecting the Best Tourism Village," *J. Tek. Inform. CIT Medicom*, vol. 13, no. 1, pp. 23–32, 2021, [Online]. Available: https://www.medikom.iocspublisher.org/index.php/JTI/article/view/51

[5] S. R. Yulistina, T. Nurmala, R. M. A. T. Supriawan, S. H. I. Juni, and A. Saifudin, "Penerapan Teknik Boundary Value Analysis untuk Pengujian Aplikasi Penjualan Menggunakan Metode Black Box Testing," *J. Inform. Univ. Pamulang*, vol. 5, no. 2, p. 129, 2020, doi: 10.32493/informatika.v5i2.5366.

This page is intentionally left blank.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Rancang Model Ontologi untuk Representasi Pengetahuan Rumah Tradisional di Indonesia

Kadek Diah Pramesti[a1], Luh Gede Astuti[a2]

[a]Informatics Study Program, Faculty of Math and Natural Science, Udayana University
Bali, Indonesia
[1]dpanbers@gmail.com
[2]lg.astuti@unud.ac.id

Abstract

Indonesia is famous for its diversity from ethnicity, religion, culture, customs, to traditional buildings or commonly called traditional houses. There are many traditional houses in Indonesia. Each region or place has its own traditional house or building. To preserve this traditional house, a model such as Ontology is needed. The ontology knowledge base is an appropriate method used to represent information. Ontology is the basis of the semantic web that will be utilized by computer applications to manipulate data for user needs. Ontology describes several concepts from a domain and the relationship between these concepts formally. In this project, the ontology model was built using the Protégé ontology development tool. We use the method of METHONTOLOGY in the development of the ontology model where this method describes each step-in detail. The ontology model built has 9 classes, 4 object properties, 4 data properties, and 80 individuals. The testing process in the development of the ontology model by performing SPARQL queries.

Keywords: *Traditional House, Ontologi, Methontologi, Query SPARQL, Protégé*

1. Pendahuluan

Indonesia terkenal akan keberagamannya dari suku, agama, budaya, adat istiadat, sampai bangunan tradisional atau biasa disebut rumah tradisional. Rumah tradisional merupakan suatu bangunan dengan struktur, cara pembuatan, bentuk dan fungsi serta ragam hias yang memilki ciri khas tersendiri, diwariskan secara turun – temurun dan dapat digunakan untuk melakukan kegiatan kehidupan oleh penduduk sekitarnya [1]. Rumah tradisional di Indonesia sangatlah banyak. Setiap daerah maupun tempat memiliki rumah tradisional atau bangunan khas tersendiri. Bangunan dari masing-masing daerah tersebut mencerminkan identitas dan kebudayaan dari daerah tersebut. Ciri khas tersebutlah yang menjadikan suatu bangunan atau rumah dapat disebut sebagai bangunan tradisional atau rumah adat. Keberagaman rumah tradisional di Indonesia membuat masyarakat kebingungan sehingga harus didokumentasikan dan tergambarkan dengan baik agar kebudayaan ini tidaklah luntur. Oleh karena itu diperlukan suatu model yang dapat merepresentasikan pengetahuan tentang rumah adat di Indonesia

Ontologi menjadi salah satu solusi untuk mengelola data sehingga dapat memberikan informasi yang bernilai semantik. Ontologi adalah dasar dari web semantik yang akan dimanfaatkan oleh aplikasi komputer untuk memanipulasi data yang data untuk kebutuhan pengguna. Ontologi mendeskripsikan beberapa konsep dari suatu domain dan keterkatitan antara konsep tersebut secara formal. Konsep dari suatu domain akan saling berkaitan sehingga nantinya akan membentuk suatu kesatuan data. Penelitian ini akan mengembangkan sebuah model ontologi pada domain rumah tradisional di Indonesia [2].

Metode yang digunakan dalam penelitian ini untuk pembangunan model ontologi yaitu Methontologi. Methontologi adalah suatu metode untuk mengembangkan model ontologi. Metode ini memiliki keunggulan dalam penggambaran setiap informasi dan aktivitas dengan jelas. Usulan

penelitian ini adalah merancang model ontologi yang merepresentasikan rumah tradisional di Indonesia dan diharapkan dapat membangun model ontologi dengan kualitas yang baik.

1.1. Rumah Tradisional

Rumah merupakan aktualisasi diri yang diejawantahkan dalam bentuk kreativitas dan pemberian makna bagi kehidupan penghuninya. Selain itu rumah juga bentuk cerminan diri. Rumat adat adalah bangunan yang memiliki ciri khas khusus didalamnya, biasa digunakan untuk tempat hunian suatu suku bangsa tertentu. Rumah adat juga biasa disebut rumah tradisional ini merupakan representasi kebudayaan pada suatu daerah atau suku. Rumah tradisional di Indonesia sangatlah beragam dan memiliki makna penting sebagai perspektif sejarah, warisan, serta kemajuan masyarakat di Indonesia dalam suatu peradaban [3].

1.2. Ontology

Ontologi adalah model konseptual dari beberapa aspek dari alam semesta wacana tertentu. Ontologi merupakan suatu cara untuk merepresentasikan pengetahuan tentang seperangkat konsep dalam domain informasi dan hubungan antar konsep-konsep ini [4]. Ontologi dapat digunakan untuk menyajikan informasi secara semantik, mengatur dan memetakan kumpulan sumber daya informasi dengan cara teroganisir dan terstruktur. Ontologi mendeskripsikan suatu teori mengenai objek dan keterkaitan antar mereka.

Ontologi berbentuk struktur jaringan yang terdiri atas: [5]

a. Kumpulan kelas, biasanya kelas digambarkan sebagai simpul dalam struktur jaringan.
b. Kumpulan relasi yang menghubungkan kelas-kelas, relasi dalam struktur jaringan biasanya digambarkan sebagai garis berarah.
c. Kumpulan *instance*s yang terdapat pada kelas-kelas tertentu.

1.3. Methontology

Methontology adalah metode terstruktur dengan baik untuk membangun ontologi dari awal [6]. Secara umum, metode memberi seperangkat pedoman tentang bagaimana seseorang harus melakukan aktivitas yang diidentifikasi dalam proses pengembangan ontologi, jenis teknik apa yang terbaik untuk setiap aktivitas, dan produk apa yang dihasilkan masing-masing.

1.4. SPARQL Query

SPARQL adalah bahasa dan protokol query standar untuk database *Linked Open Data* dan RDF [7]. Setelah dirancang untuk menanyakan berbagai macam data, itu dapat secara efisien mengekstrak informasi yang tersembunyi dalam data yang tidak seragam dan disimpan dalam berbagai format dan sumber. Dikembangkan dan didukung oleh W3C, standar SPARQL membantu pengguna dan pengembang fokus pada apa yang ingin mereka ketahui daripada bagaimana database diatur.

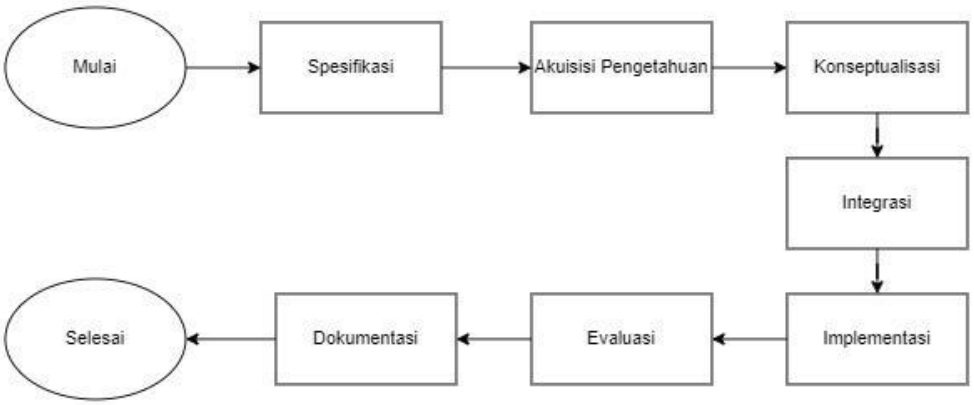
1.5. Protégé

Protégé adalah sebuah tool yang dapat digunakan untuk membangun domain ontologi serta dapat melakukan *query* dengan menggunakan SPARQL. Protégé dibuat dengan menggunakan bahasa pemrograman Java dan memiliki format penyimpanan seperti OWL, RDF, XML, Turtle, Manchester OWL, JSON-LD, LaTeX, dan OBO. Fungsi dalam tool Protégé dapat digunakan melalui *Graphical User Interface* (GUI) dengan menampilkan tab untuk masing-masing bagian dan fungsi standar [8].

2. Metode Penelitian

Metode yang digunakan pada penelitian ini adalah Methontologi. Methontologi adalah metode terstruktur yang dapat digunakan untuk membangun ontologi dari awal. Metode ini mencakup serangkaian aktivitas, teknik, dan hasil yang diproduksi oleh eksekusi dari setiap aktivitas menggunakan tekniknya masing-masing. Methontologi sangat merekomendasikan penggunaan ontologi yang telah ada. Adapun tahapan dari Methontologi sebagai berikut.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 1. Alur Metode Penelitian

2.1. Spesifikasi

Tujuan dari tahap spesifikasi adalah untuk menghasilkan dokumen spesifikasi ontologi informal, semi-formal atau formal yang ditulis dalam bahasa alami, masing-masing menggunakan seperangkat representasi perantara atau menggunakan pertanyaan kompetensi.

2.2. Akuisisi Pengetahuan

Tahap akuisisi pengetahuan merupakan kegiatan independen dalam proses pengembangan ontologi. Sumber pengetahuan diperoleh melalui pakar, buku, gambar, dan ontologi lainnya dengan menggunakan teknik seperti brainstorming, analisis teks formal dan informal, dan alat akuisisi pengetahuan.

2.3. Konseptualisasi

Pada fase ini, penulis akan membangun struktur domain pengetahuan dalam model konseptual yang menggambarkan masalah dan solusinya dalam bentuk kosakata domain yang diidentifikasi dalam aktivitas spesifikasi ontologi. Hal pertama yang penulis lakukan adalah menyusun Glosarium Istilah, memasukkan konsep, contoh, kata kerja, dan property.

2.4. Integrasi

Fase integrasi mengacu pada tujuan mempercepat pembangunan ontologi, mungkin mempertimbangkan penggunaan kembali definisi yang sudah dibangun ke dalam ontologi lain daripada memulai dari awal.

2.5. Implementasi

Pada fase ini akan diterapkan konsep ontologi yang telah dibangun. Implementasi ontologi membutuhkan penggunaan lingkungan yang mendukung meta-ontologi dan ontoiogies yang dipilih pada fase integrasi.

2.6. Evaluasi

Fase evaluasi berarti melakukan penilaian teknis dari ontologi, lingkungan perangkat lunak, dan dokumentasi sehubungan dengan kerangka acuan selama setiap fase dan di antara fase siklus hidup mereka.

2.7. Dokumentasi

Pada fase ini, tidak ada pedoman yang disepakati tentang bagaimana mendokumentasikan ontologi. Tahap dokumentasi biasanya mencakup kode untuk ontologi, bahasa alami, makalah yang akan diterbitkan dalam prosiding dan jurnal yang digunakan untuk menentukan pertanyaan penting dari ontologi yang sedang dibangun.

3. Hasil dan Pembahasan

Pada penelitian ini dibangun sebuah ontologi yang berdomain Rumat Tradisional. Berikut merupakan hasil yang diperoleh dari setiap tahapan metode penelitian yang telah dilakukan.

3.1. Spesifikasi

Tahap ini akan memberikan spesifikasi terkait ontologi yang telah dibangun berikut merupakan deskripsi dari ontologi "Rumat Tradisional":

a. Domain: Rumat Tradisional
b. Tanggal: 30 September 2022
c. Dirancang Oleh: Kadek Diah Pramesti
d. Diimplementasikan Oleh: Kadek Diah Pramesti
e. Level Formalitas: Formal
f. Ruang Lingkup: Rumah Tradisional di Indonesia
g. Sumber Pengetahuan: Internet, Buku

3.2. Akuisisi Pengetahuan

Tahap ini ditujukan untuk memperoleh pengetahuan yang dapat berguna pada ontologi Rumah Tradisional yang dibangun. Pada penelitian ini, tahapan akuisisi pengetahuan adalah sebagai berikut:

a. Diskusi dengan ahli Ontologi untuk mempersiapkan draf awal untuk merancang dan mengembangkan Ontologi.
b. Analisis teks formal maupun informal.
c. Mengidentifikasi pengetahuan dan struktur yang dirancang seperti konsep, atribut, nilai, dan hubungan.

Data yang digunakan dalam penelitian ini adalah Rumah Tradisional di Indonesia. Data yang digunakan merupakan data latih yang diperoleh dari buku pegangan dan sumber internet yang dapat dipercaya.

3.3. Konseptualisasi

Pada tahap ini yang ditujukan untuk merancang konsep yang digunakan untuk mendeskripsikan masalah dan solusi yang akan digunakan. Pada tahap ini dibangun daftar istilah lengkap yang mencakup konsep, *instance*, kata kerja, dan *property* yang berkaitan dengan domain Rumah Tradisional.

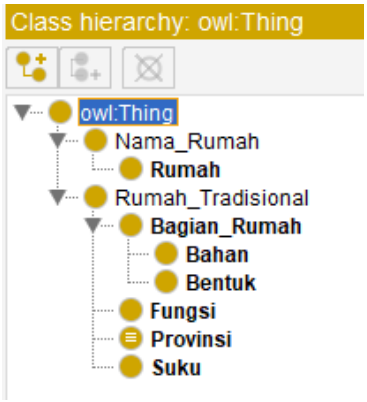
3.4. Integrasi

Tahap ini digunakan untuk menggabungkan atau mengintegrasikan ontologi yang sudah ada dengan ontologi yang akan dibangun. Dengan segala pertimbangan agar dapat sesuai dengan domain Rumah Tradisional. Pemilihan ontologi yang sesuai dapat memudahkan mendapatkan hasil yang diharapkan.

3.5. Implementasi

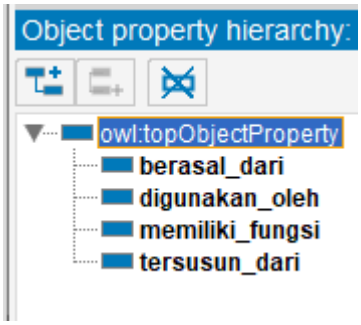
Pada tahap implementasi ontologi Rumah Tradisional ini menggunakan aplikasi Protégé 5.5.0. Perangkat lunak Protégé merupakan salah satu *tool* atau alat yang digunakan seorang *ontologi developer* untuk mengembangkan ontologi. Berdasarkan hasil implementasi ini didapatkan konsep *class* yang digunakan pada ontologi terlihat pada Gambar 1, hubungan antara *class* atau *relationships* yang ada dalam ontologi yang didefinisikan pada *object properties* dapat dilihat pada Gambar 2. *Instance* pada masing-masing *class* yang didefinisikan pada bagian individual dapat dilihat pada Gambar 3. Atribut pada masing-masing *class* atau *instance* dapat dilihat pada Gambar 4. Untuk hasil dan struktur hubungan antar *class* dapat dilihat pada Ontograf yang ada pada Gambar 5.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



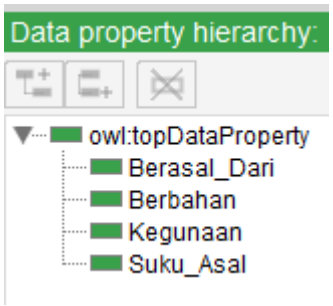
Gambar 2. *Class* dari Ontologi Rumah Tradisional

Pada Gambar 2 terdapat 9 *class* yang ada pada ontologi Rumah Tradisional. *Class* Rumah_Tradisional memiliki 4 *subclass* yaitu Bagian_Rumah dengan *subclass* Bahan dan Bentuk, Provinsi, Suku, Fungsi. Lalu *class* Nama_Rumah memiliki 1 *subclass* yaitu Rumah. Dimana nanti dari ke-9 *class* ini akan menyimpan *instance* yang sesuai dengan nama *class.*



Gambar 3. *Object Properties* dari Ontologi Rumah Tradisional

Pada Gambar 3 terdapat 4 *Object Properties* yang ada pada ontologi Rumah Tradisional. Adapun *Object Properties* tersebut yaitu berasal_dari yang nantinya akan menghubungkan antara instance dengan class Provinsi, digunakan_dari nantinya akan menghubungkan antara instance dengan class Suku, memiliki_fungsi nantinya akan menghubungkan antara instance dengan class Fungsi, lalu tersusun_dari nantinya akan menghubungkan antara instance dengan class Bahan dan Bentuk.

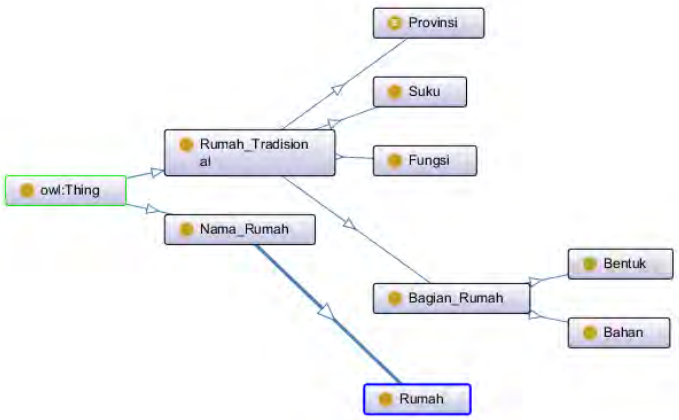


Gambar 4. *Data Properties* dari Ontologi Rumah Tradisional

Pada Gambar 4 terdapat 4 Data property yaitu Berasal_dari, Berbahan, Kegunaan, Suku_Asal. *Data Properties* digunakan untuk menghubungkan *instance* dengan *datatype value* seperti *text, string* atau *number*. Berasal_dari, Berbahan, Kegunaan, Suku_Asal akan menghubungkan instance dengan datatype string.

Gambar 5. Individu yang digunakan dalam ontologi Rumah Tradisional. Individu dalam kelas yang diperluas disebut *instance*.

Pada Gambar 5 terdapat beberapa individual yang dihasilkan pada setiap *class* yang sudah dibuat di dalam Ontologi Rumah Tradisional. Ada 30 individual untuk *class* Nama_Rumah, 2 individual untuk *class* Bahan, 2 individual untuk *class* Bentuk, 3 individual untuk *class* Fungsi, 34 individual untuk *class* Provinsi, 9 individual untuk *class* Suku.



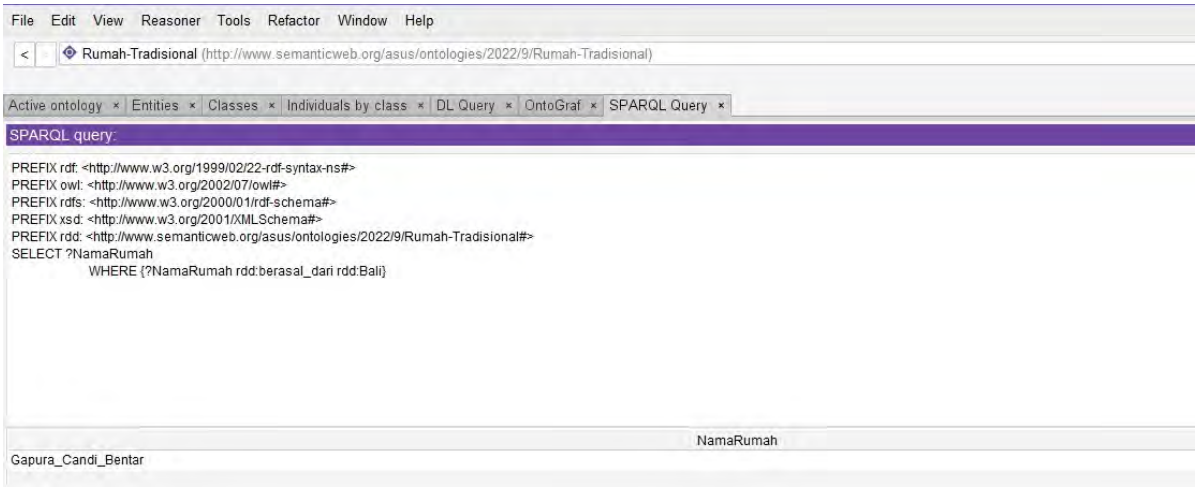
Gambar 6. *Ontograf* dari Ontologi Rumah Tradisional

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Gambar 6 adalah contoh hubungan semantik yang menggambarkan masing-masing *class, object property*, dan individual yang dibangun pada ontologi Rumah Tradisional. Hubungan tersebut direpresentasikan ke dalam bentuk gambar oleh ontograf.

3.6. Evaluasi

Pada tahap ini akan dilakukan pengujian pada ontologi yang telah dibuat. Tahap evaluasi ini dilakukan dengan cara melakukan query menggunakan SPARQL query yang terdapat pada aplikasi Protégé. Beberapa pertanyaan yang sudah disiapkan kemudian pertanyaan tersebut dapat diubah kedalam bentuk SPARQL query, sehingga dapat ditampilkan hasil yang ada pada ontologi yang sudah dibuat. Hasil query dapat dilihat pada Gambar 6. Pada Gambar 6 SPARQL query yang dijalankan adalah sebagai berikut:



Gambar 7. Hasil SPARQL query

3.7. Dokumentasi

Pada tahapan ini, dokumentasi dari model ontologi rumah tradisional yang telah dibangun berupa tulisan yang tertuang di jurnal ini.

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka ontologi terkait domain Rumah Tradisional sudah selesai dibangun. Pembangunan ontologi ini menggunakan aplikasi Protégé 5.5.0 dengan menggunakan metode Methontology dan menghasilkan 9 class, 4 Object Property, 1 data properties, dan 80 individual atau instance pada setiap class. Pada penelitian ini juga dilakukan evaluasi atau pengujian terkait model yang diajukan dengan mengajukan beberapa pertanyaan yang digunakan pengguna untuk mengakses informasi mengenai Rumah Tradisional. Sehingga, diharapkan model ontologi yang dihasilkan mampu memberikan informasi terkait lagu Rumah Tradisional secara sistematis. Dari pembangunan ontologi Rumah Tradisional selanjutnya dapat digunakan sebagai dasar dalam mengembangkan sistem terkait dengan Rumah Tradisional.

Referensi

[1] Abdul Aziz Said, 2004. “Toraja Simbolisme Unsur Visual Rumah Tradisional,” Penerbit: Ombak, Jogyakarta.

[2] Sinaga, Arnaldo Marulitua., Sipahutar. Rini Juliana., Hutasoit. Dian Ira Putri., “PENERAPAN ONTOLOGY WEB LANGUAGE PADA DOMAIN ULOS BATAK TOBA,”. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). Vol. 5, No. 4, p. 493-502, 2018, DOI: 10.25126/jtiik.201854903.

[3] Eka. Rahmawati. “KARAKTERISTIK FISIK RUMAH ADAT GORONTALO (DULOHUPA DAN BANTAYO POBO’IDE)),”. *Jurnal Arsitektur, Kota dan Permukiman (LOSARI).*

[4] [H. R. Badron, Yunizar; Agus, Fahrul; Hatta, “Studi Tentang Pemodelan Ontologi Web Semantik dan Prospek Penerapan pada Bibliografi Artikel Jurnal Ilmiah,” Pros. Semin. Ilmu Komput. dan Teknol. Inf., vol. 2, no. 1, pp. 164–169, 2017, [Online]. Available: http://journal.student.uny.ac.id/ojs/ojs/index.php/pgsd/article/viewFile/135/130.]

[5] A. Satria, A. Herdiani, and V. Effendy, “Analisis Keterhubungan Ontologi Pada Web Semantik Menggunakan Semantic-Based Ontologi Matching,” e-Proceeding Eng., vol. 3, no. 3, pp. 5345–5352, 2016.

[6] M.Fernández, A. Gómez-Pérez, and N. Juristo, “METHONTOLOGI: From Ontological Art Towards Ontological Engineering,” Proc. Ontol. Eng. AAAI-97 Spring Symp. Ser., pp. 33–40, 1997, [Online]. Available: http://speech.inesc.pt/~joana/prc/artigos/06c METHONTOLOGY from Ontological Art towards Ontological Engineering -Fernandez, Perez, Juristo -AAAI -1997.pdf.

[7] P. R. Ganeswara and C. R. A. Pramartha, “Ontology-based Approach for Klungkung Royal Family,” JELIKU (Jurnal Elektron. Ilmu Komput. Udayana), vol. 8, no. 4, p. 497, 2020, doi: 10.24843/jlk.2020.v08.i04.p16.

[8] P. I. Nugroho, B. Priyambadha, and N. Y. Setiawan, “Sistem Pencarian Koleksi Laporan Skripsi Dan PKL dengan Teknologi Web Semantik (Studi Kasus: Ruang Baca Fakultas Ilmu Komputer Universitas Brawijaya),” J. Pengemb. Teknol. Inf. dan Ilmu Komputer, Vol. 2 No.9, vol. 2, no. 9, pp. 3440–3444, 2018.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Perancangan Smart Home untuk Keamanan Rumah Berbasis Jaringan Sensor Nirkabel

Kadek Yoga Vidya Pradnyaditha[a1], AAIN Eka Karyawati[a2]

[a]Fakultas Matematika Dan Ilmu Pengetahuan Alam, Program Studi Informatika
Kuta Selatan,Badung,Bali Indonesia
[1]vidyayoga7@gmail.com
[2]eka.karyawati@unud.ac.id

Abstract_

Pada zaman kemajuan teknologi dan ilmu pengetahuan yang telah menghasilkan beberapa teknologi terbaru[1]*. Salah satu teknologi nya adalah teknologi nirkabel. Teknologi ini merupakan teknologi yang banyak digunakan di dunia dan telah digunakan pada sebagian besar kegiatan manusia. Teknologi Smart Home ini diciptakan dikarenakan adanya banyak masalah yang terjadi dalam hal kebakaran dan pencurian yang terjadi pada rumah-rumah yang ditinggalkan pemiliknya dikarenakan sesuatu dan lain hal. Solusi alternatif untuk masalah ini adalah dengan menggunakan teknologi Smart Home. Di negara maju prinsip Smart Home atau rumah pintar telah diadopsi sejak puluhan tahun yang lalu, Smart Home dalam penerapan nya memiliki fungsi sebagai remote kontrol, baik sistem pencahayaan, monitoring dan juga otomatis.*

Rancangan sistem Smart Home berbasis Wireless Sensor Networks telah dapat digunakan pada rumah model. Sistem Smart Home berbasis WSN dapat diimplementasikan dengan berbagai data sensor multimedia, yakni pada satu node sensor bisa membawa 4 sensor. Selama pengujian bagian kinerja sistem ini dapat dimulai dengan menguji akurasi atau kebenaran data masing-masing sensor sehingga waktu komunikasi antara sensor dan pusat pemantauan dan sebaliknya. Sensor PIR berhasil menemukan keberadaan seseorang di suatu ruangan dengan delay 2,8 (dua koma delapan) detik.

Keywords: *Jaringan Sensor Nirkabel, Keamanan, Smart Home, SMS*

1. Pendahuluan

Pada perkembangan kemajuan ilmu pengetahuan dan teknologi telah menghasilkan beberapa teknologi terbaru. Salah satu teknologi nya adalah teknologi nirkabel. Teknologi ini merupakan salah satu teknologi yang banyak digunakan di seluruh dunia. Teknologi nirkabel ini telah digunakan pada sebagian besar kegiatan manusia. Teknologi Smart Home ini diciptakan dikarenakan adanya banyak masalah yang terjadi dalam hal kebakaran dan pencurian yang terjadi pada rumah-rumah yang ditinggalkan pemiliknya dikarenakan sesuatu dan lain hal. Solusi alternatif untuk masalah ini adalah dengan menggunakan teknologi Smart Home. Di negara maju prinsip Smart Home atau rumah pintar telah diadopsi sejak puluhan tahun yang lalu, Smart Home dalam penerapan nya memiliki fungsi sebagai remote kontrol, baik sistem pencahayaan, monitoring dan juga otomatis.

Dalam melakukan perancangan Smart Home, menggunakan beberapa sistem *Smart Home diantaranya Smart Home berbasis SMS gateway. Smart Home berbasis SMS gateway adalah suatu fasilitas untuk mengontrol beberapa peralatan elektronik rumah menggunakan layanan SMS melalui jaringan GSM. Namun, teknologi tersebut dianggap kurang efisien dan alasannya karena dengan pesatnya perkembangan teknologi, masyarakat beralih dari penggunaan SMS ke aplikasi berbasis Internet*[2]. Hingga saat ini, orang hanya dapat mengontrol perangkat dengan remote dan sakelar inframerah yang terhubung dengan kabel, tetapi kontrol dibatasi oleh jarak jangkau. Solusi smartphone sebagai operator remote control adalah dengan dan memperluas jangkauan kontrol tersebut serta menerapkan sistem operasi mobile yang saat ini cepat, untuk mengetahui komunikasi Android, tetapi juga dikembangkan mengikuti tren dan kebutuhan Salah satu hasil perkembangan teknologi adalah tercipta nya WiFi atau teknologi Wireless Fidelity, WiFi adalah teknologi yang menggunakan peralatan elektronik untuk bertukar

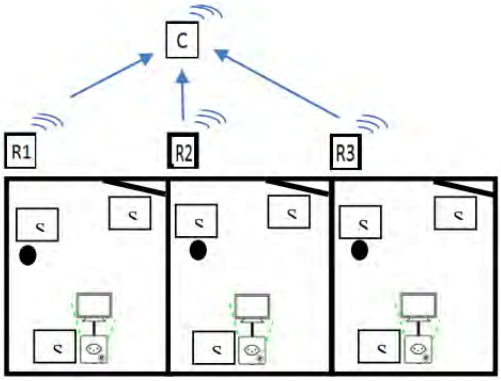
nirkabel melalui jaringan komputer, termasuk internet yang memiliki kecepatan tinggi . WiFi merupakan "produk jaringan wilayah lokal nirkabel (WLAN) apapun yang didasarkan pada standar Institute of Electrical and Electronics Engineers (IEEE) 802.11". salah satu modul WiFi yang mendukung modul Arduino ESP8266, modul ESP8266 adalah modul yang banyak digunakan untuk Internet Aplikasi seperti kontrol aktuator dan sensor membaca. Sistem kontrol dapat berbentuk server web MQTT atau yang tertanam dalam memori IC ESP8266. Komputer, ponsel, dan tablet yang dapat mengakses web, aktuator, membaca sensor. Penerapan kontrol ini dapat dilakukan pada peralatan rumah tangga. Alat rumah tangga ini dapat dimatikan dan menyala dengan WiFi dan dapat di kontrol secara otomatis dengan ponsel Situs web ini menunjukkan waktu aktif dan nonaktif. Proses ini akan menghemat listrik di rumah karena peralatan rumah tangga sesuai dengan kebutuhan pemiliknya.[3].

Penelitian dari [4] menggunakan sistem penguncian berbasis internet berbasis Bluetooth yang dapat menyesuaikan dengan bergantung apa yang akan digunakan. Sistem smart lock ini juga menggunakan face detection untuk keamanan ganda, sehingga kunci ini akan mengenali wajah yang telah terdaftar pada sistem. Selain itu, pengguna dapat membuka pintu di mana saja secara nirkabel menggunakan Internet. Konsep ini dapat menjadi langkah besar menuju rumah yang lebih cerdas dan aman. Fungsi Bluetooth dalam penelitian ini adalah menghubungkan ke jarak ter pendek untuk membuka kunci pintu, sedangkan Internet adalah remote control nirkabel. Penelitian yang dilakukan oleh Aman Pathak ini terkait dengan penelitian ini, khususnya untuk sistem penguncian pintu.

[5] yang berjudul “Smart Security System For Home Appliances Control Based on Internet Of Things”. Kajian tersebut menjelaskan bahwa ada beberapa perangkat yang terhubung membentuk suatu sistem, yaitu sistem keamanan. Perangkat yang terhubung menggabungkan sistem deteksi dalam bentuk sensor PIR , sistem ini akan mendeteksi gerakan di tempat yang ingin anda deteksi. Kemudian sistem pendeteksi suhu dan kelembaban menggunakan sensor DHT11, sistem ini akan mendeteksi kelembaban dan suhu di dalam ruangan yang akan ditanamkan. Sistem pendeteksi asap atau gas menggunakan sensor MQ2, sistem ini akan mendeteksi jika ada asap atau kebocoran gas di dalam ruangan. Sistem pendeteksi kunci pintu yang menggunakan sensor jarak dan magnet untuk membuat aula dan yang terakhir adalah relay yang menggunakan modul relay 4 saluran untuk mematikan listrik. Dalam penelitian ini, konsep kontrol yang digunakan dalam sistem Smart Home terkait , yaitu relay dan kontrol kunci pintu.

2. Metode Penelitian

Dalam merancang sistem Smart Home dapat menggunakan teknologi WSN untuk menyediakan fungsi keamanan rumah. Dalam penelitian yang dilakukan oleh [6] menggunakan stimulasi prototype pada sistem Smart Home yang nantinya akan ditempatkan 3 (tiga) sensor yaitu R1,R2 dan R3 pada setiap ruangan yang berbeda satu coordinator node (C). Untuk ilustrasi nya dapat ditampilkan pada gambar berikut :



Gambar 1 Rancangan dengan WSN [6]

Pada suatu frame data yang dikirimkan oleh node sensor ke koordinator akan berisi tiga data sensor, yaitu sensor suhu, sensor peralatan rumah tangga, dan sensor PIR.

1. (S1) Sensor suhu untuk memantau suhu tiap-tiap ruangan.
2. (S2) Sensor peralatan rumah untuk memantau apakah ada penghuni rumah yang belum padam saat pemiliknya tidak ada di rumah.
3. (S3) Sensor PIR digunakan untuk mengetahui keberadaan orang di salah satu rumah ketika semua pemiliknya sedang berada jauh dari rumah.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Saat rumah dapat di pantau pemiliknya dari jarak jauh dengan terhubung internet. Perancangan perangkat keras terdiri dari perangkat model dan perangkat pada setiap node sensor, antara lain:

a. Sensor dalam perancangan ini akan menggunakan sensor dari DHT11, limit switch dan PIR guna memindai keberadaan seseorang.
b. Tansceiver, Xbee Pro Series 2, digunakan untuk menampung dan menyampaikan menggunakan standar 802.15.4 Zigbee/IEEE.
c. Mikrokontroler Arduino UNO Rev 3, bekerja di salah satu pengontrol perangkat yang tersambung dengan mikrokontroler.
d. Sumber Listrik sebagai sumber daya agar sistem bekerja secara keseluruhan.

3. Hasil dan Pembahasan

Saat merancang Smart Home, terdapat tiga sensor pada setiap node sensor yaitu :

Gambar 2 Rangkain Sensor Smart Home

a. Sensor suhu, menggunakan sensor DHT11 dan sensor LM35 yang berfungsi sebagai pendeteksi suhu pada sekitar rumah untuk mengantisipasi terjadi kebakaran dalam rumah.
b. Sensor PIR, sebagai mendeteksi keberadaan seseorang agar terhindar dari pencurian.
c. Limit switch, berfungsi untuk mematikan dan menghidupkan peralatan rumah tangga, seperti TV, AC, dan lain-lain.

Ketika mengamati suhu di ruangan menggunakan sensor suhu DHT11 dan LM35. Dalam pengujian nya, masing-masing sensor suhu diuji dengan cara melihat perbandingan pembaca suhu digital antar sensor.

Tabel 1 Data Pengukuran Suhu [6]

No	Rata-Rata DHT11	Rata-Rata LM35	Termo Digital	Error (%) Dht11	Error (%) Lm35
1	27	27.125	27.3	1.24	3.21
2	27.1	27.235	27.6	1.23	1.56
3	27.4	27.385	28	1.56	3.45
4	28	26.576	28.2	1.2	5.45
5	28.4	27.867	28.8	0.83	4.47
6	28.8	28.91	30.6	1.8	5.15
7	30	29.357	32.8	1.65	10.65
8	33	31.125	35	0.56	3.76
9	28	32.875	30	1.34	0.77
10	32	30.235	35	1.8	4.85
		Rata-Rata Error		1.321	4.332

Dari pengujian di atas, maka didapatkan bahwa dengan menggunakan sensor DHT11 perekaman suhu yang lebih akurat dengan persentase kesalahan sekitar 1.321%. Jika dengan LM35 terjadinya persentase kesalahan sebesar 4,332%. Sensor suhu ini bekerja dengan periode sensing setiap 500mili ss, sehingga sistem Smart Home ini dapat merekam perubahan suhu dalam waktu yang singkat atau kurang dari 1 detik. Pengujian sensor PIR dilakukan dengan menguji keberadaan orang dalam suatu ruangan. Tabel 2 di bawah ini adalah hasil pengujian sensor PIR.

Tabel 2 Hasil Pengujian respons PIR [6]

Percobaan Ke-	Waktu respon (detik)	
	tidak ada > ad ada > tidak ada	
1	2.82	2.8
2	2.8	2.56
3	2.9	2.65
4	2.85	2.78
5	2.77	2.46
6	2.9	2.89
7	2.85	2.75
8	2.75	2.76
9	2.89	2.45
10	2.9	2.54
Rata-Rata	2.843	2.664

Tabel 3 Hasil Pengujian jarak jangkauan sensor PIR [6]

No	Range (m)	Status
1	0.5	Detected
2	1	Detected
3	1.5	Detected
4	2	Detected
5	2.5	Detected
6	3	Detected
7	3.5	Detected
8	4	Detected
9	4.5	Detected
10	5	Detected

Dari hasil pengujian sensor PIR, dapat disimpulkan bahwa sensor bekerja dengan baik untuk mendeteksi orang di ruangan dengan waktu 2,82 detik, dengan jangkauan maksimum 5 Sensor limit switch digunakan untuk fungsi switching beberapa perangkat elektronik, seperti lampu, televisi, AC, pompa dan lain-lain.

Tabel 4 Hasil Pengujian Limit Switch [6]

No	Keadaan	R1		R2		R3		Waktu respon (rata rata)
		L1	L2	L1	L2	L1	L2	
1	L1,R1 on	on	-	-	-	-	-	1s
2	L1,R2 on	-	-	on	-	-	-	1s
3	L1,R3 on	-	-	-	-	on	-	1s
4	L2,R1 on	-	on	-	-	-	-	1s
5	L2,R2 on	-	-	-	on	-	-	1s
6	L2,R3 on	-	-	-	-	-	on	1s
7	L2,R2 on, L2,R3	-	-	-	on	-	on	1s
8	L2,R1 on, L2,R2 on, L2,R3 on	-	on	-	on	-	on	2s
9	L2,R1 on, L2,R2 on, L2,R3 on, L1, R1 on	on	on	-	on	-	on	2s

Berdasarkan hasil di atas, dapat disimpulkan bahwa alat pemantauan dan kontrol berhasil sehingga respons kurang dari 1 detik Namun, jika perintah diberikan secara bersamaan, sistem memerlukan sekitar 2 detik guna mengolah perintah tersebut.

Tabel 5 Hasil Pengujian Komunikasi nirkabel [6]

p-ISSN: 2301-5373
e-ISSN: 2654-5101

No	Jarak antar node	Hasil	Waktu respon (rata-rata)
1	1 meter	OK	1s
2	5 meter	OK	1s
3	10 meter	OK	1s
4	20 meter	OK	1s
5	25 meter	-	1.5s

Berdasarkan data pengujian komunikasi, di mana pada prototype ini digunakan tipe Xbee series2, diperoleh jarak maksimum antar node yakni sekitar 20meter dengan waktu respons komunikasi sekitar 1-1.5 detik. Hal ini dipengaruhi oleh keadaan/letak ruangan di dalam rumah (indoor), di mana sinyal lebih sulit mendapat kondisi line of sight sehingga jarak jangkauan menjadi lebih terbatas.

Tujuan Pengujian sistem sensor yaitu berfungsi untuk menguji komunikasi nirkabel, untuk mengetahui respons status pemantauan dan kontrol dan jarak antara node penerima pemancar. Sistem ini disimulasikan seperti yang pada Gambar 2.

Gambar 3 Tampilan Sistem Smart Home Model [6]

4. Kesimpulan

Rancangan sistem Smart Home berbasis Wireless Sensor Networks telah dapat digunakan pada rumah model. sistem Smart Home dengan berbasis WSN dapat diimplementasikan dengan berbagai multimedia data sensor, yakni untuk satu node sensor dapat membawa 4 data sensor. Dalam pengujian pada bagian kinerja Smart Home dapat dimulai dengan pengujian keakurasian atau keakuratan pada masing-masing data sensor sehingga waktu respons komunikasi antara sensor ke pusat monitoring dan sebaliknya. Sensor PIR berhasil mendeteksi keberadaan orang di suatu ruangan dengan waktu delay 2.8 detik.

Referensi

[1] S. Nurulita, “Pengaruh Perkembangan Ilmu Pengetahuan dan Teknologi Terhadap Masyarakat dan Lingkungan,” *Humaniora*, vol. 11, no. 1, pp. 37–48, 2012, [Online]. Available: https://journal.ugm.ac.id/jurnal-humaniora/article/view/623.

[2] O. N. Samijayani and I. Fauzi, “Perancangan Smart Home Berbasis Jaringan Sensor Nirkabel,” *J. Al-AZHAR Indones. SERI SAINS DAN Teknol.*, vol. 3, no. 2, p. 76, 2017, doi: 10.36722/sst.v3i2.188.

[3] A. Aziz, M. H. A. Wahab, A. Mustapha, and M. F. M. Mohsin, “Design and development of Smart Home security system for disabled and elderly people,” *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 3–7, pp. 135–138, 2017.

[4] A. Pathak and U. Sundaram, “Smart Locking System For Home,” *Smart Locking Syst. Home* , vol. 9, no. 21, pp. 83–86, 2016, doi: 10.14257/ijsia.2015.98.05.

[5] O. Taiwo, A. E. Ezugwu, O. N. Oyelade, and M. S. Almutairi, “Enhanced Intelligent Smart Home Control and Security System Based on Deep Learning Model,” *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9307961.

[6] O. Nur Samijayani and I. Fauzi, “Perancangan Smart Home Berbasis Jaringan Sensor Nirkabel,” 2015.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pencarian Layout Keyboard dengan Travel Distance Terkecil Menggunakan Algoritma Genetik

I Nengah Oka Darmayasa[a1], Ngurah Agus Sanjaya ER[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]okadarmayasa10@gmail.com
[2] agus_sanjaya@unud.ac.id

Abstract

Mengetik merupakan kegiatan yang selalu dilakukan beberapa kalangan masyarakat. Salah satu parameter untuk mengetik dengan efisien adalah meminimalkan pergerakan jari untuk mengetik suatu huruf, yang disebut sebagai *travel distance*. Metode yang dapat digunakan untuk mencari *layout keyboard* dengan *travel distance* terkecil secara efisien adalah dengan men-*generate layout* menggunakan algoritma genetika. Dengan menggunakan algoritma genetika, setiap huruf pada *keyboard* akan berperan sebagai gen, *layout keyboard* berperan sebagai kromosom, dan kumpulan beberapa *layout keyboard* sebagai populasi. Setiap kromosom akan dilakukan kawin silang dengan kromosom lain dengan mengambil beberapa gen dari masing-masing kromosom lalu menggabungkannya menjadi kromosom baru. Cara yang digunakan untuk menentukan kromosom mana yang digunakan untuk melakukan kawin silang dapat menggunakan *travel distance* dari tiap kromosom. Semakin kecil *travel distance,* maka semakin besar kemungkinan kromosom tersebut untuk dikawinsilangkan. *Keyboard* yang paling banyak digunakan saat ini masihlah *keyboard* dengan *layout* bernama "QWERTY". Saat dibandingkan untuk mengetik data kalimat berbahasa indonesia dari TED-Multilingual-Parallel-Corpus, *keyboard* QWERTY memiliki *travel distance* yang lebih besar, sekitar 9.311,735, jika dibandingkan dengan *layout keyboard* yang didapat dari metode algoritma genetika yang memiliki *travel distance* sebesar 5.747,246, sekitar 62% lebih besar.

Keywords: *Genetic Algorithm, Keyboard Layout, Keyboard, Machine Learning, Algoritma Genetik*

1. Introduction

Mengetik merupakan kegiatan yang semakin banyak dilakukan orang-orang terutama setelah adanya pandemi Covid-19 yang membuat kebanyakan orang beraktivitas secara daring [1]. Dengan bertambahnya waktu yang digunakan untuk beraktivitas secara daring, tentu akan meningkatkan kesempatan untuk mengetik [2]. Bagi orang-orang yang menghabiskan kebanyakan waktunya mengetik, mengetahui teknik baru untuk mengetik dengan lebih efisien tentunya merupakan sesuatu yang diinginkan. Salah satu cara yang dapat dilakukan untuk mengefisienkan kegiatan mengetik adalah dengan mengganti *layout* dari *keyboard* yang digunakan menjadi *layout* yang memiliki *travel distance* terkecil, yaitu sebuah skor yang menandakan panjangnya pergerakan jari untuk mengetik suatu teks.

Banyaknya karakter yang ada di dalam sebuah *keyboard* dapat menyebabkan ketidakefisienan proses mengetik apabila penempatan setiap karakter berada pada posisi yang tidak optimal. Misalkan saja huruf-huruf yang sering digunakan untuk mengetik diletakkan pada tempat yang sulit terjangkau akan menyebabkan performa mengetik yang lebih rendah dibandingkan apabila seandainya karakter-karakter tersebut diletakkan di tempat yang mudah dijangkau.

Pada penelitian [3], mencari *layout keyboard* yang ergonomik dengan melihat frekuensi munculnya 3.000 kata paling umum dalam bahasa Inggris menggunakan algoritma genetika. Pada penelitian [4], mencari *layout keyboard* untuk bahasa Portugis menggunakan algoritma genetika. Dari kedua penelitian tersebut, hasil yang didapatkan adalah *layout keyboard* QWERTY memiliki performa, dengan parameter tertentu pada masing-masing penelitian, yang lebih rendah kualitasnya dibandingkan dengan *layout keyboard* hasil pencariannya.

Penelitian [5] dan [6] merupakan penelitian yang mencari sebuah *layout keyboard* lalu membandingkan *travel distance*-nya terhadap *layout keyboard* QWERTY dan Dvorak. Metode yang digunakan pada [6] adalah dengan algoritma genetika. Penghitungan *travel distance* yang digunakan adalah dengan

menghitung Euclidean Distance dari posisi jari sekarang menuju posisi huruf yang jari tersebut perlu tekan. Dari percobaannya, didapat hasil sebuah *layout keyboard* yang 32,66% lebih baik dibanding *layout keyboard* QWERTY dan 15,78% lebih baik dibandingkan *layout keyboard* Dvorak.

2. Reseach Methods

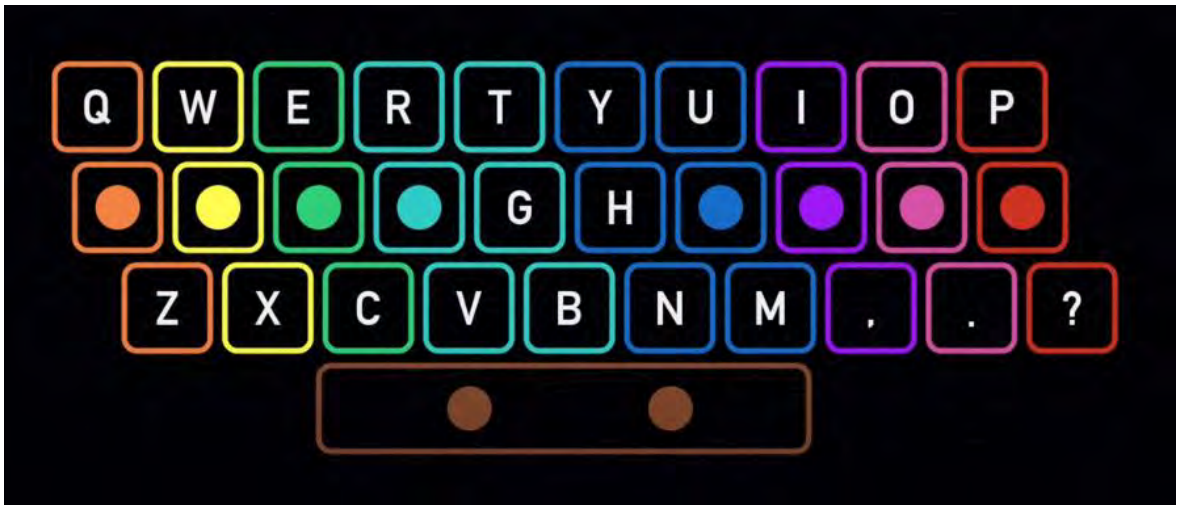
2.1 Identifikasi Masalah

Dalam penelitian ini, masalah yang ingin diselesaikan adalah mengenai pencarian sebuah *layout keyboard* yang cocok digunakan untuk mengetik kalimat-kalimat dalam bahasa Indonesia. Cara yang digunakan untuk mencari *layout keyboard* tersebut adalah dengan menggunakan algoritma genetika.

Mengetik 10 jari merupakan sebuah teknik mengetik dengan menggunakan 10 jari yang setiap jarinya berperan menekan huruf tertentu. Pada *keyboard* QWERTY, setiap jari memiliki area jangkauannya masing-masing, yaitu:

a. Kelingking kiri = q, a, z
b. Manis kiri = w, s, x
c. Tengah kiri = e, d, c
d. Telunjuk kiri = r, f, v, t, g, b
e. Jempol kiri = spasi
f. Jempol kanan = spasi
g. Telunjuk kanan = y, h, n, u, j, m
h. Tengah kanan = i, k, ,
i. Manis kanan = o, l, .
j. Kelingking kanan = p, ;, /

Setiap jari ini memiliki posisi awalnya masing-masing, misalnya pada *layout* QWERTY, seperti jari kelingking kiri, manis kiri, tengah kiri, telunjuk kiri, jempol kiri, jempol kanan, telunjuk kanan, tengah kanan, manis kanan, dan kelingking kanan terletak ada huruf a, s, d, f, spasi, spasi, j, k, l, dan ; secara terurut. Artinya, saat ingin mengetik kata “makan”, maka jari-jari yang perlu bergerak adalah jari-jari yang area jangkauannya ada pada setiap huruf pada kata “makan”, yaitu jari telunjuk kanan untuk huruf “m” yang bergerak ke bawah, jari kelingking kiri untuk huruf “a” dengan tidak bergerak, jari tengah kanan untuk huruf “k” dengan tidak bergerak, dan jari telunjuk kanan untuk huruf “n” yang bergerak ke kiri karena sebelumnya sedang berada di huruf “m” yang posisinya di sebelah kanan huruf “n”.



Gambar 1. Area Jangkauan Tiap Jari Pada *Layout* QWERTY

Namun, tidak semua *layout keyboard* mengharuskan posisi tiap jari pada huruf tertentu. Yang dijadikan patokan adalah posisinya pada *layout* tersebut, bukan hurufnya. Jadi, pada dasarnya, setiap jari akan memiliki area jangkauan sesuai dengan warna-warna yang berbeda seperti pada gambar 1.

Setiap pergerakan akan dikenai bobot yang berbeda-beda tergantung gerakannya. Nantinya, jumlah dari semua bobot untuk mengetik suatu teks dengan *layout keyboard* tertentu tersebut yang akan dijadikan *fitness value* dalam algoritma genetika.

Algoritma genetika dapat digunakan untuk mencari *layout keyboard*. Dalam algoritma genetika, akan ada beberapa komponen, yaitu gen, kromosom, populasi, dan *fitness value*. Kumpulan gen akan membentuk sebuah kromosom. Dalam hal ini, gen merupakan setiap huruf pada *layout* dan kromosom merupakan *layout keyboard*. Kumpulan dari beberapa kromosom dinamakan sebagai populasi.

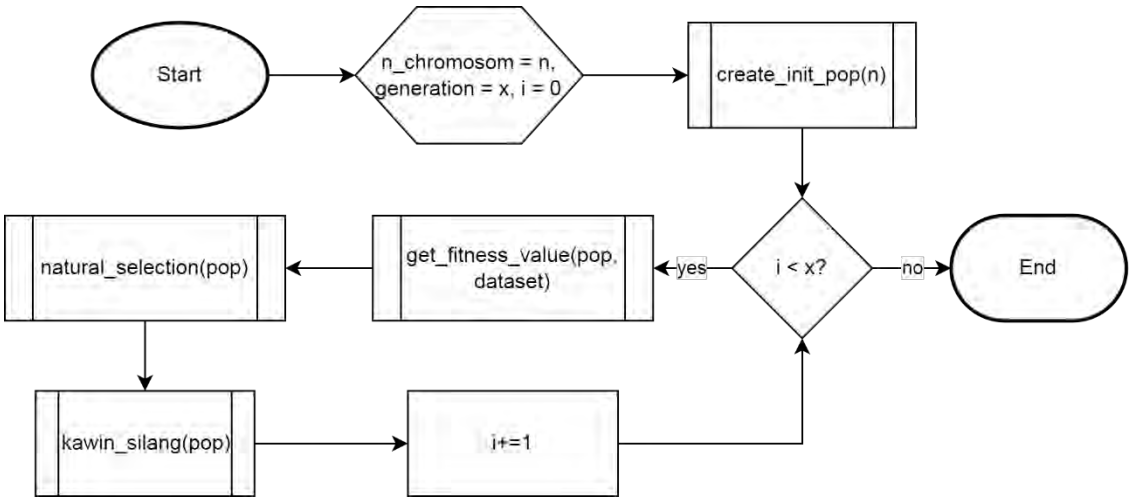
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Sedangkan *fitness value* merupakan sebuah skor yang dimiliki oleh sebuah kromosom yang digunakan untuk menentukan kemungkinan kromosom tersebut untuk tetap "hidup" dalam rangkaian alur algoritma genetik ini. Dalam kasus ini, *fitness value* yang digunakan adalah *travel distance* dari setiap kromosom saat digunakan untuk mengetik sebuah teks tertentu.

2.2 Data Penelitian

Penelitian ini akan menggunakan data berupa kumpulan kalimat dalam bahasa Indonesia yang didapat dari TED-Multilingual-Parallel-Corpus. Dataset yang digunakan terdiri dari 2.000 kata.

2.3 Desain Penelitian



Gambar 2. Flowchart Algoritma Genetik

Penelitian akan dimulai dari pengambilan data. Data yang telah didapat lalu dilakukan *preprocessing* seperti menghilangkan beberapa karakter-karakter yang tidak penting seperti angka, "-", "+", "=", "_", "[", "]", "{", "}", "\", dan "|". Setelah dilakukan *preprocessing,* maka selanjutnya adalah menentukan jumlah kromosom yang ada di dalam sebuah populasi dan juga menentukan banyak generasi. Selanjutnya adalah membuat populasi pertama secara *random* yang terdiri dari n kromosom. Lalu akan dilakukan pengulangan sebanyak generasi dengan pertama mencari *fitness value* tiap kromosom-nya dengan menggunakan data yang telah disiapkan sebelumnya. Setelah itu, 10% kromosom yang memiliki *fitness value* terkecil akan tetap berada di dalam populasi tersebut. Sedangkan, 90% kromosom sisanya akan dilakukan kawin silang dengan membagi gen dalam kromosom menjadi 2 bagian lalu menggabungkan masing-masing bagian dengan bagian lain dari kromosom yang lain. Dari proses ini akan menghasilkan kromosom baru. Hasil dari kromosom baru ini akan memiliki *trait* bagus yang didapat dari *parent*-nya, yaitu susunan huruf-huruf dalam *layout keyboard* yang berakibat pada *fitness value* dari kromosom-kromosom hasil kawin silang ini cenderung bagus, bahkan lebih bagus dari *parent*-nya. Selain kawin silang ini, pada penelitian ini juga menerapkan sebuah konsep mutasi, yaitu dalam persentase 10%, akan terjadi mutasi yang akan menukarkan satu pasang gen dari sebuah kromosom. Jumlah kromosom dalam suatu populasi akan selalu tetap. Karena dari kawin silang ini menghasilkan kromosom baru, maka ada kromosom yang perlu dibuang, yaitu kromosom yang tidak dikawinsilangkan yang akan digantikan posisinya dengan kromosom hasil kawin silang yang memiliki *fitness value* terbaik.

2.4 Algoritma Genetika Untuk Pencarian *Layout Keyboard*

Algoritma genetika dapat digunakan untuk melakukan pencarian *layout keyboard* yang efektif. Konsep utama yang ada pada algoritma ini adalah dengan melakukan kawin silang dan seleksi alam terhadap suatu komponen yang disebut sebagai kromosom [7]. Di dalam setiap kromosom terdiri dari beberapa bagian kecil yang dinamakan gen. Nantinya, setiap kromosom ini akan memiliki sebuah *fitness value*, yaitu sebuah skor yang akan menentukan kemungkinan kromosom ini untuk diseleksi. Yang dimaksud dengan seleksi adalah pengeluaran sebuah kromosom dari suatu kumpulan kromosom yang dinamakan populasi. Semakin kecil *fitness value* sebuah kromosom, maka semakin besar peluang kromosom tersebut untuk keluar dari sebuah populasi. Dari kromosom yang berhasil untuk hidup, akan dilakukan kawin silang antar kromosom tersebut. Perkawinan silang akan menggunakan beberapa gen yang diambil dari 2 kromosom, yang disebut *parent,* yang nantinya akan digabung untuk membentuk suatu kromosom baru, yang disebut *child*.

2.5 Gen

Gen merupakan bagian yang membentuk sebuah kromosom. Dalam teori evolusi, gen ini yang membawa *trait* yang dimiliki oleh sebuah spesies. Sama halnya pada kasus ini, gen akan membawa kualitas yang memungkinkannya untuk tetap berada di dalam populasi, yaitu memiliki kualitas untuk memiliki *fitness value* yang paling baik (kecil). Dalam kasus ini, yang menjadi gen adalah setiap huruf yang ada di dalam suatu *layout keyboard*.

2.6 Kromosom

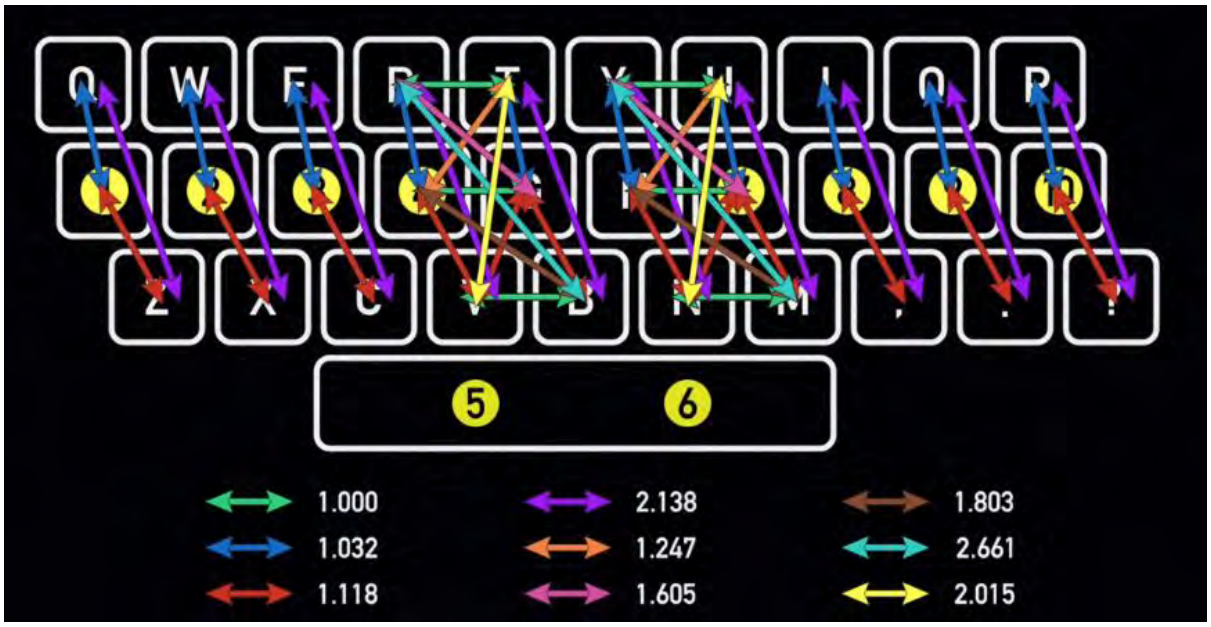
Kromosom merupakan kumpulan dari beberapa gen. Dalam kasus ini, yang menjadi kromosom dari sebuah *layout keyboard*. Nantinya, setiap kromosom ini akan dihitung *fitness value*-nya lalu dibandingkan dengan *fitness value* dari kromosom lain di dalam populasi yang sama. Setelah itu, kromosom yang memiliki *fitness value* yang baik akan dikawin silang dengan kromosom lain dengan membagi gennya menjadi 2 bagian. Hal ini akan membentuk kromosom baru dengan *trait* berkualitas baik yang didapat dari gen *parent*-nya.

2.7 Populasi

Populasi merupakan kumpulan kromosom. Jumlah kromosom di dalam suatu populasi akan selalu tetap. Hal ini akan memungkinkan terjadinya seleksi alam, yaitu keluarnya kromosom yang tidak mampu bersaing dari populasi dan digantikan dengan kromosom lain yang mampu bersaing. Indikator mampu bersaing yang digunakan adalah dengan membandingkan kromosom mana yang memiliki *fitness value* yang lebih baik. Semakin kecil *fitness value* maka semakin baik kualitasnya yang berujung pada semakin kecil kemungkinan kromosom tersebut untuk keluar dari populasi.

2.8 Perhitungan *fitness value*

Fitness value yang akan digunakan adalah besarnya bobot pergerakan jari tangan untuk mengetik dengan 10 jari. Untuk mengetik dengan 10 jari, setiap jari bertugas untuk menekan tombol-tombol tertentu. Karena jumlah tombol yang perlu ditekan lebih banyak dibandingkan dengan jumlah jari yang ada, maka tentunya jari-jari akan bergerak untuk menekan tombol-tombol yang menjadi area tugasnya.



Gambar 3. Bobot Pergerakan Tiap Jari

Posisi awal setiap jari ditandakan pada lingkaran kuning. Setiap jari, kecuali jempol dan telunjuk, memiliki area jangkauan vertikal, yaitu ke atas, ke bawah, atau diam di tempat. Untuk jempol selalu bertugas untuk menekan tombol spasi. Sedangkan jari telunjuk memiliki area jangkauan yang lebih luah dibandingkan jari lainnya. Hal ini karena posisinya yang selalu berada di tengah dan juga yang paling fleksibel. Jari telunjuk bertugas untuk menekan tombol dengan area yang lebih luas yang menyebabkan jari telunjuk memiliki *range of motion* yang lebih beragam dibandingkan jari lainnya.

Gambar di atas merupakan bobot yang akan digunakan dalam setiap pergerakan jari. Baris paling bawah *keyboard* biasanya sedikit lebih ke kanan dibandingkan dengan baris paling atas relatif terhadap baris tengah, maka bobot untuk bergerak secara vertikal dari baris paling bawah akan memiliki bobot yang lebih besar dibandingkan dengan baris paling atas.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Misalkan $l1, l2, l3, l4, l5$ merupakan jari kelingking kiri, manis kiri, tengah kiri, telunjuk kiri, dan jempol kiri, begitu juga $k1, k2, k3, k4, k5$ untuk tangan kanan, maka untuk mencari *travel distance* sebuah *layout keyboard* saat digunakan untuk mengetik suatu teks F adalah seperti berikut:

$$F = f(l1) + f(l2) + f(l3) + f(l4) + f(l5) + f(k1) + f(k2) + f(k3) + f(k4) + f(k5)$$

Dengan f merupakan fungsi yang menghitung jumlah dari bobot pergerakan setiap jari selama mengetik sebuah teks.

3. Result and Discussion

3.1 Hasil

Dalam pengimplementasiannya, penelitian ini menggunakan populasi dengan jumlah kromosom sebanyak 10 dan iterasi generasinya sebanyak 10.000 kali. Percobaan ini dilakukan sebanyak 2 kali.

```
w   j   b   /   ;   p   y   m   k   .
n   e   i   h   l   d   u   r   s   a
z   c   o   t   g   ,   f   x   v   q
bobot =  5769.054
q   w   e   r   t   y   u   i   o   p
a   s   d   f   g   h   j   k   l   ;
z   x   c   v   b   n   m   ,   .   /
bobot =  9311.735
```

Gambar 4. Hasil Percobaan Pertama

Seperti yang dapat dilihat pada gambar 3, didapatkan hasil sebuah *layout keyboard* yang memiliki bobot sekita 5.769,054 saat digunakan untuk mengetik teks yang ada dalam dataset yang digunakan. Sedangkan, di bawahnya merupaakn perbandingan pada *layout keyboard* QWERTY yang memiliki bobot sekitar 9.311,735 untuk mengetik dataset yang sama.

```
n   k   i   g   p   y   d   o   m   .
,   s   b   t   h   u   l   e   r   a
x   v   ;   w   z   f   /   j   q   c
bobot =  5747.246

q   w   e   r   t   y   u   i   o   p
a   s   d   f   g   h   j   k   l   ;
z   x   c   v   b   n   m   ,   .   /
bobot =  9311.735
```

Gambar 5. Hasil Percobaan Kedua

Pada percobaan kedua, yang dapat dilihat pada gambar 4, terlihat hasil *layout keyboard* yang didapatkan memiliki perbedaan yang sangat sedikit dari segi bobot dibandingkan dengan hasil *layout keyboard* pada percobaan pertama. Hal ini karena dari algoritmanya sendiri bersifat konvergen, yaitu akan mengerucut ke suatu hasil tertentu, dalam hal ini *layout keyboard*.

3.2 *Layout Keyboard*

```
w   j   b   /   ;   p   y   m   k   .
n   e   i   h   l   d   u   r   s   a
z   c   o   t   g   ,   f   x   v   q
bobot =  5769.054
n   k   i   g   p   y   d   o   m   .
,   s   b   t   h   u   l   e   r   a
x   v   ;   w   z   f   /   j   q   c
bobot =  5747.246
```

Gambar 6. *Layout Keyboard* yang Dihasilkan

Gambar di atas merupakan 2 *layout keyboard* yang didapat menggunakan algoritma genetika.

3.3 Analisis *Layout Keyboard*

Jika diperhatikan, kedua *layout keyboard* tersebut tidak memiliki kesamaan yang jelas terlihat, seperti huruf-huruf yang sama berada pada posisi yang sama. Hal ini terjadi karena alur dari program yang dibuat yang tidak melakukan pengembalian posisi jari ke posisi awal disaat mengetik. Artinya, dalam hasil tersebut beranggapan bahwa pengetikan setiap kata dilakukan terus menerus dengan posisi jari akan selalu berada di posisi di mana jari tersebut terakhir menekan sebuah tombol.

Selain itu, penyabab lain adalah karena indikator *fitness value* yang dipakai dalam penelitian ini hanya menggunakan faktor *travel distance*. Hal ini menyebabkan bahwa kedua *layout keyboard* tersebut saat digunakan untuk mengetik, akan memiliki *travel distance* yang lebih kecil dibandingkan dengan *layout keyboard* QWERTY.

3.4 Diskusi

Pada penelitian ini hanya berfokus pada pencarian *layout keyboard* yang memiliki *travel distance* terkecil untuk mengetik kalimat berbahasa Indonesia. Masih banyak faktor lain yang dapat digunakan untuk mengefisienkan proses mengetik, seperti bobot pada setiap jari yang berbeda tergantung dari seberapa mudah jari tersebut digerakkan, metode mengetik yang memanfaatkan penggunaan 2 tangan secara bergantian, atau bisa juga dengan meminimalkan *type jamming*, yaitu keadaan saat mengetik yang menyebabkan posisi jari tangan berada di posisi yang tabrakan atau posisi yang tidak nyaman. Faktor-faktor tersebut dapat dijadikan pertimbangan pada penelitian-penelitian selanjutnya.

4. Conclusion

Setelah didapat *layout keyboard* dengan menggunakan metode algoritma genetika, performa *layout keyboard* QWERTY dalam hal *travel distance* ternyata bernilai 62% lebih besar dibandingkan dengan *layout keyboard* yang dihasilkan. Hal ini berarti bahwa *layout keyboard* yang dihasilkan lebih efisien dalam hal *travel distance* dibandingkan dengan *layout* QWERTY.

References

[1] O. B. Adedoyin, “Covid-19 pandemic and online learning: the challenges and opportunities,” *Interactive Learning Environments,* pp. 1-13, 2020.

[2] S. Fung Lau dan C. B. H. Chan, “A special virtual keyboard for disabled computer users,” *Proceedings of the International MultiConference of Engineers and Computer Scientists 2017, IMECS 2017,* pp. 539-544, 2017.

[3] A. Onsorodi dan O. Korhan, “Application of a genetic algorithm to the keyboard layout problem,” *PLOS ONE,* vol. 15, p. e0226611, 2020.

[4] G. Pacheco, E. Palmeira dan K. Yamanaka, “Using Genetic Algorithms to Design an Optimized Keyboard Layout for Brazilian Portuguese,” *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2020),* pp. 437-448, 2020.

[5] M. Mortenson dan C. S. Sweeten, “Novel Keyboard Layout Optimization,” 2020.

[6] A. Fadel, I. Tuffaha, M. Al-Ayyoub dan Y. Jararwch, “QWERTY Keyboard? }.?BZQ is Better!,” *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA),* pp. 81-86, 2020.

[7] M. Mitchell, An Introduction to Genetic Algorithms, Cambridge, MA, USA: MIT Press, 1998.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Perbandingan Kriptografi Klasik Hill Cipher dengan Affine Cipher dalam Pengamanan Data Citra

I Nyoman Dwi Pradnyana Putra[a1], I Gede Santi Astawa[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Badung, Indonesia
[1]dwipradnyana68@gmail.com
[2]santi.astawa@unud.ac.id

Abstract

The development of technology affects several aspects of life, especially in securing confidential data and information. Because of this, there is a way to secure data, namely cryptography, many cryptography techniques have been implemented to hide data information, one of which image data, the purpose of securing image data is to prevent unwanted things. In this research, two classical cryptography algorithms will be tested, namely hill cipher and affine cipher by comparing the MSE and PSNR Image values when encryption is carried out. The encryption process is done 6 times with different image in each algorithm, the results of the encryption are compared between the results of the hill cipher encryption image with the affine cipher. After test that two of algorithms can change the encrypted image into a random pattern, but seen from the MSE and PSNR image values, the hill cipher algorithm has a value that matcher the requirements, with the best value being MSE 7073.40 and PSNR image 9.63 dB.

Keywords: *Cryptography, Encryption, Hill Cipher, Affine Cipher, MSE, PSNR Image*

1. Pendahuluan

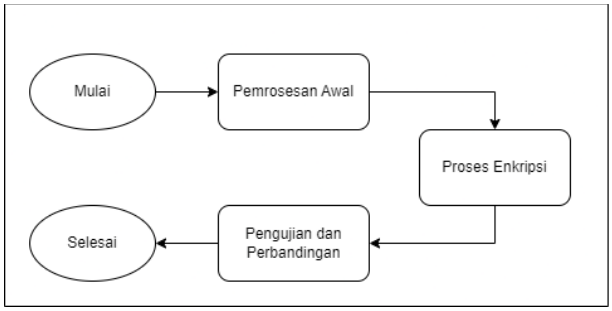
Perkembangan teknologi informasi saat ini sangat berpengaruh dalam segala aspek kehidupan, terutama dalam pengamanan data dan informasi yang bersifat rahasia. Informasi yang bersifat rahasia seharusnya perlu disembunyikan agar tidak diketahui oleh orang-orang yang tidak berhak agar tidak disalahgunakan. Karena hal tersebut diperlukan sebuah cara yang dilakukan untuk mengamankan data, salah satunya dengan menerapkan kriptografi. Kriptografi merupakan ilmu yang mempelajari teknik-teknik matematika yang berhubungan dengan aspek keamanan informasi seperti kerahasiaan, integritas data, serta autentikasi dimana proses penggunaanya memiliki berbagai teknik dan atau ilmu dan seni untuk menjaga keamanan data[1].

Banyak teknik kriptografi yang telah diimplementasikan untuk menyembunyikan informasi data, salah satunya data citra, tujuan dari pengamanan data citra yaitu agar tidak terjadi hal yang tidak diinginkan seperti penipuan dengan menggunakan identitas orang lain yang didukung dengan foto pribadi. Masih ada yang harus dipertimbangkan ketika menggunakan suatu algoritma baik itu dari tingkat kecepatan enkripsi sampai dengan seberapa baik algoritma dalam menyembunyikan informasi.

Dalam penelitian ini, peneliti ingin membandingkan dua buah algoritma yaitu hill cipher dan affine cipher. Algoritma hill cipher merupakan algoritma yang menggunakan matriks m x m yang menggunakan operasi perkalian matriks untuk proses enkripsi maupun dekripsi, sedangkan algoritma affine cipher merupakan perluasan dari algoritma caesar cipher dengan perkalian dan pergeseran nilai sebagai kunci. Kedua algoritma ini merupakan algoritma kriptografi klasik simetris dimana menggunakan kunci yang sama untuk proses enkripsi dan dekripsi, peneliti menggunakan kriptografi klasik untuk mengetahui seberapa baik kedua algoritma ini dalam menyembunyikan informasi.

Kedua algoritma tersebut dibandingkan dari nilai pengujian MSE dan PSNR. Mean Square Error (MSE) adalah nilai error kuadrat rata-rata antara citra asli dengan citra enkripsi, MSE didapatkan dengan membandingkan nilai selisih *pixel* citra asli dengan citra enkripsi. Peak Signal to Noise Ratio *Image* adalah perbandingan antara nilai maksimum dari sinyal yang diukur berdasarkan derau yang dinyatakan dalam desibel (dB).

2. Metode Penelitian

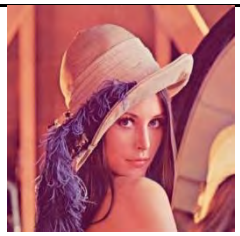


Gambar 1. Desain Metode

2.1. Pemrosesan Awal

Algoritma kriptografi yang digunakan dalam pengujian yaitu algoritma hill cipher dan affine cipher, dimana kedua algoritma ini merupakan algoritma simetris yang berarti menggunakan kunci yang sama dalam proses enkripsi dan dekripsi. Citra yang akan digunakan untuk pengujian dalam penelitian ini yaitu citra digital RGB yang berjumlah 5 citra dengan dengan ukuran yang dapat dilihat pada **Tabel 1**. Citra yang digunakan. Ukuran citra sengaja dibedakan agar dapat terlihat berapa rata-rata *pixel* yang mampu ditampung oleh citra tersebut[2].

Tabel 1. Citra yang digunakan

Nama Citra Uji	Gambar	Ukuran Citra (*pixel*)
windows.png		720 x 720
lenna.png		512 x 512

p-ISSN: 2301-5373
e-ISSN: 2654-5101

bunga.png		500 x 500
buah.png		480 x 480
hewan.png		400 x 400

2.2. Enkripsi Citra

a. Algoritma Hill Cipher

Pada tahap ini citra di enkripsi menggunakan algoritma Hill Cipher. Pada tahap ini dibutuhkan sebuah kunci dimana kunci tersebut terdiri dari 42 karakter yaitu 23 yang akan diubah menjadi nilai ASCII. Lalu ubah kunci menjadi matriks 2x2[3]. Berikut adalah langkah detil dari proses enkripsi citra dengan algoritma Hill Cipher.

1. Baca citra yang akan dienkripsi.
2. Ubah citra menjadi nilai double.
3. Inputkan kunci berupa 2 karakter.
4. Ubah kunci menjadi matriks 2x2.
5. Lakukan proses enkripsi dengan persamaan 1.
6. Dapatkan citra terenkripsi.

$$C = (K \times P) \bmod 256 \quad (1)$$

b. Algoritma Affine Cipher

Affine cipher pada metode affine adalah perluasan dari metode Caesar Cipher, yang mengalihkan plainteks dengan sebuah nilai dan menambahkannya dengan sebuah pergeseran P menghasilkan cipherteks C dinyatakan dengan fungsi kongruen[4]:

$$C \equiv m\,P + b \pmod n \quad (2)$$

Yang mana n adalah ukuran alphabet, m adalah bilangan bulat yang harus relatif prima dengan n (jika tidak relatif prima, maka dekripsi tidak bisa dilakukan) dan b adalah jumlah pergeseran (Caesar cipher adalah khusus dari affine cipher dengan m=1). Untuk melakukan deskripsi, persamaan tersebut harus dipecahkan untuk memperoleh P. Solusi kekongruenan tersebut hanya ada jika inver m (mod n), dinyatakan dengan m -1. Jika m -1 ada maka dekripsi dilakukan dengan persamaan sebagai berikut:

$$P \equiv m\ -1(C - b) \pmod{n} \quad (3)$$

2.3. Parameter Pengujian

a. Mean Square Error (MSE)

MSE merupakan tolak ukur analisis kuantitatif yang digunakan untuk menilai kualitas sebuah citra keluaran dan keunggulan sebuah metode yang digunakan. Ukuran matriks citra m * n, B1 dan B2 merupakan matriks citra. Dengan kata lain Mean Square Error (MSE) adalah kesalahan kuadrat rata- rata sinyal-sinyal *pixel* citra hasil pemrosesan sinyal terhadap sinyal asli[2]. Untuk nilai terbaik MSE dalam segi enkripsi adalah semakin besar nilai dari MSE. Untuk menentukan nilai MSR digunakan persamaa berikut ini:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \quad \sum_{j=0}^{n-1} \quad (B_1(i,j) - (B_2(i,j)) \quad (4)$$

Dimana:

m : baris matriks citra hasil pemrosesan
n : kolom matriks citra hasil pemrosesan
B_1 : *pixel* citra asli
B_2 : *pixel* citra hasil pemrosesan

b. Peak Signal to Noise Ratio (PSNR) Image

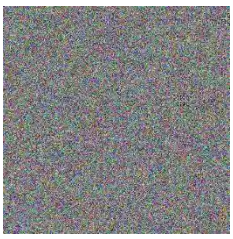
Peak Signal to Noise Ratio (PSNR) adalah perbandingan antara nilai maksimum dari sinyal yang diukur dengan besarnya derau yang berpengaruh pada sinyal tersebut, dalam satuan desibel (dB)[2]. Untuk nilai PSNR terbaik dalam segi enkripsi adalah semakin kecil nilai dari PSNR yang mana semakin tidak mirip dengan citra asli. Untuk menentukan nilai PSNR digunakan persamaan berikut ini:

$$PSNR = \frac{255^2}{MSE} \quad (5)$$

3. Hasil dan Pembahasan

3.1. Uji Coba Sistem

Dalam proses enkripsi citra menggunakan sistem yang sudah ada dengan menggunakan Bahasa pemrograman python, uji coba proses enkripsi dilakukan sebanyak 10 kali dengan algoritma hill cipher dan affine cipher yang menggunakan 5 citra yang berbeda pada setiap algoritma. Berikut merupakan hasil uji sistem proses enkripsi hill cipher dan affine cipher.


(a) encHill-windows.png


(b) encHill-lenna.png


(c) encHill-bunga.png

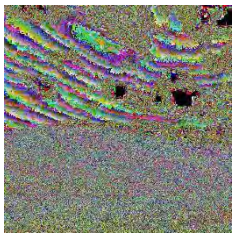
p-ISSN: 2301-5373
e-ISSN: 2654-5101

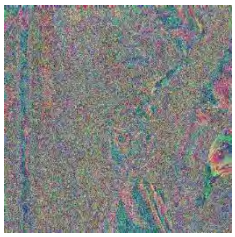

(d) encHill-buah.png

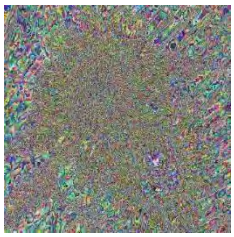

(e) encHill-hewan.png

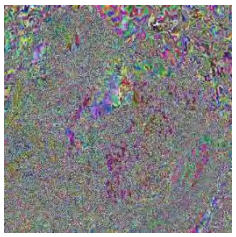
Gambar 2. Hasil Enkripsi Citra dengan Algoritma Hill Cipher

Dapat dilihat pada gambar diatas terjadi perubahan pola menjadi acak dan tidak terlihat bagaimana bentuk citra aslinya, algoritma hill cipher memiliki bentuk kemiripan yang berbeda dengan citra aslinya.


(a) encAffine-windows.png


(b) encAffine-lenna.png


(c) encAffine-bunga.png


(d) encAffine-buah.png


(e) encAffine-hewan.png

Gambar 3. Hasil Enkripsi Citra dengan Algoritma Affine Cipher

Citra hasil enkripsi menggunakan algoritma affine cipher memiliki perubahan pola yang acak, namun masih terlihat beberapa kesamaan antara citra terenkripsi dengan citra aslinya, dimana adanya sedikit kemiripan pada beberapa citra.

3.2. Perbandingan Hasil Nilai MSE dan PSNR

Untuk mendapatkan nilai MSE dan PSNR *Image* menggunakan sistem yang sudah ada dengan menggunakan bahasa pemrograman python, dimana program dalam sistem tersebut menggunakan menyesuaikan dengan perhitungan MSE pada persamaan (4) dan PSNR pada persamaan (5).

Tabel 2. Nilai MSE dan PSNR dengan Algoritma Hill Cipher

Nama Citra	Uji MSE	Uji PSNR (dB)
encHill-windows.png	6064.21	10.30
encHill-lenna.png	3715.03	12.43

encHill-bunga.png	5538.38	10.69
encHill-buah.png	7073.40	9.63
encHill-hewan.png	6891.99	9.74

Tabel 3. Nilai MSE dan PSNR dengan Algoritma Affine Cipher

Nama Citra	Uji MSE	Uji PSNR (dB)
encAffine-windows	5538.38	10.69
encAffine-lenna.png	4572.87	11.52
encAffine-bunga.png	5373.84	10.82
encAffine-buah.png	6917.35	9.73
encAffine-hewan.png	7029.37	9.66

Pada tabel diatas menunjukkan bahwa nilai MSE dan PSNR dari masing-masing algoritma memiliki perbedaan. Perbedaan nilai tersebut dipengaruhi oleh *pixel* yang ditampung oleh citra tersebut. Nilai MSE terbaik dalam proses enkripsi adalah semakin besar nilainya dan sedangkan nilai PSNR terbaik dalam proses enkripsi adalah semakin kecil nilainya. Pada tabel tersebut dapat dilihat bahwa dari semua nilai pengujian tersebut, algoritma hill cipher rata-rata mencapai nilai MSE dan PSNR yang sesuai dengan syarat pengujian dengan nilai terbaik MSE 7073.40 dan PSNR *image* 9.63 dB, sedangkan nilai yang paling jauh dengan syarat pengujian yaitu dengan MSE 3715.03 dan PSNR *image* 12.43 dB.

4. Kesimpulan

Berdasarkan hasil pembahasan dan penelitian yang telah dilakukan. dapat disimpulkan bahwa secara visual algoritma hill cipher menghasilkan citra enkripsi berupa pola acak dan tidak terlihat bentuk citra aslinya, sedangkan pada algoritma affine cipher menghasilkan pola acak namun pada beberapa citra terdapat beberapa kemiripan antara citra enkripsi dengan citra aslinya. Pada pengujian hasil enkripsi, algoritma yang mencapai syarat dari pengujian yaitu algoritma hill cipher dengan nilai terbaik MSE 7073.40 dan PSNR *image* 9.63 dB.

Referensi

[1] W. Sari *et al.*, “Analisa Algoritma Elgamal Dalam Penyandian Data,” vol. 2, no. 1, pp. 60–70, 2018.

[2] S. Y. Doo, S. Tena, and V. M. Ndolu, “Implementasi Pengamanan Data Menggunakan Metode Kriptografi Hill Cipher Dan Steganografi Least Significant Bit (Lsb) Pada Media Citra Digital,” *J. Media Elektro*, vol. VIII, no. 2, pp. 93–99, 2019, doi: 10.35508/jme.v0i0.1778.

[3] B. Firmanto, D. Putri, K. Ningrum, A. Bramanto, and W. Putra, “Perbandingan Hasil Performa Optimasi Transposisi Hill Cipher dan Vigenere Cipher pada Citra Digital,” *SMARTICS J.*, vol. 7, no. 2, pp. 65–71, 2021, [Online]. Available: https://doi.org/10.21067/smartics.v7i2.5931

[4] D. S. Ginting, J. Simanulang, and M. Zarlis, “Kriptografi Simetris Dengan Kombinasi Hill cipher Dan Affine Cipher Di Dalam Matriks Cipher Transposisi Dengan Menerapkan Pola Alur Bajak Sawah,” *Semin. Nas. Teknol. Inf.*, no. The Future of Computer Vision, pp. 191–198, 2017.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Long-Short Term Memory (LSTM) pada Klasifikasi Kategori Berita

Anak Agung Ngurah Andhika Satrya Nugraha[a1], Ida Bagus Made Mahendra[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bali, Indonesia
[1]andhikangurah@gmail.com
[2]ibm.mahendra@unud.ac.id

Abstrak

Pada era informasi ini, berita dan informasi baru menyebar dengan sangat cepat, terlalu banyak berita baru yang muncul setiap harinya. Oleh karena itu, diperlukan sebuah cara untuk memilah berita yang ingin dilihat. Salah satu cara untuk memilah berita adalah dengan membagi berita ke dalam beberapa kategori. Pembagian kategori pada berita masih dilakukan secara manual dengan memberi label kategori pada berita yang ingin diunggah. Penelitian ini membahas tentang implementasi metode Long-Short Term Memory (LSTM) untuk mengklasifikasikan berita berdasarkan judulnya ke dalam 7 kategori, yaitu finance, food, health, internet, otomotive, sport, dan travel. Terdapat dua model yang diimplementasikan pada penelitian ini, yaitu model LSTM dan model Bidirectional LSTM. Pengujian dilakukan dengan menggunakan nilai akurasi, yaitu perbandingan antara judul berita yang berhasil diklasifikasikan dengan keseluruhan judul berita. Berdasarkan hasil pengujian, didapatkan bahwa model LSTM yang dibuat berhasil mengklasifikasikan berita dengan akurasi sebesar 85.36%, model Bidirectional LSTM juga berhasil mengklasifikasikan berita dengan akurasi sebesar 84.15%.

Kata kunci: *LSTM, Bidirectional LSTM, Klasifikasi, Natural Language Processing, Berita*

1. Pendahuluan

Saat ini, informasi tidak dapat dipisahkan dari kehidupan manusia, sangat banyak berita dan informasi baru yang muncul setiap harinya. Karena terlalu banyaknya berita yang ada, diperlukan sebuah cara untuk memilah berita yang ingin dilihat setiap harinya, salah satu cara yang dapat digunakan adalah dengan membagi berita yang ada ke dalam beberapa kategori, sehingga pembaca dapat memilih kategori berita yang ingin dibaca.

Pembagian kategori berita biasanya dilakukan secara manual oleh pembuat berita saat berita akan diunggah, berita yang sudah diberikan kategori kemudian dapat dikelompokkan pada aplikasi dan pengguna aplikasi tersebut dapat memilih kategori berita yang ingin dibaca. Pemberian kategori berita secara manual adalah cara yang digunakan untuk mengelompokkan berita saat ini, namun terdapat kekurangan pada cara pengelompokan ini. Pengunggah berita perlu mengetahui kategori berita yang akan diunggah, pengunggah berita juga perlu mengetahui kategori berita apa saja yang tersedia pada aplikasi tersebut, ini akan membuat berita tidak dapat langsung diunggah karena kategori berita perlu ditentukan terlebih dahulu secara manual.

Oleh karena itu, diperlukan suatu cara untuk dapat mengelompokkan berita secara otomatis. Salah satu cara yang dapat digunakan untuk mengelompokkan berita secara otomatis adalah menggunakan *natural language processing* (NLP), NLP digunakan untuk memproses judul berita dan menentukan kategori berita berdasarkan judulnya. Terdapat berbagai macam metode pada NLP untuk melakukan klasifikasi teks, namun pada penelitian ini, metode *long-short term memory* (LSTM) digunakan untuk membuat model klasifikasi berita.

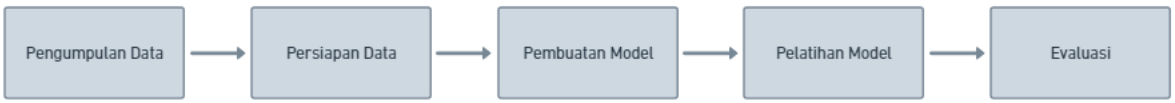
Long-short term memory (LSTM) merupakan salah satu metode yang digunakan pada *recurrent neural network* (RNN) untuk mengatasi masalah *vanishing error* [1]. Metode ini menggunakan blok memori khusus yang dapat mengingat urutan kronologis pada data input [2], dalam konteks NLP, metode ini digunakan untuk mengingat urutan kronologis dari kata-kata pada kalimat. Salah satu kelebihan dari metode ini adalah memungkinkan model *neural network* untuk mengingat urutan kata, sehingga model tidak hanya menentukan makna kalimat berdasarkan banyaknya kata yang muncul, namun juga

berdasarkan kata apa yang muncul terlebih dahulu. Bidirectional LSTM merupakan metode lanjutan dari LSTM, pada metode ini, data dimasukkan dalam urutan maju dan mundur, sehingga model juga dapat mengingat kata-kata setelah kata yang sedang diproses. Model yang sudah dilatih selanjutnya dapat diuji menggunakan *testing data*, di mana nilai akurasi dapat ditentukan dengan membandingkan jumlah berita yang berhasil diklasifikasikan dan jumlah total berita pada data *testing.*

Terdapat banyak penelitian terkait yang sudah dilakukan sebelumnya tentang klasifikasi berita, namun, kebanyakan dari penelitian yang ada membahas tentang klasifikasi berita palsu. Salah satu penelitian yang ada berhasil mengklasifikasikan berita palsu menggunakan metode Random Forest dengan akurasi sebesar 84% [3]. Penelitian ini akan berfokus pada klasifikasi berita berdasarkan kategori.

2. Metode Penelitian

Metode yang digunakan pada penelitian ini terdiri dari 5 bagian, yaitu pengumpulan data, persiapan data, pembuatan model, pelatihan model, dan evaluasi. Model yang dibuat mengimplementasikan *layer Word Embedding* untuk merepresentasikan judul berita sebagai vektor dan *layer* LSTM untuk mengklasifikasikan judul berita.



Gambar 1. Diagram Metode Penelitian

2.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah dataset judul dan kategori berita yang didapatkan melalui platform Github [4]. Dataset ini berisi 91017 data yang terdiri dari tanggal *posting* dari berita. URL berita, judul berita dalam Bahasa Indonesia, dan kategori berita.

2.2. Persiapan Data

Adapun prosedur-prosedur yang dilakukan setelah mengumpulkan data yaitu:

a. Menghapus kategori “news” dan “hot”, penghapusan ini dilakukan karena kategori tersebut merupakan kategori umum yang berisi berita dari berbagai kategori
b. Mengubah semua huruf pada judul berita menjadi *lowercase*
c. Menghapus *stop words* pada judul berita
d. Mengubah data kategori menjadi vektor dengan metode *one-hot encoding*
e. Memisahkan dataset menjadi 80% *training data* dan 20% *testing data*

2.3. Tokenisasi

Setelah dataset dibagi menjadi *training data* dan *testing data*, dilakukan tokenisasi untuk mengubah setiap kata pada judul berita ke dalam bentuk numerik. 10.000 kata terbanyak pada dataset akan diberi representasi angka, sedangkan kata-kata lainnya diberikan representasi khusus (*out-of-value token*). Kemudian, judul berita dapat direpresentasikan sebagai *sequence* numerik. Panjang *sequence* yang digunakan adalah 20, untuk *sequence* dengan jumlah token kurang dari 20, nilai 0 akan ditambahkan sebagai *padding* di awal *sequence*, sedangkan untuk *sequence* dengan jumlah token lebih dari 20, dilakukan pemotongan di awal *sequence* sehingga hanya 20 token terakhir yang digunakan.

2.4. Word Embedding

Word Embedding adalah sebuah metode yang digunakan untuk mengubah *sequence* kata menjadi vektor [5]. Pada penelitian ini, *Word Embedding* digunakan sebagai layer awal pada model, setiap judul berita yang berupa sequence 20 token diubah menjadi vektor berukuran 64 nilai [6]. Vektor ini kemudian akan digunakan untuk klasifikasi di *layer* selanjutnya.

2.5. Long-Short Term Memory (LSTM)

Long-Short Term Memory merupakan metode berbasis gradient yang diciptakan untuk mengatasi masalah *vanishing error* pada model *recurrent neural network* (RNN) [1]. Pada penelitian ini, LSTM digunakan sebagai layer pada model setelah dilakukan *Word Embedding*. banyaknya unit LSTM yang digunakan adalah sebanyak 64 [6]. Setiap *unit* pada *layer* LSTM

p-ISSN: 2301-5373
e-ISSN: 2654-5101

terdiri dari 3 bagian, yaitu *forget gate*, *input gate*, dan *output gate*. *Forget gate* digunakan untuk me-*reset state internal* dari *unit* sehingga *vanishing gradient* dapat dihindari. *Input gate* digunakan untuk memasukkan data baru ke dalam *state internal* dari *unit*. *Output gate* menggunakan *state internal* dari *unit* dan nilai *input* untuk menghasilkan nilai *output* dari *unit* [1]. Ketiga bagian ini memungkinkan *unit* LSTM untuk menghasilkan prediksi dan mengingat data yang diberikan sebagai *state*.

2.6. Bidirectional LSTM

Bidirectional LSTM merupakan metode yang digunakan untuk mengatasi kelemahan dari LSTM, yaitu ketidakmampuan LSTM dalam memproses data dari masa depan [7]. Pada *Bidirectional LSTM*, dilakukan proses input secara maju dan mundur (*forward* dan *backward*) menggunakan dua *layer* LSTM terpisah [1]. Pada penelitian ini, kedua *layer* LSTM yang digunakan memiliki unit yang sama, yaitu sebanyak 64 [6]. Model *Bidirectional LSTM* pada penelitian ini digunakan sebagai perbandingan dengan model LSTM untuk menentukan metode yang lebih baik dalam klasifikasi kategori berita.

2.7. Pelatihan Model

Pada tahap ini, model yang sudah dibuat akan dilatih menggunakan *training data*, *training data* terdiri dari 80% dataset total, sedangkan 20% sisanya akan digunakan pada proses evaluasi. Proses *training* dilakukan sebanyak 15 *epoch*.

2.8. Metode Evaluasi

Nilai-nilai yang digunakan untuk menguji dan mengevaluasi model adalah nilai akurasi, nilai *precision*, nilai *recall*, dan nilai F1. Nilai akurasi adalah perbandingan antara jumlah berita yang berhasil diklasifikasikan dengan jumlah total berita. Metrik evaluasi selain akurasi dihitung per kategori (kelas) berita, di mana nilai yang digunakan adalah nilai rata-rata dari evaluasi masing-masing kategori. Cara menentukan nilai metrik evaluasi dapat dilihat pada persamaan berikut.

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

Di mana TP (*true positive*) adalah banyaknya berita yang berhasil diklasifikasikan sebagai kategori yang dievaluasi, FP (*false positive*) adalah banyaknya berita yang tidak berhasil diklasifikasikan sebagai selain kategori yang dievaluasi, TN (*true negative*) adalah banyaknya berita yang berhasil diklasifikasikan sebagai kategori selain kategori yang dievaluasi, dan FN

(*false negative*) adalah banyaknya berita yang tidak berhasil diklasifikasikan sebagai kategori yang dievaluasi.

3. Hasil dan Pembahasan

Penelitian ini menggunakan 2 model untuk klasifikasi kategori berita, yaitu model LSTM dan Bidirectional LSTM. Model dibuat menggunakan Jupyter Notebook dengan library Tensorflow dan Keras. Layer yang digunakan pada masing-masing model dapat dilihat pada tabel 1.

Tabel 1. Layer yang Digunakan pada Model

LSTM	Bidirectional LSTM
Embedding Layer (Ukuran vektor output = 64)	Embedding Layer (Ukuran vektor output = 64)
LSTM Layer (64 unit, *forward*)	LSTM Layer (64 unit, *backward*)
-	LSTM Layer (64 unit, *forward*)
Hidden Layer (64 Layer)	Hidden Layer (64 unit)
Output Layer (7 unit/kelas)	Output Layer (7 unit/kelas)

Kedua model yang sudah dibuat kemudian dilatih menggunakan data *training* sebanyak 15 *epoch* dengan *optimizer* Adam. Hasil evaluasi model LSTM dan Bidirectional LSTM dapat dilihat pada tabel 2.

Tabel 2. Hasil Evaluasi Model LSTM

Metrics	LSTM	Bidirectional LSTM
Accuracy	0.8536	0.8415
Average Precision	0.85	0.84
Average Recall	0.85	0.84
Average F1-Score	0.85	0.84

Berdasarkan tabel 2, model LSTM sedikit lebih unggul dibandingkan dengan model Bidirectional LSTM dengan akurasi sebesar 85.36%. Kedua model ini memiliki akurasi yang hampir sama.

4. Kesimpulan

Model LSTM yang dibuat dapat mengklasifikasikan kategori berita berdasarkan judulnya dengan akurasi yang cukup tinggi, yaitu sebesar 85.36%. Selanjutnya, akurasi dari model dapat ditingkatkan

menggunakan metode lain atau dengan *hyperparameter tuning*. Model yang sudah dilatih juga dapat diimplementasikan pada aplikasi lain seperti aplikasi web dan android.

Referensi

[1] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks," pp. 1–42, 2019, [Online]. Available: http://arxiv.org/abs/1909.09586.
[2] Trivusi, "Mengenal Algoritma Long Short Term Memory (LSTM)," 2022. https://www.trivusi.web.id/2022/07/algoritma-lstm.html (accessed Nov. 20, 2022).
[3] N. Ghaniaviyanto Ramadhan, F. Dharma Adhinata, A. Jala, T. Segara, P. Rakhmadani, and F. Informatika, "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression," *J. Ris. Komputer)*, vol. 9, no. 2, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i2.3979.
[4] Ibamibrahim, "Dataset Judul Berita Indonesia." https://github.com/ibamibrahim/dataset-judul-berita-indonesia.
[5] K. S. Witanto, N. A. S. ER, and A. E. Karyawati, "Implementasi LSTM pada Analisis Sentimen Review Film Menggunakan Adam dan RMSprop Optimizer," vol. 10, no. 4, pp. 351–362, 2022.
[6] Keras, "Keras layers API." https://keras.io/api/layers/.
[7] M. Ilmiah, *Implementasi Metode Bidirectional Long Short-Term Memory (Bi-LSTM) untuk Prediksi Kasus Positif Covid-19 di Indonesia*. 2022.

Halaman ini sengaja dibiarkan kosong

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Algoritma Naïve Bayes Terhadap Klasifikasi Ulasan Aplikasi Tokopedia

Yauw James Fang Dwiputra Harta[a1], I Gede Arta Wibawa[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Badung, Indonesia
[1]jamesfangyauw@gmail.com
[2]gede.arta@unud.ac.id

Abstract

Tokopedia has an average number of visitors per month which is 149.67 visitors. In addition, the Tokopedia application has been downloaded by more than 100 million users with 6.34 million reviews and received a rating of 4.7 out of 5. This study aims to classify the positive and negative reviews of the Tokopedia application on the Google Playstore using Naive Bayes Algorithm. The results of the 2000 test data testing obtained 842 positive reviews and 1158 negative reviews. This means the percentage for positive reviews is only 42.1%, in contrast to negative reviews of 57.1%. The performance generated in this test against 2000 testing data is 96.19% accuracy value, with a precision value of 1. While in Class Recall the resulting value is 93.45% (positive class: negative). Then for the AUC value is 0.975. It is hoped that these results can help Tokopedia management to improve existing deficiencies so that they can get more positive reviews.

Keywords : *Tokopedia, classifier, analisyst sentiment, naive bayes, preprocessing*

1. Introduction

Perkembangan *e-commerce* di Indonesia saat ini sangatlah pesat. Jika pada tahun sebelum 2010, kebanyakan masyarakat datang ke toko-toko atau supermarket untuk belanja/membeli barang-barang yang dibutuhkan dan diinginkannya. Tetapi semenjak adanya *e-commerce* ini, masyarakat tidak perlu untuk datang ke toko secara langsung. Cukup dengan memesannya lewat aplikasi *e-commerce* yang dikembangkan oleh perusahaan *e-commerce* tersebut. Bahkan hingga kuartal 1 tahun 2022, menurut hasil statistik yang diterbitkan pada artikel liputan 6, bahwa pemerintah telah mencatatkan statistik jumlah transaksi di *e-commerce* seluruh Indonesia sebesar Rp 108,54 triliun[1].

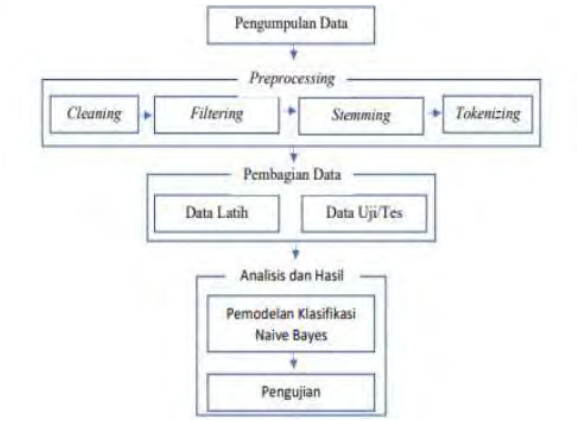
Tentu ini merupakan suatu hal yang positif, karena semakin tinggi angka transaksi pada *e-commerce* tentu ini menandakan bahwa masyarakat Indonesia sudah lebih beradaptasi dengan perkembangan teknologi saat ini khususnya dalam menggunakan aplikasi *e-commerce*. Salah satu e-commerce dengan jumlah pengguna terbanyak di Indonesia adalah Tokopedia. Menurut sumber yang dituliskan oleh Vika Azkiya Dihni dikatakan bahwa Tokopedia menjadi yang tertinggi dibandingkan kompetitor lainnya pada tahun 2021 dengan jumlah pengunjung rata-rata tiap bulan adalah 149,67 pengunjung[2] Selain itu, aplikasi tokopedia sudah didownload oleh lebih dari 100 juta pengguna dengan 6,34 juta ulasan dan mendapatkan rating 4,7 dari 5. Meski memiliki rating yang tinggi di google play store, tetapi tidak semua ulasan dari 6,34 juta ulasan tersebut memiliki arti yang positif, tidak sedikit juga yang bermakna negatif. Ulasan-ulasan negatif tersebut tentu akan bisa menjadi tolak ukur kita mengenai peforma dan layanan aplikasi tokopedia, selain tolak ukur juga bisa menjadi pacuan perusahaan tokopeda untuk mengembangkan agar aplikasinya menjadi lebih baik lagi. Maka dari itu lah, penulis merasa perlu untuk melakukan analisis ulasan aplikasi tokopedia dan mengklasifikasikannya kedalam kelompok positif dan negatif.

Pada tahun 2020, hal ini sudah pernah diteliti oleh Afifah Faadilah dengan jurnal yang berjudul "Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Di Google Play Store Dengan Menggunakan Metode Long Short Term Memory"[3]. Penelitian tersebut mengambil ulasan data dengan jumlah 3067 dengan jumlah kelas negatif 2591 dan positif 476. Tentunya data

ulasan tahun 2020 tersebut telah tidak relevan lagi saat ini sehingga penulis mengambil topik ini. Perbedaan penelitian ini dengan penelitian sebelumnya adalah terletak pada metode algoritma yang digunakan, dimana pada penelitian ini menggunakan metode Naive Bayes dengan dataset yang berbeda juga. Hasil akhir dari penelitian ini adalah tingkt prosentase / jumlah ulasan aplikasi tokopedia yang telah diklasifikasikan menjadi kelompok positif dan negatif. Hasil ini akan dapat menjadi manfaat untuk manajemen Tokopedia kedepannya untuk lebih mengembangkan aplikasi tokopedia ini menjadi lebih baik sehingga masyarakat Indonesia tetap memberikan kepercayaannya kepada pihak Tokopedia yang tentu ini menjadi keuntungan juga bagi negara karena tokopedia merupakan perusahaan asli buatan warga Indonesia yang akan bisa memberikan income kepada negara maupun UMKM di *marketplace* Tokopedia.

2. Research Methods

Metode penelitian yang digunakan pada penelitian ini adalah dengan menggunakan metode eksperimen dengan urutan langkah-langkah seperti gambar dibawah ini.



Gambar 1. Langkah Langkah Penelitian

Berikut adalah penjelasan lebih lanjut mengenai diagram alur metode penelitian diatas :

1. Pengumpulan Data
   Pengumpulan data yang digunakan dalam penelitian ini dilakukan dengan dua tahap, tahap pertama adalah mengumpulkan data yang akan dijadikan data latih/training dan tahap kedua mengumpulkan data yang akan dijadikan data uji/testing. Data yang diambil adalah ulasan ulasan bahasa indonesia dari aplikasi tokopedia yang ada di playstore yang dimulai dari 1 Maret 2022 hingga 29 September 2022 dengan total ulasan sebanyak 2000 untuk data testing dan 1000 untuk data latih/training. Data latih/training yang telah dikumpulkan ini diberikan label positif/negatif secara manual. Dalam pengumpulan data ini menggunakan metode web scrapping pada browser chrome dengan menggunakan bantuan extension yaitu dataminer[4].
   Contoh :

Redy Erwin 27 September 2022 di android redmi note 7, Kenapa ya aplikasi Tokopedia berat bangat, Kadang2 berenti, mau masuk ke apk susah, mau geser cari produk jg gak bisa, aku kira sudah ada perbaikan, ternyata masih belum, tolong dong Tokopedia,, biar bisa aku pertahankan aplikasi nya.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2. Preprocessing
   Setelah didapatkan dataset, maka langkah selanjutnya adalah melakukan preprocessing dataset tersebut yang diimplementasikan dengan bahasa pemrograman python. Adapun pada preprocessing ini terdiri dari 4 tahapan yaitu sebagai berikut[5] :
   a. Cleaning
      Pada proses ini kita melakukan pembersihan pada dataset yang telah kita dapatkan tadi, dimana pembersihan ini meliputi menghapus karakter-karakter selain huruf, menghapus nama/email penguna, menghapus emoticon, menghapus hashtag (#), menghapus url, dan mengubah semua huruf menjadi huruf kecil.
      Contoh :

september di android redmi note kenapa ya aplikasi tokopedia berat bangat kadang berenti mau masuk ke apk susah mau geser cari produk jg gak bisa aku kira sudah ada perbaikan ternyata masih belum tolong dong tokopedia biar bisa aku pertahankan aplikasi nya

   b. Filtering
      Pada tahap ini dilakukan penyaringan dengan menerapkan proses untuk mendapatkan kata kata dan melakukan pembuangan terhadap kata kata yang tidak penting. Setelah itu dilakukan juga penghapusan tanda baca yang ada dan juga kata kata yang termasuk *stopwords*. Kriteria dari stopwords ini yaitu tidak memiliki makna penting, sering keluar dalam teks . Contohnya seperti kata kata penghubung, waktu, dan lain sebagainya.
      Contoh :

android redmi note kenapa aplikasi tokopedia berat berenti masuk apk susah geser cari produk gak bisa aku kira ada perbaikan ternyata belum tolong pertahankan

   c. Stemming
      Setelah didapatkan hasil kata kata dari filttering, maka dilakukanlah stemming. Pada stemming ini kita mengubah kata kata yang memiliki imbuhan menjadi kata kata dasar sesuai EYD.
      Contoh :

android redmi note kenapa aplikasi tokopedia berat henti masuk apk susah geser cari produk tidak bisa aku kira ada baik ternyata belum tolong tahan

   d. Tokenizing
      Setelah didapatkan data berupa kata kata dasar, maka tahap selanjutnya adalah memisahkan kata kata tersebut sesuai dengan spasinya. Proses inilah yang dinamakan sebagai tokenizing.
      Contoh :

[‘android’, ‘redmi’, ‘note’, ‘kenapa’, ‘aplikasi’, ‘tokopedia’, ‘berat’, ‘henti’, ‘masuk’, ‘apk’, ‘susah’, ‘geser’, ‘cari’, ‘produk’, ‘tidak’, ‘bisa’, ‘aku’, ‘kira’, ‘ada’, ‘baik’, ‘ternyata’, ‘belum’, ‘tolong’, ‘tahan’]

3. Pembagian Data
   Mula-mula data diberikan pelabelan secara manual dan setelah itu dilakukanlah pembagian data menjadi 2 yaitu data latih dan data uji. Dimana terdapat 2 kelompok dalam penelitian ini yaitu kelompok data yang pertama adalah jumlah data latih lebih sedikit dari data uji. Sedangkan kelompok data kedua adalah data laih lebih banyak dari data uji.
4. Analisis dan Hasil
   Pada tahap ini, peneliti melakukan analisis dengan menerapkan metode naive bayes yang diimplementasikan dengan bahasa pemrograman python. Definisi dari naive bayes yaitu suatu metode untuk mengklasifikasi probalistik yang berdasarkan Teorema Bayes[6]. Rumus dari naive bayes seperti dibawah ini :

$$P(B) = \frac{P(A)\,.\;P(A)}{P(B)}$$

Keterangan :
B : Data dengan *class* yang masih belum diketahui
A : Hipotesis data A yang merupakan suatu *class* bersifat spesifik
P(A|B) : Probabilitan dari hipotesis A bedasarkan kondisi kasus B (posteriori probability)
P(A) : Probabilitas dari hipotesis A (prior probability)
P(B|A) : Probabilitas kasus B berdasarkan kondisi hipotesis A
P(B) : Probabilitas hipotesis B

Hasil dari ini nanti akan digunakan untuk menentukan ulasan yang termasuk positif dan negatif.

3. Result and Discussion

Dataset yang telah dikumpukan tadi yaitu data latih dan data testing akan dimasukan ke dalam filenya masing masing dengan nama training_ulasan.xlsx dan testing_ulasan.xlsx yang dimana 2 data set ini akan nilai untuk pembobotan kata, dalam memberikan nilai tersebut dilakukan secara otomatis. Selain pemberian nilai tersebut, juga dilakukan pemberian label yang dilakukan secara otomatis oleh aplikasi pada penelitian ini menggunakan algoritma naive bayes. Untuk data latih/training telah didapatkan label di masing masing datanya karena pada tahapan ekperimen tadi penulis melakukannya secara manual untuk pemasangan label pada data training. Dimana terdapat 500 data yang dilabel positif dan 500 data yang dilabel negatif. Hal ini dilakukan agar data training mempunya kategori atau label sehingga dapat dilakukan proses pelatihan dan pengujian data.

Pada tahapan klasifikasi dengan algoritma berupa naïve bayes, dimana proses ini dimulai dengan melakukan pengukuran kata kata. Pengukuran kata kata dilakukan pada dataset untuk data latih/training. Hasil pada pengukuran ini adalah nilai probabilitas/kemungkinan suatu data positif dan probabilitas suatu data termasuk negatif. Kemudian terkait hasil dari pengukuran pada data training tadi, selanjutnya digunakan dalam menentukan nilai acuan atau patokan ketika mengklasifikasikan ulasan positif dan ulasa/sentimen negatif dari data testing yang menjadi data utama dari penelitian ini. Penentuan sentimen positif atau negatif dilakukan dengan menghitung nilai probabilitas dari data testing, setelah dihitung kemudian nilai tersebut dibandingkan dengan nilai probabilitas yang telah didapatkan dari data training pada proses yang sebelumnya.

Pada pelaksanaan proses ini, metode yang digunakan adalah algoritma naïve bayes yang dapat melakukan klasifikasi ulasan tokopedia secara otomatis oleh aplikasi berbasis bahasa pemrograman python yang dibuat dengan mengimplementasikan algoritma naïve bayes. Pada konsepnya, proses ini dilakukan dengan membandingkan nilai dari data testing

p-ISSN: 2301-5373
e-ISSN: 2654-5101

dengan nilai dari data training yang dijadikan patokan untuk klasifikasi. Setelah dilakukan proses tersebut hingga semua kata telah berhasil dibandingkan maka akan didapatkan hasil berupa dataset untuk data testing telah terklafikasi menjadi ulasan positif dan ulasan negatif. Ulasan yang termasuk klasifikasi positif karena hasil dari algoritma naive bayes pada proses sebelumnya memiliki nilai probabilitas positif yang lebih besar dibandingkan probabilitas negatif. Begitupun sebaliknya untuk ulasan yang termasuk negatif karena hasil dari probabilitas/kemungkinan negatif yang didapatkan pada proses ini lebih besar daripada probabilitas positif.

Hasil dari analisis ulasan-ulasan pada data testing dengan menggunakan aplikasi yang dikembangkan oleh penulis berbasis python yaitu 842 ulasan positif dan 1158 ulasan negatif. Ini berarti persentase untuk ulasan positif hanya 42,1%, berbanding terbalik dengan ulasan negatif sebesar 57,1%.

Data yang telah didapatkan ini perlu dilakukan suatu evaluasi dengan menerapkan suatu pengujian dan untuk mendapatkan hasil dari pengujian dengan menggunakan bantuan *tools* Rapidminer terhadap 2000 data testing ulasan pengguna aplikasi Tokopedia. Pengujian yang dilakukan berdasarkan nilai Class Recal, accuracy, dan terakhir Class Precision pada analisis klasifikasi ulasan pengguna aplikasi tokopedia ini. Dari hasil pengujian ini didapatkan nilai accuracy sebesar 96,19%, kemudian untuk nilai class precision tergolong pada kelas 1. Kemudian terkait class recall didapatkan nilai 93,45%. Kemudian yang terakhir untuk nilai AUC adalah 0,975.

4. Conclusion and Suggestion

Bedasarkan hasil pembahasan diatas, maka terbukti bahwa metode naive bayes dapat digunakan untuk proses klasifikasi ulasan pada aplikasi tokopedia yang terdapat di google play store. Dalam ujicobanya menggunakan data training dan data testing. Data training telah dilabeli secara manual sebelumnya dan data ini yang akan menjadi acuan dari metode untuk mengklasifikasikan data testing termasuk ulasan positif ataupun ulasan negatif. Dalam mengklasifikasikan itu, peneliti menggunakan bantuan aplikasi rapidminer agar ujicoba dapat dilakukan secara *realtime.* Peforma yang dihasilkan pada pengujian diatas terhadapat 2000 data testing adalah 96,19% nilai accuracynya dan untuk nilai precision tergolong pada kelas

dengan nilai 1. Kemudian pada kelas class recall didapatkan nilai 93,45%. Kemudian untuk nilai AUC adalah 0,975. Dari 2000 data testing diperoleh 842 ulasan positif dan 1158 ulasan negatif. Ini berarti persentase untuk ulasan positif hanya 42,1%, berbanding terbalik dengan ulasan negatif sebesar 57,1%. Hasil ini tentu berbanding terbalik dengan penilaian pada aplikasi tokopedia yang mendapatkan rating 4,7 dari 5. Tapi ini bisa disebabkan karena data testing yang digunakan juga masih termasuk sedikit. Maka dari itu peneliti mengharap kepada peneliti lain bisa menggunakan jumlah data testing yang lebih banyak juga bisa dengan menggunakan metode algoritma yang lain agar mendapatkan nilai peforma yang lebih baik dari naive bayes.

References

[1] "Transaksi E-Commerce Indonesia Rp 108,54 Triliun di Kuartal I-2022," *liputan 6*, 2022. https://www.liputan6.com/bisnis/read/5032523/transaksi-e-commerce-indonesia-rp-10854-triliun-di-kuartal-i-2022 (accessed Sep. 20, 2022).

[2] Vika Azkiya Dihni, "Tokopedia, E-Commerce dengan Pengunjung Terbanyak pada 2021," *Databooks*, 2022. https://databoks.katadata.co.id/datapublish/2022/04/12/tokopedia-e-commerce-dengan-pengunjung-terbanyak-pada-2021 (accessed Sep. 20, 2022).

[3] A. Faadilah, "ANALISIS SENTIMEN PADA ULASAN APLIKASI TOKOPEDIA DI GOOGLE PLAY STORE MENGGUNAKAN METODE LONG SHORT TERM MEMORY," 2020.

[4] Gede Putra Aditya Brahmantha and I Wayan Santiyasa, "Sentiment Analysis of the Enforcement of PSBB Part II ," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. Volume 9 No. 2, Sep. 2020.

[5] "Teknik pre-processing dan classification dalam data science," 2022. https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/ (accessed Sep. 29, 2022).

[6] Fitri Handayani and Feddy Setio Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *Jurnal Teknik Elektro Vol. 7 No. 1* , 2018.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Gaya Musik *Pop Punk* Berdasarkan Era dengan Metode K-Nearest Neighbor

I Kadek Riski Ari Putra[a1], Luh Arida Ayu Rahning Putri[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]fxdeki@gmail.com
[2]rahningputri@unud.ac.id

Abstrak

Pop Punk – merupakan salah satu subgenre dari musik punk yang menggabungkan elemen-elemen dari musik pop dengan musik punk. Musik pop punk identik dengan tangga nada yang disadur atau turunan dari musik pop, namun untuk tempo dan gaya bermain mengikuti musik punk yang identik dengan tempo cepat dan suara gitar yang ramai. Pop punk telah mengalami naik turun popularitas seiring dengan berjalannya waktu, namun musik pop punk selalu dapat berkembang dari waktu ke waktu sehingga musik ini dapat diklasifikasikan dalam dua era secara garis besar yaitu oldschool dan modern berdasarkan gaya permainannya. Penelitian ini membahas klasifikasi musik pop punk berdasarkan kedua era tersebut (oldschool dan modern) dengan metode K-Nearest Neighbor menggunakan 5 fitur musik (danceability, energy, key, speechiness, dan valence). Fitur musik tersebut dipilih karena key dan speechiness dapat melambangkan perbedaan tangga nada dan gaya vokal yang digunakan pada kedua gaya musik tersebut, serta danceability, energy, dan valence dapat melambangkan suasana dari musik. Nilai akurasi klasifikasi yang dihasilkan penelitian ini adalah 76%.

Keywords: *klasifikasi musik, subgenre musik, gaya musik, k-nearest neighbor, dataset.*

1. Pendahuluan

Perkembangan musik pada era digital ini sangatlah pesat, masyarakat dapat mendengarkan dan menyebarkan musik secara digital dengan mudah. Seiring bertambahnya jumlah musik digital yang tersebar maka muncul masalah baru dalam cara organisasi musik tersebut. Musik dapat diorganisir berdasarkan artis, album, atau *genre/subgenre*. *Genre/Subgenre* merupakan cara organisasi musik yang paling umum[3], hal ini disebabkan mudahnya pendengar mencari musik dengan gaya yang sama atau sejenis untuk didengarkan sesuai dengan preferensi masing-masing.

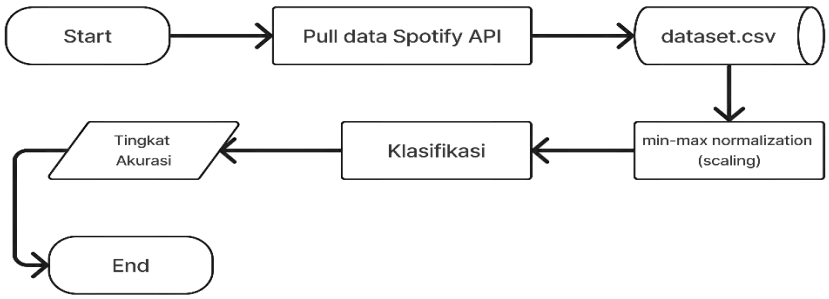
Banyaknya jumlah musik digital menyebabkan pengklasifikasian *genre* perlu diotomatisasi dengan algoritma kecerdasan buatan yang dapat membantu manusia agar lebih efektif dan tidak memakan banyak waktu [2]. Selain itu, banyaknya perbedaan gaya permainan dalam sebuah *genre/subgenre* membuat cara organisasi musik semakin beragam Klasifikasi gaya dalam *genre/subgenre* musik menggunakan kecerdasan buatan dapat dilakukan dengan menggunakan fitur-fitur yang ada dalam sebuah musik seperti *loudness, energy, tempo,* dsb.

Berbagai penelitian telah dilakukan mengenai teknik klasifikasi musik, salah satunya adalah klasifikasi genre musik yang dilakukan oleh [1]. Pada penelitian tersebut digunakan metode deep learning convolutional neural network dan mel-spektogram untuk klasifikasi genre musik dengan menggunakan spektogram dari musik.

Penelitian ini membahas klasifikasi gaya musik *genre* pop punk berdasarkan era menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan beberapa fitur musik (*danceability, energy, key, speechability,* dan *valence*) yang didapatkan melalui Spotify API. Algoritma *K-Nearest Neighbor* dipilih karena merupakan algoritma paling sederhana untuk memecahkan masalah klasifikasi dengan hasil yang cukup baik dan kompetitif.

2. Metode Penelitian

Dalam penelitian ini, data yang digunakan adalah dataset yang dibangun dari dua *playlist* spotify menggunakan layanan Spotify API. Spotify API dapat memberikan data artis, album, fitur musik, dan sumber daya lain yang ada pada spotify kepada pengguna [5]. Proses berjalannya sistem dapat dilihat pada gambar 1.



Gambar 1. Flowchart Klasifikasi Gaya Musik *Pop Punk*

Proses dimulai dengan melakukan pengambilan data musik dalam dua *playlist* jenis musik *pop punk* yang telah disiapkan, kemudian disimpan kedalam sebuah dataset. Fitur-fitur musik yang didapatkan kemudian dinormalisasi menggunakan teknik *min-max normalization* atau *scaling.* Setelah data dinormalisasi, data akan diklasifikasi menggunakan algoritma *K-Nearest Neighbor.* Kemudian tingkat akurasi data akan diukur.

2.1 Dataset

Musik yang digunakan dalam dataset ini berasal dari Spotify yang berjumlah 100 musik dari gaya musik *pop punk oldschool* dan *pop punk modern*.
Sejumlah 5 fitur dari musik digunakan dalam penelitian ini yang berupa *danceability, energy, key, speechiness,* dan *valence.* Fitur dan data musik tersebut didapatkan melalui Spotify API.

2.2 Min-Max Normalization

Fitur-fitur musik yang didapatkan dari Spotify API dapat memiliki data yang tidak berdistribusi normal sehingga normalisasi data dilakukan sebelum masuk ke tahap klasifikasi. Tahap *min-max normalization* adalah sebagai berikut

a. Ambil nilai tertinggi dari setiap fitur musik
b. Ambil nilai terendah dari setiap fitur musik
c. Hitung persamaan (1) untuk setiap data tersebut

$$norm(x) = \frac{x - minValue}{maxValue - minValue} norm(x) = \frac{x - minValue}{maxValue - minValue} \quad (1)$$

norm(x) merupakan data yang telah dinormalisasi. minValue dan maxValue merupakan nilai terkecil dan terbesar dari fitur ke-i dalam data yang belum dinormalisasi. x merupakan data yang belum dinormalisasi.

2.3 Klasifikasi

Tahapan klasifikasi dilakukan dengan algoritma *K-Nearest Neighbor* dimana algoritma ini bekerja dengan mengklasifikasikan data tes berdasarkan jaraknya terhadap data *training* [4]. Kelas dari data diperoleh dari kelas data mayoritas dalam nilai K. Nilai K merupakan jumlah data dengan jarak terdekat atau dapat disebut sebagai tetangga terdekat. Algoritma dari *K-Nearest Neighbor* adalah sebagai berikut.

a. Diketahui $x_i x_i$ merupakan data *training* dan $y_k y_k$ merupakan data *testing*
b. for setiap data *testing* $y_k y_k$
c. for setiap data *training* $x_i x_i$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

d. Hitunglah jarak euclidean y_k dengan x_i
e. end for
f. Urutkan hasil secara *ascending*
g. Pilih hasil teratas sejumlah-k
h. Klasifikasi y_k berdasarkan mayoritas hasil teratas
i. end for

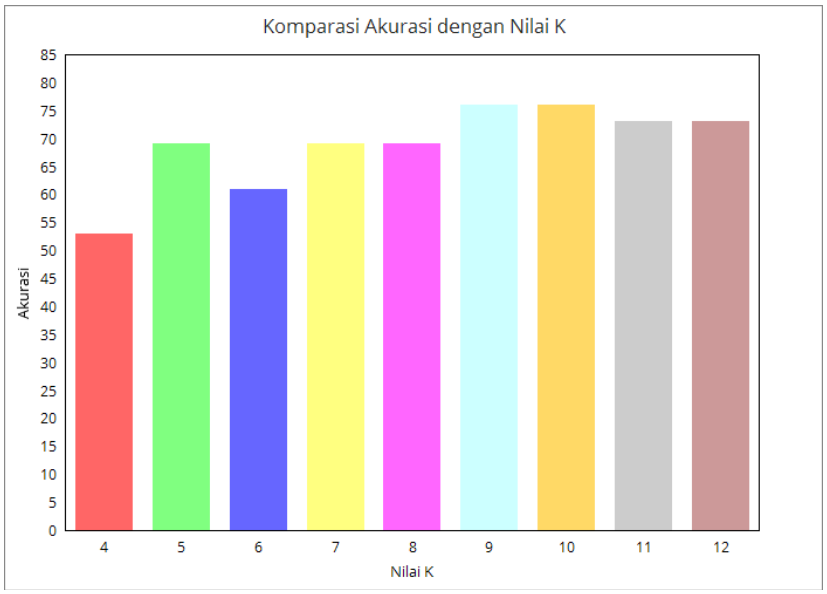
Jarak euclidean dapat dihitung dengan persamaan (2).

$$d(xi, yi) = \sqrt{\sum_{i=0}^{n} (xi - yi)^2} \quad (2)$$

Pada penelitian ini nilai K yang digunakan bervariasi dari 4 sampai 12 untuk mendapatkan nilai K dengan akurasi klarifikasi tertinggi.

3. Hasil dan Diskusi

3.1. Komparasi Akurasi Klasifikasi Berdasarkan Nilai K



Gambar 2. Grafik Hasil Akurasi Klasifikasi Berdasarkan Nilai K

Berdasarkan tahap klasifikasi yang telah dilakukan menggunakan *K-Nearest Neighbor* dengan nilai K yang bervariasi didapatkan hasil yang ditampilkan pada gambar 2. Pada Gambar 2 ditunjukkan bahwa nilai akurasi terendah atau *error* tertinggi terdapat pada nilai K=4 sebesar 53%. Pada nilai K=4 sampai K=7 terlihat bahwa nilai akurasi bersifat fluktuatif dan pada K=8 hingga K=10 nilai akurasi naik lebih tinggi. Nilai akurasi terbaik atau tingkat *error* terkecil terdapat pada nilai K=9 dan K=10 sebesar 76%. Menggunakan algoritma *K-Nearest Neighbor* dengan nilai K=9 menghasilkan akurasi sebesar 76% yang berarti gaya musik *pop punk* yang diklasifikasikan dengan benar sebanyak 76% dan sebanyak 24% diklasifikasikan dengan salah.

4. Kesimpulan

Nilai akurasi klasifikasi yang didapatkan menggunakan algoritma *K-Nearest Neighbor* adalah sebesar 76%, dimana nilai tersebut merupakan jumlah yang signifikan. Tingginya nilai akurasi pada penelitian ini dapat terjadi karena fitur musik yang dipilih melambangkan perbedaan dari gaya musik *pop punk* pada era *oldschool* dengan era *modern*.

Pada penelitian selanjutnya dapat dilakukan seleksi fitur musik yang dapat melambangkan *genre/subgenre* musik maupun gaya permainan dalam suatu *genre/subgenre* tersebut dengan baik. Serta dapat dilakukan penerapan algoritma lainnya yang dapat menghasilkan tingkat akurasi yang lebih tinggi.

Referensi

[1] D. Lionel, R. Adipranata and E. Setyati, “Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan Mel-Spektrogram” Jurnal Infra, vol.7, no. 1, p. 51-55, 2019.
[2] Gst. A. V. M. Giri, “Klasifikasi dan Retrieval Musik Berdasarkan Genre (Sebuah Studi Pustaka)” Jurnal Ilmiah Ilmu Komputer Universitas Udayana, 2017.
[3] M. Sarofi, I. Irhamah and A. Mukarromah, "Identifikasi Genre Musik dengan Menggunakan Metode Random Forest" Jurnal Sains & Seni ITS, vol. 9, no. 1, p. 79-86, 2020.
[4] I. Nikmatun, and I. Waspada, “Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor” Jurnal Simetris, vol. 10, no. 2, p. 421-432, 2019.
[5] R. Rachmandany, A. Kharisma, and I. Arwani, “Pengembangan Aplikasi Autoplay dengan Konsep Context-Aware menggunakan Spotify API berbasis Android” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 3, no. 7, p. 6616-6623, 2019.

Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 12, No 3. Februari 2024 e-ISSN: 2654-5101

Implementasi Methontologi Untuk Pembangunan Model Ontologi Pada Sistem Informasi Produk Bodycare

Ida Ayu Taria Putri Mahadewi[a1], Ida Bagus Gede Dwidasmara[a2]
[a]Informatics Department, Udayana Univercity
Bali, Indonesia
[1]gek.taria@gmail.com
[2]dwidasmara@unud.ac.id

Abstrak

Pada zaman sekarang, wanita khususnya remaja sedang gencar melakukan perawatan, baik perawatan wajah, rambut, maupun tubuh. Perawatan tubuh ini dilakukan untuk menjaga kesehatan, kebersihan, kecantikan, dan kelembaban kulit tubuh. Produk bodycare yang beredar di pasaran tentunya sangat beragam. Hal tersebut tentunya menyebabkan banyaknya pertimbangan untuk memilih produk bodycare yang tepat dan sesuai dengan kebutuhan atau permasalahan kulit masing-masing individu karena kandungan pada setiap produk bodycare tentunya berbeda-beda. Untuk memperoleh informasi mengenai kandungan yang terdapat pada bodycare dapat memanfaatkan perkembangan teknologi. Namun, sering kali situs web memberikan informasi-informasi yang tidak relevan. Sehingga, pada penelitian ini akan dikembangkan model ontologi semantik dengan domain produk bodycare menggunakan metode Methontologi. Model pengembangan ontologi produk bodycare ini menghasilkan 10 class, 5 Object Properties, 5 Data Properties, dan 45 individual. Pada tahap evaluasi menggunakan SPARQL query yang digunakan untuk merepresentasikan pertanyaan-pertanyaan yang diajukan terkait dengan domain produk bodycare. Dengan penelitian ini, diharapkan dapat membantu masyarakat dalam menentukan produk bodycare yang tepat sesuai dengan kebutuhan dan permasalahan kulit pada masing-masing individu.

Kata Kunci: *Bodycare, Ontology, Methontology, SPARQL, Protégé.*

1. Pendahuluan

Wanita tentunya tidak dapat terlepas dari perawatan tubuh dari zaman dahulu hingga saat ini. Perawatan tubuh wanita ini perlu untuk dilakukan agar dapat menjaga kesehatan, kebersihan, kecantikan, dan kelembaban kulit tubuh. Namun, kandungan yang terdapat pada produk bodycare tentunya berbeda-beda. Kandungan yang terdapat di dalam bodycare tentunya harus disesuaikan dengan kebutuhan atau permasalahan pada kulit. Gunakan bodycare dengan kandungan bahan yang aman dan sesuai dengan tipe jenis kulit masing-masing individu.

Produk bodycare yang beredar di pasaran tentunya sangat beragam. Hal tersebut tentunya menyebabkan banyak pertimbangan untuk memilih produk bodycare yang tepat dan sesuai dengan kebutuhan atau permasalahan kulit masing-masing individu. Perawatan kulit dengan menggunakan bodycare secara rutin agar kulit tetap terawat dan terhindar dari kondisi kulit kering dan kusam. Untuk memperoleh informasi mengenai kandungan bodycare, dapat memanfaatkan perkembangan teknologi berbasis internet. Namun, seringkali pada situs web memberikan informasi-informasi yang terpecah yang menyebabkan konsumen kesulitan untuk memperoleh informasi mengenai kandungan yang terdapat pada bodycare yang dicari. Konsumen juga harus memastikan bahwa informasi-informasi yang didapat mengenai produk tersebut sudah relevan. Maka, untuk mengatasi hal tersebut dapat menggunakan teknologi web semantik.

Sehingga, pada penelitian ini akan dikembangkan model ontologi pada domain produk bodycare serta dilakukan pengujian dengan mengajukan beberapa pertanyaan yang dapat

digunakan untuk mengakses informasi mengenai bodycare yang diperlukan. Ontologi merupakan suatu model fundamental yang berasal dari web semantik yang dapat dimanfaatkan untuk memanipulasi informasi yang ada untuk kebutuhan pengguna[1]. Pada penelitian ini, akan dilakukan pengujian model dengan mengajukan pertanyaan-pertanyaan yang dapat digunakan oleh pengguna pada saat mengakses informasi-informasi yang diinginkan mengenai produk bodycare yang dicari. Model ontologi ini diharapkan dapat membantu untuk menentukan produk bodycare yang tepat sesuai dengan kebutuhan dan permasalahan pada kulit masing-masing individu. Metode yang digunakan pada model ontologi ini yaitu Methontologi. Methontologi adalah suatu metode pengembangan ontologi yang mengusulkan pengekspresian ide[2].

1.1 Ontologi

Ontologi adalah suatu teknik yang digunakan untuk merepresentasikan pengetahuan yang diimplementasikan dengan web semantic yang secara teknik direpresentasikan dalam bentuk class, property, facet, dan instance. Komponen ontologi terdiri dari konsep, relasi, fungsi, aksiom, dan intances[1].

- Konsep (Concept)
  Sebuah konsep terdiri dari obyek-obyek yang merupakan penjelasan dari tugas, fungsi, aksi, strategi, dan sebagainya. Sebuah kelas juga bisa memiliki subkelas yang akan mempresentasikan konsep yang lebih spesifik daripada superkelasnya.
- Relasi (Relation)
  Relasi merupakan representasi sebuah tipe dari interaksi antara konsep dari sebuah domain. Sebagai contoh dari relasi binari termasuk subclass-of dan connected-to. Relasi harus mampu mendefinisikan hubungan dari entitas yang ada.
- Fungsi (Functions)
  Fungsi adalah relasi khusus dimana elemen ke n dari relasi.
- Aksiom (Axioms)
  Aksiom digunakan untuk memodelkan sebuah sentence yang selalau benar.
- *Instances* (Individual)
  Instances adalah komponen dasar dari suatu ontologi. *Instances* atau Individual menyatakan obyek-obyek dalam suatu domain yang diteliti yang digunakan untuk merepresentasikan elemen nyata seperti hewan, tanaman, dan manusia, maupun elemen abstrak seperti bilangan dan huruf.

1.2 Bodycare

Bodycare merupakan perawatan kulit yang perlu untuk dilakukan agar dapat menjaga kesehatan, kebersihan, kecantikan, dan kelembaban kulit tubuh. Namun, kandungan yang terdapat pada produk bodycare tentunya berbeda-beda. Kandungan yang terdapat di dalam bodycare tentunya harus disesuaikan dengan kebutuhan atau permasalahan pada kulit masing-masing individu.

1.3 Protégé

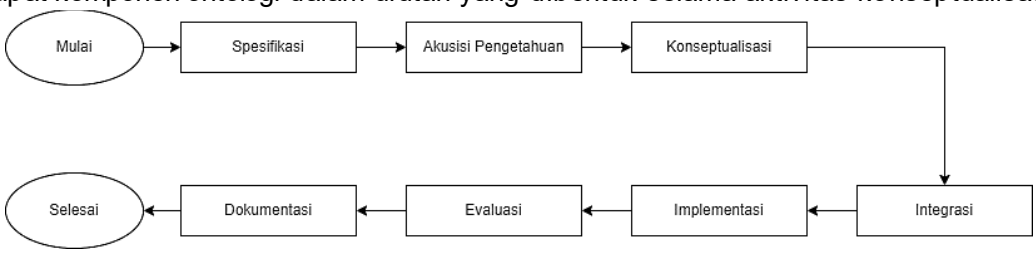
Protégé merupakan suatu alat yang digunakan untuk membantu membuat sebuah domain ontologi, menyesuaikan form untuk entry data, dan memasukan data. Protégé dapat melakukan query dengan menggunakan SPARQL. Protégé mendukung berbagai format penyimpanan seperti OWL, RDF, XML, dan HTML yang dimana Protégé ini dibuat dengan menggunakan bahasa pemrograman Java. Tools yang terdapat pada Protégé ini dogunakan melalui Graphical User Interface (GUI) dengan menyediakan Tab untuk masingmasing bagian dan fungsi standar[3].

1.4 SPARQL

SPARQL adalah perintah atau bahasa yang digunakan untuk mengakses query pada sebuah model data semantik pada format data RDF. SPARGL dapat disebut juga dengan bahasa untuk mengakses *linked data* dengan menggunakan *end point* untuk dapat menghasilkan relasi antara satu informasi dengan yang lainnya. Bahasa SPARQL dianggap setara dengan bahasa SQL yang mempunyai format sintak yang serupa, hanya saja berbeda dalam penggunaannya[4].

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu metode Methontologi. Methontologi merupakan suatu metodologi pembangunan model ontologi yang memiliki kelebihan mengenai dengan deskripsi setiap aktivitas yang harus dilakukan secara mendetail. Methontologi juga memiliki kemampuan untuk membangun ontologi yang dapat digunakan kembali untuk pengembangan sistem lebih lanjut. Adapun tahapan-tahapan dari metode Methontologi ini yaitu, sebagai berikut[5]. Methontologi merupakan suatu metode pengembangan model ontolgi yang dapat mengusulkan pengekspresian ide sebagai suatu himpunan dari Intermediate Representations (IR) dan menghasilkan ontologi menggunakan translators. Methontology memasukkan dalam aktivitas konseptualisasi beberapa tugas (task) untuk menstrukturkan pengetahuan. Pada setiap tugasnya, terdapat komponen ontologi dalam urutan yang dibentuk selama aktivitas konseptualisasi [2].



Gambar 1. Alur Metode Penelitian

2.1 Spesifikasi

Pada tahap spesifikasi ini memiliki tujuan yaitu untuk melakukan produksi dokumentasi spesifikasi ontologi formal, semi formal, ataupun informal yang ditulis dengan *natural language*. Pada tahap spesifikasi ini menggunakan satu set representasi menengah atau dapat juga menggunakan pertanyaan kompetensi.

2.2 Akusisi Pengetahuan

Pada tahap akusisi pengetahuan ini merupakan tahapan yang independen dalam melakukan pembangunan ontologi. Tahap akusisi ini sebagian besar telah diselesaikan bersamaan dengan tahap spesifikasi dan terus menurun seiring berjalannya proses pengembangan ontologi.

2.3 Konseptualisasi

Pada tahap konseptualisasi ini akan membangun model konseptual dari pengetahuan domain yang akan menggambarkan masalah beserta dengan solusinya dalam kosakata domain yang telah diidentifikasi sebelumnya pada tahap spesifikasi. Hal yang perlu untuk dilakukan yaitu membangun Glossary of Terms (GT) yang meliputi konsep, *instances*, kata kerja, dan properti. Glossary of Terms (GT) digunakan untuk mencari dan mengumpulkan semua yang berpotensi untuk pengetahuan domain berserta dengan artinya.

2.4 Integrasi

Pada tahap integrasi merupakan tahap yang digunakan untuk membuat pertimbangan dalam menggunakan definisi ontologi yang sudah ada dan kemudian dibangun ke dalam ontologi lain. Hal tersebut menyebabkan pembangunan ontologi tidak perlu dimulai dari awal lagi.

2.5 Implementasi

Pada tahap implementasi ini merupakan proses dilakukannya penerapan dari perancangan ontologi yang telah dibuat sebelumnya pada tahap spesifikasi sampai dengan integrasi. Hasil yang diperoleh pada tahap ini yaitu pendefinisian kembali serta implementasi dari rancangan ontologi dengan menggunakan software Protégé.

2.6 Evaluasi

Pada tahap evaluasi ini dilakukan penilaian teknis dari ontologi, lingkungan software, serta dokumentasi mengenai kerangka referensi pada setiap tahap dan diantara tahap lifecycle. Tahap evaluasi terdiri dari dua proses yaitu proses verifikasi dan proses validasi. Proses verisifkasi ini mengacu pada proses teknis yang menjamin kebenaran dari suatu ontologi, lingkungan software, dan dokumentasi yang berkaitan dengan kerangka acuan pada setiap tahap dan diantara tahap lifecycle. Sedangkan, pada proses validasi ini menjamin model ontologi, lingkungan software, dan dokumentasi yang sesuai dengan sistem yang ingin diwakili.

2.7 Dokumentasi

Pada tahap dokumentasi ini akan dilakukan proses dokumentasi dalam kode ontologi, teks *natural language* yang dilampirkann pada definisi formal, ataupun makalah yang akan diterbitkan dalam proses konferensi dan jurnal yang mengatur mengenai pertanyaan-pertanyaan yang penting dari ontologi yang sudah ada atau sudah dibangun.

3. Hasil dan Pembahasan

Pada penelitian ini, akan dibangun suatu model ontologi yang memiliki domain Produk Bodycare. Di bawah ini merupakan hasil yang diperoleh dari setiap tahapan-tahapan pada penelitian yang sudah dilakukan.

3.1 Spesifikasi

Pada tahap spesifikasi ini akan memberikan suatu spesifikasi mengenai ontologi yang telah dibangun. Berikut ini merupakan deskripsi dari model ontologi pada “Produk Bodycare”.

a. Domain : Produk Bodycare
b. Tanggal : 1 Oktober 2022
c. Diimplementasikan Oleh : Ida Ayu Taria Putri Mahadewi
d. Level Formalitas : Formal
e. Ruang Lingkup : Produk Bodycare
f. Sumber Pengetahuan : Internet (website resmi produsen bodycare lokal)

3.2 Akusisi Pengetahuan

Pada tahap akusisi ini dibuat untuk memperoleh pengetahuan yang dapat dimanfaatkan pada model ontologi Produk Bodycare yang dibangun. Pada tahapan akusisi di penelitian ini yaitu sebagai berikut :

a. Melakukan wawancara dengan admin produsen bodycare lokal untuk mendapatkan informasi mengenai domain produk bodycare.
b. Melakukan analisis teks informal untuk mempelajari konsep-konsep utama yang diberikan dalam buku dan studi pegangan.
c. Melakukan analisis teks formal dengan melakukan identifikasi pada struktur yang akan dideteksi (definisi, penegasan, dan lain sebagainya) dan jenis pengetahuan yang akan dikontribusikan oleh masing-masing (konsep, atribut, nilai, dan hubungan).

Data yang akan digunakan untuk membangun model ontologi ini yaitu data produk bodycare dari beberapa produsen bodycare lokal. Data yang diperoleh berasal dari web resmi dari beberapa produsen bodycare tersebut. Jumlah data yang diperoleh yaitu sebanyak 25 produk bodycare dari 5 merek produk bodycare lokal.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Merek Bodycare	Nama Produk	Jenis Bodycare	Masalah Kulit	Waktu Penggunaan	Usia
Scarlett Whitening	Fragrance Body Cream	Body Cream	Kulit kering	Pagi dan malam	≥ 13 tahun
Marina	Healthy & Glow	Body Lotion	Kulit gelap dan kulit kusam	Pagi dan Malam	≥ 12 tahun
Scarlett Whitening	Body Serum Happy	Body Serum	Warna kulit tidak merata, kulit gelap, dan kulit kusam.	Pagi dan malam	≥ 13 tahun
Sensatia Botanicals	Calming Body Wash	Body Wash	Kulit kering, kulit sensifit, dan iritasi	Pagi dan malam	≥ 13 tahun
Sensatia Botanicals	Lemongrass & Mandarin Sea Salt Scrub	Body Scrub	Kulit kering	1-2 kali seminggu (pagi atau malam)	≥ 13 tahun

Tabel 1. Contoh Data Produk Bodycare

3.3 Konseptualisasi

Pada tahap konseptualisasi ini ditujukan untuk merancang konsep yang akan digunakan untuk mendeskripsikan masalah dan solusi yang akan digunakan. Pada tahap konseptualisasi ini akan dibangun daftar istilah lengkap yang mencakup konsep, *instance*, kata kerja, dan properti yang berhubungan dengan domain Produk Bodycare.

3.4 Integrasi

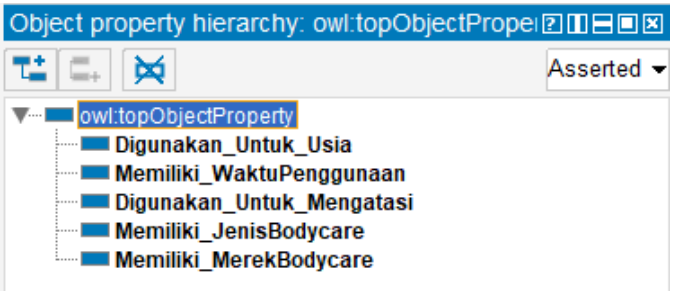
Pada tahap integrasi ini digunakan untuk menggabungkan atau mengintegrasikan model ontologi yang sudah ada dengan model ontologi yang akan dibangun atau dibuat. Dengan pertimbangan yang ada agar dapat sesuai dengan domain Produk Bodycare, pemilihan model ontologi yang sesuai dapat memudahkan hasil yang diharapkan.

3.5 Implementasi

Pada tahap implementasi model ontologi pada Produk Bodycare ini menggunakan aplikasi Protégé versi 5.5.0. Protégé merupakan suatu alat yang digunakan untuk membantu membuat sebuah domain ontologi, menyesuaikan form untuk entry data, dan memasukan data. Berdasarkan hasil pada tahap implementasi ini maka, dihasilkan konsep *class* yang digunakan pada model ontologi yang terlihat pada Gambar 2, hubungan antara *class* atau *relationships* yang terdapat pada model ontologi yang didefinisikan pada *object properties* dapat dilihat pada Gambar 3. *Instance* yang terdapat pada masing-masing *class* yang didefinisikan pada bagian individual dapat terlihat pada Gambar 4. Atribut pada masing-masing *class* atau *instance* dapat dilihat pada Gambar 5. Lalu, untuk hasil dan struktur hubungan antar *class* tersebut dapat dilihat pada bagian Ontograf yang terdapat pada Gambar 6 dan 7.

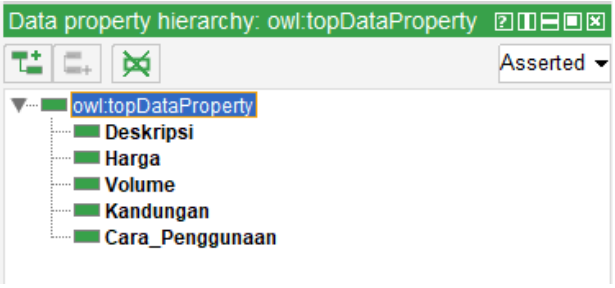
Gambar 2. *Class* dari Ontologi Produk Bodycare

Pada Gambar 2 di atas, terdapat 10 *class* yang ada pada model ontologi produk bodycare. *Class* Bodycare memiliki *subclass* yaitu Nama_Produk. Lalu, *subclass* Nama_Produk tersebut memiliki tiga *subclass* yaitu Jenis_Bodycare, Merek_Bodycare, dan Waktu_Penggunaan. Kemudian, pada *class* Kulit memiliki *subclass* yaitu Kulit_Tubuh dan *subclass* Kulit_Tubuh ini memiliki *subclass* lagi yaitu Masalah_Kulit. Lalu, *class* Pengguna memiliki *subclass* yaitu Usia.



Gambar 3. *Object Properties* dari Ontologi Produk Bodycare

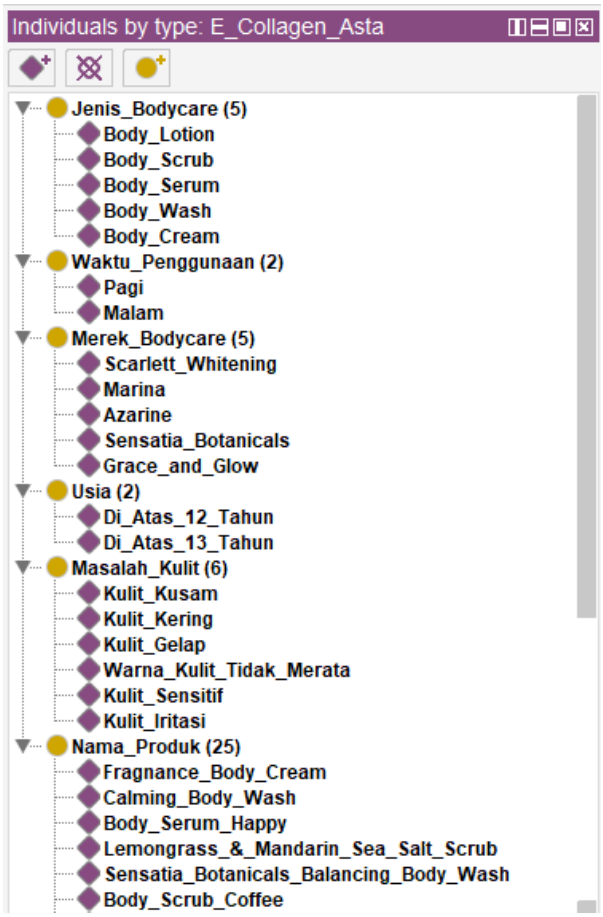
Pada Gambar 3 di atas, terdapat 5 *Object Properties* yang ada pada model ontologi produk bodycare. Pada masing-masing *Object Properties* akan menghubungkan antar *instance* atau individu.



Gambar 4. *Data Properties* dari Ontologi Produk Bodycare

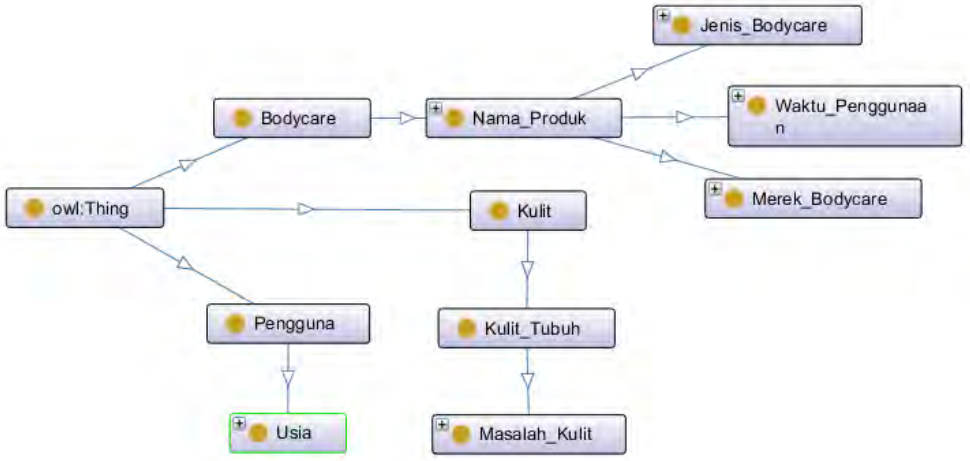
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pada Gambar 4 di atas, terdapat 5 *Data Properties* yang ada pada model ontologi produk bodycare. *Data Properties* ini digunakan untuk menghubungkan *instance* dengan *datatype value* seperti *text, string,* atau *number*.

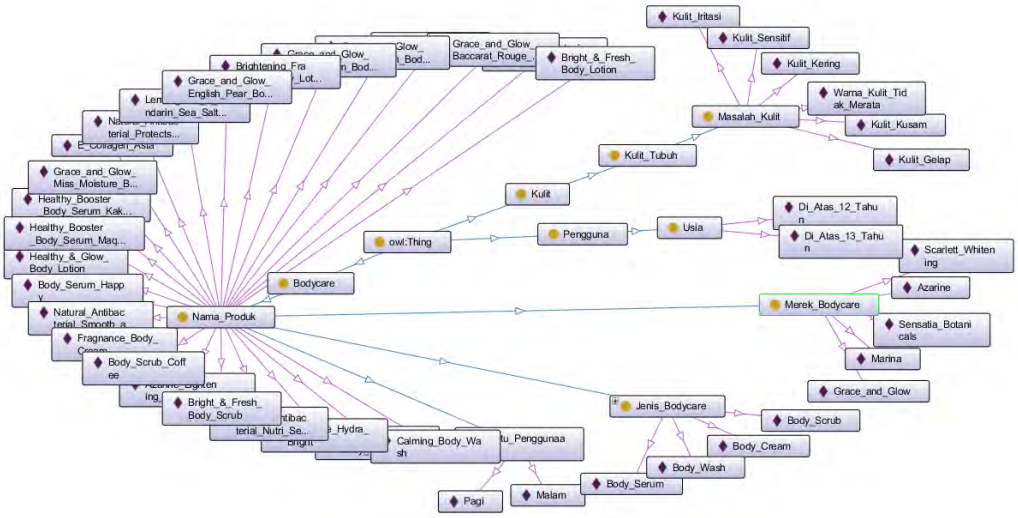


Gambar 5. Individual dari Ontologi Produk Bodycare

Pada Gambar 5 di atas, terdapat beberapa individual yang telah dihasilkan pada setiap *class* yang telah dibuat pada ontologi produk bodycare. Terdapat 5 individual untuk *class* Jenis_Bodycare, 2 individual untuk *class* Waktu_Penggunaan, 5 individual untuk *class* Merek_Bodycare, 2 individual untuk *class* Usia, 6 individual untuk *class* Masalah_Kulit, dan 25 individual untuk *class* Nama_Produk.



Gambar 6. Ontograf dari Ontologi Produk Bodycare



Gambar 7. Ontograf Dengan Individual dari Ontologi Produk Bodycare

Pada Gambar 6 dan Gambar 7 di atas merupakan contoh hubungan semantik yang menggambarkan masing-masing *class, object properties,* dan individual yang dibuat atau dibangun pada ontologi produk bodycare yang dimana pada hubungan tersebut direpresentasikan ke dalam ontograf dalam bentuk gambar. Pada ontograf dari ontologi produk bodycare di atas, terdapat 10 *class* utama serta hubungan diantaranya. Hubungan antara *class* dengan *subclass* ditandai dengan tanda panah berwarna biru. Sedangkan, tanda panah yang berwarna ungu tersebut menandakan adanya relasi yang dihubungkan oleh *object properties*.

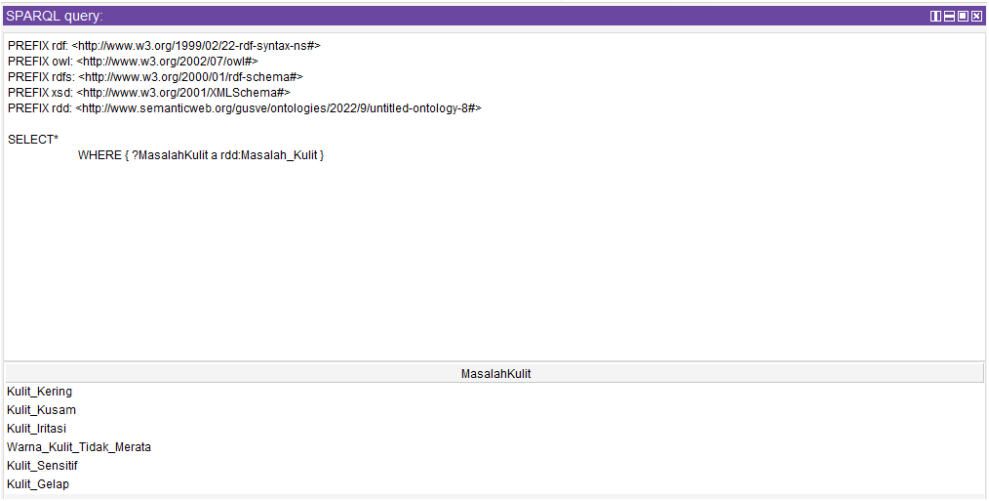
3.6 Evaluasi

Pada tahap evaluasi ini dapat dilakukan dengan cara melakukan *query* dengan menggunakan SPARQL *query* yang ada pada aplikasi Protégé versi 5.5.0. Pertanyaan-pertanyaan yang diinginkan dapat diubah ke dalam bentuk query SPARQL. Kemudian,

p-ISSN: 2301-5373
e-ISSN: 2654-5101

akan ditampilkan hasil yang terdapat pada model ontologi yang telah dibuat atau dibangun. Hasil dari *query* tersebut dapat dilihat pada Gambar 6. Pada Gambar 6 tersebut, SPARQL *query* yang dijalankan yaitu sebagai berikut ini :

a. Pertanyaan : Masalah kulit apa saja yang dapat terjadi pada kulit tubuh?
   Hasil *query* dapat dilihat pada Gambar 7 di bawah yang menampilkan jawaban dari pertanyaan masalah kulit apa saja yang dapat terjadi pada kulit tubuh. Dari pertanyaan tersebut akan menampilkan satu aspek sebagai hasilnya yaitu “MasalahKulit”.



Gambar 7. Hasil SPARQL Query

3.7 Dokumentasi

Pada tahap dokumentasi ini memiliki tujuan untuk menghasilkan dokumentasi dari pembangunan model ontologii Produk Bodycare. Adapun dokumentasi yang telah dilakukan yaitu berupa hasil dari laporan jurnal ini.

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan dengan tahapan-tahapan di atas, maka model ontologi yang terkait dengan domain produk bodycare telah selesai dibangun atau dibuat. Pembangunan model ontologi ini menggunakan aplikasi Protégé versi 5.5.0 dengan metode yang digunakan yaitu Methontologi dan menghasilkan 10 *class*, 5 *Object Properties,* 5 *Data Properties*, dan 45 individual. Pada tahap evaluasi dilakukan dengan menggunakan SPARQL *query*, model ontologi yang dibangun dapat merepresentasikan pertanyaan-pertanyaan yang diajukan dengan baik terkait dengan doamin produk bodycare. Model yang telah dihasilkan melalui pembangunan ontologi produk bodycare dapat digunakan sebagai dasar pengembangan sistem manajemen pengetahuan mengenai produk bodycare.

Daftar Pustaka

[1] C. Pramartha, “Pengembangan Ontologi Tujuan Wisata Bali Dengan Pendekatan Kulkul Knowledge Framework,” *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 77–89, 2020, doi: 10.31598/sintechjournal.v3i2.592.

[2] K. W. Triyoga, D. E. Cahyani, S. W. Sihwi, F. Matematika, and U. S. Maret, “Pembangunan Ontology Berbasis Metode Methontology Untuk Domain Tuberculosis,” pp. 47–54, 2019.

[3] A. Nugroho, “Membangun Ontologi Jurnal Menggunakan Protege,” *J. Transform.*, vol. 10, no. 1, p. 20, 2012, doi: 10.26623/transformatika.v10i1.66.

[4] P. D. Bangsa and I. Hermawan, “Jurnal Teknologi Terpadu,” *J. Teknol. Terpadu*, vol. 7, no. 1, pp.

15–22, 2021, [Online]. Available: https://media.neliti.com/media/publications/493730-water-ph-and-turbidity-control-system-in-0a553e14.pdf.

[5] K. D. P. Novianti, “Implementasi Methontology Untuk Pembangunan Model,” *J. TEKNOIF*, vol. 4, no. 1, pp. 40–47, 2016, [Online]. Available: https://ejournal.itp.ac.id/index.php/tinformatika/article/view/588/424.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Jenis Sampah Menggunakan Metode *Transfer Learning* Pada *Convolutional Neural Network* (CNN)

[1]Wahyu Vidiadivani, [2]I Ketut Gede Suhartana

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]vany0141@gmail.com
[2]ikg.suhartana@unud.ac.id

Abstract

Sampah merupakan buangan atau sisa hasil produksi. Berdasarkan sumbernya, sampah dapat berasal dari beberapa sumber atau tempat, yaitu sampah yang berasal dari masyarakat atau pemukiman yang biasanya sampahnya bersifat basah dan berjenis organik, sedangkan sampah dari tempat umum atau perdagangan sampahnya bersifat kering dan berjenis anorganik. Jumlah sampah yang sangat besar dan beragamnya jenis sampah yang tersebar di masyakarat, perlu adanya klasifikasi yang dapat mengidentifikasi jenis-jenis sampah ke beberapa kategori sehingga mudah untuk didaur ulang kembali. Klasifikasi jenis sampah pada penelitian ini dibagi menjadi 12 jenis, yaitu *trash, clothes, green-glass, meal, paper, battery, biological, shoes brown-glass, cardboard plastic,* dan *white-glass* menggunakan metode *Transfer Learning* pada *Convolutional Neural Network* (CNN). CNN (Convolutional Neural Network) merupakan salah satu algoritma deep learning yang populer digunakan untuk klasifikasi citra dan dinilai memiliki performa yang bagus. Pada penelitian ini, arsitektur yang digunakan adalah EffecienNetB0. Dataset yang digunakan dengan total data sebanyak 12412 data, data yang tervalidasi sebanyak 1552 data, dan data yang digunakan pada proses testing sebanyak 1552 data yang terbagi ke 12 kelas. Hasil dari penelitian ini mengunakan model EfficientNetB0 menggunakan Metode Transfer Learning dengan tingkat akurasi sebesar 88,53 % dengan loss sebesar 41,41%

Keywords: *Waste, Computer Vision, CNN, Transfer Learning, EffecientNetB0*

1. Introduction

Sampah merupakan permasalahan setiap negara dari tahun ke tahun yang diyakini bahwa kehidupan manusia sehari-hari tidak akan pernah terlepas dari sampah dan aktivitas-aktivitas manusia menghasilkan sampah. Terdapat berbagai jenis sampah yang tersebar di masyarakat sehingga memakan waktu dan tenaga yang sangat besar untuk mengolah sampah-sampah tersebut. Pada penelitian ini mengkhususkan jenis-jenis sampah ke dalam 12 kelas, yaitu *trash, clothes, green-glass, meal, paper, battery, biological, shoes brown-glass, cardboard plastic, dan white-glass.*

CNN (Convolutional Neural Network) merupakan salah satu algoritma *Deep Learning* yang umum digunakan untuk klasifikasi citra dimana terdapat beberapa arsitektur yang digunakan pada penelitian ini, yaitu EfficientNetB0 yang merupakan salah satu pre-trained model CNN (Convolutional Neural Network) untuk metode Transfer Learning pada permasalahan klasifikasi citra dimana arsitektur ini memiliki fokus pada keefesienan waktu. Menurut Tan & Le (2019), arsitektus EffecientNetB0 sudah menunjukkan akurasi tinggi dengan jumlah parameter rendah dan cepat dibandingkan model pre-trained lainnya. Pada penelitian sebelumnya yang dilakukan oleh Kusumaningrum T. F. (2018) yaitu implementasi CNN (Convolutional Neural Network) dalam klasifikasi jamur menggunakan metode Keras menunjukkan akurasi sebesar 81,667 %. Lalu, pada penelitian yang dilakukan oleh Naufal M. F. dan Kusuma S. F. (2021) melakukan pendeteksian citra masker wajah menggunakan CNN (Convolutional Neural Network) menghasilkan akurasi sebesar 98% menggunakan arsitektur Xception. Dari beberapa penelitian yang dilakukan, dapat disimpulkan bahwa CNN (Convolutional Neural Network) merupakan salah satu algoritma terbaik yang populer digunakan untuk klasifikasi data citra dan perlu adanya

penelitian lebih lanjut untuk dapat membuktikan kesesuaian dalam pemilihan arsitektur CNN (Convolutional Neural Network).

Tujuan utama dari penelitian ini adalah mengklasifikasikan 12 jenis yaitu *battery, biological, brown-glass, cardboard, clothes, green-glass, meal, paper, plastic, shoes, trash,* dan *white-glass*, yang tersebar di masyarakat menggunakan dataset yang diperoleh dari situs kaggle yang terdiri dari 12412 data. Penggunaan CNN (Convolutional Neural Network) dengan Metode *Transfer Learning* yang diharapkan menghasilkan nilai akurasi yang lebih baik. Penelitian ini juga dilakukan untuk menganalisis kinerja masing-masing arsitektur dan memberikan informasi yang bermanfaat bagi peneliti memilih arsitektur yang tepat.

2. Reseach Methods

Metode yang digunakan pada penelitan ini adalah CNN CNN (Convolutional Neural Network) dengan metode Transfer Learning menggunakan arsitektur EffecientNetB0 untuk klasifikasi jenis sampah yang dibagi menjadi 12 kelas, yaitu *trash, clothes, green-glass, meal, paper, battery, biological, shoes brown-glass, cardboard plastic,* dan *white-glass* yang diperoleh dari situs website kaggle yang menyediakan dataset untuk penelitian. Langkah-langkah yang dilaksanakan pada penelitian ini adalah sebagai berikut:

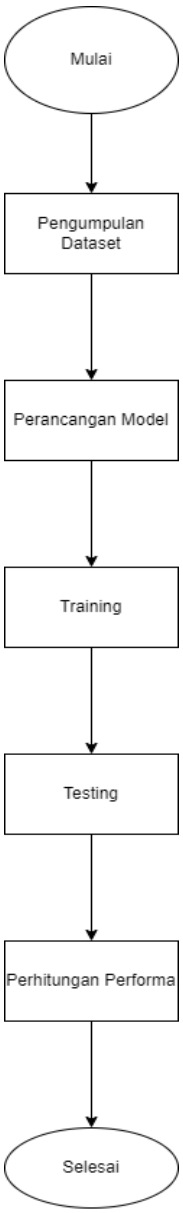


Figure 1. Tahapan Penelitian

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Data yang digunakan pada penelitian ini adalah dataset yang diperoleh dari situs website kaggle yang terdiri dari 12 jenis sampah, yaitu *battery (945), biological (985), brown-glass (607), cardboard (891), clothes (5325), green-glass (629), metal (769), paper (1050), plastic (865), shoes (1977), trash (697), dan white-glass (775).*

Table 1. Data Jenis-Jenis Sampah

Jenis Sampah	Jumlah
Metal	769
White-Glass	775
Biological	985
Paper	1050
Brown-glass	607
Battery	945
Trash	697
Cardboard	891
Shoes	1977
Clothes	5325
Plastic	865
Green-glass	629

Selanjutnya dilakukan tahapan perancangan model dengan metode Transfer Learning (Pre-Trained Model) menggunakan arsitektur EffecientNetB0 yang harapannya dengan menggunakan arsitektur terebut dapat menghasilkan hasil yang sesuai dengan menggunakan parameter yang tepat.

```
base_model = EfficientNetB0(weights='imagenet', include_top=False,
input_shape=(im_shape[0], im_shape[1], 3))

x = base_model.output
x = Flatten()(x)
x = Dense(100, activation='relu')(x)
predictions     =     Dense(num_classes,     activation='softmax',
kernel_initializer='random_uniform')(x)

model = Model(inputs=base_model.input, outputs=predictions)

# Freezing pretrained layers
for layer in base_model.layers:
      layer.trainable=False
# model.summary()

optimizer = Adam()
model.compile(optimizer=optimizer,loss='categorical_crossentropy',m
etrics=['accuracy'])
```

Saatnya melakukan proses Training menggunakan subset dataset yang digunakan untuk melatih model yang sudah dirancang sebelumnya dan proses Testing untuk melakukan pengecekan akurasi model yang telah dirancang.

```
score = model.evaluate(val_generator)
print('Val loss:', score[0])
print('Val accuracy:', score[1])

score = model.evaluate(test_generator)
print('Test loss:', score[0])
print('Test accuracy:', score[1])
```

3. Result and Discussion

3.1. Pengumpulan Data

Dataset yang diambil dalam format .jpg di situs Kaggle dengan total gambar yang terkumpul sebanyak 12412 gambar yang kemudian dibagi menjadi 12 kelas, yaitu *battery (945), biological (985), brown-glass (607), cardboard (891), clothes (5325), green-glass (629), metal (769), paper (1050), plastic (865), shoes (1977), trash (697), dan white-glass (775).* Pada data train terdapat 7557 gambar dengan 12 kelas berbeda. Sedangkan, pada data test ditemukan 1940 gambar dengan 12 kelas berbeda.

3.2. Pelatihan Data

Data training yang digunakan adalah 12 % dari total data gambar keseluruhan yang terbagi di 12 kelas. Berikut hasil training data adalah sebagai berikut:

```
Epoch 1/5
121/121 [==============================] - 319s 3s/step - loss: 0.0179 - accuracy:
0.9951 - val_loss: 0.5203 - val_accuracy: 0.9309

Epoch 00001: val_loss improved from inf to 0.52026, saving model to model_Efficien
tnetB0.h5
Epoch 2/5
121/121 [==============================] - 316s 3s/step - loss: 0.0273 - accuracy:
0.9935 - val_loss: 0.5270 - val_accuracy: 0.9314

Epoch 00002: val_loss did not improve from 0.52026
Epoch 3/5
121/121 [==============================] - 316s 3s/step - loss: 0.0122 - accuracy:
0.9965 - val_loss: 0.6267 - val_accuracy: 0.9264

Epoch 00003: val_loss did not improve from 0.52026
Epoch 4/5
121/121 [==============================] - 319s 3s/step - loss: 0.0203 - accuracy:
0.9953 - val_loss: 0.5152 - val_accuracy: 0.9392

Epoch 00004: val_loss improved from 0.52026 to 0.51520, saving model to model_Effi
cientnetB0.h5
Epoch 5/5
121/121 [==============================] - 323s 3s/step - loss: 0.0184 - accuracy:
0.9960 - val_loss: 0.5603 - val_accuracy: 0.9297

Epoch 00005: val_loss did not improve from 0.51520
```

Figure 2. Hasil Training

Hasil training dari data di atas adalah 0,01 % loss dan 99,6 % akurasi kemudian dengan validation loss sebesar 5 % dan validation accuracy sebesar 92,97 %.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

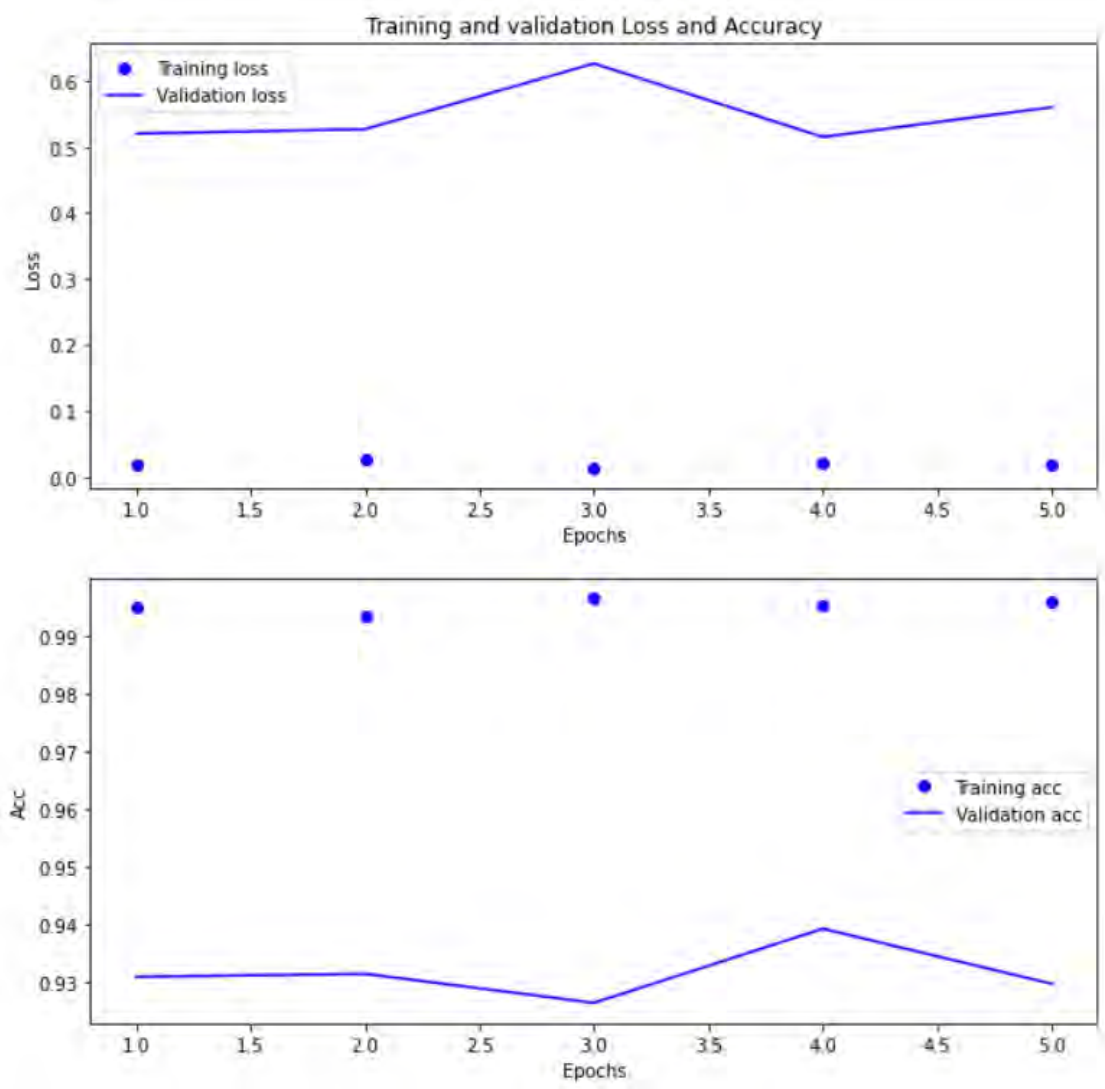


Figure 3. Grafik Training dan Validation Accuracy

Grafik berikut menggambarkan Training dan Validation Accuracy dan Training dan Validation Loss yang dapat disimpulkan bahwa akurasi yang dihasilkan sangat baik, namun loss yang dihasilkan juga terbilang cukup besar. Hal tersebut dikarenakan kurangnya sampel data gambar yang digunakan.

3.3. Hasil Klasifikasi

Table 3. Classification Report

Jenis Sampah	Precission	Recall	F1-Score	Support
White-Glass	0.69	0.83	0.75	98
Metal	0.88	0.76	0.81	123
Biological	0.92	0.96	0.94	118
Paper	0.89	0.92	0.90	123
Brown-glass	0.96	0.79	0.93	62
Battery	0.92	0.98	0.95	99
Trash	0.99	0.83	0.75	100
Cardboard	0.96	0.91	0.93	120
Shoes	0.95	0.99	0.97	257
Clothes	0.99	0.99	0.99	629
Plastic	0.82	0.78	0.80	123
Green-glass	0.88	0.95	0.92	88
Accuracy			0.93	1940
Macro avg	0.90	0.89	0.90	1940
Weighted avg	0.93	0.93	0.93	1950

4. Conclusion

Penelitian ini memberikan solusi tentang pengidentifikasian dan klasifikasi jenis-jenis sampah yang ada di masyarakat yang menggunakan algoritma CNN (Convolutional Neural Network) dengan metode Transfer Learning dan model arsitektur CNN, yaitu EffecientB0. Hasil dari penelitian ini mengunakan model EfficientNetB0 menggunakan Metode Transfer Learning dengan tingkat akurasi sebesar 88,53 % dengan loss sebesar 5% dimana loss yang dihasilkan masih cukup dibilang besar. Hal ini dikarenakan kurangnya sampel data gambar, yaitu hanya menggunakan 12% dari total gambar keseluruhan. Solusi yang dapat diberikan adalah menambahkan sampel data gambar pada validation, menambahkan beberapa epoches, dan pemodifikasian model CNN.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

References

Azahro Choirunisa, N., Karlita, T., & Asmara, R. (2021). Deteksi Ras Kucing Menggunakan Compound Model Scaling Convolutional Neural Network. *Technomedia Journal*, *6*(2), 236–251. https://doi.org/10.33050/tmj.v6i2.1704

Alamgunawan, S., & Kristian, Y. (2021). Klasifikasi Tekstur Serat Kayu pada Citra Mikroskopik Veneer Memanfaatkan Deep Convolutional Neural Network. *Journal of Intelligent System and Computation*, *2*(1), 06–11. https://doi.org/10.52985/insyst.v2i1.152

Ibnul Rasidi, A., Pasaribu, Y. A. H., Ziqri, A., & Adhinata, F. D. (2022). Klasifikasi Sampah Organik dan Non-Organik Menggunakan Convolutional Neural Network. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1), 142–149. https://doi.org/10.28932/jutisi.v8i1.4314

Leonardo, L., Yohannes, Y., & Hartati, E. (2020). Klasifikasi Sampah Daur Ulang Menggunakan Support Vector Machine Dengan Fitur Local Binary Pattern. *Jurnal Algoritme*, *1*(1), 78–90. https://doi.org/10.35957/algoritme.v1i1.440

Rais, I. L., & Jondri, J. (2020). Klasifikasi Data Kuesioner dengan Metode Recurrent Neural Network. *EProceedings of Engineering*, *7*(1), 2817–2826.

Rima Dias Ramadhani, Nur Aziz Thohari, A., Kartiko, C., Junaidi, A., Ginanjar Laksana, T., & Alim Setya Nugraha, N. (2021). Optimasi Akurasi Metode Convolutional Neural Network untuk Identifikasi Jenis Sampah. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 312–318. https://doi.org/10.29207/resti.v5i2.2754

Solihin, A., Mulyana, D. I., & Yel, M. B. (2022). Klasifikasi Jenis Alat Musik Tradisional Papua menggunakan Metode Transfer Learning dan Data Augmentasi. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, *5*(2), 36–44. https://doi.org/10.47970/siskom-kb.v5i2.279

Wonohadidjojo, D. M. (2021). Perbandingan Convolutional Neural Network pada Transfer Learning Method untuk Mengklasifikasikan Sel Darah Putih. *Ultimatics : Jurnal Teknik Informatika*, *13*(1), 51–57. https://doi.org/10.31937/ti.v13i1.2040

Zayd, M. H., Oktavian, M. W., Meranggi, D. G. T., Figo, J. A., & Yudistira, N. (2022). Improvement of garbage classification using pretrained Convolutional Neural Network. *Teknologi*, *12*(1), 1–8. https://doi.org/10.26594/teknologi.v12i1.2403

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

CI/CD Implementation On Microservices For Increasing Availability On Big Data Processing

Kompiang Gede Sukadharma[a1], I Putu Gede Hendra Suputra[a2]

[a]Informatics, Udayana University
Jl. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]kompiang.sukadharma@gmail.com
[2] hendra.suputra@unud.ac.id

Abstract

Big Data Processing needed a reliable system that not let the data loss. But sometimes we need to make the system down for a while because we need to push the newest changes of the system. The automation will help us achieve that. Continuous Integration and Continuous Deployment help us to reduce the downtime and increase the availability of the system. Thus, the implication will be led to reduce of data loss. This research focusses on the implementation of CI/CD Pipeline on single-point-of-failure service on Microservices Architecture. This research use Load-Testing to measure data loss on certain amount of time. The result on this research show that implementing CI/CD Pipeline on the Microservices that we made, make the down time will be less than 45 Second with 20 Virtual user who send the data.

Keywords: *Microservices, CI/CD, Load Test, GitHub Actions, Single-Point-of-Failure*

1. Pendahuluan

Seiring dengan meningkatnya kebutuhan dari industri teknologi, hal tersebut juga membuat banyak perubahan untuk beradaptasi terhadap perubahan yang perlu dilakukan. Tingginya perubahan yang terjadi membuat waktu yang digunakan untuk mengirimkan perubahan terbaru ke pengguna menjadi hal yang krusial. Perubahan yang terjadi memiliki kemungkinan membuat sebuah layanan tidak dapat diakses. Dalam pemrosesan Big Data, diperlukan arsitektur yang andal dan tingkat *availability* yang tinggi agar data-data dari pengguna tidak hilang. Oleh karena itu, diperlukan metode yang tepat untuk mengirimkan perubahan ke pengguna agar tidak mengganggu layanan yang sedang berjalan. Dalam pemrosesan Big Data, kita tidak bisa melakukannya secara konvensional [1]. Hal tersebut didukung dengan beberapa variabel yang telah teridentifikasi, yaitu (1) *volume*, (2) *velocity*, (3) *variety*, (4) *veracity* [2]. Variabel tersebut selanjutnya disebut sebagai sifat-sifat yang dimiliki oleh Big Data.

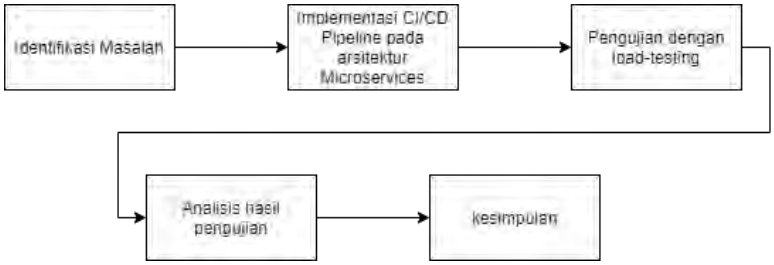
Tingginya tingkat *velocity* dan besarnya *volume* ketika melakukan pemrosesan Big Data menjadi tantangan untuk mengirimkan rilis terbaru ke pengguna. Untuk itu, praktik *continous* diharapkan menyediakan beberapa kelebihan, seperti: (1) Mendapatkan tanggapan lebih banyak dan lebih cepat dalam proses pengembangan dan juga dari pengguna, karena perubahan terkirim lebih cepat; (2) Merilis perubahan secara cepat dan reliabel yang mana akan meningkatkan kualitas produk; (3) *Continuous Deployment* atau CD akan mempercepat proses perubahan lingkungan aplikasi dari *development* dan *production* [3]; (4) Layanan yang sedang berjalan tidak mengalami *downtime* yang lama, jadi data yang sedang terkirim tidak banyak yang hilang.

Grab, salah satu *startup* yang memiliki layanan *superapp* yang bergerak di bidang transportasi, pengantaran makanan, dan *payment solutions*. Pada Agustus tahun 2022 mencatat 23.6 Juta transaksi pengguna [4]. Dengan banyaknya transaksi tersebut perusahaan ini menggunakan praktik *Continous* untuk mengirimkan perubahan rilis dengan lebih cepat agar operasi ini tidak memengaruhi kecepatan dan stabilitas mereka. Pada artikel yang dirilis oleh Grab, terdapat peningkatan banyaknya perubahan yang dirilis dari *staging environment* ke *production environment* sebanyak 14,6% dari hanya 10% perubahan yang diimplementasikan menggunakan metode konvensional pada tahun 2018 menjadi 24.6% perubahan yang diimplementasikan dari *staging* ke *production* menggunakan teknik CI/CD. Pada artikel disebutkan bahwa mereka menghemat 5000 hari kerja pda tahun 2020 sendiri. Hal tersebut setara dengan 20 tahun kerja [5].

2. Metodologi Penelitian

2.1 Alur Penelitian

Penulis mengikuti alur penelitian yang terdapat pada **Gambar 1** agar penelitian dapat berjalan secara lancar dan runtut



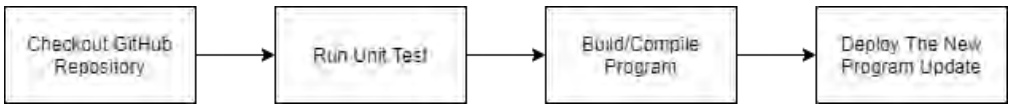
Gambar 1. Alur penelitian

Penjelasan dari alur penelitian di atas sebagai berikut:

a. Identifikasi Masalah
   Pada awal penelitian, penulis melakukan identifikasi masalah yang ada di lapangan dan juga melakukan riset dengan menggunakan literatur yang dapat membantu penelitian. Lalu, masalah yang ditemukan oleh penulis adalah besarnya kemungkinan *data loss* pada layanan *single point of failure* yang terjadi ketika suatu layanan dalam *microservices* dimatikan untuk mengirimkan perubahan yang terbaru.
b. Implementasi CI/CD Pipeline Pada Arsitektur Microservices
   Pada tahapan ini, penulis membuat sebuah sistem dengan arsitektur *microservices* dan juga *pipeline* CI/CD untuk menyimulasikan bagaimana CI/CD dapat meningkatkan *availibility* pada sebuah sistem. Implementasi menggunakan beberapa teknologi, untuk membuat sistem dengan arsitektur *microservices* penulis menggunakan bahasa pemrograman Go dan juga VS Code. Lalu, untuk membuat pipeline CI/CD, penulis menggunakan GitHub Actions dan juga menggunakan *compute service* dari AWS sebagai *instance* untuk menerapkan sistem ini.
c. Pengujian Dengan Load Testing
   Pengujian dilakukan dengan metode *Load-Testing* menggunakan K6 dan memperhatikan beberapa metriks.
d. Analisis Hasil pengujian
   Peneliti melakukan analisis terhadap beberapa metriks untuk menarik sebuah kesimpulan
e. Kesimpulan
   Pada tahap ini, penulis menarik kesimpulan dari beberapa metriks untuk menarik kesimpulkan dari hasil penelitian ini

2.2 CI/CD Pipeline

Pengujian pada metode ini berada pada lingkup proses pengintegrasian dan pengiriman perubahan terbaru pada sistem arsitektur *microservices*. Proses pengintegrasian dan pengiriman perubahan terbaru dapat dibuat ke dalam sebuah alur kerja yang kita konfigurasikan agar prosesnya berjalan secara otomatis serta dapat meminimalkan *downtime* dari sebuah service. Selain itu, pada penelitian ini penulis juga menggunakan metode ini untuk meneliti *data loss* yang diharapkan metode ini dapat meminimalkan hal tersebut. Alat yang digunakan dalam lingkup penelitian ini untuk menjalankan CI/CD Pipeline pada penelitian kali ini adalah GitHub Actions. **Gambar 2** menunjukkan alur kerja yang akan kita implementasikan menggunakan CI/CD Pipeline.



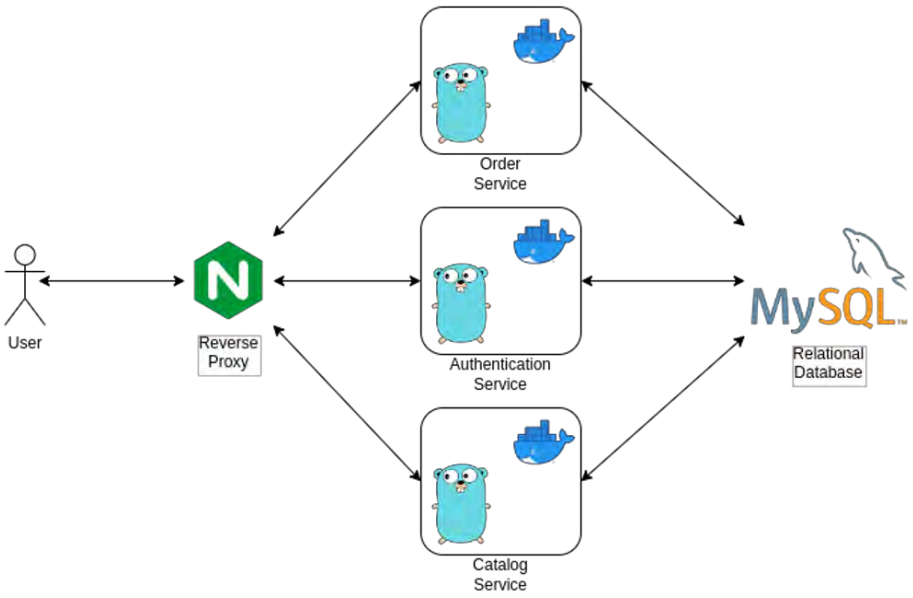
Gambar 2. Workflow CI/CD Pipeline

2.3 Arsitektur Microservices

Penelitian ini menggunakan arsitektur *microservices* untuk melihat dampak sebuah *service* yang memiliki sifat *single point of failure* atau sebuah *service* yang apabila terjadi kegagalan, kegagalan tersebut akan mengganggu kinerja dari sebuah sistem. Hal tersebut merupakan hal yang fatal pada arsitektur *microservices*. *Microservices* merupakan sebuah model arsitektur [6]. Ide utama dari

p-ISSN: 2301-5373
e-ISSN: 2654-5101

arsitektur jenis ini adalah memecah sistem yang lebih besar menjadi sistem-sistem yang lebih kecil. Hal ini dapat mengurangi kompleksitas dari sistem [3], membuat sistem yang lebih kecil menjadi lebih independen dan dapat membuat hubungan antar sistem atau *service* menjadi lebih renggang. Selain itu, sistem yang independen berimplikasi terhadap proses pengembangan dan *deploy* yang lebih mandiri. Lalu, kelebihan lain yang dimiliki oleh *microservices* adalah ketika sebuah *service* mengalami kegagalan, kegagalan tersebut hanya terisolasi dan tidak mengganggu *service lain*, kecuali *service* tersebut adalah *single point of failure.* **Gambar 3** merepresentasikan arsitektur yang penulis gunakan pada penelitian ini dengan mengimplementasikan *microservices* serta menjadikan Authentication Service menjadi sebuah *single point of failure.*



Gambar 3. Arsitektur *Microservices*

2.4 Docker

Penelitian ini menggunakan teknologi kontainerisasi untuk melakukan isolasi pada setiap *service* yang digunakan. Teknologi kontainerisasi memungkinkan kita untuk membuat sebuah virtualisasi pada tingkat aplikasi yang akan berjalan secara konsisten pada setiap infrastruktur. Kontainerisasi juga hanya akan membungkus aplikasi dengan aplikasi yang kita perlukan saja, hal tersebut akan membuat kontainer ini menjadi lebih ringan dan lebih cepat dimulai daripada virtualisasi pada tingkat sistem operasi. Salah satu *container engine* dan juga yang digunakan pada penelitian ini adalah Docker.

2.5 Load-Testing

Pengujian yang dilakukan pada penelitian ini menggunakan metode *Load-Testing* dengan bantuan K6. Metode *Load-Testing* akan menyimulasikan pengguna yang mengirim *request* sesuai dengan parameter yang diberikan ke program [7]. K6 merupakan alat bantuan untuk melakukan *Load-Testing* yang bersifat *open-source.* Dengan ini kita bisa mengetes performa dan keandalan sistem untuk mencari tahu masalah lebih awal. K6 akan menghasilkan kembalian berupa beberapa metris yang berkaitan tentang *request* yang telah dikirim. Metris ini akan membantu penulis untuk mengukur *data loss* yang terjadi pada penelitian ini. Beberapa metris yang digunakan pada penelitian dapat dilihat pada **Tabel 1**

Tabel 1. Metris Load-Testing

Key	Deskripsi
http_req_failed	Metris yang merepresentasikan banyaknya data yang berhasil dikirim dan banyaknya data yang gagal dikirim pada HTTP Request

3. Hasil dan Pembahasan

Berdasarkan alur penelitian yang dijelaskan pada bab sebelumnya. Maka hasil penelitian yang peneliti lakukan dapat dijabarkan sebagai beriku

3.1. Identifikasi Masalah

Pada penelitian ini penulis melakukan identifikasi masalah pada sistem yang menangani pemrosesan Big Data. Dengan mengidentifikasi sifat-sifat yang dimiliki oleh Big Data, seperti volume dan *velocity* yang dimiliki oleh Big Data serta juga mengidentifikasi perubahan yang terjadi pada industri yang sangat cepat. Penulis mengidentifikasi adanya masalah apabila terdapat perubahan yang diimplementasikan ketika sebuah layanan sedang melakukan pemrosesan data. Ketika mengimplementasikan perubahan pada sebuah layanan, layanan tersebut harus dimatikan terlebih dahulu. Pada layanan yang memproses Big Data, hal tersebut sangat fatal, karena ketika layanan dimatikan, data-data yang dikirimkan dari pengguna layanan akan hilang. Oleh karena itu, penulis memikirkan solusi untuk mempercepat proses pengimplementasian tersebut menggunakan CI/CD *pipeline*. CI/CD *pipeline*, akan membantu untuk meningkatkan *availibility time* dari sebuah layanan ketika dilakukan implementasi dari sebuah perubahan. Oleh karena itu, perubahan yang diimplementasikan akan lebih cepat sampai ke pengguna dengan waktu *down time* yang sedikit.

3.2. Implementasi CI/CD Pada Arsitektur Microservice

Sesuai dengan arsitektur *microservices* yang sudah penulis jelaskan pada bab sebelumnya, penulis akan melakukan implementasi perubahan terbaru terhadap Authentication Service. Authentication Service pada arsitektur di atas bersifat *single point of failure*. Oleh karena itu, Authentication Service perlu memiliki *down time* yang kecil agar layanan ini tidak mengganggu layanan lain bekerja. Untuk itu, penulis mengimplementasikan CI/CD *pipeline* menggunakan GitHub Actions. Intisari dari implementasi CI/CD *pipeline* dapat dilihat pada kode di **Tabel 2**.

Tabel 2. Implementasi CI/CD

	Kode
1	name: Deploy Auth Service
2	
3	on:
4	push:
5	tags:
6	- auth*
7	
8	env:
9	DOCKER_IMAGE_REPOSITORY_NAME: snatia-ci-cd
10	DOCKER_SERVICE_NAME: auth_service
11	
12	jobs:
13	build:
14	runs-on: ubuntu-latest
15	steps:
16	- name: Checkout Repository
17	uses: actions/checkout@v2
18	- name: Get Tag Name
19	run: echo "RELEASE_VERSION=${GITHUB_REF#refs/tags/}" >> $GITHUB_ENV
20	- name: Set up Go
21	uses: actions/setup-go@v2
22	with:
23	go-version: 1.18
24	- name: Run the unit test for this project
25	run: go test ./... -cover
26	- name: Login to Docker Hub
27	uses: docker/login-action@v1
28	with:
29	username: ${{ secrets.DOCKER_HUB_USERNAME }}
30	password: ${{ secrets.DOCKER_HUB_ACCESS_TOKEN }}
31	- name: Set Up Docker Buildx
32	uses: docker/setup-buildx-action@v1
33	- name: Build and push
34	uses: docker/build-push-action@v2
35	with:

p-ISSN: 2301-5373
e-ISSN: 2654-5101

```
36          context: .
37          file: ./auth.Dockerfile
38          push: true
39          tags: ${{ secrets.DOCKER_HUB_USERNAME }}/${{
    env.DOCKER_IMAGE_REPOSITORY_NAME }}:${{ env.RELEASE_VERSION }},${{
    secrets.DOCKER_HUB_USERNAME }}/${{ env.DOCKER_IMAGE_REPOSITORY_NAME
    }}:auth-latest
40      - name: Executing remote command using ssh
41        uses: appleboy/ssh-action@master
42        with:
43          host: ${{ secrets.SSH_SERVER }}
44          username: ${{ secrets.SSH_USERNAME }}
45          password: ${{ secrets.SSH_PASSWORD }}
46          port: 22
47          script: |
48            cd deploy/
49            docker compose rm -sf ${{ env.DOCKER_SERVICE_NAME }}
50            docker rmi ${{ secrets.DOCKER_HUB_USERNAME }}/${{
    env.DOCKER_IMAGE_REPOSITORY_NAME }}:auth-latest
51            docker compose up -d ${{ env.DOCKER_SERVICE_NAME }}
52  name: Deploy Auth Service
```

Implementasi dari CI/CD dapat di lihat pada **Tabel 2**. Penulis akan mengimplementasikan *pipeline* yang dijelaskan juga pada bab 2.2. Pada baris pertama, penulis memberikan nama terhadap *pipeline* ini, hanya sebuah metadata. Setelah itu, pada baris ke-3 hingga ke-6, penulis memberitahukan tentang *trigger* yang perlu didengarkan oleh GitHub Actions. Jadi, ketika *trigger* atau *event* tersebut terjadi, *workflows* ini akan berjalan. Selanjutnya, baris 16 hingga baris 17, penulis melakukan *checkout* terhadap *repository* dari GitHub yang terbaru. Baris 18 hingga 19, kita mengambil *tag* yang akan kita gunakan nantinya. Kemudian, baris ke-20 hingga baris ke-25 *hosted runner* mempersiapkan lingkungan untuk menjalan bahasa pemrograman Go, kemudian Go akan menjalankan semua test yang ada di *project* tersebut. Kemudian baris ke-26 hingga baris ke-39, penulis mempersiapkan lingkungan Docker untuk melakukan *compile program* dan mengirimkannya ke Docker Registry yang akan diambil oleh server nantinya. Lalu, baris ke-40, hingga barus ke-46 berguna untuk mempersiapkan koneksi SSH. Setelah itu, baris ke-47 hingga ke-52 akan melakukan *deployment* terhadap perubahan yang terbaru.

3.3. Tahap Load-Testing

Pengujian pada penelitian ini dilakukan dengan menggunakan *Load-Testing* menggunakan K6. K6 menggunakan JavaScript untuk menuliskan test yang akan digunakan. Penggalan kode dari *Load-Testing* menggunakan K6 sudah penulis lampirkan pada Tabel 3

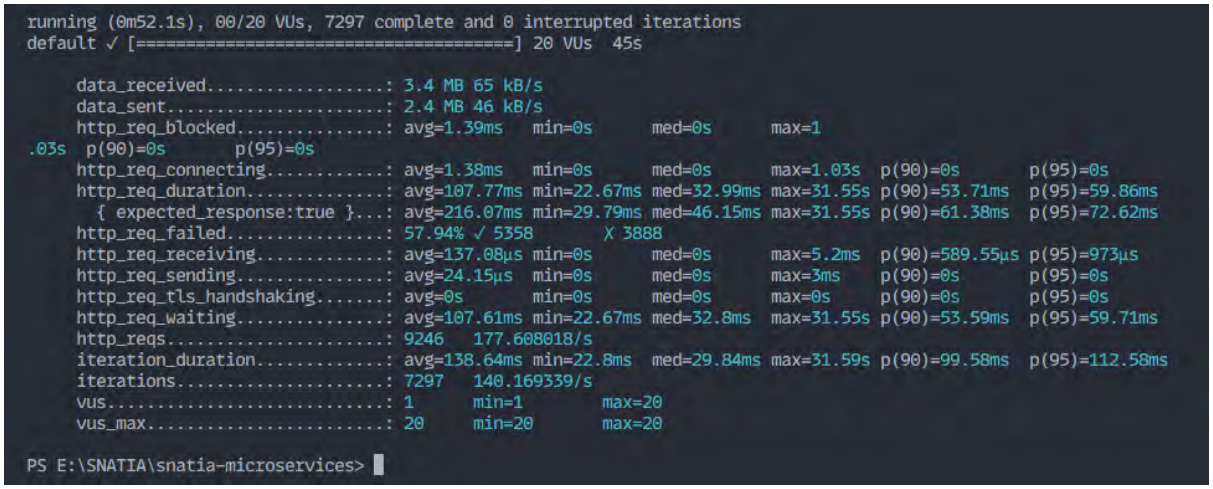
Tabel 3. Penggalan Kode Load Testing

	Kode
1	import http from 'k6/http';
2	
3	export const options = {
4	vus: 20,
5	duration: '45s'
6	}
7	
8	export default function () {
9	const base = {
10	baseUrl: "http://108.136.231.153:8000",
11	auth: "/auth",
12	order: "/order"
13	}
14	
15	const payload = {
16	authPayload: {

```
      "username": "kompiangg",
      "password": "12345678"
    },
    orderPayload: {
      "order": [
        {
          "item_id": 1,
          "price": 100000,
          "quantity": 2,
          "total_price": 200000
        }
      ]
    }
  }

  const res = http.post(
      base.baseUrl + base.auth + '/login',
      JSON.stringify(payload.authPayload), {
      headers: {'Content-Type': 'application/json'}
    }
  );

  const accessToken = res.json().data.access_token;
  http.post(
      base.baseUrl + base.order + '/orders',
      JSON.stringify(payload.orderPayload), {
      headers: {
        'Content-Type': 'application/json',
        'Authorization': 'Bearer ' + accessToken
      }
    }
  );
}
```

Penggalan kode di atas merupakan *Load-Test* yang penulis gunakan. Pada baris ke-3 hingga ke-7 dijelaskan parameter yang akan digunakan untuk melakukan *Load-Test*, yaitu ‘vus’ yang berarti banyaknya user yang akan mengirimkan paket request. Lalu, parameter ‘durations’, yaitu lamanya user akan mengirimkan paket data. Penulis menggunakan 45 detik, karena dari beberapa pengujian, proses yang dihabiskan untuk menjalakan pipeline, yaitu 1 menit dan proses deploy sendiri menghabiskan waktu 30 detik, karena keterbatasan *resource* yang dimiliki penulis, maka durasi yang digunakan hanya 45 detik. Lalu, di dalam default function terdapat dua buah *service* yang di-test, yaitu Login dan Order yang berada pada service yang berbeda tapi Order *service* memiliki ketergantungan terhadap *service Login*.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Hasil Load-Testing CI/CD Pipeline

3.4. Analisis Hasil Test

Dalam pengukuran *availibility* ketika diimplementasikan CI/CD Pipeline pada penelitian ini, dapat kita buatkan sebuah tabel perbandingan data yang berhasil terkirim dan kembali (*response*) dan yang gagal kembali dalam waktu 45 Detik

Tabel 4. Test Result

Key	Berhasil	Gagal	Durasi	Virtual User
http_req_failed	3888	5358	45 Detik	20

4. Kesimpulan

Berdasarkan penelitian dan hasil analisis dari penelitian, dapat dilihat pada tabel 4, bahwa ketika CI/CD Pipeline berjalan ketika masih terdapat *data stream* yang mengakses server, dalam 45 detik dengan 20 user didapatkan bahwa paket yang masih berhasil kembali sebanyak 3.888 dari 9.246 atau sebesar 42%. Hal ini menunjukkan bahwa CI/CD Pipeline dapat melakukan *deployment* pada sistem yang dibuat untuk penelitian ini kurang dari 45 detik. Lalu, CI/CD Pipeline dapat memperkecil waktu *downtime* ketika melakukan deployment daripada metode manual.

Referensi

[1] N. Shakhovska, N. Boyko, Y. Zasoba, and E. Benova, “Big data processing technologies in distributed information systems,” *Procedia Comput. Sci.*, vol. 160, no. 2018, pp. 561–566, 2019, doi: 10.1016/j.procs.2019.11.047.

[2] P. Muthulakshmi and S. Udhayapriya, “a Survey on Big Data Issues and Challenges,” *Int. J. Comput. Sci. Eng.*, vol. 6, no. 6, pp. 1238–1244, 2018, doi: 10.26438/ijcse/v6i6.12381244.

[3] M. Shahin, M. Ali Babar, and L. Zhu, “Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices,” *IEEE Access*, vol. 5, no. March, pp. 3909–3943, 2017, doi: 10.1109/ACCESS.2017.2685629.

[4] DMR, “Grab Statistics and Facts (2022),” 2022. https://expandedramblings.com/index.php/grab-facts-statistics/.

[5] S. Bougerel, “Our Journey to Continuous Delivery at Grab (Part 2),” *engineering.grab.com*, 2021. https://engineering.grab.com/our-journey-to-continuous-delivery-at-grab-part2.

[6] F. Rademacher, S. Sachweh, and A. Zundorf, “Differences between model-driven development of service-oriented and microservice architecture,” *Proc. - 2017 IEEE Int. Conf. Softw. Archit. Work. ICSAW 2017 Side Track Proc.*, pp. 38–45, 2017, doi: 10.1109/ICSAW.2017.32.

[7] H. Schulz, T. Angerstein, and A. Van Hoorn, “Towards automating representative load testing in continuous software engineering,” *ICPE 2018 - Companion 2018 ACM/SPEC Int. Conf. Perform. Eng.*, vol. 2018-January, pp. 123–126, 2018, doi: 10.1145/3185768.3186288.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pemanfaatan Google Analytics Sebagai Upaya Optimasi UX Untuk Meningkatkan Konversi Pada Website

I Gusti Agung Istri Bianca Githa Saraswati[a1], I Gusti Ngurah Anom Cahyadi Putra[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana
Jalan Raya Kampus Unud, Badung, 08361, Bali, Indonesia
[1]biancasaraswati@gmail.com
[2]anom.cp@unud.ac.id

Abstract

In a world driven by advances in technology, society uses technology more than ever to help support them in many ways. A few of the common ones are to have a business with information systems offline to help achieve their business goals and to have a website so that their business would have a presence online and can be reached globally by the international audience. With every aspect of modern technology, be it offline or online, the user interface and user experience play a critical role in how the new technology can be usable and adaptable by users of businesses, and their customers. In the online space, the user experience should be perfected in order to make the website more valuable, accessible, usable, desirable, trustable, and credible to help the user feel more comfortable using the website. This journal will cover how Google analytics can be used to optimize the user experience (UX) to increase the conversion of a website. In order to analyze & enhance the UI and UX of a website, one needs to use both a heuristic approach or known as the rule of thumb approach coupled with a data analysis/data-driven approach. For the data analysis approach to work we would need to sample user traffic data of the website, and identify behavior trends, findings and make conclusions. There are many ways to attract user traffic to a website. Email channel, direct, organic social, organic search, paid search, and paid social are just a few of those many ways. The fastest way out of these options would be to attack traffic via ads placements or paid search ads and paid social ads which can be used as a source of the traffic for the data analyses to function. that will be covered in this study. Once traffic is available, Google Analytics can be used to analyze traffic in order to improve the user experience of the website. This journal has proven that google analytics can help the business to optimize the UX of our website. The goal of this business is to get the user to click the WhatsApp button and communicate with the owner. We get 11 users that click the WhatsApp button in January which is before the website UX is upgraded, 16 users in February when the new version design is launched, and the new one in March which has 25 users clicking the WhatsApp button on the website. This has been proven that the conversion rate of users clicking the WhatsApp button after the UX design has been launched is higher than before the website used the UX design.

Keywords: *User Experience, Google Analytics, Conversions, Business, Website*

1. Latar Belakang

Perkembangan teknologi informasi dan juga komunikasi mengiringkan para pengguna untuk selalu memutakhirkan penerapan teknologinya demi meningkatkan kualitas dari bidang-bidang yang mereka lakukan.[1] Hampir pada sebagian besar masyarakat di dunia dalam berbagai kalangan sangat memerlukan teknologi untuk mengembangkan bidang mereka. Terutama pada bidang bisnis yang sedang maraknya berkembang dan penuh persaingan di zaman ini.

Website menjadi teknologi yang ideal dimiliki oleh setiap pengusaha bisnis sebagai media informasi kepada masyarakat umum yang memerlukan informasi dari bisnis tersebut.

Penggunaan website dapat digunakan sebagai pendukung strategi pemasaran karena perusahaan akan lebih memiliki jangkauan wilayah pemasaran yang tak terbatas. Adanya otomatisasi penerimaan order lewat email, yang menghemat waktu, penghematan biaya promosi, karena website perusahaan dapat diakses selama 24 jam dari segala penjuru dunia. Adanya penghematan biaya tenaga kerja tanpa perlu ruang pameran/display. [2]

Dalam rangka mencapai tujuan bisnis tersebut *website* perlu dibuat demi mengembangkan dan memperluas jangkauan pelanggan pada masyarakat luas, dan *website* diperlukan untuk menjadi media penghubung *user* dengan usaha yang dituju demi mencapai tujuan mereka pada bisnis tersebut. Salah satu contoh bisnis tersebut adalah pada bisnis *e-commerce* yang memerlukan peningkatan pada proses konversi user pada website tersebut demi mencapai titik tuju mereka yaitu membeli barang tersebut.

Prediksi untuk *Conversion Rate (CVR)* pada dunia industry *e-commerce platforms* sangatlah penting karena dapat langsung berkontribusi pada pendapatan dan goals yang dituju. [3] *Conversion Rate* atau konversi dalam *analytics* adalah proses dari perilaku users dalam menjalankan langkah-langkah pada *website*. Konversi juga dapat dikatakan langkah untuk mencapai sasaran dalam menyelesaikan tindakan dari awal sampai tujuan *website*. Konversi terdiri dari dua bagian yaitu makro konversi dan mikro konversi. Makro Konversi merupakan *goals* yang ingin dituju oleh pemilik *website* dan juga *user*. Sedangkan mikro Konversi merupakan beberapa langkah yang dilakukan *user* dalam mencapai Makro Konversi atau *goals*.[3]

Penelitian dari *website* sebuah perusahaan bisnis, yaitu memiliki beberapa mikro konversi dan satu makro konversi. Mikro konversi pada *website* tersebut adalah dalam proses melihat dan interaksi pada produk yang ditampilkan pada *website*, namun makro konversi pada perusahaan tersebut adalah untuk menghubungi pemilik perusahaan bisnis tersebut melalui *whatsapp*. Pada penelitian ini difokuskan pada *UX* yang dimiliki oleh bagian *contact us*. Jika diilustrasikan proses konversi akan seperti gambar 1.

Gambar 1. Proses konversi pada website bisnis X

Dalam meningkatkan proses Konversi, diperlukan *user experience* (*UX)* yang baik pada *website*. *user experience* (*UX)* adalah persepsi dan respon dari user dari penggunaan atau antisipasi dari produk. *User experience* (*UX)* dapat dikatakan bagaimana perasaan user dari setiap interaksi yang dihadapi saat menggunakan produk atau *website* tersebut. [4]

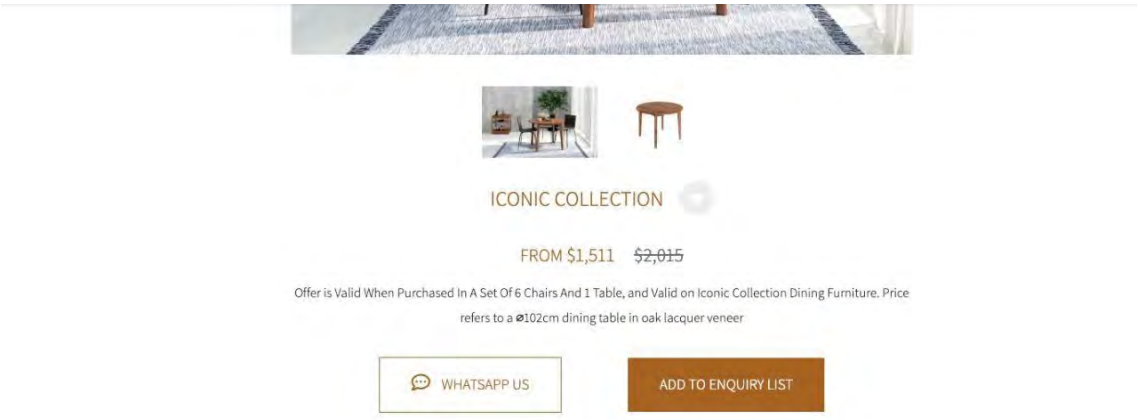
Website yang memiliki *UX* yang baik adalah yang merancang design dengan konsep *heuristic evaluation*, ditemukan oleh yaitu metode penilaian kegunaan suatu produk digital yang bertujuan untuk membuat *user experience* lebih baik dan nyaman untuk digunakan oleh *user* sesuai dengan gambar 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Gambar 2. Proses *Heuristic Rules*

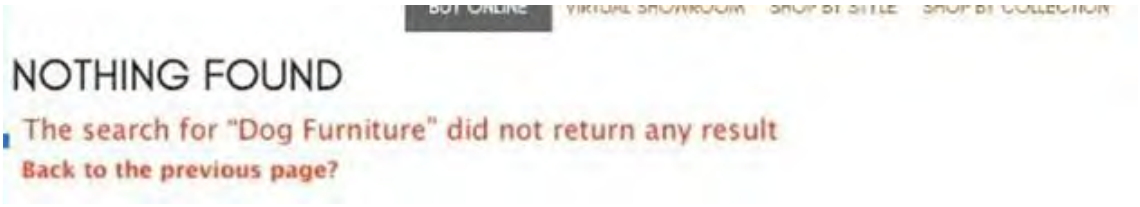
Namun *user experience (UX)* pada *website* tersebut masih belum *usable* untuk mencapai makro conversion yang dapat dilihat dari gambar interface website sebelum didevelop design UX yang baik dalam bagian *contact us*.
Perbandingan pertama adalah pada halaman pemesanan, dibawah harga barang yang belum berisi bagian *contact* yaitu *whatsapp,* sehingga *user* bingung setelah melihat barang kemana mereka harus menghubungi jika masih ada kebingungan. Maka untuk *usability user* ditambahkan tombol *whatsapp* dibawah *list* dari produk sehingga mencapai makro konversi yaitu menghubungi pembisnis yaitu melalui tombol *whatsapp*.

Gambar 3. Bagian halaman pemesanan yang belum berisi tombol *whatsapp*



Gambar 4. Bagian halaman pemesanan yang sudah berisi tombol *whatsapp*

Perbandingan kedua adalah ketika *user search furniture* yang tidak ada pada *list product* di *website.* Ditambahkan fitur *whatsapp* untuk menghubungi pemilik bisnis jika ingin bertanya mengenai produk tersebut, sehingga dapat mencapai makro konversi.



Gambar 5. Bagian *Contact us* yang belum berisi tombol *whatsapp*

Gambar 6. Bagian *Contact us* yang sudah berisi tombol *whatsapp*

2. Metode Penelitian

Setelah *UX* diperbaiki dan rancangan *website* terbaru di *launch*, perlunya *traffic* untuk menjadi upaya optimasi *website* dengan membuktikan apakah pencapaian interaksi menuju makro konversi. *Google analytics* adalah *digital analytics* yang akan dipakai untuk meneliti konversi pada *website* tersebut. *Google analytics* akan membantu dalam proses *analysis of quantitative research.*

Optimasi *user experience* (*UX)* pada *website* diperlukan *Digital Analytics Method* untuk mempermudah dan mempercepat proses *objective tracking, measurement, reporting,* dan *analysis of quantitative research* dari sebuah *website*. Salah satu digital analytics yang dapat dipakai adalah *Google analytics* yang merupakan produk yang dimiliki oleh salah satu perusahaan besar yang sudah dipercaya oleh dunia yaitu *Google*. [5]

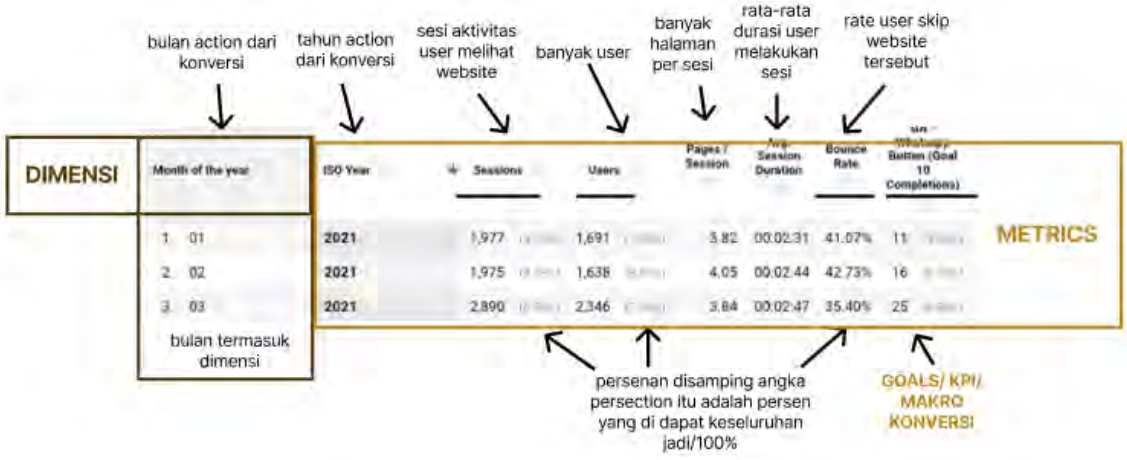
Gambar 7. Proses *Google Analytics*

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tahap pertama adalah tahap pengumpulan data, yaitu bagaimana cara *google analytics* mendapatkan data ke dalam *google analytics*. Untuk mendapatkan data, harus ditambahkan *google analytics code* ke dalam *website*, kode ini akan merekam semua aktivitas *user* di dalam *website*. Kode ini ada beberapa jenis dan tergantung untuk mengukur apa, sebagai contoh untuk *website* kode ini berupa *javascript* yang diletakkan diantara *tag* <head></head>.

Tahap kedua adalah konfigurasi dan pemrosesan data, dimana selama pemrosesan data, *google analytics* merubah data mentah dari proses *collection* tadi menggunakan pengaturan dalam *google analytics account.*

Tahap ketiga adalah reporting, dimana pada proses ini dapat diakses dan data dapat dianalisa oleh pemilik *website* menggunakan *reporting interface* dan mudah digunakan untuk visualisasi data.
Pada *google analytics* mampu menganalisis konversi dengan mengsetup *goals* pada *google analytics* tersebut. Dibawah ini terdapat reporting interface data yang terlihat pada *google analytics.*



Gambar 8. Fitur-fitur yang terdapat pada *Google Analytics*

Gambar 8 adalah penjelasan dari beberapa fitur custom report yang membantu analisa proses konversi. Pada *google analytics* terdapat 2 bagian untuk menjelaskan data yaitu dimensi dan metriks. Dimensi menjelaskan data yang tidak berupa angka seperti bulan dan judul dari fitur. Sedangkan metriks menjelaskan data tentang angka seperti jumlah *user*, jumlah sesi, dan lainnya.
Pada jurnal ini, yang penting untuk diperhatikan dalam analisa website bisnis yang diteliti adalah membuat *goals/KPI (Key Performance Indicator)* tinggi dengan memperhatikan kenaikan *whatsapp clicking rate* yang merupakan *goals* dari bisnis yang diteliti. Lalu berapa banyak *User* yang diharapkan lebih tinggi dari sebelumnya. Selain itu, *bounch rate* yang seharusnya menurun agar proses penekanan tombol *Whatsapp* meningkat.

	Month of the year	ISO Year	Sessions	Users	Pages / Session	Avg. Session Duration	Bounce Rate	Goals/KPI: GA - Whatsapp Button (Goal 10 Completions)	
1.	01	2021	1,977 (4.76%)	1,691 (5.18%)	3.82	00:02:31	41.07%	11 (8.94%)	Fase 2
2.	02	2021	1,975 (4.75%)	1,638 (5.01%)	4.05	00:02:44	42.73%	16 (13.01%)	
3.	03	2021	2,890 (6.96%)	2,346 (7.18%)	3.84	00:02:47	35.40%	25 (20.33%)	Fase 3
4.	04	2021	2,350 (5.66%)	1,920 (5.88%)	3.98	00:02:42	39.74%	21 (17.07%)	
5.	05	2021	2,674 (6.44%)	2,142 (6.56%)	3.79	00:02:45	42.60%	13 (10.57%)	
6.	06	2021	3,323 (8.00%)	2,603 (7.97%)	4.55	00:04:23	37.89%	15 (12.20%)	
7.	07	2021	2,528 (6.08%)	1,990 (6.09%)	3.94	00:03:06	39.20%	26 (21.14%)	
8.	08	2021	2,391 (5.76%)	1,876 (5.74%)	3.52	00:02:43	43.08%	30 (24.39%)	
9.	09	2021	2,317 (5.58%)	1,754 (5.37%)	3.22	00:02:18	44.50%	22 (17.89%)	
10.	01	2020	251 (0.60%)	223 (0.68%)	4.48	00:02:47	30.28%	1 (0.81%)	
11.	07	2020	3,885 (9.35%)	2,823 (8.64%)	3.69	00:02:32	46.95%	25 (8.96%)	
12.	08	2020	3,525 (8.48%)	2,668 (8.17%)	3.71	00:02:41	50.13%	15 (5.38%)	
13.	09	2020	3,069 (7.39%)	2,362 (7.23%)	3.45	00:02:29	48.71%	15 (5.38%)	Fase 1
14.	10	2020	3,488 (8.40%)	2,787 (8.53%)	3.24	00:02:23	54.70%	16 (5.73%)	
15.	11	2020	2,744 (6.60%)	2,035 (6.23%)	3.87	00:02:53	45.88%	14 (5.02%)	Fase 2
16.	12	2020	2,159 (5.20%)	1,811 (5.54%)	4.05	00:02:38	37.38%	14 (5.02%)	

Gambar 10. *Custom Report Whatsapp Google Analytics Universal (UA)*

Gambar 10 yaitu *Custom Report Whatsapp Google Analytics Universal (UA)* yang dimiliki oleh bisnis. Pada jurnal ini akan difokuskan pada makro konversi bisnis tersebut yaitu *click* pada *whatsapp website* dimana terlihat pada baris yang paling kanan. Proses dari pengerjaan *website* yang diteliti terdiri dari beberapa fase.

Fase 3 adalah dimana *design UX* yang sudah diperbaiki akan melakukan *deploy*. Maka yang akan dijadikan perbandingan adalah pada bulan Januari (01) yaitu sebelum *deploy* dan Bulan Februari (02) serta Maret (03) yaitu *UX design* yang sudah diperbaiki melakukan *deploy*.

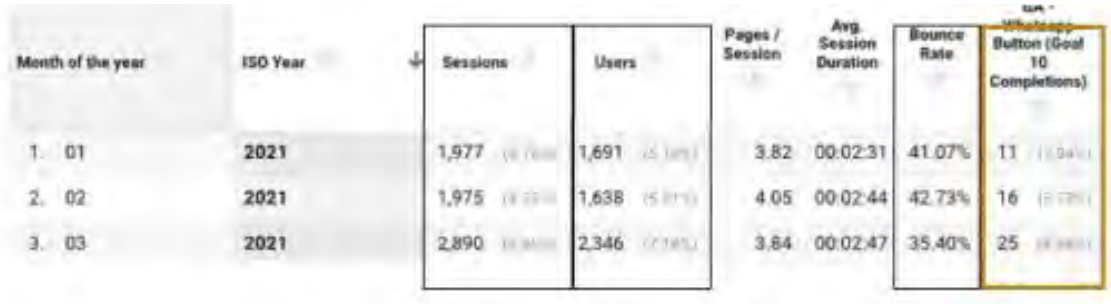
Selain melihat konversi ke *whatsapp*, pada *google analytics* juga dapat melihat *session, users*, dan *bounce rate* tujuannya untuk melihat berapa persen ketika *user* masuk langsung *bounce* menuju halaman lain.

3. Hasil dan Diskusi

Hasil yang didapat pada penelitian ini, *Google analytics* mampu menjadi upaya optimasi pada dengan bentuk report analysis interface.

3.1. Hasil

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Month of the year	ISO Year	Sessions	Users	Pages / Session	Avg. Session Duration	Bounce Rate	Button (Goal 10 Completions)
1. 01	2021	1,977	1,691	3.82	00:02:31	41.07%	11
2. 02	2021	1,975	1,638	4.05	00:02:44	42.73%	16
3. 03	2021	2,890	2,346	3.84	00:02:47	35.40%	25

Gambar 11. *Custom Report* bulan Januari, Februari, dan Maret

a. Pada gambar nomor 11 menjelaskan laporan analysis bisnis dengan *google analytic universal* pada bulan Januari, Februari, dan Maret.
   1. *Interface* ini akan ditemukan pada *custom report* yang merupakan fitur yang dimiliki *google analytics* universal yaitu memilih fitur yang diperlukan untuk *report.*
   2. Dapat berupa tahun *report* tersebut, bulan *report,* waktu pada *website*, banyaknya pelanggan yang masuk ke *website*, rata-rata waktu yang ada pada halaman tersebut, rata-rata pelanggan meloncat atau meninggalkan *website* tersebut, dan juga tujuan bisnis tersebut yaitu *click* pada *whatsapp* yang merupakan *goals* bisnis tersebut.
   3. Dapat membantu mempercepat proses analisa manual dengan data yang sudah disediakan *google analytics* yaitu dapat dengan mudah melihat *conversion rate* dari waktu ke waktu. Seperti gambar 11, disimpulkan bahwa pemilik *website* dapat membandingkan *UX* yang memiliki design yang sudah diperbaiki dan yang belum diperbaiki, sehingga lebih mudah untuk menyimpulkan optimasi dari *UX* yang meningkat pada *website* yang diteliti.

Tabel 1. Tabel perbandingan data konversi *Whatsapp*

Bulan	Pengguna *(Users)*	*Whatsapp*
01 (Januari)	1,691	11 (3.94%)
02 (Februari)	1,638	16 (5.73%)
03 (Maret)	2,346	25 (8,96%)

b. Pada table nomor 1 menjelaskan fokus dari permasalahan jurnal ini yaitu analisis optimasi *UX* pada *website* dari analisis konversi. Dimana pada bulan Februari *website* tersebut di-*launch/deploy* maka di analisis 2 bulan terdekat dengan Februari yaitu Januari dan Maret.
   1. Dari analisis tabel 1, terlihat bahwa user pada bulan maret adalah yang paling banyak diantra semuanya, maka disimpukan bahwa *UX* yang sudah dibuat telah berhasil melancarkan konversi.
   2. Dilihat persen disamping angka pelanggan memasuki *whatsapp* paing besar, juga menjadi salah satu bukti bahwa *UX* tersebut adalah *UX* yang baik.
   3. Sudah terbukti bahwa fitur *conversion rate* pada Google Anlytics dapat membantu pemilik menyimpulkan,
      a. *Design* tersebut sudah mencapai makro konversi dengan angka kenaikan *user* yang menekan tombol whatsapp, sesi *user* yang bertambah, dan *bounce rate* yang menurun sehingga user terbukti tidak meninggalkan website ke halaman lain.
      b. Dengan penjelasan pada research metode dapat terlihat mudahnya menggunakan *google analytics*, tidak berbayar, dan praktis. Maka dari itu direkomendasikan untuk memakai *Google analytics.*

3.2. Diskusi

Dengan pembahasan yang sudah dijelaskan diatas bahwa *google analytics* memiliki fitur-fitur penting digunakan dalam mengolah data dan menjadi optimasi dari pembuatan *UX. UX* dapat dikoreksi dengan mudah menggunakan konvesi untuk mencapai tujuan dari pengguna ataupun

pemilik *website* tersebut. Sehingga seluruh pemakai dapat merasa bahwa menganalisa data bisnis dapat dilakukan dengan teratur dan nyaman.

4. Conclusion

Dengan penelitian ini disimpulkan bahwa google analytics mampu menganalisa konversi untuk membantu optimasi dari *UX* yang dibuat oleh *UX researcher*, pembisnis, ataupun setiap orang yang menggunakan *google analytics*. Kedepannya diharapkan pembaca dapat mengimplementasikan *google analytics* conversion rate sebagai acuan data demi meningkatkan bisnis yang dimiliki dan membantu mencapai tujuan pengguna dan pemilik.

References

[1] R. Raafi'udin, B. Hananto, and C. Nugrahaeni PD, "Analisa Trafik Pengunjung Website dalam Pengembangan UI dan UX," *JURNAL INFORMATIK* , vol. 15, 2019.

[2] A. Susanto, H. S. Riyadi, "PENGGUNAAN WEB SEBAGAI SALAH SATU PENDUKUNG STRATEGI PEMASARAN PRODUK OLEH PERUSAHAAN KUSUMA AGRO INDUSTRI BATU."

[3] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, "Hierarchically Modeling Micro and Macro Behaviors via Multi-Task Learning for Conversion Rate Prediction," in *IEEE International Conference on Program Comprehension*, 2022, vol. 2022-March, pp. 36–47. doi: 10.1145/nnnnnnn.nnnnnnn.

[4] D. Orlando Putra, A. Setiawan, "The Importance of User Experience Analysis in the Design of an Education Information System Application," BIS-HESS, 2019. [Online]. Available: https://wearesocial.com/blog/2019/01/digital-2019-

[5] C. L. Teng, "Google analytics," *Malaysian Family Physician*, vol. 3, no. 2. p. 71, 2008. doi: 10.5596/c13-022.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Cryptocurrency Prediction With Social Media

Darryl Patrick Matheuw Kurniawan[a1], Drs. I Wayan Santiyasa[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia
[1]Darrylpatrck82@gmail.com
[2]Santiyasa@unud.ac.id

Abstract

Social Media adalah suatu wadah bagi kita manusia dalam berkomunikasi, tidak jarang juga kita banyak membagikan kehidupan kita didalam social media. Baik berupa hiburan, bisnis atau hal lainnya.maka dari itu penelitian ini akan menggunakan social media konten untuk memprediksi masa depan , contohnya adalah memprediksi harga dari cryptocurrency. Dengan menggunakan twitter sebagai media social, jurnal ini akan menunjukan suatu model yang dapat kita gunakan untuk memprediksi harga cryptocurrency.

Keywords: *Coin, Predict, Social Media*

1. Introduction

Seiring dengan berkembangnya ekonomi dan teknologi,cryptocurrency atau coin adalah alternatif lain mata uang dimana coin ini dijadikan sebagai uang digital yang dapat diperjualbelikan dan digunakan [1].coin digunakan untuk mendecentralisasi nilai mata uang yang beredar di dunia agar lebih mudah dilakukan, coin pertama yang dikeluarkan adalah bitcoin dimana bitcoin ini dikeluarkan pada tahun 2009 oleh Satoshi Nakamoto sebagai software open-source[2].kemudian coin ini menarik banyak perhatian orang-orang dan mengakibatkan munculnya coin alternatif sekitar 10.000 lebih coin hingga December 2021 terakhir[3]. Saat ini coin – coin yang menjadi coin utama adalah bitcoin (BTC), Ethereum(ETH), Litecoin(LTC) dan Cardano (ADA). Yang berhasil bertumbuh secara eksponensial hingga mencapai capitalization market lebih dari 2 triliun dollar dimana bitcoin memegang 40% total , ehthereum 21%,Litecoin sekitar 0,5% dan cardano hampir 2%[4].

Dengan memperhatikan aktifitas – aktifitas yang dilakukan oleh masyarakat di media social penelitian ini akan memprediksi harga coin dengan mengandalkan aktifitas masyarakat. Media yang digunakan nantinya adalah twitter dimana twitter adalah salah satu platform yang sangat besar dengan pengguna sekitar 400 juta orang lebih dan 200 juta pengguna aktif setiap harinya[5]. Twitter digunakan karena jika dibandingkan dengan platform media social lainnya yang lebih banyak menggunakan gambar dan visual lainnya mengakibatkan sulitnya dilakukan analisis dan fakta bahwa twitter memiliki textual context terbesar sehingga kita dapat menganalisis lebih mudah dan akurat[6].

Metode yang akan digunakan nantinya adalah menggunakan tweet volumes dan sentiment analysis,sentiment analysis adalah opini mining atau emosi ai dimana sentiment analysis ini menggunakan Natural Language Processing (NLP) untuk mengesktrak data twitter yang berupa text nantinya[7]. Sementara tweet volumes adalah volume dari tweet yang ada pada saat waktu – waktu yang ditentukan[8].

2. Reseach Methods

2.1 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan twitter API dan google trends untuk mendapatkan data. Antara lain Tweepy adalah suatu python library yang digunakan untuk mengakses twitter API dan mengumpulkan data. Library ini dapat membantu kita memfilter data berdasarkan hastags atau text yang ada pada twitter. Dalam penelitian ini menggunakan "#bitcoin" dan "#ethereum". Selain itu google juga memberikan akses untuk mengakses data melalui google trends. Google adalah salah satu mesin pencari yang paling popular didunia dimana google digunakan lebih dari 74% dari setiap pencarian yang dilakukan pada web, jadi tidak diragukan lagi bahwa google adalah yang terbaik dalam memberikan informasi. Data yang diberikan oleh google adalah Search Volume Index (SVI) yang menunjukkan seberapa sering istilah penelusuran tertentu ditelusuri relatif terhadap total volume penelusuran di seluruh dunia, selama rentang tanggal tertentu yang dimasukkan pengguna.

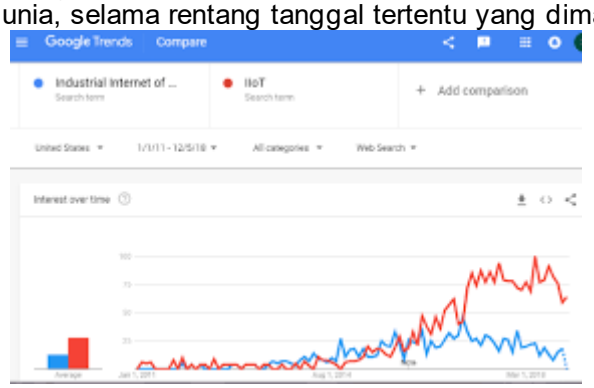


Figure 1. contoh penggunaan google trends

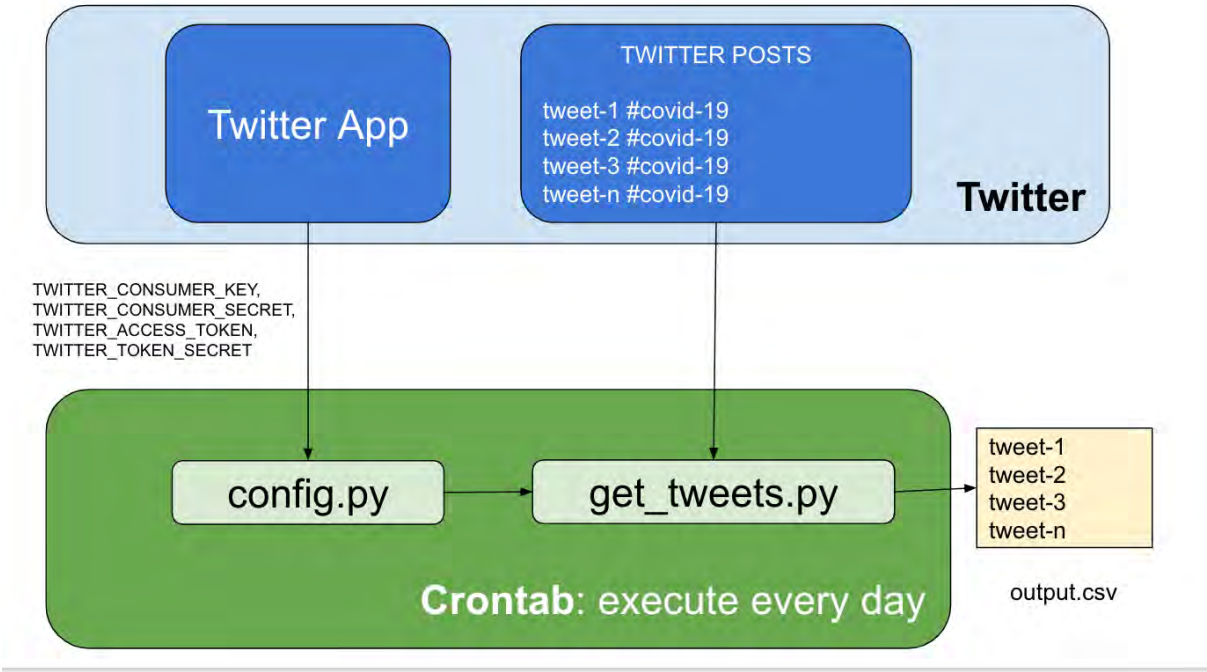


Figure 2. cara kerja tweepy

2.2 Cleaning data

Tahap setelah pengumpulan data adalah pembersihan data dimana pada proses ini data dibersihkan dengan menghilangkan hashtags ,petik dan tanda – tanda lainnya sehingga data hanya tersisa text utuh agar lebih mudah diproses nantinya.

2.3 Sentiment Analysis

Pada tahap ini menggunakan VADER(Valence Aware Dictionary for Sentiment Reasoning) adalah suatu python library untuk memproses NLP dan subjektifitas dan polaritas.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

```
1st statement :
Overall sentiment dictionary is :  {'neg': 0.165, 'neu': 0.588, 'pos': 0.247, 'compound': 0.5267}
sentence was rated as  16.5 % Negative
sentence was rated as  58.8 % Neutral
sentence was rated as  24.7 % Positive
Sentence Overall Rated As Positive

2nd Statement :
Overall sentiment dictionary is :  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
sentence was rated as  0.0 % Negative
sentence was rated as  100.0 % Neutral
sentence was rated as  0.0 % Positive
Sentence Overall Rated As Neutral

3rd Statement :
Overall sentiment dictionary is :  {'neg': 0.508, 'neu': 0.492, 'pos': 0.0, 'compound': -0.4767}
sentence was rated as  50.8 % Negative
sentence was rated as  49.2 % Neutral
sentence was rated as  0.0 % Positive
Sentence Overall Rated As Negative
```

Figure 3. contoh penggunaan sentiment analysis

3. Result and Discussion

Berikut adalah hasil dari Vader figure 1 bitcoin dan figure 2 ethereum

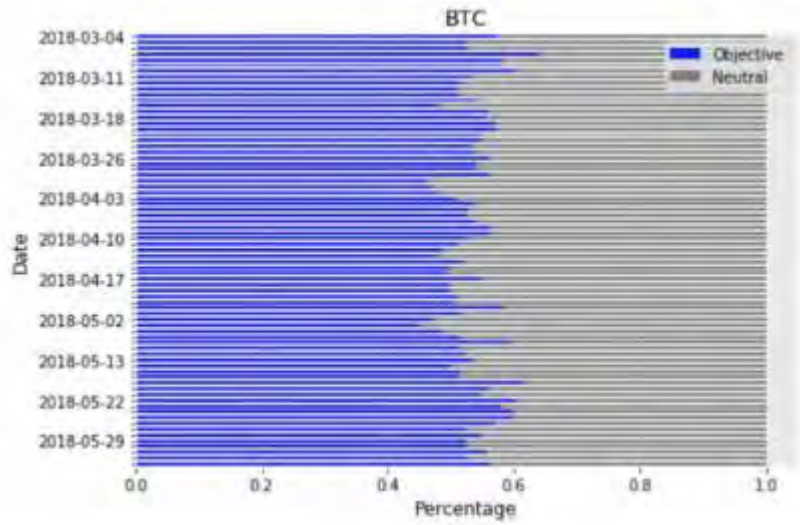


Figure 1. persentase objektif bitcoin (memiliki nilai positif atau negatif) dengan neutral

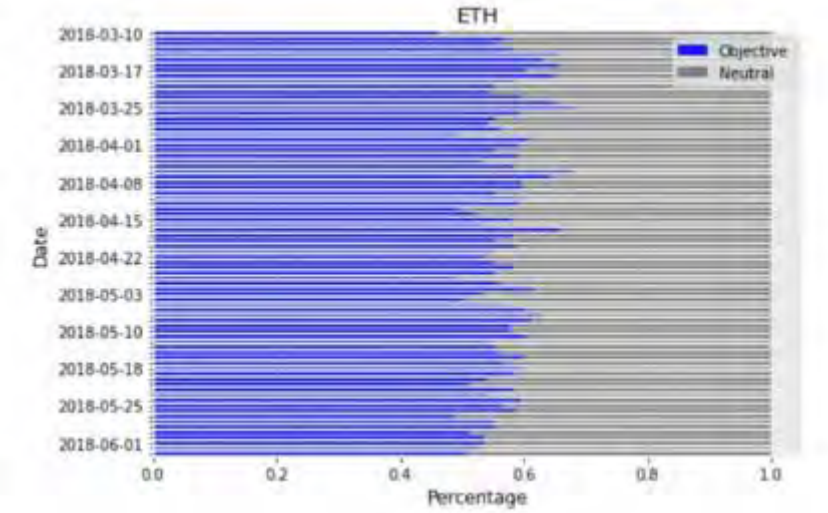


Figure 2. persentase objektif ethereum (memiliki nilai positif atau negatif) dengan neutral

Figure 3 dan 4 menunjukan sentiment tweet lebih neutral daripada objektif

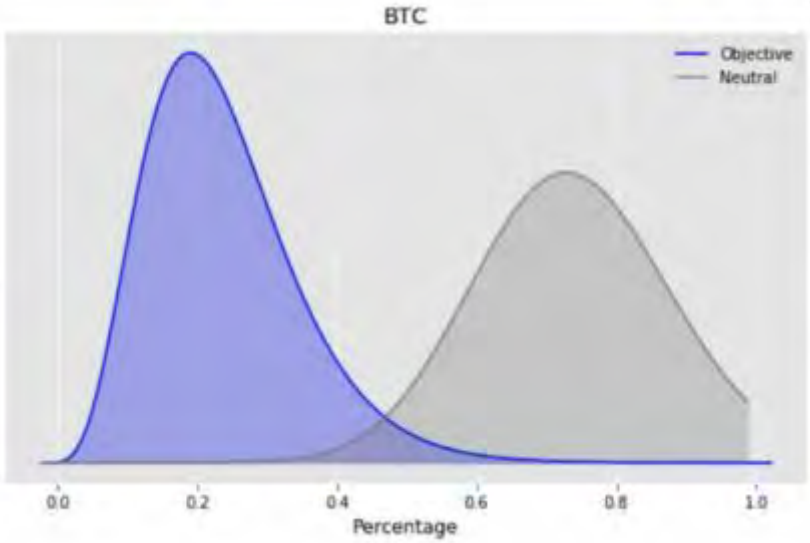


Figure 3. distribusi objektif bitcoin

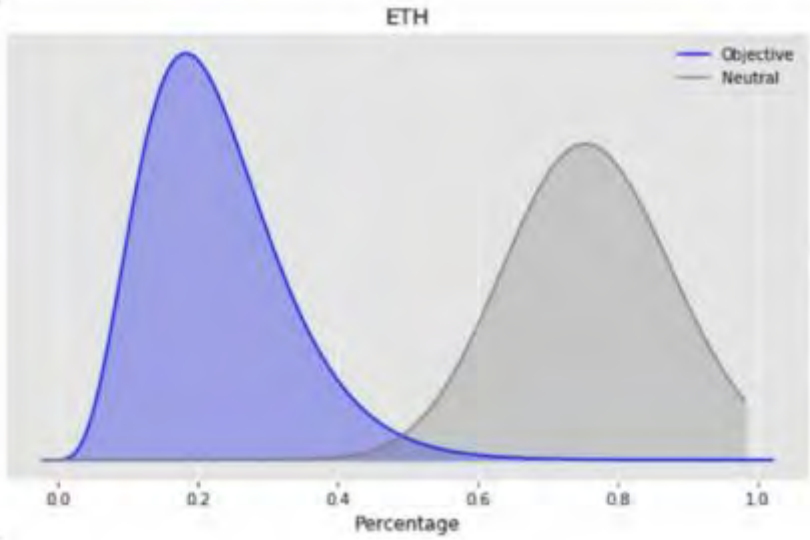


Figure 4. distribusi objektif Ethereum

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Fakta bahwa setengah dari tweets memiliki sentiment netral, hal ini memungkinkan sentiment positif atau negative dapat memberikan insight bahwa model ini berhubungan perubahan harga. Hal tersebut dapat kita lihat pada fluktuasi harga, berikut figure 5 dan 6 adalah perubahan harga dan polarity pada tweet.

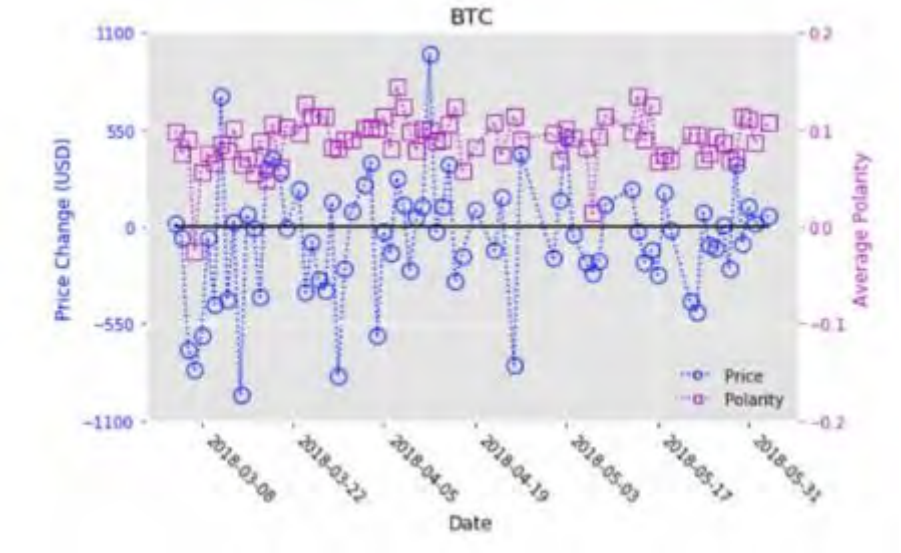


Figure 5. perubahan harga bitcoin dan polarity tweet

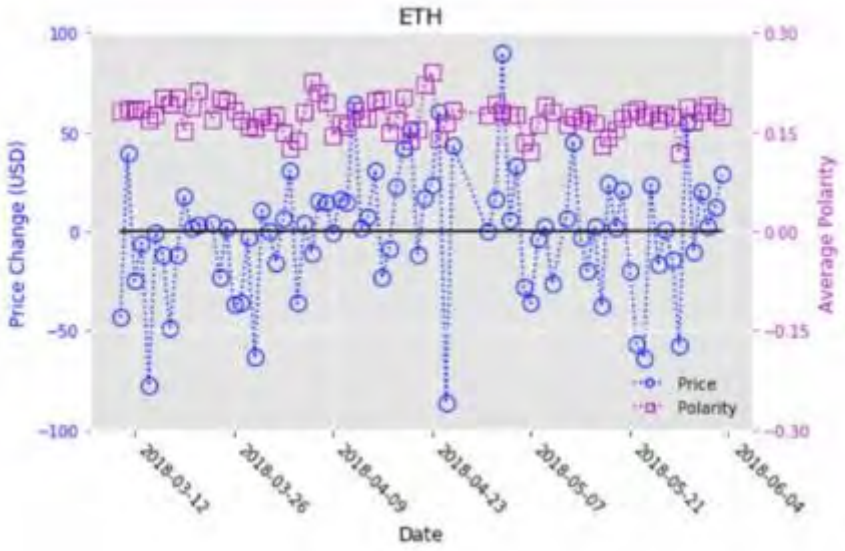


Figure 6. perubahan harga Ethereum dan polarity tweet

Dapat kita lihat polarity dan harganya cukup berbeda walaupun berhubungan tetapi Ketika harga mengalami penurunan polaritynya tetapi bagus, itu berarti orang-orang yang ada di twitter tidak membicarakan positif atau negative membelinya tetapi mereka suka dengan hal – hal lainnya seperti keamanan atau lain sebagainya.karena perbedaan yang drastic berarti sentiment analysis tidak dapat kita gunakan dan tweet volume dapat menjadi indicator yang lebih baik dibandingkan sentiment.

Karena sentiment sudah dikeluarkan maka tweet volume dan google trends menjadi indicator yang akan digunakan berikut SVI dan tweet volume sebagai indicator yang memiliki korelasi dengan naik turunnya harga crypto. Dengan menggunakan multiple linear regression berikut hasil prediksinya pada figure 7 dan 8.

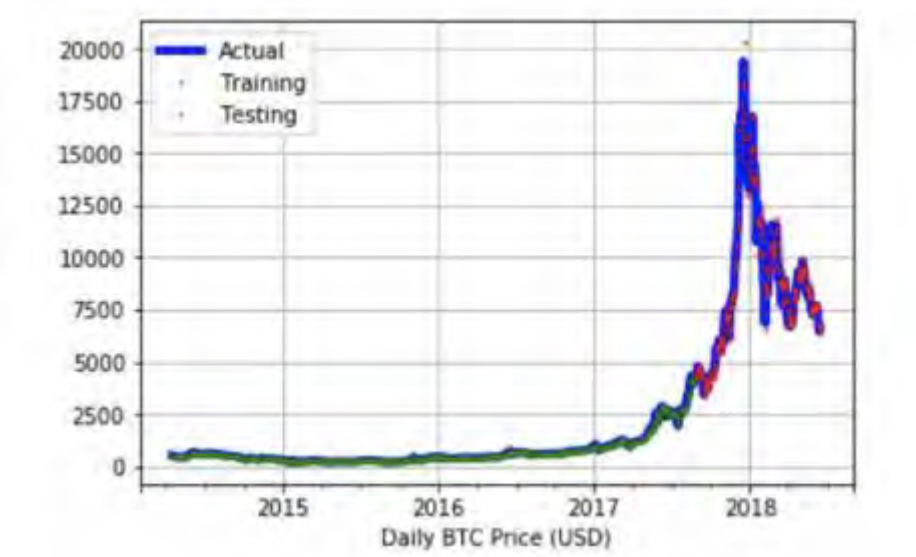


Figure 7. model menunjukan kecocokan dengan harga asli bitcoin dengan warna hijau sebagai train data dan hasil tes adalah titik merah

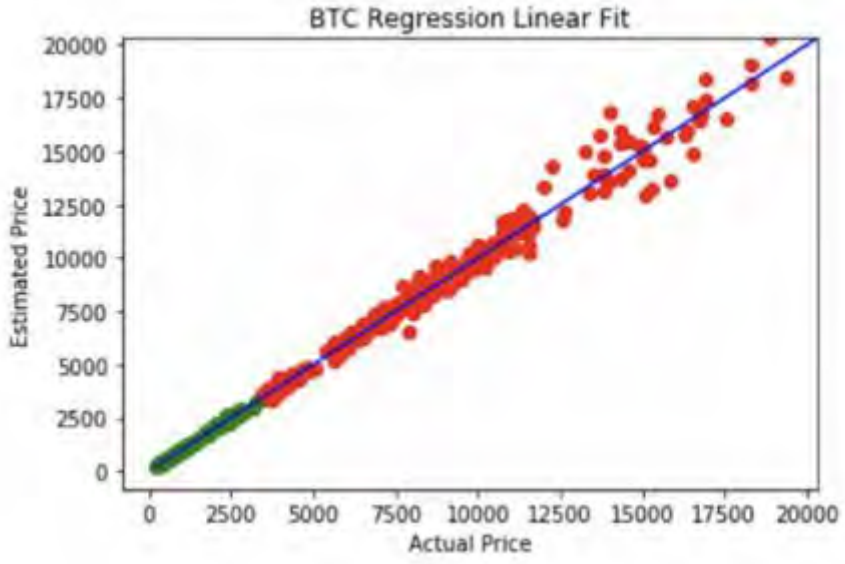


Figure 8. bitcoin regression menunjukan kecocokan dengan perkiraan harga y-axis, dan x-axis adalah harga asli.titik hijau adalah train data dan titik merah adalah hasil

Figure 7 merupakan model machine learning yang digunakan yakni multiple linear regression dimana data dari Google Trends dan tweet volume dibagi dengan 80% dari total data adalah data training dan 20% adalah data test dan dapat kita lihat dari hasilnya model ini memiliki akurasi yang baik dimana data training dan test cocok dengan harga asli yang ada. Sementara figure 8 adalah tampilan dengan linear regression dimana graphic menunjukan bahwa garis positif dan hubungan antara harga asli dan perkiraan harga mayoritas cocok, dimana titik hijau adalah data train dan titik merah adalah hasil test.

Hal ini dapat terjadi karena hubungan antara Google Trends data dengan perubahan harga crypto cukup baik. Dengan membandingkan berdasarkan R-pearson dan P-value metrik. R-pearsonnya adalah 0.817 yang berarti hubungan positif dengan Google Trend data, dan P-valuenya adalah 0.000, yang berarti hasilnya signifikan secara statistic.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

4. Conclusion

Hasil prediksi pada model menunjukan hasil yang baik hanya saja terdapat beberapa indicator yang tidak dapat digunakan dalam memprediksi harga dengan social media seperti sentiment analysis karena Ketika sentiment menunjukan positif harga tidak lah turun mengakibatkan ketidak cocokan sentiment analysis sebagai indicator serta google trends dan volume dari tweet dapat menjadi indicator karena aktifitas yang masyarakat lakukan mempengaruhi pasar yang ada hal ini membuktikan bahwa kita dapat menggunakan media social untuk memprediksi masa depan.

References

[1] Connor Lamon, Eric Nielsen ,Eric Redondo, Cryptocurrency Price Prediction Using News and Social Media Sentiment, 2017.
[2] Michail Vlachos and Giovnopoulus, Forecasting Cryptocurrency Price Movements with Predictive Social Media Analytics,2022
[3] Oikonomopoulos Sotirios, Cryptocurrency price prediction using social media sentiment analysis,2022.
[4] Wolfgang Karl Härdle, Campbell R. Harvey and Raphael C. G. Reule, Understanding Cryptocurrencies,2018.
[5] Sitaram Asur and Bernardo A. Huberman ,Predicting the Future With Social Media,2010

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Fast Fourier Transform (FFT) Dalam Analisis Frekuensi Alat Musik Harmonika

Ida Bagus Made Surya Widnyana[a1],
Ngurah Agus Sanjaya ER[a2]
[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,Universitas Udayana
Bali, Indonesia
[1]odesuryawidnyana@email.com
[2]agus_sanjaya@unud.ac.id

Abstract

Music is an art of entertainment that is attached to the world community. Many types of music and instruments are used to perform the song/instrument. Harmonica is a musical instrument that is played by blowing. How to use the Harmonica musical instrument is to suck and blow the holes contained in the harmonica so that it can produce a beautiful sound or melody. Harmonica uses a sound source in the form of a reed mounted on a reed plate. When air passes, the reed responds by vibrating back and forth through the holes (slits) that have been made in the reed plate and produces sound. In this case, it is processing the sound of the harmonica to analyze how the frequency of the harmonica music instrument is, of course, each hole on the harmonica has a different magnitude based on the frequency density of the sound produced. Fast Fourier Transform (FFT) is a transformation model that is used to convert signals from time domain form to frequency domain form. harmonica has a different frequency - different based on the amount of density / Amplitude resulting from audio recording. From a sampling rate of 22050, each slot has a different sample rate and it can be seen that each hole has a fundamental and harmonic frequency in the range 0 – 1500 Hz. This study was conducted to analyze the frequency of the harmonica per hole (slot) using the Fast Fourier Transform (FFT) method, the results of which can be used to reduce the frequency characteristics of the harmonics of each hole.

Keywords: *Harmonica, Fast Fourier Transform, Music, Audio Processing, Frequency*

1. Pendahuluan

Musik merupakan seni hiburan yang melekat pada masyarakat dunia. Banyak jenis musik dan instrumen yang digunakan untuk membawakan lagu/instrumen tersebut. Harmonika adalah alat musik yang sering digunakan untuk mengiringi suatu pementasan musik. Harmonika merupakan salah satu alat musik yang dimainkan dengan cara ditiup. Cara menggunakan alat musik Harmonika adalah dengan menyedot dan meniup lubang yang terdapat pada harmonika sehingga dapat menghasilkan suara atau melodi yang indah. Harmonika lebih dari sekedar instrumen kecil dan portabel. Harmonika menggunakan sumber suara berupa buluh yang dipasang pada pelat buluh. Ketika udara melewati, buluh merespon dengan bergetar bolak-balik melalui lubang (celah) yang sudah dibuat di pelat buluh dan menghasilkan suara [5].

Pengolahan suara adalah suatu aktivitas yang menghasilkan suatu suara yang sangat kuat baik dengan lirik maupun tidak[4]. Dalam hal ini adalah mengolahan suara harmonika untuk dianalisis bagaimana frekuensi pada alat music harmonika yang tentunya masing masing lubang pada harmonika memiliki magnitude yang berbeda beda berdasarkan kerapatan frekuensi dari suara yang dihasilkan.

Pada penelitian sebelumnya yang membahas tentang penggunaan metode *Fast Fourier Transform* (FFT) dalam Analisis Frekuensi namun yang dianalisis adalah nada pada alat Musik Piano. Digunakannya metode ini pada penelitian sebelumnya yaitu untuk mengetahui persentase nilai standart suatu piano tipe keyboard berdasarkan teori musik sehingga dapat mengetahui seberapa tingkat keteraturan frekuensi tangga nada piano berdasarkan standart teori musik[6].

Dalam penelitan ini, akan menerapkan *Fast Fourier Transform* (FFT) merupakan sebuah metode/model transformasi yang biasanya digunakan untuk merepresentasikan sinyal suara dalam domain waktu diskrit menjadi sinyal suara dalam domain frekuensi atau memindahkan sinyal domain waktu menjadi domain frekuensi.[2] dengan menggunakan metode tersebut maka akan menghasilkan frekuensi dasar sebagai pembeda dari masing-masing *lubang* pada alat musik harmonika.

2. Metode Penelitian

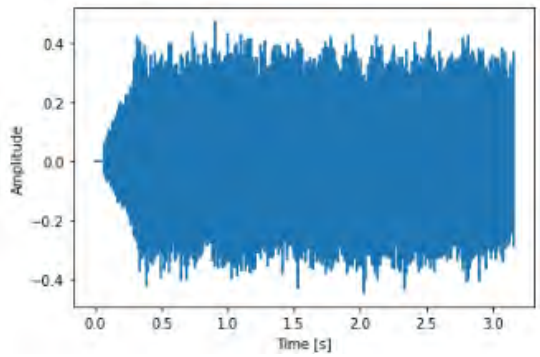
2.1 Pengumpulan Data

Data yang dikumpulkan berupa data rekaman suara dari masing-masing lubang pada alat musik Harmonika. Masing-masing *lubang* pada harmonika ditiup satu per satu selama 3 detik lalu direkam menggunakan audio ponsel lalu dipindahkan ke laptop menggunakan jalur google drive dan diubah menjadi format .wav untuk diekstrasi fiturnya yang akan menghasilkan frekuensi dasar sebagai pembeda dari masing-masing *lubang* pada alat musik harmonika. Hasil dari rekaman tersebut akan diproses menggunakan metode *Fast Fourier Transform*.

2.2 Metode

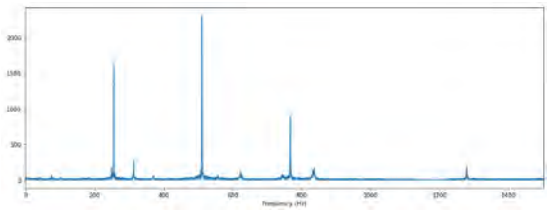
2.2.1 Fast Fourier Transform

Fast Fourier Transform adalah model transformasi yang diterapkan untuk mengkonversikan sinyal dari bentuk domain waktu ke bentuk domain frekuensi, sinyal dalam bentuk domain waktu dapat dilihat pada contoh gambar dibawah ini dengan menggunakan sample audio lubang 1 harmonika.



Gambar 1. Domain waktu

Fast Fourier Transform merupakan suatu algoritma yang dapat diterapkan untuk menghitung Discrete Fourier Transform (DFT), FFT merupakan perhitungan diskrit dengan domain frekuensi. FFT dibutuhkan untuk mengkonversi sinyal yang ditangkap oleh sensor menjadi domain frekuensi. Sebagai contoh dapat dilihat pada gambar dibawah ini dengan menggunakan sample audio lubang 1 harmonika



Gambar 2. Domain frekuensi

Perhitungan FFT dapat membagi suatu sinyal menjadi frekuensi yang berbeda – beda dalam fungsi eksponensial yang kompleks[7]. pemilihan Algoritma FFT karena FFT merupakan prosedur penghitungan DFT yang efisien sehingga akan mempercepat proses penghitungan DFT yang secara substansial dapat lebih menghemat waktu dari pada metode yang konvensional[3].

Fast Fourier Transform dapat didefinisikan melalui persamaan (1) sebagai berikut[7]:

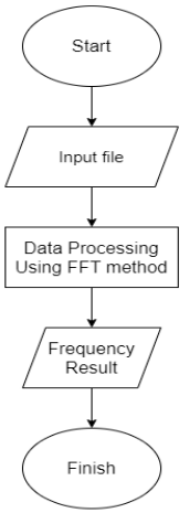
$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft}\, dt \qquad (1)$$

$S(f)$ = sinyal pada domain frekuensi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$s(t)$ = sinyal pada domain waktu
$s(t)e^{-j2\pi ft}$ = konstanta sinyal
f = frekuensi
t = waktu

2.3 FlowChart



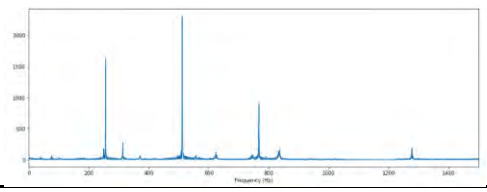
Gambar 3. flowchart

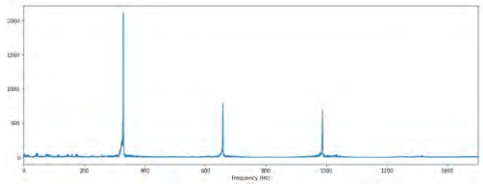
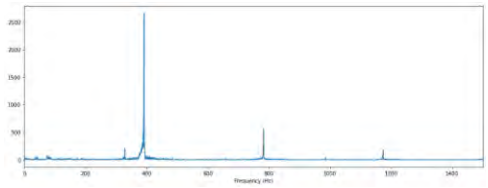
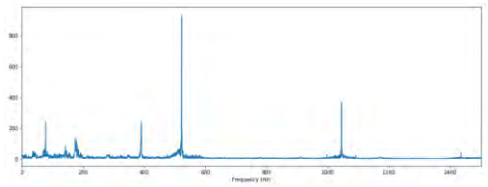
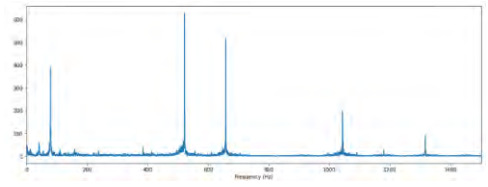
Pada gambar flowchart di atas pada studi ini, penelitian ini berfokus untuk menganalisis fitur frekuensi pada harmonika. *Input File* merupakan audio yang berasal dari rekaman suara harmonika dengan format .wav. Hasil rekaman akan diproses menggunakan FFT untuk menganalisis audio yang semula berbentuk sinyal menjadi bentuk frekuensi Hasil dari penelitian dapat dilihat pada tabel 1.

3. Hasil dan Pembahasan

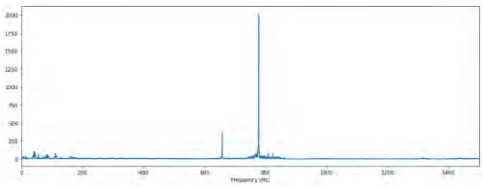
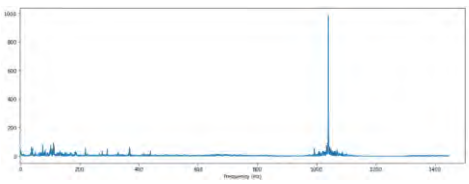
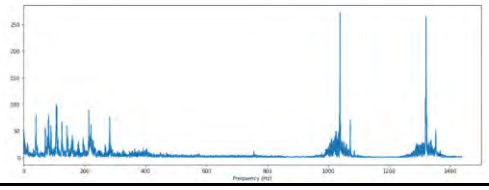
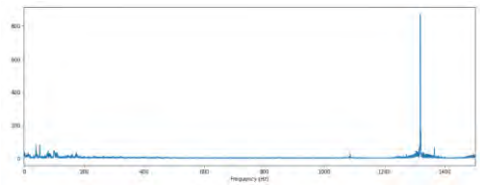
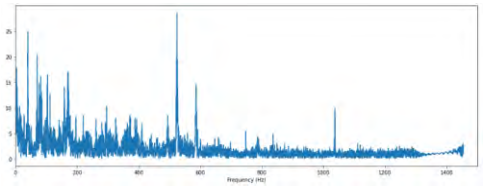
Dari hasil penelitian yang telah dilakukan, tiap lubang memiliki fundamental frequency (frekuensi dasar) dan peak frequency (*harmony*). Berikut tabel dari hasil penelitian yang telah dilakukan.

Tabel 1. Hasil Penelitian

No	Nama	Frekuensi	Gambar
1	Lubang 1	514 Hz	

2	Lubang 2	328 Hz	
3	Lubang 3	391 Hz	
4	Lubang 4	522 Hz	
5	Lubang 5	524 Hz	

p-ISSN: 2301-5373
e-ISSN: 2654-5101

6	Lubang 6	775Hz	
7	Lubang 7	1038 Hz	
8	Lubang 8	1040Hz	
9	Lubang 9	1318Hz	
10	Lubang 10	524 Hz	

Berdasarkan hasil pada tabel diperlihatkan bahwa *sample audio* setiap slot pada harmonika memiliki frekuensi yang berbeda – beda berdasarkan jumlah kerapatan/Amplitude yang dihasilkan dari perekaman audio. Dari sampling rate sebanyak 22050, masing masing slot memiliki sample rate yang

berbeda dan bisa dilihat juga setiap lubang memiliki frequency fundamental dan harmoni di rentang 0 – 1500 Hz.

4. Kesimpulan

Penelitian ini dilakukan untuk menganalisis frekuensi alat musik harmonika per lubang (slot) menggunakan metode Fast Fourier Transform (FFT) yang hasilnya dapat digunakan untuk menurunkan karakteristik frekuensi harmonika tiap lubang. Pada penelitian ini tidak menggunakan fitur yang langsung menghasilkan frekuensi secara langsung, tetapi frekuensi yang diterima akan membedakan setiap lubang. Diharapkan penelitian ini dapat dilaksanakan kembali dengan mendapatkan frekuensi setiap lubang dengan metode yang sama atau dengan menggunakan metode lain dan mencari karakteristik lain.

References

[1] Sutara, F. A., Situmorang, M. & Dian, W., 2014. ANALISIS DAN IMPLEMENTASI SONG RECOGNITION MENGGUNAKAN ALGORITMA FAST FOURIER TRANSFORM

[2] Irtawaty, A. S. (2019). Implementasi Metode Fast Fourier Transform (Fft) Dalam Mengklasifikasikan Suara Pria Dan Wanita Di Laboratorium Jurusan Teknik Elektro Politeknik Negeri Balikpapan. *JTT (Jurnal Teknologi Terpadu)*, *7*(2), 70–75. https://doi.org/10.32487/jtt.v7i2.661

[3] Abdillah, F. N. (2017). IMPLEMENTASI ALGORITMA FAST FOURIER TRANSFORM (FFT) DAN ALGORITMA HARMONIC PRODUCT SPECTRUM (HPS) Pada TUNER GITAR BERBASIS ANDROID Firdaus Noor Abdillah Fakultas Ilmu Komputer Universitas Kuningan Jalan Tjut Nyak Dhien Cijoho Kuningan Telepon (0232. *Jurnal Nuansa Informatika*, *11*(2), 18–25. https://www.journal.uniku.ac.id/index.php/ilkom/article/viewFile/1326/982

[4] Apsari, M. S. A., & Widiartha, I. M. (2021). Classification of Women's Voices Using Fast Fourier Transform (FFT) Method. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *10*(1), 59. https://doi.org/10.24843/jlk.2021.v10.i01.p08

[5] Kiranawati, S. (2021). Pengenalan Nada Harmonika Menggunakan Windowing Koefisien DST dan Jarak Matusita. *Repository.Usd.Ac.Id*, 1–85. https://repository.usd.ac.id/25510/2/084114001_Full%5B1%5D.pdf

[6] Palondongan, H. C., Budiman, E., & Taruk, M. (2019). Analisis Frekuensi Nada Piano Menggunakan Algoritma Fast Fourier Transform. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *3*(1), 81. https://doi.org/10.30872/jurti.v3i1.2467

[7] Kusuma, D. T. (2020). Fast Fourier Transform (FFT) Dalam Transformasi Sinyal Frekuensi Suara Sebagai Upaya Perolehan Average Energy (AE) Musik. *Petir*, *14*(1), 28–35. https://doi.org/10.33322/petir.v14i1.1022

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Database Performance Optimization using Lazy Loading with Redis on Online Marketplace Website

Albertus Ivan Suryawan[a1], Agus Muliantara[a2]

[a]Informatics Department, Udayana University
Bali, Indonesia
[1]albertusivan15@gmail.com
[2]muliantara@unud.ac.id (Corresponding author)

Abstract

With the pandemic lasting over 2 years, many businesses start to adapting a digital approach of their business to stay alive. While in the same time, a number of users of digital platform also skyrocketed due to physical contact restriction policy. This causes a performance hit toward several online services such as an online marketplace due to high network traffic from many users accessing it in the same time, including latency issues. In this research, the authors try to implement an application-level caching with an in-memory database, Redis, using Lazy Loading approach. Beside implementing caching, authors also compare the performance of using cache and not using cache by load testing both implementation using similarly built application. Based on the result, there is a performance gain of 38-65% based on the load and scenario by using application-level caching.

Keywords: *Cache, Cache-Aside Pattern, Lazy Loading, performance optimization, application-level caching, in-memory database*

1. Introduction

With the pandemic lasting over 2 years since March 2020, many businesses start to migrating their operation to digital platform, in order to survive due to physical contact restriction policy. Some start to use social media as an e-commerce platform, while others making their own online services from which accessible through online. The latter is the case of a certain business called Business X, which name cannot be disclosed, in which they start building their own online marketplace business to continue selling several products. Over a year after starting an online marketplace, they start noticed a high latency issue while navigating through the website, due to high network traffic. The authors try to analyze the infrastructure used, and found out there is a bottleneck on the database due to high traffic mention earlier. To address this issue, the authors try to implement caching into the application, in form of application-level caching.

Caching is often being used to reduce load coming into the main memory, which commonly known to be slow, by storing data that frequently accessed in the cache itself. Application-level is a type of caching in which the implementation of the cache is done by the developer itself based on the condition. Usually, this type of cache is implemented in the server-side of the application, and mostly relied on in-memory database such as Redis[1], [2]. In order to avoid additional strain on in-memory database due to excessive data stored inside, Cache-aside pattern or Lazy-Loading, will be used to limit when data should be stored. Cache-aside pattern is a type of data synchronization where data is being stored in the cache on demand[3].Consequently, instead of storing all data directly on the cache, only previously requested data will be stored inside until its expired or an update for the that data occurred in main database.

There are several research that has been done regarding caching such as on research done by Saldhi et al. [1] which analyze performance difference between two Cache System namely, InfiniSpan and Hazelcast. In that research, the author suggests to do a comparative performance analysis across different cache pattern. Another research with similar interest that has been done is [2] in which a performance between MySQL database were compared with Redis as cache database. Based on the result, there are some degree of performance gains by using Write-Through pattern that the writer has implemented. Although [2]may be similar to this research, there is a difference in the type of data

synchronization pattern being used, where author used Cache-Aside pattern instead of Write-Through pattern.

2. Reseach Methods

Due to the nature of the subject, the methodology used on this research mainly consist of 5 stages as shown on Figure 1, including system requirement analysis, database modeling, cache strategy modeling, application development, and performance analysis.

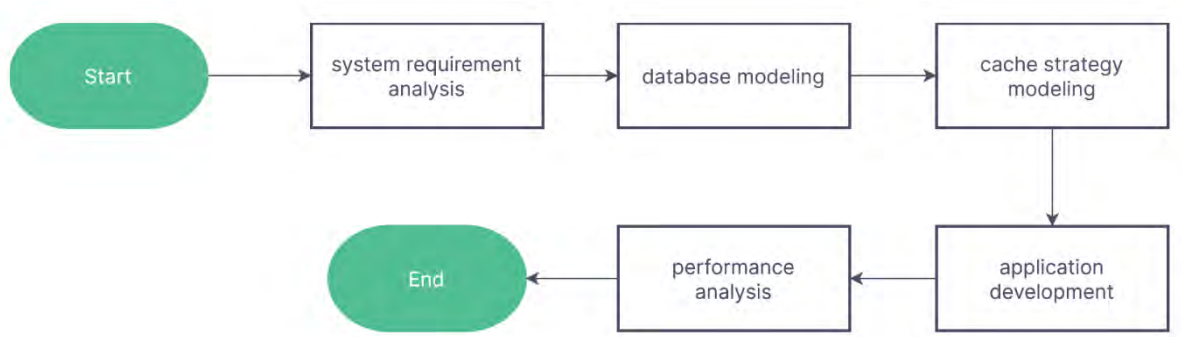


Figure 1. Research Methodology

2.1. System Requirement Analysis

System Requirement Analysis is the first stage of System Development Life Cycle (SDLC) where the developer will conduct an analysis to list any requirement needed for the system. On this research, there are 3 main feature that needed to be existed including, user registration and login; get list of available products and get product's detail; and checkout an order.

2.2. Database Modeling

In this research, there are 2 types of databases being used for the application, a relational database using MySQL, and an in-memory database namely Redis. Redis is chosen due to its popularity as a cache management system with high flexibility of usage such as shown in [4]. MySQL will be used as the main database, where all of the data will be stored, and Redis will be used as a cache layer, where all cached data will be store based on the cache strategy modeling being used. MySQL database consist of 4 tables as shown in **Figure 2** and **Figure 3**.

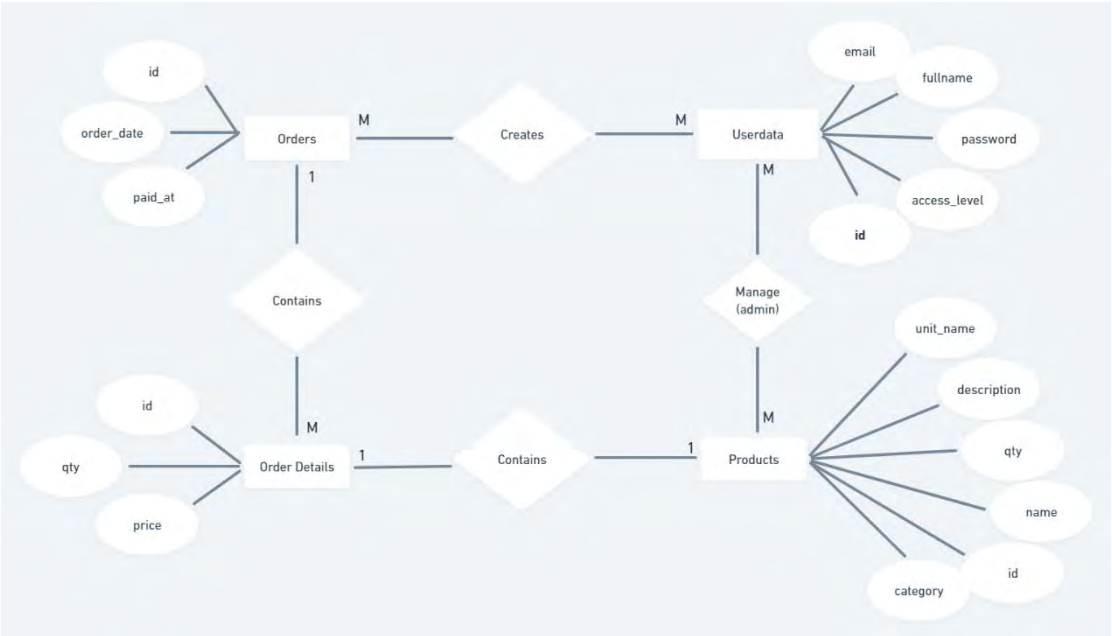


Figure 2. Entity Relational Diagram

p-ISSN: 2301-5373
e-ISSN: 2654-5101

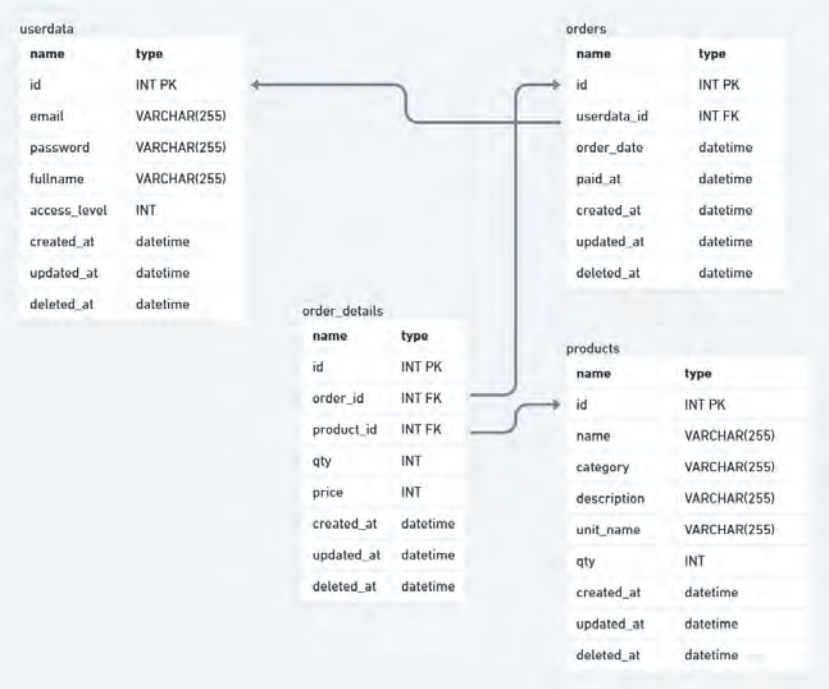


Figure 3. Relational Database Model

Table *userdata* contains all user data including email, password, full name, and access privilege level. Table *orders* and *order_details* contain all order data such as product bought, quantity, price at the time bought, and user who ordered it. Table *products* contains all product data such as product name, price, and remaining stock. While MySQL can be model after their table, Redis in the other hand is a key-value database in which described as shown in Table 1. Standardized naming scheme for Redis' key is being implemented to keep the caching logic simple and avoid several problems in the future such as name collisions [5]

Table 1. Redis Naming Scheme

Field	Naming Scheme	Expiry Duration
Userdata Cache	user:<userdata_id>	1 Hour
Session Cache	session:<session_id>	5 Minutes
User-Session Index Key	user-session:<user_id>	5 Minutes
Product Detail Cache	product:<product_id>	1 Hour
All Product Summary	product:summary	1 Hour

2.3. Cache Strategy Modeling

There are several ways to implement data synchronization between main database and the cache. As mention in the introduction, the author chooses to implemented a Cache-Aside pattern in which data is stored in the cache when there is a request for it, and retain in the cache until its expiry time reached or the corresponding data is updated in the database. In **Figure 4**, the overall flow of retrieving data using Cache-Aside pattern is shown.

First, application will try to fetch the data from the cache, if the data existed, it immediately returned. Otherwise, application will try to fetch the requested data from the main database, and store its value to the cache if data found in the database, and then its value returned. In **Figure 5**, the overall flow of updating data is shown, where data will be discarded from the cache after data successfully updated in the main database. In this application, deleted data is a data in the main database that has field *deleted_at* set to non-*NULL* value (i.e., *timestamp* of data being deleted).

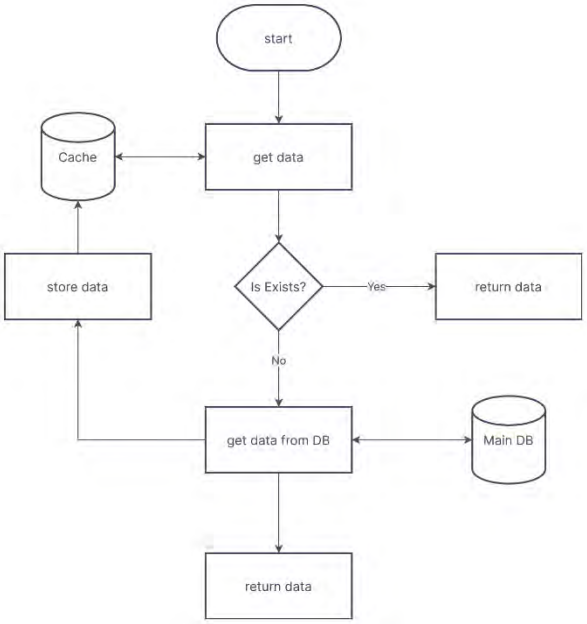


Figure 4. Cache-Aside Data Retrieval

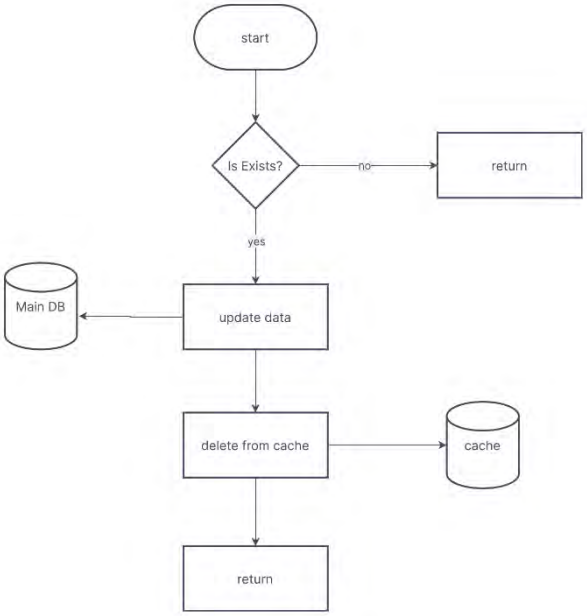


Figure 5. Data Update Flowchart

2.4. Application Development

In this stage, application is developed in Go Language. This language is chosen because it's known for the performance especially as a server-side application. There are 3 main components needed, *controller, service,* and *repository*. Controller is part of application in which incoming requests processed before passed to *service* to be consumed. Service is part of application in which the business logic lies. In *service,* program will also call *repository* to fetch data whether it's from cache or main database. Repository is part of application which process all database-related operation, whether from main database or from cache layer. In *repository,* all database connections are handled by each officially supported connector driver for Go Language.

3. Result and Discussion

In this research, the performance of the proposed application is measured by running a load test between proposed application and a similarly built application without cache being used. Load test is

p-ISSN: 2301-5373
e-ISSN: 2654-5101

done using k6 with fixed time of 300 seconds and virtual users, which denoted a concurrent connection to the application in a time, of {100, 200, 500}. This load test measures the time from sending requests to first byte received for 3 types of requests scenario. The overall result is shown in **Figure 6**

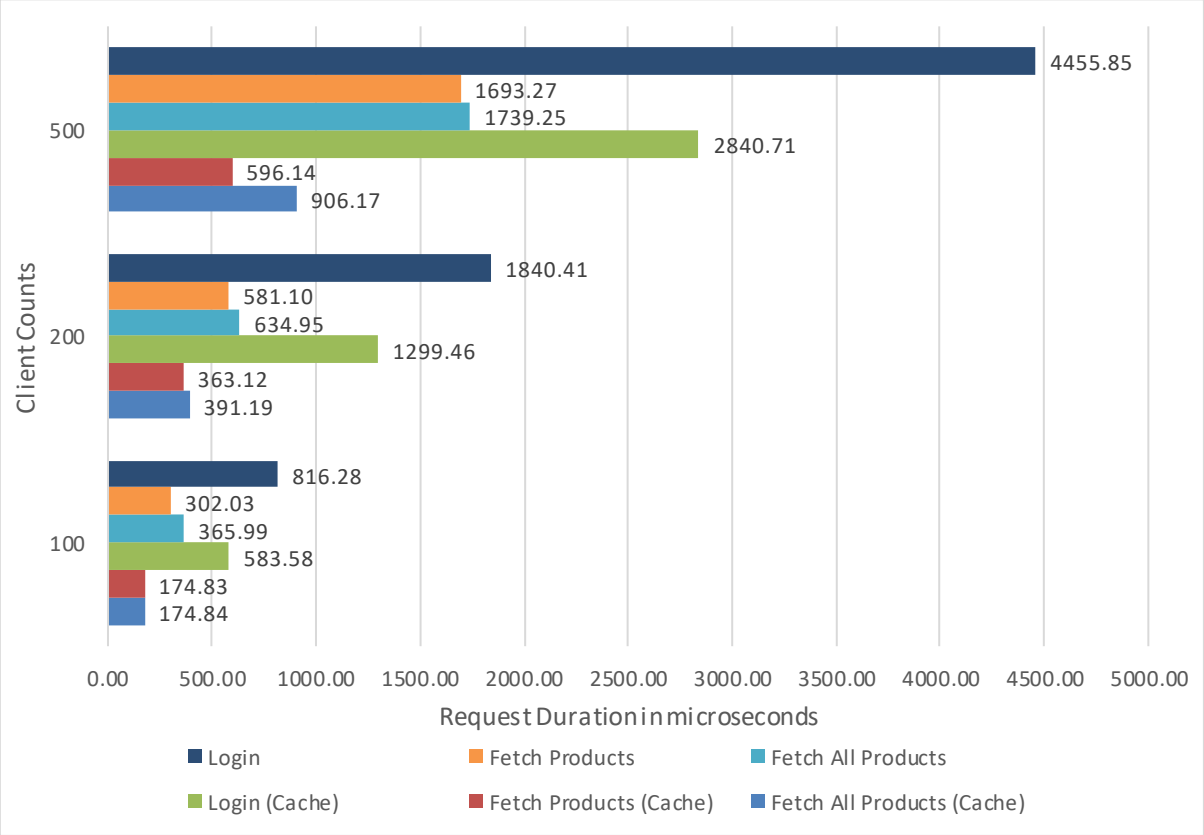


Figure 6. Overall performance of each scenario in term of request duration

3.1. Login Scenario

This scenario simulates users' action of login with their account which simulate cold data, data that has not been stored in the cache, being fetch and stored in the cache for future use (i.e., re-login to get new *sessionID* for API authentication). Result of this interaction is shown in Table 2. Based on the result, by implementing cache on login endpoint, results in performance gain about 28-36% depending on the load in average. There are some performance degradations as shown by the minimum value of using cache when interacting with 500 concurrent users of about -343% relative to using no cache.

Client Count	Request Duration (*ms*)					
	Avg	Min	Max	Avg	Min	Max
	Cache			No Cache		
100	583.58	1.68	3633.21	816.28	3.20	4859.88
200	1299.46	2.10	6024.22	1840.41	7.22	9693.84
500	2840.71	25.43	16179.62	4455.85	6.00	22780.09

Table 2. Login Scenario Load Test Result

3.2. Fetch All Products Scenario

This scenario simulates users' action of browsing through the products catalogue which simulate repetition of fetching the same data over and over. Result of this interaction is shown in Table 3. Based on the result, by implementing cache on fetch all product endpoint, results in performance gain about 38-53% depending on the load in average.

Client Count	Request Duration (*ms*)					
	Avg	Min	Max	Avg	Min	Max
	Cache			No Cache		
100	174.84	0.51	2767.48	365.99	1.04	1837.42
200	391.19	1.00	4810.10	634.95	0.77	6466.26
500	906.17	1.16	9499.17	1739.25	1.00	12837.18

Table 3. Fetch All Products Scenario Load Test Result

3.3. Fetch Product's Detail Scenario

This scenario simulates users' action of browsing details for product they have interest with, which simulates random repetition of fetching the same data. Result of this interaction is shown in Table 4. Based on the result, by implementing cache on fetch product's detail endpoint, results in performance gain about 38-65% depending on the load in average.

Client Count	Request Duration (*ms*)					
	Avg	Min	Max	Avg	Min	Max
	Cache			No Cache		
100	174.83	0.52	1976.53	302.03	0.54	1810.73
200	363.12	0.53	2762.30	581.10	0.63	7013.00
500	596.14	1.12	12169.89	1693.27	1.00	13770.91

Table 4. Fetch Product's Detail Load Test Result

4. Conclusion

There are many ways to improve performance of a web-based application such as an online marketplace, including implementing a cache layer. Based on the result shown in the previous section, the implementation of cache in the application can be considered as a successful attempt. Implementation of cache do increase the application request performance by 38-65% based on the scenario, with repetition of fetching the same data (i.e., all products summary, and product's detail) gained at least 38% on high load. Besides performance gains, the authors also found out that number of concurrent users also affect the performance of the application as shown in the results, whereas the client count increases, the duration it takes to complete the request also increases. For future work, authors suggest make another approach of cache implementation.

References

[1] H. Salhi, F. Odeh, R. Nasser, and A. Taweel, "Benchmarking and Performance Analysis for Distributed Cache Systems: A Comparative Case Study," 2018, pp. 147–163. doi: 10.1007/978-3-319-72401-0_11.

[2] M. I. Zulfa, A. Fadli, and A. W. Wardhana, "Application caching strategy based on in-memory using Redis server to accelerate relational data access," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 157–163, Apr. 2020, doi: 10.14710/jtsiskom.8.2.2020.157-163.

[3] Microsoft, "Cache-Aside pattern." https://learn.microsoft.com/en-us/azure/architecture/patterns/cache-aside (accessed Oct. 01, 2022).

[4] R. K. Singh and H. K. Verma, "Effective Parallel Processing Social Media Analytics Framework," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2860–2870, Jun. 2022, doi: 10.1016/j.jksuci.2020.04.019.

[5] J. Mertz and I. Nunes, "A Qualitative Study of Application-level Caching," Oct. 2020, doi: 10.1109/TSE.2016.2633992.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Analisis Sentimen Pengguna Aplikasi MyPertamina Dengan Menggunakan Algoritma Naïve Bayes

I Gusti Bgs Darmika Putra[a1], Cokorda Rai Adi Pramarthar[a2]

[a]Program Studi Informatika, Universitas Udayana
Bali, Indonesia
[1]gstbgsdarmika@gmail.com
[2]cokorda@unud.ac.id

Abstrak

MyPertamina adalah aplikasi layanan keuangan digital yang di miliki oleh PT Pertamina dan anggota Badan Usaha Milik Negara (BUMN), dengan menggunakan aplikasi ini pengguna dapat membeli beberapa produk buatan Pertamina secara cashless atau nontunai Tujuan dari aplikasi MyPertamina yaitu untuk mendata masyarakat yang telah membeli BBM bersubsidi. Dalam membangun aplikasi, diperlukan keahlian khusus serta banyak hal yang harus dipetimbangkan, termasuk ulasan-ulasan pengguna setelah menggunakan aplikasi karena dapat menentukan kualitas dari aplikasi yang diciptakan apakah bagus atau justru sebaliknya. Dengan analisis sentiment, pengembang aplikasi dapat dengan mudah untuk mengklasifikasikan ulasan karena data ulasan akan dikelompokan menjadi ulasan positif atau ulasan negative. Tujuan dari penelitian ini, yaitu untuk mengetahui sentimen pengguna aplikasi MyPertamina yang terdapat pada ulasan Google Playstore dengan menerapkan metode Naïve Bayes untuk mengklasifikasikan sentiment dan menggunakan metode tf-idf yang digunakan untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data training (training set). Dari hasil penelitian memperoleh hasil akurasi sebesar 72% dengan menggunakan 200 data ulasan pengguna aplikasi MyPertamina yang terdiri dari 100 ulasan postif dan 100 ulasan negative.

Keywords: *Analisis Sentimen, Naïve Bayes, Multinomial Naïve Bayes,TF-IDF, Ulasan Aplikasi, Aplikasi Mypertamina,*

1. Pendahuluan

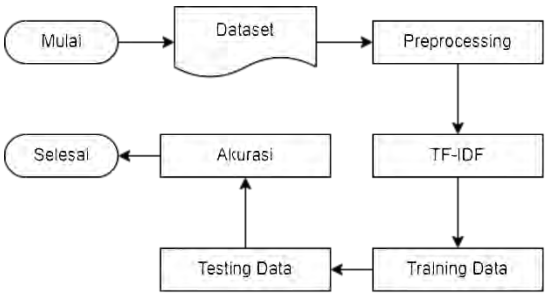
Perkembangan teknologi ini membuat semua aspek kehidupan tidak dapat terlepas dari penggunaan teknologi, baik secara langsung maupun secara tidak langsung. Perkembangan teknologi memuncul berbagai jenis aplikasi digital yang dapat menunjang kebutuhan manuasia untuk membuat pekerjaan manusia menjadi lebih mudah, cepat dan efisien. Pada tahun 2012, terdapat sekitar 700.000 aplikasi digital yang tersedia untuk pengguna android serta terdapat sekitar 25 juta aplikasi yang telah diunduh dari Google Playstore atau toko aplikasi utama Android [1]. Aplikasi yang ada di Google Playstore memiliki berbagai macam fungsi dan manfaat bagi pengguna yang dapat menunjang kebutuhan pengguna, salah satunya yaitu aplikasi MyPertamina. Aplikasi ini merupakan aplikasi layanan keuangan digital yang di miliki oleh PT Pertamina dan anggota Badan Usaha Milik Negara (BUMN) dengan menggunakan aplikasi ini pengguna dapat untuk membeli beberapa produk buatan Pertamina, termasuk BBM, secara cashless atau nontunai. Tujuan dari aplikasi MyPertamina yaitu untuk mendata masyarakat yang telah membeli BBM bersubsidi. Alih-alih untuk mempermudah pengguna dalam melakukan pembayaran pembelian BBM melalui aplikasi, malah justru kebalikannya yaitu membuat pengguna kesulitan dalam menggunakan aplikasi hal ini dapat dilihat dari ulasan yang diberikan oleh pengguna yang telah menggunakan aplikasi MyPertamina. Hal ini membuat tuntutan terhadap mutu pelayanan aplikasi MyPertamina harus di tingkatkan agar pengalaman pengguna dapat lebih baik lagi.

Klasifikasi sentiment bertujuan untuk mengatasai masalah ini dengan cara otomatis mengklasifikasikan ulasan dari pengguna dengan mengkelompokan menjadi ulasan positif atau ulasan negatif [2]. Terdapat beberapa penelitian sebelumnya mengenai ulasan aplikasi yang terdapat pada Google Playstore diantaranya, penelitian yang dilakukan oleh [3] pada tahun 2020 dengan judul Algoritma Klasifikasi Naïve Bayes untuk Analisa Sentimen Aplikasi Shopee. Hasil penelitian ini menghasilkan nilai akurasi sebesar 71.50% dan Nilai AUC (Area Under Curve) sebesar 0,500. Penelitian selanjutnya oleh [4] pada tahun 2019 dengan judul Analisis Sentimen Aplikasi Transportasi Online Krl Access Menggunakan Metode Naïve Bayes. Penelitian ini menggunakan metode NBC yang memberikan hasil akurasi hingga 84.00% untuk data uji review opini positif berbahasa Indonesia pada pemilihan aplikasi transportasi darat pada smartphone.

Merancang suatu aplikasi diperlukan keahlian khusus serta banyak hal yang harus dipetimbangkan dalam proses perancangannya, namun dalam merancang suatu aplikasi diperlukan juga untuk memperhatikan ulasan-ulasan yang diberikan dari pengguna terhadap aplikasi yang telah diciptakan karena dengan melihat ulasan tersebut pengembang aplikasi tau hal apa saja yang membuat aplikasi buruk atau sebaliknya, sehingga memudahkan pengembang aplikasi dalam menambahkan dan memperbaiki fitur dari aplikasi yang diciptakan. Dalam penelitian ini, penulis melakukan analisis sentimen pada aplikasi MyPertamina untuk mengetahui apakah ulasan yang diberikan oleh pengguna termasuk ke ulasan positif atau negatif. Hasil akhir yang akan dihasilkan adalah tingkat akurasi yang dicapai dalam melakukan analisis sentimen menggunakan metode Naïve Bayes.

2. Metode Penelitian

Pada penelitian tentang analisisa sentimen pengguna aplikasi MyPertamina ini, penulis melakukan metode eksperimen,dengan tahapan berikut:



Gambar 1. Tahap Penelitian

2.1 DataSet

Pada penelitian ini, pengguna menggunakan data-data komentar yang ada pada tahun 2022. Data yang dikumpulkan sebanyak 200 data. Data komentar yang terdiri dari data komentar positif sebanyak 100 data, dan data komentar negatif sebanyak 100 data. Data ini bersumber dari https://play.google.com.

2.2 Preprocessing

Pada tahapan preprocessing yang penulis gunakan untuk dataset yang dimiliki adalah:

a. Tokenization
   Pada proses ini, dilakukan pemotongan string input berdasarkan tiap kata yang menyusunya dan kata yang memiliki tanda baca serta simbol yang bukan huruf dihilangkan.

b. Filtering (Stopword Removal)
   Proses penghapusan atau pembuangan kata-kata yang sering ditampilkan dalam dokumen. Proses menghilangkan kata-kata tersebut adalah dengan melakukan pengecekan pada kamus *stopword.* Jika kata tersebut ada di dalam kamus stopword, maka
   kata tersebut akan dihilangkan. Stopword dapat berupa kata depan, kata sambung, kata seru, dan lain sebagainya Contoh stopword dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll.

c. Stemming

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Stemming merupakan tahapan mengubah kata menjadi bentuk dasar dengan menghilangkan semua imbuhan. Terdapat dua pendekatan dalam melakukan stemming yaitu dengan pendekatan kamus dan pendekatan aturan. Pada penelitian ini menggunakan bantuan Sastrawi Python untuk mendapatkan kata dasar.merupakan library python yang dapat digunakan untuk mendapatkan kata dasar dari kata yang kita inputkan. Sastrawi Python sendiri bergantung pada kamus kata dasar dari www.kateglo.com. Algoritma yang digunakan oleh library ini adalah algoritma Nazief dan Andriani, di mana algoritma ini merupakan salah satu algoritma yang cukup populer untuk melakukan stemming kata dalam bahasa Indonesia. [5].

Tabel 1. Hasil Preprocessing

Data	Aplikasi bagus, cuman kok gak ad jawaban Pertamina ya,
Tokenization	aplikasi, bagus, cuman, kok, gak, ad, jawaban, pertamina, ya
Stopwords Removal	aplikasi, bagus, cuman, kok, gak, ad, jawaban, pertamina, ya
Stemming	aplikasi, bagus, cuman, gak, ad, pertamina, ya

2.3 Term Frequency Invers Document Frequency (TF-IDF)

TF-IDF merupakan metode yang merupakan integrasi antar term frequency (TF), dan inverse document frequency (IDF). Fungsi metode tf-idf adalah untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data training *(training* set) dimana nantinya dibentuk suatu vektor Antara dokumen dengan kata (documents *with terms*) yang kemudian untuk kesamaan antar dokumen dengan *cluster* akan ditentukan oleh sebuah *prototype* vektor yang disebut juga dengan *cluster centroid* [6].

2.4 Naïve Bayes

Naïve Bayes adalah salah satu algoritma klasifikasi yang sederhana dan sering digunakan untuk klasikasi dokumen. Dalam proses membangun sistem pengklasifikasi menggunakan NB terdapat 2 tahapan yang dilakukan. Tahap pertama adalah proses pelatihan (training) dan tahap yang kedua adalah proses pengujian (testing) [7].

a. Multinomial Naïve Bayes
Multinomial Naïve Bayes adalah salah satu metode khusus dari metode Naïve Bayes yang menggunakan probabilitas bersyarat. Probabilitas bersyarat dapat dilakukan dengan menggunakan frekuensi kemunculan suatu kata dalam suatu kelas (*raw term frequency*) [8].
$P(c|term\ \ dokumen\ \ d) = \ P(c)x\ P(t1|c)x\ \ P(t2|c)x\ P(t3|c)x. . . x\ P(tn|c)$

Probabilitas kelas c sebelumnya ditentukan oleh rumus:
$\mathrm{P(c)} = \frac{Nc}{N}$

Multinomial Naïve Bayes untuk nilai input x ditunjukkan pada persamaan berikut:
$\mathrm{P(tn|c)} = \frac{W(c,tn)+1}{(\sum \mathrm{W\prime < VW\prime (c,tn)} + \mathrm{B`})}$

Keterangan :
$\mathrm{W}(c, tn)$: Nilai pembobotan tfidf atau W dari term t di kategori c
$\sum \mathrm{W'<} VW'ct$: Jumlah total W dari keseluruhan term yang berada di kategori c.
B' : Jumlah W kata unik (nilai idf tidak dikali dengan tf) pada seluruh dokumen.

b. Skenario Eksperimen
Perhitungan akurasi dari sistem yang berguna untuk mengukur kinerja sistem yang telah lakukan dengan menggunakan rumus:

$\mathrm{Akurasi} = \frac{Db}{N}\ x\ 100\%$

3. Hasil dan Pembahasan

Metode yang diusulkan menggunakan 200 data ulasan dari pengguna aplikasi MyPertamina dengan 100 data ulasan positif dan 100 data ulasan negative, data ini bersumber dari

https://play.google.com. Terdapat 20 % dataset digunakan untuk data uji dan sisanya digunakan sebagai data latih. Terdapat tahapan preprocessing data sebelum data ulasan pengguna digunakan analisis data. Proses ini berguna untuk memastikan kualitas data baik dan lebih bersih sehingga bisa digunakan untuk pengolahan selanjutnya. Terdapat beberapa tahapan dalam proses preprocessing yaitu: (1) Pertama, memisahkan kalimat ulasan menjadi beberapa kata dan mengubah seluruh data menjadi huruf kecil, (2) Menghilangkan tanda baca yang ada di setiap ulasan pengguna. (3) Kemudian penghapusan atau pembuangan kata-kata yang sering ditampilkan atau kata-kata yang tidak relevan menggunakan Sastrawi. Setelah melalui tahap preprocessing, hasilnya dari proses preprocessing akan dibobot menggunakan rumus *Term Frequency Invers Document Frequency* (TFIDF). Setelah menentukan bobot dari suatu term (kata) menggunakan tfi-df maka Langkah yang terakhir yaitu n klasifikasi menggunakan metode Naïve Bayes dan Multinomial Naïve Bayes.

Tabel 2. Hasil akurasi Naïve Bayes dan Multinomial Naïve Bayes

Accuracy	72 %
Precision	88 %
Recall	64 %
f1_score	74 %

Pada percobaan pertama, kami menggunakan 100 data ulasan yang diberikan oleh pengguna aplikasi MyPertamina yang terdiri dari 50 ulasan postif dan 50 ulasan negative namun akurasi dari percobaan pertama ini masih kurang baik yaitu 65 % dikarenakan data ulasan tersebut masih kurang bersih pada saat tahapan preprocessing data yang mengakibatkan akurasi dari mesin kurang maksimal. Kemudian untuk percobaan kedua kami menggunakan 200 data ulasan pengguna aplikasi yang terdiri dari 100 ulasan postif dan 100 ulasan negative, dari hasil percobaan kedua di dapatkan akurasi dari algoritma naïve bayes sebesar 72% sesuai dengan tabel 2 yang tertera diatas.

4. Kesimpulan

Metode Naïve Bayes dan Multinomial Naïve Bayes memberikan hasil akurasi hingga 72% untuk data uji ulasan pengguna terhadap aplikasi MyPertamina yang bersumber dari https://play.google.com. Metode Naïve Bayes dan Multinomial Naïve Bayes memberikan hasil yang tepat dalam mengklasifikasikan ulasan apakah termasuk dalam sentiment positif atau termasuk dalam sentiment negatif. Metode tf-idf digunakan untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data training *(training* set).

Dari penelitian ini, kedua metode tersebut dapat memberikan hasil yang tepat dalam mengklasifikasikan ulasan pada aplikasi MyPertamina yang menentukan apakah ulasan tersebut termasuk kedalam sentiment postif atau negatif. Kedepannya dapat dilakukan perbaikan dengan melakukan tahapan preprocessing data yang lebih baik lagi agar data yang digunakan lebih bersih sehingga mendapatkan akurasi yang lebih baik lagi.

Referensi

[1] R. Ridjalaludin, I. A. Ratnamulyani, and A. A. Kusumadinata, “Pengaruh Penggunaan Layanan Aplikasi Digital Google Play Dalam Smartphone Terhadap Pemenuhan Kebutuhan Informasi Mahasiswa,” *J. Komun.*, vol. 2, no. 2, pp. 135–146, 2017, doi: 10.30997/jk.v2i2.229.

[2] R. Sari, “Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization,” *J. Inform.*, vol. 6, no. 1, pp. 23–28, 2019, doi: 10.31311/ji.v6i1.4695.

[3] S. Masripah and L. D. Utami, “Algoritma Klasifikasi Naïve Bayes untuk Analisa Sentimen Aplikasi Shopee,” *Swabumi*, vol. 8, no. 2, pp. 114–117, 2020, doi: 10.31294/swabumi.v8i2.8444.

[4] S. Nurwahyuni, “Analisis Sentimen Aplikasi Transportasi Online Krl Access Menggunakan Metode Naive Bayes,” *Swabumi*, vol. 7, no. 1, pp. 31–36, 2019, doi: 10.31294/swabumi.v7i1.5575.

[5] G. Setiawan, H. N. Palit, and E. Setyati, “Aspect Based Sentiment Analysis pada Layanan Umpan Balik Universitas dengan Menggunakan Metode Naïve Bayes dan Latent Semantic Analysis,” *J. Infra*, vol. 7, no. 1, pp. 170–174, 2019.

[6] N. K. Widyasanti, I. K. G. Darma Putra, and N. K. Dwi Rusjayanthi, “Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia,” *J. Ilm. Merpati (Menara Penelit.*

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Akad. Teknol. Informasi), vol. 6, no. 2, p. 119, 2018, doi: 10.24843/jim.2018.v06.i02.p06.

[7] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.

[8] I. G. C. P. Yasa, N. A. Sanjaya ER, and L. A. A. Rahning Putri, "Sentiment Analysis of Snack Review Using the Naïve Bayes Method," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 3, p. 333, 2020, doi: 10.24843/jlk.2020.v08.i03.p16.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Perancangan Sistem Aplikasi Penyewaan Mobil Berbasis Mobile

I Wayan Agus Juniartha[a1], I Gede Santi Astawa[a2]

[a]Informatics Department, Udayana University
Bali, Indonesia
[1]agusjuniartha16@email.com

[2]santi.astawa@unud.ac.id

Abstract

Nowadays, people tend to search for goods or services through their mobile device using an electronic marketplace. In Bali, a rental business for tourist, weekend events, and other eventsgrow rapidly. This business provides services that rents various kinds of car such as innova, alphard, alya, and others that are commonly used in weekend events. However, unfortunate for rental service entrepreneurs who have high prospects, many of them are using the traditional method of marketing and sales such as word-of- mouth. Moreover, many of them not catching up with the current development of the rental car. By listing the services on the rental, easier for the customer to find the car they want that fit their needs and budget. Also, by utilizing rental car will save the customer time, because the customer can just search, compare, and book torent car without leaving home. The purpose of this research project is to create a mobile- based application of market place that accommodates entrepreneurs renting car. Our purposed mobile-based rental uses the prototyping method to build this mobile-based application.
Keywords: *Rental, Find Car, Application, Mobile,Tourist*

1. Introduction

Pariwisata dinilai oleh banyak pihak memiliki arti penting sebagai salah satu alternatif pembangunan, terutama bagi negara atau daerah yang memiliki keterbatasan sumber daya alam. Untuk memaksimumkan dampak positif dari pembangunan pariwisata dan sekaligus menekan serendah mungkin dampak negatif yang ditimbulkan, diperlukan perencanaan yang bersifat menyeluruh dan terpadu. Rencana pengembangan pariwisata diperlukan oleh berbagai pihak sebagai pedoman dalam mengembangkan aktivitas dibidang masing-masing. Bahkan, rencana pengembangan dimaksud harus bersinergi dengan rencana-rencana pembangunan pada sektor-sektor lain dan tetap konsisten dengan rencana pembangunan kepariwisataan nasional secara keseluruhan. Pariwisata merupakan kegiatan yang kompleks, bersifat multi sektoral dan terfragmentsikan, karena itu koordinasi antar berbagai sektor terkait melalui proses perencanaan yang tepat sangat penting artinya. Perencanaan juga diharapkan dapat membantu tercapainya kesesuaian (match) antara ekspektasi pasar dengan produk wisata yang dikembangkan tanpa harus mengorbankan kepentingan masing-masing pihak. Mengingat masa depan penuh perubahan, maka perencanaan diharapkan dapat mengantisipasi perubahan-perubahan lingkungan strategis yang dimaksud dan menghindari sejauh mungkin dampak negatif yang ditimbulkan oleh perubahan-perubahan lingkungan tersebut.

.

A. Penyewaan.

Penyewaan adalah sebuah persetujuan di mana sebuah pembayaran dilakukan atas penggunaan suatu barang atau properti secara sementara oleh orang lain. Barang yang dapat disewa bermacam-macam, tarif dan lama sewa juga bermacam-macam. Rumah umumnya disewa dalam satuan tahun, mobil dalam satuan hari, permainan komputer seperti PlayStation disewa dalam satuan jam. Untuk sewa mobil, biasanya perusahaan jasa penyewaa mobil menerapkan tarif per 12 jam atau per 24 jam

B. Transportasi

Transportasi adalah kebutuhan utama untuk setiap orang di masa modern ini. Kesibukan, baik itu untuk pekerjaan maupun untuk wisata, sangat bergantung kepada kemampuan kita untuk bergerak dari satu tempat ke tempat lain. Salah satu moda transportasi yang paling sering ditemukan di jalan-jalan kota adalah mobil. Banyak orang memiliki mobil pribadi untuk kemudahan bepergian, karena fleksibilitasnya untuk dapat dibawa ke mana saja. Tetapi, tidak semua orang memiliki mobil, atau mungkin saja pendatang dari luar kota atau luar negeri untuk keperluan bisnis atau wisata membutuhkan mobil.

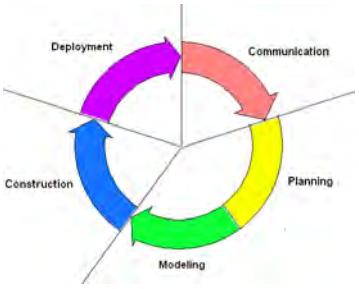
C. Android.

Android adalah sistem operasi berbasis Linux yang dirancang untuk perangkat seluler lapisan sentuh seperti telepon pintar dan komputer tablet. Android pada awalnya dikembangkan oleh Android, Inc. Android pertama kali dirilis secara resmi pada tahun 2007. Antarmuka pengguna android umumnya berupa manipulasi langsung, menggunakan gerakan sentuh yang meliputi menggeser, mengetuk, dan mencubit untuk memanipulasi objek pada layer. Untuk memasukkan teks di Android, ada keyboard virtual untuk menulis teks. Saat ini Android merupakan sistem operasi smartphone yang paling banyak digunakan di dunia. Android menjadi pilihan bagi perusahaan atau pengembang teknologi yang memiliki biaya murah, dapat dimodifikasi dengan mudah dan ringan untuk perangkat berteknologi tinggi tanpa harus mengembangkannya dari awal. Selain itu beberapa keunggulan sistem operasi android dibandingkan dengan sistem operasi lainnya diantaranya adalah user friendly, yang dimaksud disini adalah sistem android sangat mudah untuk dijalankan. Android sangat mudah dijalankan bahkan untuk orang yang tidak familiar dengan smartphone. Selain itu, tampilan yang lebih menarik juga menjadi salah satu alasan mengapa Android lebih unggul dari sistem operasi lainnya.

2. Research Method

A. Prototyping Method

Proses pengembangan sistem sering menggunakan pendekatan prototipe (prototyping). Metode ini paling baik digunakan untuk menyelesaikan masalah kesalahpahaman antara pengguna dan analis yang timbul dari pengguna yang tidak dapat dengan jelas mendefinisikan kebutuhan mereka [1]



Picture 1. Metode *Prototyping*

Tahapan dalam Model Prototyping dirangkum sebagai berikut:

1) Langkah 1 : Pengumpulan dan Analisis Persyaratan

Sebuah model prototyping dimulai dengan analisis kebutuhan. Pada fase ini, kebutuhan sistem didefinisikan secara rinci. Selama proses, pengguna sistem diwawancarai untuk mengetahui apa harapan mereka dari sistem.

2) Langkah 2 : Desain Pendahuluan

Tahap kedua adalah desain pendahuluan atau quick design. Pada tahap ini, desain sederhana dari sistem dibuat. Namun, itu bukan desain yang lengkap. Ini memberikan gambaran singkat tentang sistem kepada pengguna. Desain cepat membantu dalam mengembangkan prototipe.

3) Langkah 3 : Membangun Prototipe

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pada fase ini, prototipe sebenarnya dirancang berdasarkan informasi yang dikumpulkan dari desain cepat. Ini adalah model kerja kecil dari sistem yang diperlukan.

4) Langkah 4 : Evaluasi Awal User

Pada tahap ini, sistem yang diusulkan disajikan kepada klien untuk evaluasi awal. Ini membantu untuk mengetahui kekuatan dan kelemahan model kerja. Komentar dan saran dikumpulkan dari pelanggan dan diberikan kepada pengembang.

5) Langkah 5 : Memperbaiki Prototipe
Jika pengguna tidak senang dengan prototipe saat ini, Anda perlu memperbaiki prototipe sesuai dengan umpan balik dan saran pengguna.
Fase ini tidak akan berakhir sampai semua persyaratan yang ditentukan oleh pengguna terpenuhi. Setelah pengguna puas dengan prototipe yang dikembangkan, sistem akhir dikembangkan berdasarkan prototipe akhir yang disetujui.

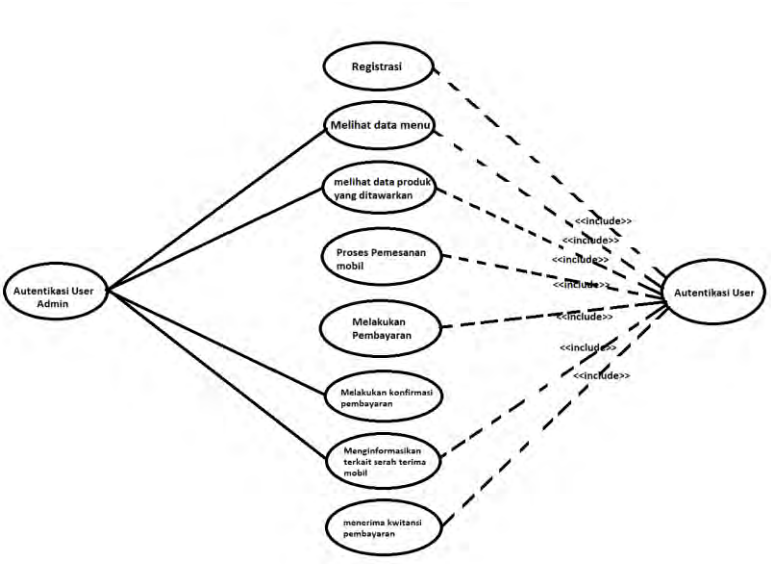
6) Langkah 6 : Menerapkan dan Perawatan rutin Produk

Setelah sistem final dikembangkan berdasarkan prototipe akhir, sistem tersebut diuji secara menyeluruh dan disebarkan ke produksi. Sistem menjalani perawatan rutin untuk meminimalkan waktu henti dan mencegah kegagalan skala besar.

B. Modeling Methods

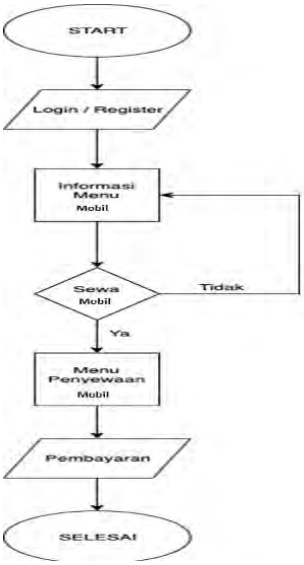
Unified Modeling Language (UML) adalah salah satu alat yang paling dapat diandalkan di dunia pengembangan sistem berorientasi objek. Hal ini karena UML menyediakan bahasa pemodelan visual yang memungkinkan pengembang sistem untuk membuat cetak biru visi mereka dalam bentuk standar, mudah dipahami dan dilengkapi dengan mekanisme yang efektif untuk berbagi dan mengkomunikasikan desain mereka dengan orang lain. Alur perancangan sistem yang digunakan adalah use case diagram dan activity diagram.

3. Result and Discussion



Picture 2. *Use Case* Diagram

Pada use case diagram di atas dapat dijelaskan bahwa, admin login terlebih dahulu setelah itu admin dapat mengakses data menu, data produk yang ditawarkan mengkonfirmasi pembayaran, menginformasikan tentang penyewaan mobil. Kemudian pengguna dapat mengakses semua menu yang ada di aplikasi setelah melakukan login terlebih dahulu..



Picture 3. System Flowchart

A. Testing

- Proses pertama pada tahapan Tes atau pengujian ini adalah dengan memberikan task atau tugas yang harus dilakukan oleh pengguna pada prototype.

Task Pengujian Prototype	
Task 1	Melakukan Register Pada Aplikasi
Task_2	Melakukan Login Pada Aplikasi
Task 3	Melihat List Mobil yang disewa
Task 4	Melakukan Logout Aplikasi

- Proses selanjutnya yaitu memberikan pengguna kuesioner SUS yang berisi 10 buah pertanyaan.

Daftar Pertanyaan Kuisioner	
No	Pertanyaan
1	Saya berfikir akan menggunakan sistem ini lagi
2	Saya merasa sistem ini rumit untuk di gunakan
3	Saya merasa sistem ini mudah digunakan
4	Saya membutuhkan bantuan dari orang lain atau teknisi dalam menggunakan sistem ini
5	Saya merasa fitur-fitur sistem ini berjalan dengan semestinya
6	Saya merasa ada banyak hal yang tidak konsisten (tidak serasa pada sistem ini)
7	Saya merasa orang lain akan memahami cara menggunakan sistem ini dengan cepat
8	Saya merasa sistem ini membingungkan
9	Saya merasa tidak ada hambatan dalam menggunakan sistem ini
10	Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan sistem ini

- Masing-masing pertanyaan mempunyai nilai 1 - 5 dengan keterangan sebagai berikut:
  1=Sangat Tidak Setuju,
  2=Tidak Setuju,
  3=Ragu,
  4=Setuju,
  5=Sangat Setuju
- Dari kuesioner yang diberikan kepada 5 orang pengguna, didapat data skor awal sebagai berikut:

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Responden 1	3	4	2	1	5	2	3	3	4	3
Responden 2	2	2	5	4	2	2	4	4	5	3
Responden 3	5	1	1	3	4	3	2	3	4	2
Responden 4	1	5	3	2	1	4	1	2	2	5
Responden 5	4	3	4	5	3	1	5	3	1	2

- Data skor dari pengguna selanjutnya diolah dengan menggunakan metode SUS (System Usability Scale), dimana setiap skor dari pertanyaan nomor ganjil (1,3,5,7,9) akan dikurangi dengan 1. Sedangkan pada pertanyaan bernomor genap (2,4,6,8,10) nilai 5 dikurangi dengan skor dari setiap pertanyaan

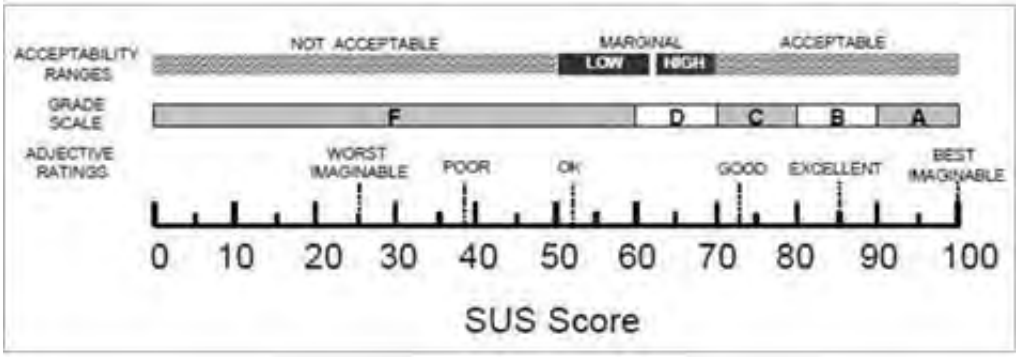
.

Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Responden 1	3	1	3	0	4	3	2	2	1	2
Responden 2	3	3	4	1	3	3	1	1	4	2
Responden 3	5	0	0	2	1	2	3	2	1	3
Responden 4	0	4	2	3	0	1	0	3	3	4
Responden 5	1	2	1	4	2	0	4	2	0	3

- Selanjutnya jumlah skor dari setiap responen akan dikalikan dengan 2,5 terlebih dahulu, sebelum dicari nilai rata-ratanya. Nilai rata-rata terdebutlah yang menjadi skor akhir dari pengujian.

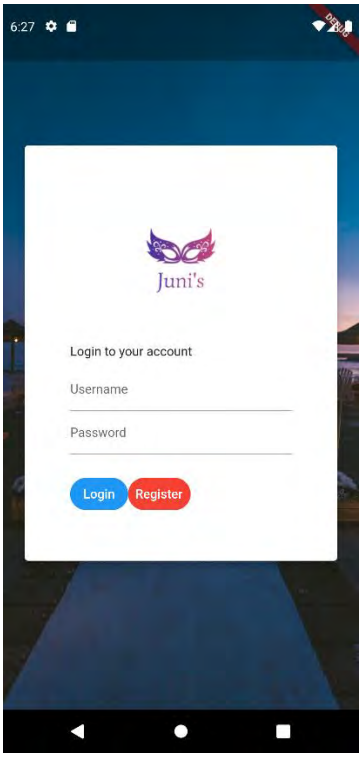
Responden	Jumlah	Jumlah x 2,5
Responden 1	21	52,5
Responden 2	25	62,5
Responden 3	19	47,5
Responden 4	20	50
Responden 5	19	47,5
SUS Skor		260/5 = 52

-
- Dengan hasil penghitungan SUS pada tabel, maka nilai akhir dari pengujian adalah 52.
- Skor akhir dari pengujian prototype yang dibuat adalah 74, merujuk dari SUS Acceptability Score **nilai diatas 70 masih berstatus Acceptable** atau dengan kata lain masih layak dilanjutkan ke tahap pengembangan lebih lanjut dengan memperhatikan masukan-masukan dari pengguna.
-



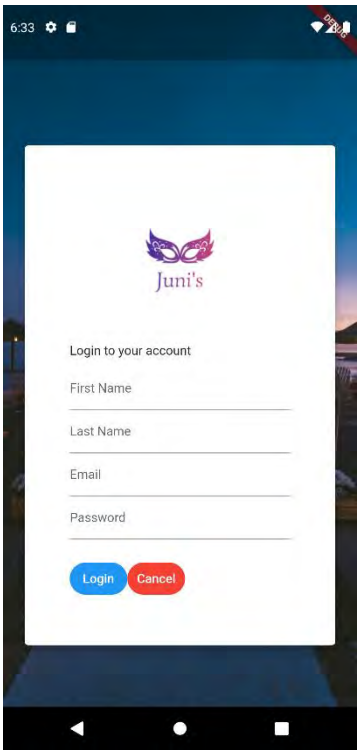
Grade Rankings of SUS Acceptability Score (researchgate.net)

4. Implementation



Picture 4. Login Page

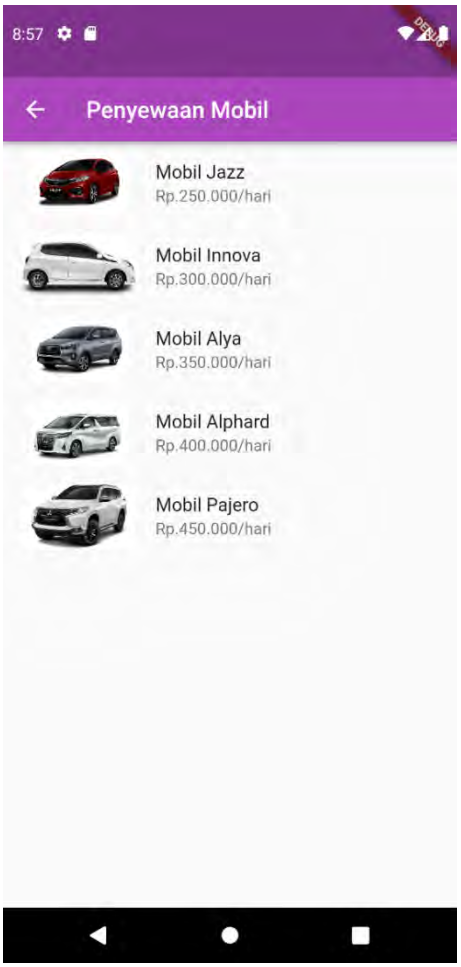
Pada login page ini user bisa menginputkan username dan password yang sudah didaftarkan pada sistem ini jika ada kesalahan maka akan keluar peringatan username atau password belum terdaftar



p-ISSN: 2301-5373
e-ISSN: 2654-5101

Picture 5. Register Page

Pada register page ini user yang belum memiliki akun bisa mendaftarkan diri dengan mengisi form register yang sudah tersedia lalu akan disimpan pada penyimpanan local untuk profil user setelah melakukan register user bisa kembali kehalaman login untuk melakukan login menggunakan akun yang sudah didaftarkan



Picture 6. Main Page

Pada halaman main page ini user bisa melihat list-list mobil yang sudah di sediakan dari database sistem untuk melihat mobil mana yang ingin disewa oleh user beserta harganya

5. Conclusion

Berdasarkan uraian dan pembahasan sebelumnya, dalam penelitian ini dapat disimpulkan bahwa dengan adanya Penyewaan mobil ini, pengguna dalam mencari dan memesan mobil menjadi lebih mudah. Pengguna yang ingin mencari jasa sewa mobil dimudahkan dengan banyaknya pilihan..

References

[1] Ogedebe, P.M.,& Jacob, B.P. , 2012, Software Prototyping: A Strategy to Use When User Lacks Data Processing Experience. ARPN Journal of Systems and Software. VOL. 2, NO.6 , 2012

[2] Axel, Najoan, dkk. “Rancang Bangun Aplikasi Berbasis Android Untuk Informasi Kegiatan dan Pelayanan Gereja”, UNSRAT, 2017.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Default Risk Prediction Using Decision Tree Study Case of Home Credit

Dewa Nyoman Agung Adipurwa Mahandiri[a1], Agus Muliantara[a2]

[a]Informatics Department, Faculty of Math and Natural Sciences, Udayana University
Bali, Indonesia
[1]adiagung707@email.com
[2]muliantara@unud.ac.id

Abstract

Consuming loans with any service has become a trend in modern society. However, that trend gives some risk for the loan company such as Home Credit. Home Credit needs to create an automation analytic for predicting customers that might be default in future. So, we build a machine learning model using the Decision Tree algorithm to resolve that risk. The Decision Tree model can give mean score 85% accuracy, 91% precision ,and 92% recall score for Home Credit study case.

Keywords: *Credit, Default Risk Prediction, Machine Learning, Decision Tree, CRIPS DM*

1. Introduction

Seiring berjalannya perkembangan zaman, kualitas hidup masyarakat bergerak ke tingkat yang lebih tinggi disertai dengan perubahan gaya hidup yang semakin modern dan terbuka. Dengan hal tersebut, konsumsi pinjaman dengan berbagai metode seperti kredit ataupun *pay later* secara bertahap diterima oleh masyarakat. Berdasarkan data dari Statistik Fintech Lending Indonesia oleh Otoritas Jasa Keuangan (OJK), peningkatan entitas peminjam terus terjadi selama dua tahun terkahir. Peningkatan sebesar 19,8 juta peminjam tercatat dari Agustus 2021 hingga Agustus 2022. Kemudian peningkatan lebih besar terjadi pada rentang Agustus 2020 hingga Agustus 2021 sebesar 41 juta peminjam. Sejalan dengan peningkatan tersebut, pinjaman pribadi dengan layanan *peer-to-peer lending* cenderung lebih diminati oleh masyarakat terutama anak muda karena minimnya persyaratan dan prosedur sehingga waktu dibutuhkan lebih pendek, tetapi tetap aman sebab telah diawasi oleh OJK. Kemudahan tersebut menyebabkan banyak masyarakat yang mencoba untuk mendapatkan pinjaman. Namun, setiap lembaga atau perusahaan yang memberikan layanan peer-to-peer lending harus tetap membatasi dan memilah peminjam berdasarkan analisis kemampuan peminjam membayar kembali pinjaman. Pembatasan ini bertujuan untuk mengurangi kemungkinan risiko kegagalan kredit. Risiko kegagalan kredit yang paling sering terjadi pada layanan peer-to-peer lending adalah risiko kegagalan awal atau *default risk* [1].

Default risk adalah risiko yang diambil oleh pemberi pinjaman bila peminjam tidak dapat melakukan pembayaran yang diperlukan dari kewajiban utang pinjamannya [2]. Analisis terhadap default risk merupakan faktor penting dalam lembaga keuangan karena memungkinkan untuk memberikan pinjaman hanya kepada konsumen yang memiliki kredit baik. Analisis default risk secara konvensional dilakukan oleh lembangan dengan kartu penilain kredit pada setiap konsumen. Kartu penilaian kredit menganalisis secara statistik kelayakan kredit konsumen dan membantu lembaga keuangan untuk menolak atau memperpanjang kredit [3]. Dengan nilai kredit yang diperoleh, konsumen akan diklasifikasikan sebagai “kredit baik” hingga “kredit buruk”. Namun sebagaimana cara konvensional lainnya, hal tersebut akan sulit berlaku pada lembaga keuangan yang telah menerima ribuan bahkan ratusan juta kredit sehingga diperlukan teknologi untuk membantu otomatisasi prediksi default risk.

Machine learning adalah pendekatan yang tepat untuk data analisis dan prediksi yang terotomatisasi analisis konvensional. Sebagai subdomain dari a*rtificial intelligence, machine learning* dapat belajar dari data yang diberikan, mengidentifikasi pola, dan mengambil keputusan berdasarkan informasi yang diekstrak dengan meminimalisir intervensi manusia[4]. Dengan menggunakan *machine learning*,

proses analisis konsumen yang tepat untuk diberikan kredit dapat lebih cepat dilakukan bahkan dapat dilakukan prediksi berdasarkan catatan kredit yang pernah dilakukan sebelumnya ataupun hanya berdasarkan informasi dari latar belakang konsumen sebagai dasar data latihan untuk mesin dengan menggunakan algoritma tertentu. Berbagai algoritma dapat diuji untuk membangun model *machine learning* dengan akurasi, presisi, dan *recall* suatu model salah satunya decision tree. Banyak penelitian telah menggunakan algoritma *decision tree* untuk membangun model prediksi yang akurat. Oleh sebab itu, penelitian ini dilakukan untuk membentuk model prediksi *default risk* dari lembaga kredit Home Credit dengan menggunakan model *decision tree* dengan metode *Cross-Industry Standard Process for Data Mining* (CRISP DM).

2. Reseach Methods

Cross-Industry Standard Process for Data Mining adalah model proses data mining (data mining framework) yang diprakarsai oleh 5 perusahan yaitu Integral Solution Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Framework ini kemudian dikembangkan kembali oleh ratusan perusahaan dan organisasi sebagai metodologi terbuka (open source) bagi data mining. Model proses CRISP DM memberikan gambaran tentang siklus hidup pengembangan proyek data mining melalui 6 tahapan yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [5].

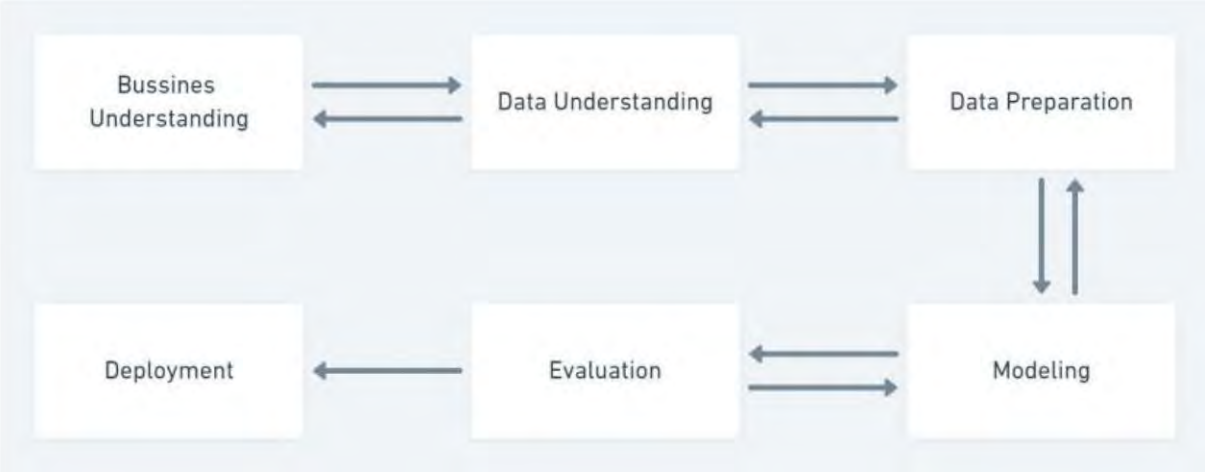


Figure 1. CRISP DM Diagram

2.1. Bussiness Understanding

Kebutuhan konsumsi terkadang datang secara insidental seperti mengganti alat rumah tangga utama, memperbaiki rumah, ataupun membayar pengeluaran yang tidak direncanakan. Dalam kondisi ini, konsumen mungkin kekurangan uang tunai untuk segera menangani kewajiban pembayaran. Oleh karena itu, layanan pinjaman hadir untuk memberi konsumen bantuan keuangan jangka pendek atau pencapaian tujuan jangka pendek.

Home Credit adalah penyedia layanan pinjaman *multi-channel* internasional untuk pembiayaan konsumen. Home Credit berupaya untuk memperluas ekspansi pelayanannya bagi masyarakat yang tidak memiliki rekening bank. Dengan penargetan lebih banyak konsumen maka kemungkinan Home Credit untuk meningkatkan dan menempati lini teratas sebagai penyedia pinjaman bukan bank. Namun, semakin luas cakupan penawaran perusahaan, semakin banyak risiko yang harus ditangani oleh Home Credit salah satunya *default risk*. Home Credit harus mampu menentukan kemungkinan konsumen yang akan mengalami *default*. Jika tingkat *default* tinggi, perusahaan akan mengalami kerugian dalam ekspansinya. Tujuan Home Credit adalah memberikan menawarkan pinjaman kepada individu yang berkemungkinan tinggi dapat membayar kembali pinjaman dan menolak permintaan pinjaman untuk yang mungkin mengalami *default* dimasa depan. Tantangan yang harus dihadapi adalah data dari latar belakang peminjam yang sangat beragam. Dalam penelitian ini, saya menggunakan *machine learning* dengan algoritma *decision tree* untuk memprediksi apakah konsumen akan mengalami default.

2.2. Data Understanding

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Kebutuhan Data terdiri atas 8 set dengan 2 data utama yaitu data latih "application_train.csv" dan data tes "application_test.csv". Data latih terdiri atas 307.551 tinjauan dari 122 variabel. Pada data latih ini terdapat variabel "TARGET" yang yang menunjukan apakan konsumen mengalami kesulitan dalam membayar kredit. Selanjutnya, data tes terdiri atas 48744 tinjaun dan 121 variabel. Data test tidak mencakup variabel "TARGET" karena data tes akan digunakan untuk menguji output dari model apakah mampu menghasilkan prediksi target yang akurat. Secara tidak langsung, variabel target pada data latih telah menyatakan hampir semua konsumen dengan nilai variabel "TARGET" sama dengan 1 (benar) akan mengalami *default*. Data set lainnya adalah pecahan dari cakupan data latih dan data tes untuk mempermudah memahami setiap bagian variabel yang serumpun seperti dataset "berau.csv", "credit_card_balance" , dan lainnya. Oleh sebab itu, untuk mempersingkat waktu penelitian maka data set hanya fokus pada data latih dan data tes sebagai data utama. Sebanyak 122 variabel pada data latih dapat dipecah menjadi 5 rumpun yaitu informasi pribadi konsumen, informasi pinjaman, informasi demografi konsumen, dokumen yang diberikan konsumen, dan pertanyaan yang diajukan ke biro kredit.

2.3. **Data Preparation**

Sebelum melakukan analisis, hal pertama yang perlu dilakukan adalah membuat data set mentah dapat digunakan. Langkah pertama dengan membersihkan data yaitu mencari cara untuk menangani data yang hilang atau bernilai Null. Penanganan dapat mempertimbangkan beberapa acuan yaitu makna keberadaan data dan besarnya data, sehingga dapat diputuskan apakah akan mengecuailkan variabel keseluruhan, menghapus satu baris selaras dengan data yang mengandung data hilang, atau melakukan imputasi pada baris data yang mengandung data hilang. Dalam analisis ini, data hilang diimputasi dengan menggunakan median dan modus. Median untuk jenis data numerik dan modus untuk jenis data kategorik atau objek, proses tersebut dilakukan untuk kedua data utama. Jumlah data hilang tertinggi pada kedua jenis data lebih dari 21.000.

Langkah kedua, mengubah data kategorik menjadi data numerik agar dapat diperhitungkan secara komputasi dengan cara melabelkan setiap variasi. Variabel "GENDER" akan diubah menjadi 0 untuk perempuan (F) dan 1 untuk laki-laki (M), serta data kategorikal lainnya.

Langkah ketiga, menemukan dan menangani outliers dengan metode IQR.

$$\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1 \quad (1)$$

$$\text{Lower Bound} = (\mathrm{Q}_1 - (1.5 \times \mathrm{IQR}) \quad (2)$$

$$\text{Upper Bound} = (\mathrm{Q}_3 + (1.5 \times \mathrm{IQR}) \quad (3)$$

$$\text{Lower Bound} < \text{Data Outliers} < \text{Upper Bound} \quad (4)$$

Data outliers dapat mengganggu model dan menyebabkan bias pada model sehingga perlu ditangani. Untuk menangani data outliers dapat digunakan teknik yang sama dengan penanganan data hilang. Namun, penangangan data outliers perlu mempertimbangkan apakah data tersebut time series atau tidak memiliki hubungan yang cukup berpengaruh antara data lainnya. Pada penelitian ini, penanganan dirasa tidak perlu dilakukan karena cukup banyak data yang bersifat times series dan tidak ada variabel dengan data outliers yang sangat tinggi hingga mendekati setengah dari data keseluruhan.

Langkah terakhir adalah visualisasi data yang telah siap untuk digunakan, proses visualisasi ini akan dapat digunakan lagi untuk melakukan data understanding atau dapat melanjutkan langkah berikutnya pada modeling.

2.4. Modeling

Berdasarkan tujuan pada pemahaman bisnis, penyelesaian masalah yang diperlukan adalah menentukan apakah konsumen dikategorikan akan *default* atau tidak. Dengan demikian, penyelesaian masalah ini berupa pembentukan model kalsifikasi.

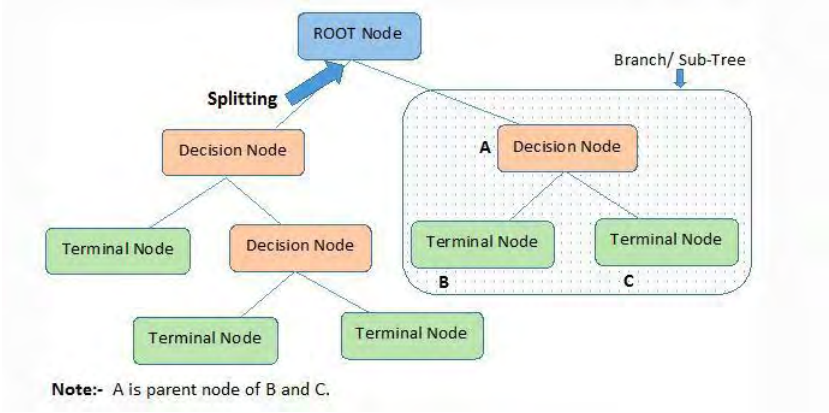


Figure 2. Decision Tree Plot

Pada penelitian ini, decision tree akan digunakan sebagai pilihan model klasifikasi untuk menentukan apakah konsumen tersebut bernilai "Ya" atau "Tidak" sebagai konsumen yang berpotensi akan default. Proses modeling terdiri atas, pembuangan variabel yang tidak dibutuhkan dalam pembentukan model pada data latih. Selanjutnya, pembagian data menjadi data untuk latih dan validasi berdasarkan data set latih dengan rasio 25%-50% untuk data validasi [6]. Berdasarkan partisi data tersebut, diperoleh 206.032 data latih dan 101.479 data validasi. Dengan menggunakan modul sklearn, model decision tree dibentuk sebagai berikut

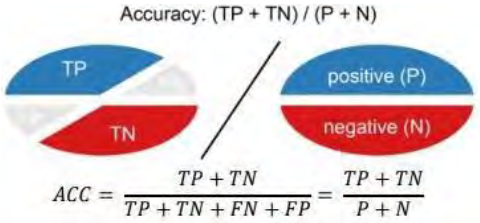
```
#Buat data x dan y
x=data_train.drop(['TARGET','SK_ID_CURR'],axis=1)

y=data_train['TARGET']

#Partisi data dengan 67:33
x_train, x_valid, y_train, y_valid = train_test_split(x,y,test_size=.33,random_state=1)

#Panggil modul untuk model Decision Tree
model = DecisionTreeClassifier()
model.fit(x_train, y_train)

#Lakukan prediksi model
y_pred_dt = model.predict(x_valid)
```

Table 1. Model Decision Tree

Setelah model berhasil dibangun, maka perlu evaluasi untuk menilai model apakah sudah baik atau belum

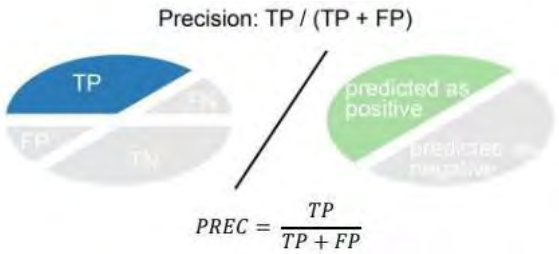
2.5. Evaluation

Evaluasi digunakan untuk menilai kinerja dari model machine learning. Ada banyak metode atau cara untuk menilai kinerja sebuah model. Secara umum, motode yang sering digunakan untuk menilai kinerja model klasifikasi yaitu accuracy, precision, recall.



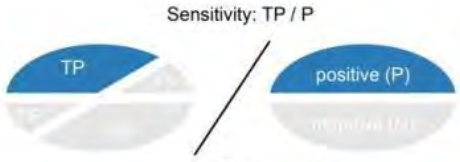
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Accuracy adalah perhitungan jumlah dari dua prediksi akurat (TP + TN) dibagi banyak data set (P+N). Akurasi memiliki rentang dari terbaik sampai terburuk (1.0 - 0.00) [7].



$$PREC = \frac{TP}{TP + FP}$$

Precision adalah perhitungan nilai prediksi benar positif (TP), dibagi dengan total nilai prediksi positif (TP + FP) [7].



$$SN = \frac{TP}{TP + FN} = \frac{TP}{P}$$

Recall adalah perhitungan nilai prediksi benar positif (TP), dibagi dengan nilai positif (P) [7].

2.6. Deployment

Tahap terakhir model akan dibentuk menjadi sebuah servis berupa API menggunakan Fast API. Sebelum dapat dihadirkan dalam API, model yang telah siap digunakan harus disimpan dalam disk dengan bentuk file. Salah satu bentuk yang umum digunakan adalah model.sav dengan menggunakan modul pickle.

3. Result and Discussion

Adapun beberapa temuan dari penelitian tentang default risk prediction berdasarkan data set Home Credit. Pada penelitian ini model yang diimplementasikan adalah Decision Tree dengan beberapa eksperimen untuk memperoleh kinerja model Decision Tree terbaik.

Eksperimen pertama dengan melakukan partisi 80% data latih dan 20% data validasi

	DecisionTree 80:20
accuracy_score	0.851259
recall_score	0.926132
precision_score	0.910831

Eksperimen pertama dengan melakukan partisi 70% data latih dan 30% data validasi

	DecisionTree 70:30
accuracy_score	0.850836
recall_score	0.925846
precision_score	0.910635

Eksperimen pertama dengan melakukan partisi 67% data latih dan 33% data validasi

	DecisionTree 67:33
accuracy_score	0.851575
recall_score	0.925877
precision_score	0.911505

Eksperimen pertama dengan melakukan partisi 60% data latih dan 40% data validasi

	DecisionTree 60:40
accuracy_score	0.850022
recall_score	0.925419
precision_score	0.910196

Eksperimen pertama dengan melakukan partisi 50% data latih dan 50% data validasi

	DecisionTree 50:50
accuracy_score	0.849736
recall_score	0.925401
precision_score	0.909899

Berdasarkan eksperimen model yang dibangun menggunakan algorimat Decision Tree dengan bergaram rasio partisi menghasilkan rata rata nilai akurasi 85%. Nilai akurasi 85% sudah tergolong nilai yang sangat baik untuk akurasi suatu model machine learning yang disertai dengan rata-rata recall sebesar 92% dan precision sebesar 91%

4. Conclusion

Pada penelitian ini, dataset dibagi menjadi berbagai rasio untuk menemukan model Decision Tree terbaik dalam mendeteksi risiko default pada konsumen berdasarkan 5 bagian dara konsumen yaitu informasi pribadi konsumen, informasi pinjaman, informasi demografi konsumen, dokumen yang diberikan konsumen, dan pertanyaan yang diajukan ke biro kredit. Dengan melakukan eksperimen tersebut diperoleh model Decision Tree dengan rata-rata akurasi yaitu 85% diikut nilai recall dan precision masing-masing sebesar 92% dan 91%.. Dengan demikian, model ini sudah cukup baik untuk digunakan memprediksi risiko *default* pada konsumen. Meskipun model ini sudah cukup baik, kami rasa masih perlu penelitian lebih lanjut terkait kemungkinan model lainnya dapat memperoleh nilai akurasi yang lebih tinggi.

References

[1] A. Krichene, “Using a naive Bayesian classifier methodology for loan risk assessment: Evidence from a Tunisian commercial bank,” *Journal of Economics, Finance and Administrative Science*, vol. 22, no. 42, pp. 3–24, 2017, doi: 10.1108/JEFAS-02-2017-0039.

[2] Y. R. Chen, J. S. Leu, S. A. Huang, J. T. Wang, and J. I. Takada, “Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3079701.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

[3] Z. Khemais, D. Nesrine, and M. Mohamed, "Credit Scoring and Default Risk Prediction: A Comparative Study between Discriminant Analysis & Logistic Regression," *Int J Econ Finance*, vol. 8, no. 4, 2016, doi: 10.5539/ijef.v8n4p39.

[4] J. Y. Seo, "Machine Learning in Consumer Credit Risk Analysis: A Review," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, 2020, doi: 10.30534/ijatcse/2020/328942020.

[5] M. F. M. Salleh and T. Suryanto, "Fraud Detection on Banking Industry in South Sumatera: A Study on the Role of Internal Auditors'," *International Journal of Shari'ah and Corporate Governance Research*, vol. 2, no. 2, 2019, doi: 10.46281/ijscgr.v2i2.399.

[6] R. R. Picard and K. N. Berk, "Data splitting," *American Statistician*, vol. 44, no. 2, 1990, doi: 10.1080/00031305.1990.10475704.

[7] Ž. Vujović, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Analisis Keamanan pada Aplikasi *Udayana Mobile* Mengacu pada *OWASP Mobile Top 10 2016*

Muhammad Arrysatrya Yusuf Putranda[a1], I Komang Ari Mogi[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jalan Raya Kampus Unud, Jimbaran, Bali, 80361, Indonesia
[1]yanda4869@gmail.com
[2]arimogi@unud.ac.id

Abstract

Udayana Mobile merupakan aplikasi yang dirilis pada smartphone oleh Universitas Udayana yang bertujuan memudahkan pegawai, dosen, serta mahasiswa mengakses beberapa fitur yang dimiliki Universitas Udayana seperti SIMAK, SIRAISA, SIPENA, dan UKT-Ku. Namun seperti semua aplikasi pada platform smartphone, terdapat kemungkinan isu kerentanan yang ada pada aplikasi Udayana Mobile. Menggunakan MobSF untuk melakukan Analisis Keamanan, dan mengacu pada OWASP *Mobile Top 10 2016.* Didapatkan hasil bahwa aplikasi ini memiliki tiga isu kerentanan. Rincian kerentanan yang ditemukan merupakan *Insufficient Cryptography* (kerentanan akibat kurang kuatnya proses hashing pada aplikasi), *Client Code Quality* (Kerentanan akibat penggunaan database SQLite dimana memungkinan penyerang melakukan serangan SQL Injection), dan *Reverse Engineering* (Kerentanan yang memungkinan penyerang mendapatkan informasi sensitif dari pengguna seperti *username, password, dan key*). Berdasarkan kerentanan tersebut, diberikan rekomendasi dalam peningkatan keamanan aplikasi berupa enkripsi data yang lebih kuat, proses validasi dalam eksekusi *SQL query*, serta teknik obfuscation pada aplikasi.

Keywords: *Udayana Mobile, MobSF, Analisis Keamanan, Owasp Mobile Top 10 2016, Vulnerabillity*

1. Pendahuluan

Perkembangan aplikasi yang berbasis smartphone, terutama android memiliki kemampuan dalam meningkatkan produktivitas seperti tersedianya layanan, fitur serta data yang mudah diakses sehingga banyak digunakan. Saat ini Android menguasai 82,8% pangsa pasar sebagai platform perangkat smartphone di dunia yang paling banyak digunakan [1].

Namun, besarnya pasar Android berbanding lurus dengan besarnya risiko kerentanan yang terdapat pada aplikasi yang tersedia di Android. Menurut laporan *Vulnerabilities and Threats in Mobile Application 2019* yang diterbitkan oleh ptsecurity.com, dilaporkan bahwa aplikasi android memiliki kerentanan yang cukup besar, yakni sebesar 43%. Kerentanan yang umum ditemukan adalah *insecure data storage* yang memiliki presentase sebesar 76%. Kerentanan tersebut sebagian besar disebabkan karena lemahnya keamanan yang dibuat ketika proses pembuatan aplikasi. [2].

Aplikasi berbasis android mulai banyak digunakan di beberapa sektor, salah satunya adalah pada sektor pendidikan. Universitas Udayana merupakan salah satu universitas di Indonesia yang menggunakan teknologi sistem informasi yang memudahkan proses organisasi informasi di kalangan civitas akademik. Salah satunya adalah aplikasi *Udayana Mobile*. Udayana Mobile merupakan aplikasi yang memungkinkan pengguna mengakses beberapa fitur seperti SIRAISA dan SIPENA bagi dosen dan pegawai, serta Sistem Informasi Manajemen Akademik (SIMAK) dan UKT-Ku bagi mahasiswa.

Berdasarkan uraian diatas, penulis melakukan penelitian ini untuk melakukan Analisis Keamanan yang ada pada aplikasi Udayana Mobile dengan menggunakan *OWASP Mobile Top 10 2016* sebagai acuan serta memberikan rekomendasi dari hasil analisis keamanan untuk meningkatkan keamanan pada aplikasi.

2. Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode kaulitatif dimana metode ini merupakan metode untuk mengumpulkan data yang berfokus kepada analisis keamanan aplikasi

Udayan Mobile dengan mengacu kepada *OWASP Mobile Top 10 2016* dan menggunakan tools Mobile Security Framework (MobSF).

2.1. Kajian Pustaka

a. OWASP *Mobile Top 10 2016*

Berkembangnya aplikasi berbasis smartphone diiringi dengan tingginya tingkat ancaman serta kerentanan yang ada. OWASP melakukan survei pada tahun 2015 yang bertujuan menganalisis serta mengkategorikan kerentanan yang terdapat pada aplikasi *mobile* yang selanjutnya dirilis sebagai OWASP *Mobile Top 10 2016* dimana ke-10 kerentanan ini berfokus pada aplikasinya, adapun daftarnya adalah [3]:

1. *Improper Platform Usage* [M1]
   Kerentanan ini merupakan kerentanan penyalahgunaan fitur yang ada pada platform serta kegagalan dalam kontrol keamanan yang memungkinkan pengguna mengakses fitur-fitur tertentu meski tidak memiliki akses ke fitur tersebut.

2. *Insecure Data Storage* [M2]
   Kerentanan ini merupakan kerentanan penyimpanan yang tidak aman serta adanya kemungkinan kebocoran data. Kerentanan ini memungkinkan penyerang mengakses berbagai informasi yang disimpan dalam aplikasi seperti informasi pribadi, *cookies* untuk login, dan sebagainya.

3. *Insecure Communication* [M3]
   Kerentanan ini merupakan kerentanan yang umum ditemukan pada aplikasi yang memiliki struktur *client-server* dimana data yang dikirimkan tidak menggunakan enkripsi SSL/TLS.

4. *Insecure Authentication* [M4]
   Kerentanan ini merupakan kerentanan yang terjadi akibat lemahnya prosedur autentikasi serta pengaturan sesi login. Autentikasi pada aplikasi *mobile* dapat bekerja ketika *offline* sehingga dapat diekploitasi oleh penyerang untuk melewati protokol autentikasi.

5. *Insufficient Cryptography* [M5]
   Kerentanan ini merupakan kerentanan akibat lemahnya kriptografi yang lemah serta tidak amannya algoritma kriptografi yang digunakan sehingga penyerang dapat dengan mudah mendekripsikan informasi yang diperoleh.

6. *Insecure Authorization* [M6]
   Kerentanan ini merupakan kerentanan akibat gagalnya suatu server dalam menerapkan identitas serta izin yang benar. Perbedaannya dengan *Inseruce Authentication* adalah jika *Insecure Authentication* mengacu pada pengguna yang mengelabui proses autentikasi, maka *Insecure Authorization* lebih ke kegagalan server pada aplikasi dalam menentukan identitas serta izin.

7. *Client Code Quality* [M7]
   Kerentanan ini merupakan kerentanan akibat kesalahan dalam pembuatan kode aplikasi, dimana hal ini dapat dimanfaatkan untuk melakukan eksploitasi serta berpotensi terjadinya *bypass* kontrol keamanan oleh penyerang.

8. *Code Tampering* [M8]
   Kerentanan in merupakan kerentanan dimana penyerang melakukan modifikasi pada kode aplikasi yang memungkinkan untuk membuat *backdoor* pada aplikasi.

9. *Reverse Engineering* [M9]
   Kerentanan ini merupakan kerentanan yang terkait dengan algoritma enkripsi yang digunakan untuk mencari informasi cara kerja server *back-end*.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

10. *Extraneous Functionality* [M10]
    Kerentanan ini merupakan kerentanan dimana ditemukannya *backdoor* atau bug pada aplikasi yang sengaja dibuat pada saat proses pengembangan namun tidak dihapus ketika masuk produksi sehingga dapat dimanfaatkan penyerang untuk masuk menggunakan *backdoor* tersebut.

b. Mobile Security Framework (MobSF)
Mobile Security Framework atau MobSF merupakan suatu framework yang biasa digunakan untuk melakukan uji penetrasi terhadap aplikasi smartphone secara otomatis, dimana framework ini dapat melakukan pengujian secara statis maupun dinamis serta malware [4]. MobSF dapat digunakan untuk melakukan pengujian keamanan suatu aplikasi binari *Android Package Kit (APK)* dan *iPhone Application (IPA)* serta kode sumber zip. Selain itu MobSF juga dapat melakukan pengujian aplikasi secara dinamis pada saat *runtime* dengan menggunakan metode *fuzzing* [5].

2.2. Analisis Kebutuhan

Kebutuhan Non-Fungsional :

a. Hardware (Perangkat-Keras)
   1. AMD Ryzen 7 4800H
   2. NVIDIA GeForce GTX 1650 Ti
   3. SSD 512GB
   4. RAM 8GB

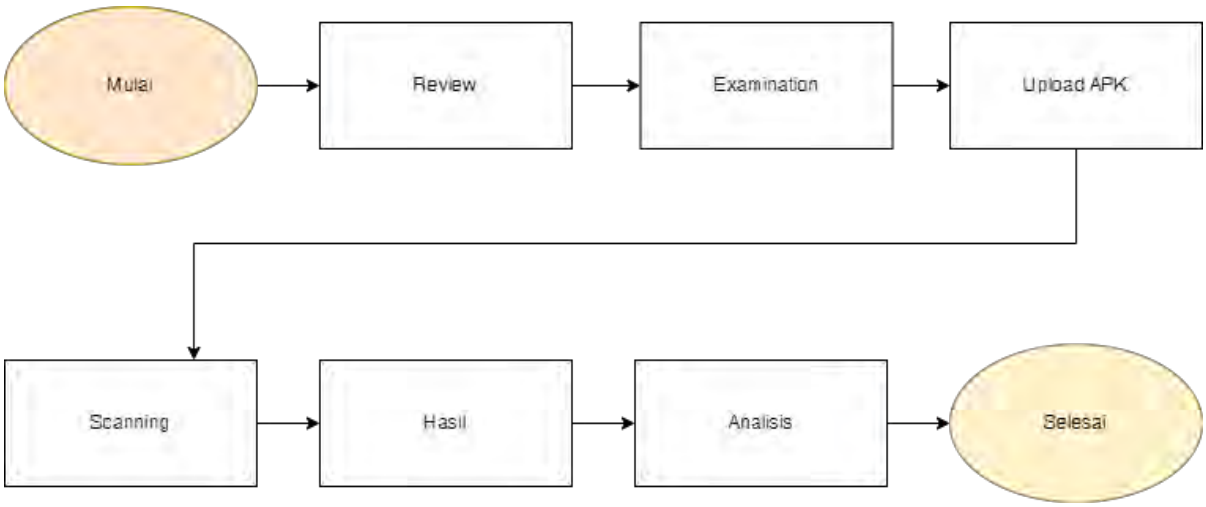
b. Software (Perangkat Lunak)
   1. Windows 10
   2. Mobile Security Framework (MobSF)
   3. Mozilla Firefox
   4. Docker Desktop

Kebutuan Fungsional :

a. Kemampuan meng-ekstraksi file .apk aplikasi
b. Kemampuan menganalisis kerentanan pada aplikasi

2.3. Rancangan Analisis Sistem

Penelitian ini menggunakan metode analisis statik yang dilakukan menggunakan aplikasi Mobile Security Framework (MobSF). Dimana pertama akan dilakukan review mengenai aplikasi seperti izin aplikasi, kemudian berlanjut ke proses examination dimana aplikasi Udayana Mobile akan diupload ke MobSF selanjutnya melalui proses scanning dan hasil scanningnya akan digunakan untuk melakukan analisis kerentanan sesuai dengan OWASP *Mobile Top 10 2016*. Adapun tahapannya dapat dilihat pada **Gambar 1.**



Gambar 1. Flowchart pengujian

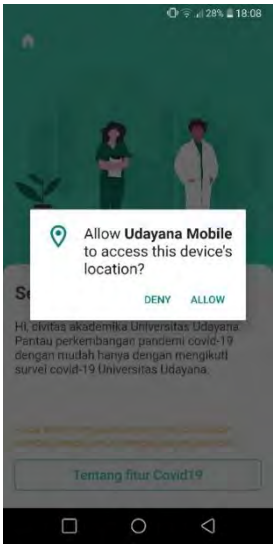
Seperti yang ditampilkan pada **Gambar 1.** Tahapan pengujuan dimulai dari review aplikasi oleh penulis, kemudian berlanjut ke tahap examination dimana aplikasi Udayana Mobile diupload pada MobSF kemudian akan dilakukan proses *scanning*. Setelah proses *scanning* selesai maka hasilnya akan muncul, terakhir akan dilakukan analisis keamanan aplikasi menggunakan hasil *scanning* tersebut dengan mengacu kepada OWASP *Mobile Top 10 2016.*

3. Hasil dan Pembahasan

3.1. Review

Aplikasi Udayana Mobile merupakan aplikasi yang dikembangkan oleh Unit Sumber Daya Informasi Universitas Udayana. Aplikasi ini dapat memudahkan pegawai, dosen, dan mahasiswa mengakses beberapa layanan yang dibutuhkan dalam menunjang kebutuhan akademik di Universitas Udayana seperti SIRAISA dan SIPENA bagi dosen dan pegawai, serta Sistem Informasi Manajemen Akademik (SIMAK) dan UKT-Ku bagi mahasiswa

Gambar 2. Dashboard Udayana Mobile



Gambar 3. Izin akses lokasi

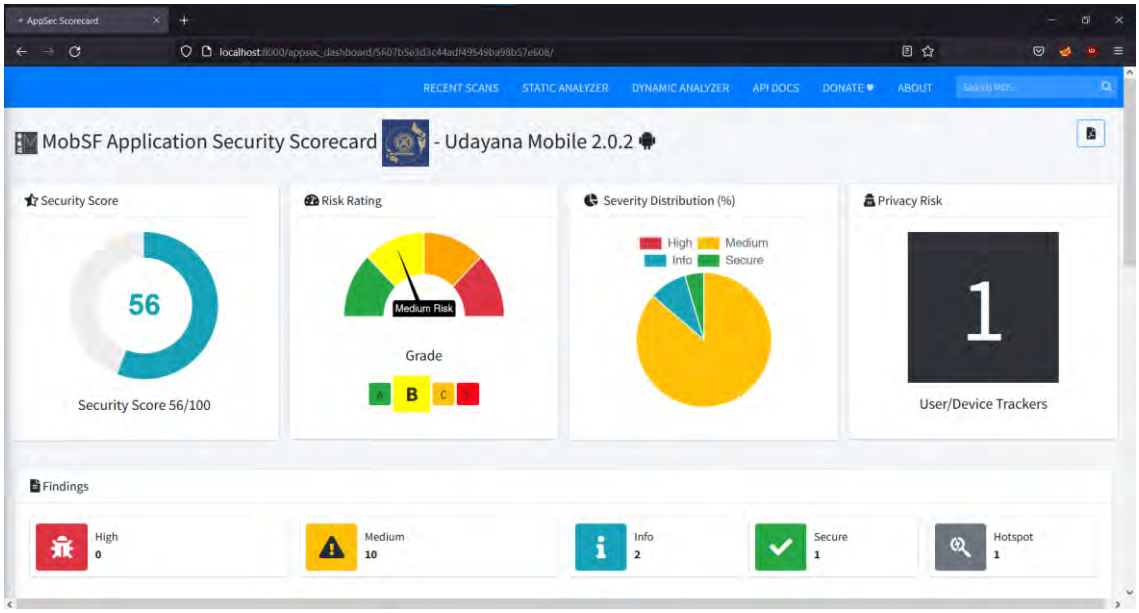
Karena aplikasi ini terintegritas dengan sistem *Integrated Management Information System, the Strategic of UNUD* (IMISSU) milik Universitas Udayana, maka data data pribadi dari pengguna juga disimpan pada aplikasi ini.

Aplikasi Mobile Udayana memiliki izin penggunaan aplikasi, namun tidak langsung meminta izin pengguna ketika diinstall, melainkan permintaan izin muncul ketika mengakses menu tertentu seperti yang ditunjukan pada **Gambar 3.**. Adapun persrtujuan terkait izin untuk lokasi, yakni Izin untuk mengakses lokasi pengguna secara akurat atau perkiraan lokasi.

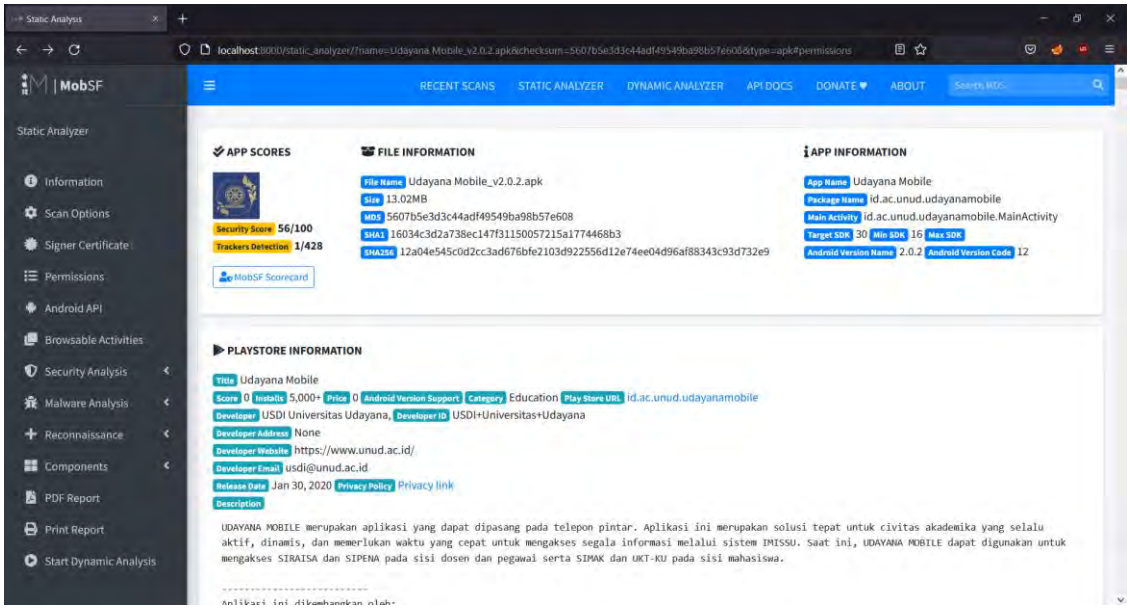
3.2. Examination

Proses *Examination* terdiri dari Upload file APK Udayana Mobile ke MobSF kemudian melalui proses scanning dan hasil dari scanning tersebut akan keluar. Berikut merupakan hasil scanning dari aplikasi Udayana Mobile pada MobSF ditunjukan pada **Gambar 4.** dan **Gambar 5.**

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Udayana Mobile Scorecard



Gambar 5. Hasil Analisis Aplikasi Udayana Mobile

Tabel 1. Hasil Kerentanan yang Ditemukan

No	OWASP *Mobile Top 10 2016*	Hasil
[M1]	*Improper Platform Usage*	-
[M2]	*Insecure Data Storage*	-
[M3]	*Insecure Communication*	-
[M4]	*Insecure Authentication*	-
[M5]	*Insufficient Cryptography*	√
[M6]	*Insecure Authorization*	-
[M7]	*Client Code Quality*	√
[M8]	*Code Tampering*	-
[M9]	*Reverse Engineering*	√
[M10]	*Extraneous Functionality*	-

3.3. Analisis

Dari **Gambar 1**. Dapat dilihat bahwa aplikasi Udayana Mobile mendapatkan Security Score 56/100 dengan Risk Rating B. Kemudian ditemukan juga 10 *Severity* pada tingkat *Medium.*

Berdasarkan hasil kerentanan yang ditemukan seperti pada **Tabel 1.** Ditemukan 3 jenis kerentanan berdasarkan OWASP *Mobile Top 10 2016* pada aplikasi ini, meliputi :

a. *Insufficient Cryptography* [M5]
Kerentanan ini muncul sebagai peringatan akan adanya kerentanan kriptografi yang terdeteksi pada aplikasi Udayana Mobile. Kerentanan ini berupa penggunaan dari *hascode java* yang digunakan lemah sehingga ada kemungkinan dapat disalahgunakan oleh penyerang.

b. *Client Code Quality* [M7]
Kerentanan ini muncul sebagai peringatan akan adanya isu kerentanan *Cilent Code Quality* dimana aplikasi ini menggunakan database SQLite dan akan mengeksekusi *SQL Query* secara lansung tanpa adanya proses validasi terlebih dahulu. *SQL Query* yang diinput dapat menjadi celah dalam melakukan serangan SQL Injection. Kemudian informasi sensitif pada database tidak dilakukan enkripsi dan hanya disimpan begitu saja pada database.

c. *Reverse Engineering* [M9]
Kerentanan ini muncul sebagai peringatan akan adanya isu kerentanan *Reserve Engineering* dimana ada kemungkinan informasi sensitif seperti *username*, *password*, *keys*, dan lain lain dapat diambil dari aplikasi ini.

4. Kesimpulan

Berdasarkan penelitian ini, dapat disimpulkan Analisis Keamanan pada Aplikasi Udayana Mobile dilakukan secara statis menggunakan MobSF dimana hasilnya aplikasi ini memiliki Security Score yang terbilang cukup, yakni 56/100 dan juga memiliki Risk Rating B. Kemudian terdapat 10 kerentanan pada tingkat medium pada aplikasi ini. Mengacu pada OWASP *Mobile Top 10 2016* ditemukan isu kerentanan pada aplikasi ini diantaranya :

a. *Insufficient Cryptography*
Kerentanan ini merupakan kerentanan akibat lemahnya proses kriptografi yang digunakan pada aplikasi, yakni *java hashcode*. Direkomendasikan untuk melakukan enkripsi ekstra pada sumber kode.

b. *Client Code Quality*
Kerentanan ini muncul akibat penggunaan database SQLite dimana eksekusi query tidak terverifikasi terlebih dahulu sehingga memungkinkan terjadinya serangan SQL Injection. Direkomendasikan melakukan validasi pada query yang dimasukan.

c. *Reverse Engineering*
Kerentanan ini merupakan kerentanan yang cukup berbahaya karena informasi sensitif seperti data login pengguna atau bahkan data pribadi dapat diambil dari aplikasi ini. Direkomendasikan pada sumber kode menggunakan teknik *obfuscation* sehingga ketika terjadi *Reverse Engineering*, *string* yang dihasilkan tidak berupa plain text.

References

[1] N. V. Duc, P. T. Giang, and P. M. Vi, “Permission Analysis for Android Malware,” *Proc. 7th VAST - AIST Work. “RESEARCH Collab. Rev. Perspect.*, no. November 2015, pp. 207–216, 2016.

[2] Candra Kurniawan and N. Trianto, “Security Assessment Pada Aplikasi Mobile Android XYZ Dengan Mengacu Pada Kerentanan OWASP Mobile Top Ten 2016,” *Info Kripto*, vol. 15, no. 1, pp. 11–18, 2021, doi: 10.56706/ik.v15i1.2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

[3] G. Basatwar, “OWASP Mobile Top 10: A Comprehensive Guide For Mobile Developers To Counter Risks,” *21 September*, 2021. https://www.appsealing.com/owasp-mobile-top-10-a-comprehensive-guide-for-mobile-developers-to-counter-risks/ (accessed Sep. 30, 2022).
[4] F. Nurindahsari and B. Parga Zen, “Analisis Statik Keamanan Aplikasi Video Streaming Berbasis Android Menggunakan Mobile Security Framework (Mobsf),” *Cyber Secur. dan Forensik Digit.*, vol. 4, no. 2, pp. 63–80, 2022, doi: 10.14421/csecurity.2021.4.2.3373.
[5] A. Abraham, Magaofei, M. Dobrushin, and V. Nadal, “Mobile Security Framework MobSF.” https://github.com/MobSF/Mobile-Security-Framework-MobSF.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Serangan Application Layer Denial of Service Menggunakan Support Vector Machine (SVM) dan Chi Square

Putu Agus Prawira Dharma Yuda[a1], Cokorda Pramartha[ab2], I Komang Ari Mogi[a3]

[a]Informatics Engineering, Faculty of Math and Science, University of Udayana
[b]Center for Interdisciplinary Research on the Humanities and Social Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia
[1]agusprawira28@email.com
[2]cokorda@unud.ac.id
[3]arimogi@unud.ac.id

Abstract

In an era marked by widespread computer usage, security emerges as a critical focal point demanding meticulous attention. The spectrum of potential threats encompasses various methods of attacking computer systems, with Denial of Service (DoS) attacks being a prominent concern. This study delves into the enhancement of cybersecurity by implementing a system capable of discerning between DoS attack data and normal data, employing the Support Vector Machine (SVM) algorithm. To optimize the efficacy of the classification system, a strategic feature selection process is imperative. This research advocates for the utilization of the Chi-square method for this purpose, aiming to eliminate irrelevant features and thereby enhance system performance. The Support Vector Machine algorithm, hinging on hyperplanes for classification, gains efficiency through judicious feature selection. The empirical findings of this research unveil that employing Chi-square feature selection significantly elevates the performance of the classification system when dealing with application layer attacks. Remarkably, this enhancement is achieved without compromising the accuracy of the system. Specifically, the classification of DoS application layer attacks using SVM in tandem with Chi-square yielded identical accuracy results compared to using SVM alone. The average accuracy reached an impressive 99.9995%, with a processing time of 6.08 minutes with chi-square selection feature. In contrast, the classification system without feature selection consumed a comparatively longer processing time of 6.85 minutes. This underscores the efficacy of Chi-square feature selection in optimizing the performance of cybersecurity systems, demonstrating a streamlined approach to safeguarding computer networks from malicious threats.

Keywords: *Denial of Service (DoS), Support Vector Machine (SVM), Chi-square, Classification, Feature Selection*

1. Pendahuluan

Di zaman sekarang, peran komputer tidak bisa terlepas dari kehidupan kita sehari-hari. Seperti contoh, kita menggunakan komputer untuk melakukan aktivitas seperti bertransaksi, membuat tugas, menjelajah internet, mengirim pesan, membuat konten, melestarikan budaya, dan masih banyak lagi [1, 2]. Dengan adanya komputer, kehidupan kita menjadi lebih mudah karena komputer mampu memproses tugas dengan cepat. Dengan begitu maka komputer memiliki peran yang sangat penting dalam kehidupan sehari-hari. Dengan banyaknya pengguna komputer maka keamanan menjadi aspek yang sangat perlu diperhatikan. Menurut Badan Siber dan Sandi Negara (BSSN) yang dilansir oleh situs berita Media Indonesia, selama periode Januari hingga Mei 2021 telah terjadi serangan siber di Indonesia sebanyak 448 juta kasus [3]. Ada berbagai cara untuk melakukan serangan ke komputer. salah satunya yaitu dengan melakukan serangan DoS (Denial of Service). DoS merupakan jenis serangan yang terjadi pada komputer maupun server pada jaringan internet yang dimana seorang intruder menyerang dengan menghabiskan resource dari komputer atau server hingga komputer atau server tersebut tidak dapat menjalankan fungsinya. Serangan tersebut sangat mempengaruhi kinerja dari layanan komputer atau server internet seperti contoh membuka webpage menjadi sangat lamban atau kinerja komputer menjadi sangat berat akibat dari serangan tersebut [4].

Untuk mengatasi masalah tersebut maka diperlukan sebuah sistem yang dapat mengklasifikasikan data serangan Denial of Service dan data normal. Dari sistem klasifikasi serangan *application layer DoS* yang dibangun bertujuan untuk mendeteksi adanya potensi serangan *application layer DoS* sehingga dapat memperkecil potensi kerusakan yang terjadi. Ada banyak metode yang dapat digunakan. Salah satunya yaitu dengan menggunakan metode Support Vector Machine. Support Vector Machine merupakan salah satu algoritma yang menggunakan hyperplane untuk melakukan klasifikasi. Dan untuk meningkatkan performa dari sistem klasifikasi maka perlu dilakukan seleksi pada fitur yang digunakan untuk menghilangkan fitur yang dianggap tidak berpengaruh pada sistem sehingga performa sistem dapat meningkat. Salah satu metode yang dapat digunakan untuk seleksi fitur yaitu metode Chi-square. Dengan menggunakan kedua metode tersebut, penulis akan melakukan klasifikasi serangan *application layer DoS* dengan menggunakan *dataset* yang telah ditentukan serta menganalisa tingkat efektifitas dari sistem yang dibangun melalui performa yang dihasilkan (akurasi, *precision*, *recall, f1-score,* dan waktu proses) dalam melakukan klasifikasi data serangan *application layer DoS*.

2. Metode Penelitian

2.1 Pengumpulan Data

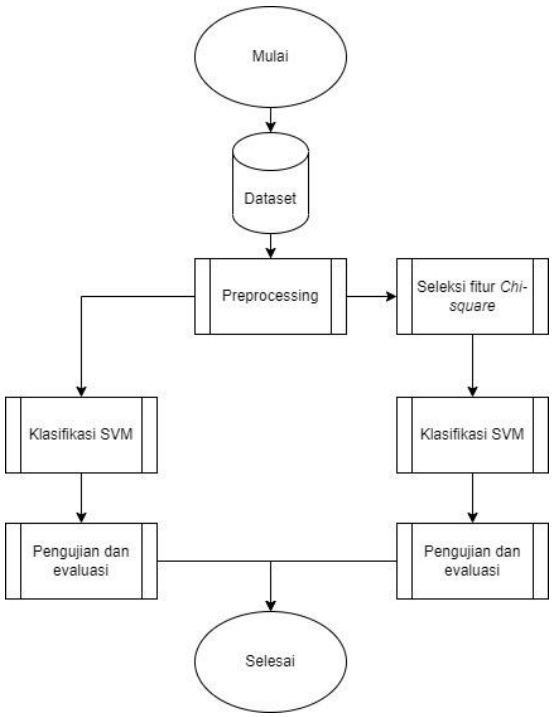
Dalam penelitian ini akan menggunakan dataset CSE-CIC-IDS2018 karena CSE-CIC-IDS2018 merupakan dataset serangan DoS paling terbaru sehingga hasil penelitian menjadi lebih relevan. Dataset CSE-CIC-IDS2018 memiliki jumlah fitur sebanyak 80 fitur yang dimana jumlah data serangan DoS yaitu sebanyak 601.802 data dari total keseluruhan data sebanyak 1.048.574. pada tabel di bawah ini merupakan rincian dari label data pada dataset.

Tabel 1. Rincian Data Target

	Normal/Benign	**DoS**	
		Hulk	**SlowHTTP**
	446772	461912	139890
Total	446772	601802	

2.2 Alur Penelitian

Alur dari penelitian dari masing-masing seleksi fitur yang akan digunakan yang digambarkan melalui flowchart berikut.



Gambar 1. Alur Penelitian

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tahap pertama pada yang akan dilakukan pada penelitian ini adalah mengumpulkan data yang digunakan untuk penelitian. Pada penelitian ini menggunakan data dari CSE-CIC-IDS2018 yang kemudian dikonversi ke format *.csv. Setelah mengumpulkan data kemudian dilanjutkan dengan tahap preprocessing yang bertujuan untuk menghilangkan data yang tidak relevan dan menyetarakan format data pada masing-masing fitur sehingga mudah diproses oleh program. Selanjutnya data secara terpisah dilakukan seleksi fitur dengan metode Chi-square yang nanti akan dipilih fitur yang memiliki nilai $\alpha < 0{,}05$. Kemudian dilanjutkan dengan proses klasifikasi menggunakan SVM pada masing-masing data yang sudah di seleksi fitur. Dan yang terakhir dilakukan proses pengujian beserta dengan evaluasi atas hasil pengujian yang didapatkan.
Untuk klasifikasi tanpa menggunakan seleksi fitur, alur penelitian hampir sama dengan alur penelitian dengan menggunakan seleksi fitur. Perbedaannya pada alur penilitian tersebut, setelah melakukan preprocessing langsung dilakukan proses klasifikasi.
Dari evaluasi yang dilakukan akan didapatkan output berupa tingkat performa dari hasil masing-masing metode yang digunakan yaitu tingkat akurasi dan waktu proses pada masing-masing metode.

2.3 Appilcation Layer Denial of Service

Denial of Service (DoS) merupakan sebuah tipe serangan pada komputer maupun server yang dimana penyerang menyiapkan komputer yang disebut dengan *botnet* yang dikendalikan oleh komputer server untuk membantu penyerangan terhadap komputer target yang menyebabkan komputer target menjadi lambat hingga sistem komputer menjadi *error* [5]. Ada berbagai tipe DoS salah satunya yaitu *Application Layer Denial of Service*. *Application Layer Denial of Service* adalah sebuah tipe serangan DoS yang mengeksploitasi semantik spesifik aplikasi dan pengetahuan domain untuk meluncurkan serangan DoS yang dimana sangat sulit bagi filter DoS untuk membedakan rangkaian permintaan serangan dengan rangkaian permintaan yang sah. Ada dua karakteristik yang membuat serangan *Application Layer Denial of Service* sangat merusak. Pertama, serangan *Application Layer Denial of Service* meniru permintaan dan karakteristik lalu lintas jaringan yang sama dengan permintaan klien yang sah, sehingga membuatnya sulit untuk dideteksi. Kedua, penyerang dapat menargetkan sumber daya yang lebih tinggi pada *server* seperti *socket, bandwidth disk, bandwidth database*, dan proses kerja [6].

2.4 Normalisasi Data

Normalisasi data adalah proses transformasi data yang dimana sebuah atribut diskalakan ke dalam rentang -1.0 hingga 1.0 atau dari 0.0 hingga 1.0. Normalisasi data memiliki berbagai jenis metode yang dapat digunakan, salah satunya dengan menggunakan *Min-max Normalization. Min-max Normalization* adalah metode normalisasi yang memetakan nilai v dari atribut A menjadi v' ke dalam rentang [new_min_A, new_max_A] berdasarkan rumus [7].

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

2.5 Seleksi Fitur Chi-square

Chi-square adalah salah satu seleksi fitur yang bertipe *supervised* yang dimana dapat menghilangkan fitur-fitur yang tidak relevan tanpa mengurangi tingkat akurasi [8]. Chi-square dapat dihitung dengan menggunakan persamaan.

$$x^2 = \sum_{i=1}^{r} \frac{(o_i - e_i)^2}{e_i} \quad (2)$$

Dimana

$$e_i = \frac{\Sigma\, baris . \Sigma kolom}{n} \quad (3)$$

Keterangan:
a. r = Jumlah baris
b. c = jumlah kolom
c. o_i = frekuensi observasi ke-i
d. e_i = frekuensi harapan ke-i
e. n = jumlah sampel

Adapun langkah-langkah dalam menggunakan metode *Chi-square* adalah sebagai berikut [9].
a. Merumuskan hipotesis H0 dan H1 yang dimana
    1. H0 = Kedua variabel tidak memiliki pengaruh yang signifikan
    2. H1 = Kedua variabel memiliki pengaruh yang signifikan
b. Tentukan nilai α (contoh $\alpha = 0.05$)
c. Hitung nilai df (*degree of freedom*) dengan persamaan.

$$df = (baris - 1)(kolom - 1) \quad (4)$$

d. Hitung nilai frekuensi harapan dengan persamaan (3)
e. Hitung nilai distribusi *chi-square* dengan persamaan (2)
f. Tentukan kriteria pengujian
   1. Jika nilai x^2 hitung ≤ x^2 tabel maka H0 diterima
   2. Jika nilai x^2 hitung > x^2 tabel maka H0 ditolak
   3. Jika nilai $\alpha \geq 0.05$ tabel maka H0 diterima
   4. Jika nilai $\alpha < 0.05$ tabel maka H0 ditolak

2.6 Support Vector Machine

Support Vector Machine (SVM) merupakan sebuah algoritma yang bertipe *supervised learning* yang digunakan untuk klasifikasi maupun regresi. SVM dikembangkan pertama kali oleh Boser, Guyon, dan Vapnik pada tahun 1992. Konsep SVM yaitu mencari *hyperplane* berukuran x-1 untuk memisahkan kedua buah kelas pada input space. Proses menemukan *hyperplane* dilakukan dengan memaksimalkan jarak antar kelas pada data yang digunakan [10]. SVM memiliki berbagai jenis kernel. Dalam penelitian ini kernel SVM yang digunakan yaitu SVM kernel linear. Adapun persamaan fungsi dari kernel SVM linear yaitu.

$$K\left(x_i,x_j\right) = x_i^T x_j \quad (5)$$

Adapun langkah-langkah dalam menggunakan metode SVM adalah sebagai berikut [11].

a. Lakukan *training* data hasil pembobotan dengan *sequential training* SVM.
b. Inisiasi parameter yang digunakan seperti λ, ε, γ, αi, dan C.
c. Hitunglah matriks Hessian dengan persamaan
   $$D_{ij} = y_i y_j (K(x_i x_j) + \lambda^2) \quad (6)$$
d. Lakukan iterasi perhitungan dari data ke-1 hingga data ke-n dengan persamaan berikut
   $$E_i = \sum_{j=1}^{n} a_j D_{ij} \quad (7)$$
   $$\delta ai = \min\{\max[\gamma(1 - E_i), -a_i], C - a_i\} \quad (8)$$
   $$\alpha i = \alpha_i + \delta\alpha_i \quad (9)$$
e. Dari perhitungan yang dilakukan sebelumnya, cari nilai αi yang terbesar dan dilanjutkan dengan melakukan perhitungan untuk menentukan bias dengan persamaan *lagrange* berikut.
   $$b = -\frac{1}{2}[(\sum_{i=1}^{n} a_i y_i K(x_i, x^-)) + (\sum_{i=1}^{n} a_i y_i K(x_i, x^+))] \quad (10)$$
f. Untuk mengetahui hasil klasifikasi, dilakukan testing dengan melakukan perhitungan fungsi f(x) yang didapatkan dengan persamaan berikut.
   $$f(x) = \sum_{i=0}^{n} a_i y_i K(x_i, x^-) + b \quad (11)$$

Keterangan:

a. αi = alfa, digunakan untuk mencari support vector
b. γ = gamma, untuk mengatur kecepatan *learning rate*
c. C = variabel slack
d. ε = epsilon, untuk pencari nilai error
e. D_{ij} = matriks *Hessian*
f. x_i = data ke-i
g. x_j = data ke-j
h. y_i = kelas data ke i
i. $K(x_i x_j)$ = fungsi kernel

2.7 K-Fold Cross Validation

K-Fold Cross Validation merupakan metode yang digunakan untuk menguji tingkat keberhasilan suatu sistem dengan melakukan pengujian berulang dengan mengacak atribut masukan [12]. *K-Fold Cross Validation* membagi data menjadi subdata yang saling bebas secara acak. Pada himpunan data yang dibangun masing-masing data memiliki (k-1) *fold* data sebagai data latih dan 1 *fold* data sebagai data uji [13]. Berikut merupakan contoh dari pembagian dataset ke dalam proses *5-Fold Cross Validation.*

Tabel 2. K-Fold Cross Validation

100 data					
	A = 20 data	B = 20 data	C = 20 data	D = 20 data	E = 20 data
Percobaan 1	A (Data testing)	B	C	D	E
Percobaan 2	A	B (Data testing)	C	D	E

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Percobaan 3	A	B	C (Data testing)	D	E
Percobaan 4	A	B	C	D (Data Testing)	E
Percobaan 5	A	B	C	D	E (Data Testing)

Adapun langkah-langkah *K-Fold Cross Validation* yaitu.

a. Data dibagi menjadi k bagian
b. Pada *fold* pertama, data bagian ke-1 menjadi data uji dan sisanya menjadi data latih yang kemudian menghitung akurasi dengan persamaan

$$Akurasi = \frac{\sum data\ uji\ yang\ benar\ diklasifikasi}{\sum total\ data} . 100\% \quad (12)$$

c. Lakukan hal yang serupa pada iterasi selanjutnya hingga mencapai fold ke-k. Lalu hitung rata-rata akurasi.

2.8 Confusion Matriks

Confusion Matriks merupakan metode yang digunakan untuk menguji tingkat akurasi sebuah sistem klasifikasi dengan menggunakan tabel matriks untuk menghitung tingkat akurasi [14].

Tabel 3. Confusion Matriks

		Sebenarnya	
		Positif	Negatif
Prediksi	Positif	TP	FN
	Negatif	FP	TN

Selain menghitung tingkat akurasi, *confusion matriks* juga dapat digunakan untuk menghitung nilai *precision, recall, dan F1-score*. Berikut merupakan rumus-rumus untuk menghitung akurasi, *precision, recall, dan F1-score*.

$$Akurasi = \frac{TP+TN}{\sum jumlah\ data} . 100\% \quad (13)$$

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

$$Fmeasure = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} \quad (16)$$

3. Hasil dan Pembahasan

Berikut merupakan hasil klasifikasi serangan *application layer* menggunakan SVM baik dengan seleksi fitur *chi-square* maupun tanpa seleksi fitur.

Tabel 4. Hasil Klasifikasi SVM Full Fitur dengan Gamma = 0.01

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**7.17**
0.1	0.999992847	0.999987538	1	0.999993769	10.04
1	0.999985695	0.999975077	1	0.999987538	16.43

Tabel 5. Hasil Klasifikasi SVM Full Fitur dengan Gamma = 0.1

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**6.89**
0.1	0.999992847	0.999987538	1	0.999993769	9.46
1	0.999985695	0.999975077	1	0.999987538	16.6

Tabel 6. Hasil Klasifikasi SVM Full Fitur dengan Gamma = 1

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**6.85**
0.1	0.999992847	0.999987538	1	0.999993769	10.31
1	0.999985695	0.999975077	1	0.999987538	19

Tabel 7. Hasil Klasifikasi SVM Chi-square dengan Gamma = 0.01

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**7.04**
0.1	0.999992847	0.999987538	1	0.999993769	8.39
1	0.999985695	0.999975077	1	0.999987538	14.98

Tabel 8. Hasil Klasifikasi SVM Chi-square dengan Gamma = 0.1

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**6.03**
0.1	0.999992847	0.999987538	1	0.999993769	8.51
1	0.999985695	0.999975077	1	0.999987538	15.54

Tabel 9. Hasil Klasifikasi SVM Chi-square dengan Gamma = 1

Lambda	Rata-rata	Precision	Recall	F1-scores	Waktu proses (menit)
0.01	**0.999995232**	**0.999991692**	**1**	**0.999995846**	**6.76**
0.1	0.999992847	0.999987538	1	0.999993769	9.91
1	0.999985695	0.999975077	1	0.999987538	15.45

Dari hasil pengujian didapatkan hasil klasifikasi yang dimana tingkat akurasi yang dihasilkan selalu sama mengikuti parameter lambda yang digunakan baik perbedaan parameter gamma maupun perbedaan pada dataset yang digunakan. Sebagai contoh hasil akurasi untuk klasifikasi data serangan DDoS dengan SVM berparameter lambda 0.01 akan selalu menghasilkan tingkat akurasi 0.999995232 atau 99.9995% meskipun parameter gamma yang digunakan berbeda dan itu berlaku pada kedua jenis dataset baik dataset yang full fitur maupun dataset hasil seleksi chi-square.
Kemudian untuk pengaruh parameter terhadap performa sistem klasifikasi serangan DoS didapatkan kesimpulan bahwa parameter gamma mempengaruhi performa system dalam hal waktu proses yang dimana semakin besar nilai gamma memiliki kecenderungan semakin lama waktu proses yang dilakukan pada penggunaan parameter lambda yang sama. Sebagai contoh pada klasifikasi serangan DoS dengan seleksi fitur pada parameter $\lambda = 0.1$ didapatkan waktu proses yaitu pada $\gamma = 0.01$ memerlukan waktu 8.39 menit, $\gamma = 0.1$ memerlukan waktu 8.51 menit, dan $\gamma = 0.01$ memerlukan waktu 9.91 menit. Sedangkan parameter lambda mempengaruhi performa system dalam hal hasil klasifikasi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

dan waktu proses yang dimana semakin besar lambda maka semakin kecil tingkat akurasi yang dihasilkan serta waktu proses yang diperlukan juga semakin melambat.
Untuk pemilihan model parameter yang akan digunakan untuk perbandingan klasifikasi dipilih berdasarkan hasil klasifikasi yang dihasilkan (nilai akurasi rata-rata, precision, recall, f1-scores) serta tingkat efisiensi pada waktu proses sistem. Pada klasifikasi SVM dengan dataset full fitur menggunakan model pengujian dengan parameter γ (gamma) sebesar 1 dan λ (lambda) sebesar 0.01. Hali itu dikarenakan model tersebut menghasilkan tingkat akurasi terbaik yaitu 99.9995% dengan waktu yang paling efisien yaitu 6.85 menit. Sementara itu Pada klasifikasi SVM dengan dataset hasil seleksi fitur menggunakan model pengujian dengan parameter γ (gamma) sebesar 0.1 dan λ (lambda) sebesar 0.01. Hali itu dikarenakan model tersebut menghasilkan tingkat akurasi terbaik yaitu 99.9995% dengan waktu yang paling efisien yaitu 6.03 menit. Berikut merupakan tabel dari hasil perbandingan klasifikasi SVM pada data serangan application layers DDoS baik menggunakan full fitur maupun hasil seleksi fitur menggunakan chi-square.

Tabel 10. Perbandingan Hasil Klasifikasi Serangan DoS dengan full fitur dan chi-square

Tipe Klasifikasi	Rata-rata	Precision	Recall	F1-scores	Waktu (menit)
SVM Full Fitur	0.999995232	0.999991692	1	0.999995846	6.85
SVM Chi Square	0.999995232	0.999991692	1	0.999995846	6.03

Dari tabel diatas, dapat disimpulkan bahwa klasifikasi SVM dengan seleksi fitur Chi-square mampu meningkatkan tingkat efisiensi waktu proses pada sistem klasifikasi data serangan application layers DoS tanpa mengurangi tingkat akurasi, precision, recall, dan f1-scores. Hal itu dapat dilihat dengan hasil klasifikasi yang sama yaitu nilai akurasi rata-rata sebesar 0.999995232 atau 99.9995%, precision sebesar 0.999991692, recall sebesar 1, serta nilai f1-scores sebesar 0.999995846. Untuk waktu proses, metode klasifikasi SVM dengan seleksi fitur chi-square menjalankan proses klasifikasi lebih cepat dari Klasifikasi SVM dengan full fitur yang dimana waktu proses SVM dengan seleksi fitur chi-square memerlukan waktu 6.03 menit sedangkan untuk SVM dengan full fitur menggunakan waktu 6.85 menit.

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan dapat dimuat kesimpulan yaitu.

a. Perbandingan performa yang dihasilkan antara kedua tipe klasifikasi serangan *DoS* baik dengan SVM full fitur maupun SVM dengan seleksi *chi-square* pada pengaruh parameter memiliki kesamaan yaitu parameter gamma mempengaruhi performa system dalam hal waktu proses yang dimana semakin besar nilai *gamma* memiliki kecenderungan semakin lama waktu proses yang dilakukan pada penggunaan parameter *lambda* yang sama. Sedangkan parameter lambda mempengaruhi performa system dalam hal hasil klasifikasi dan waktu proses yang dimana semakin besar nilai lambda yang digunakan maka semakin kecil tingkat akurasi yang dihasilkan serta waktu proses yang diperlukan juga semakin melambat.
b. Penggunaan metode seleksi fitur chi-square pada klasifikasi serangan application layer DoS tidak mempengaruhi hasil klasifikasi pada sistem. Dimana hasil klasifikasi serangan application layer DoS menggunakan SVM dan chi-square memiliki hasil akurasi yang sama dengan hasil klasifikasi serangan application layer DoS dengan hanya menggunakan SVM yaitu nilai akurasi rata-rata sebesar 0.999995232 atau 99.9995%, precision sebesar 0.999991692, recall sebesar 1, serta nilai f1-scores sebesar 0.999995846.
c. Klasifikasi serangan application layer DoS menggunakan SVM dan chi-square dapat meningkatkan tingkat efisiensi pada waktu proses pada system yang dimana waktu proses SVM dengan seleksi fitur chi-square memerlukan waktu 6.03 menit sedangkan untuk SVM dengan full fitur menggunakan waktu 6.85 menit.

Referensi

[1] C. Pramartha, I. Koten, I. G. N. A. C. Putra, I. W. Supriana, and I. W. Arka, “Pengembangan Sistem Dokumentasi Melalui Pendekatan Ontologi untuk Praktek Budaya Bali,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 259–268, Dec. 2022, doi: 10.23887/janapati.v11i3.53939.

[2] C. Pramartha, I. G. N. A. C. Putra, I. P. G. H. Suputra, and I. W. Arka, "Adopsi dan Pelatihan Penggunaan Perangkat Digital Papan Tombol Aksara Bali," *Jurnal Widya Laksmi*, vol. 3, no. 1, pp. 14–20, 2023.

[3] P. Ananda, "Serangan Siber di RI Terus Meningkat, Capai 448 Juta Kasus." Accessed: Dec. 16, 2021. [Online]. Available: https://mediaindonesia.com/politik-dan-hukum/414225/serangan-siber-di-ri-terus-meningkat-capai-448-juta-kasus

[4] Y. W. Pradipta, "Implementasi Intrusion Prevention System (Ips) Menggunakan Snort Dan Ip Tables Berbasis Linux," *Jurnal Manajemen Informatika*, vol. 7, no. 1, 2017.

[5] I. B. G. Amartya, I. M. Widiartha, I. G. A. G. A. Kadnyanan, I. G. N. A. C. Putra, I. P. G. H. Suputra, C. Pramartha, "Implementasi Algoritma Naive Bayes Classifier (NBC) Dan Information Gain Untuk Mendeteksi DDoS," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 11, no. 2, pp. 273–282, 2022, [Online]. Available: https://research.unsw.edu.au/projects/unsw-nb15-dataset.

[6] M. Srivatsa, A. Iyengar, J. Yin, and L. Liu, "Mitigating application-level denial of service attacks on Web servers: A client-transparent approach," *ACM Transactions on the Web*, vol. 2, no. 3, Jul. 2008, doi: 10.1145/1377488.1377489.

[7] H. Junaedi, H. Budianto, I. Maryati, dan Y. Melani, "Data Transformation pada Data Mining," *Prosiding Konferensi Nasional Inovasi dalam Desain dan Teknologi-IDeaTech*, vol. 7, pp. 93–99, 2011.

[8] L. Luthfiana, J. C. Young, dan A. Rusli, "Implementasi Algoritma Support Vector Machine dan Chi Square untuk Analisis Sentimen User Feedback Aplikasi," *Ultimatics : Jurnal Teknik Informatika*, vol. 12, no. 2, pp. 125–126, 2020, doi: 10.31937/ti.v12i2.1828.

[9] I. C. Negara and A. Prabowo, "Penggunaan Uji Chi–Square untuk Mengetahui Pengaruh Tingkat Pendidikan dan Umur terhadap Pengetahuan Penasun Mengenai HIV–AIDS di Provinsi DKI Jakarta," *Prosiding Seminar Nasional Matematika dan Terapannya 2018*, pp. 1–8, 2018.

[10] H. E. Wahanani, B. Nugroho, dan G. I. Prakoso, "Analisa Serangan Smurf Dan Ping of Death Dengan Metode Support Vector Machine (Svm)," *Jurnal Teknologi Informasi dan Komunikas*, vol. 11, pp. 71–76, 2016.

[11] C. C. Aggarwal, "Mining Text Data," *Data Mining*, pp. 429–455, 2015, doi: 10.1007/978-3-319-14142-8_13.

[12] M. A. Banjarsari, H. I. Budiman, dan Farmadi, "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4," *Klik - Kumpulan Jurnal Ilmu Komputer*, vol. 2, no. 2, pp. 159–173, 2015.

[13] Suyanto, "Data Mining untuk Klasifikasi dan Klusterisasi Data," Informatika Bandung. Accessed: Dec. 10, 2021. [Online]. Available: https://scholar.google.co.id/scholar?hl=id&as_sdt=0%2C5&q=data+mining+untuk+klasifikasi+dan+klusterisasi+data+suyanto&btnG=

[14] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, pp. 5–10, 2014, [Online]. Available: www.bluefame.com,

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Sistem Monitoring Indek Kualitas Udara Berbasis Open Meteo Air Quality API

I Gede Surya Rahayuda[a1], Ni Putu Linda Santiari[a2]

[a]Program Studi Informatika, Fakultas MIPA, Universitas Udayana
Jalan Kampus Bukit Jimbaran, Bali, Indonesia
[1]igedesuryarahayuda@unud.ac.id (Corresponding author)

[b]Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, ITB STIKOM Bali
Jalan Raya Puputan No. 86 Renon, Denpasar, Bali, Indonesia
[2]linda_santiari@stikom-bali.ac.id

Abstract

This research develops an air quality monitoring system utilizing the Open Meteo API and OpenStreetMap API, employing the Software Development Life Cycle (SDLC). Users input latitude and longitude to access the air quality index (AQI) at a specific location, with a default one-day hourly forecast. Configuration options include location and forecast duration. The system successfully determines AQI values based on user settings and employs color-coded visual representations following US Air Quality Index standards for easier comprehension of health risks. The SDLC ensures a streamlined development process. Additionally, the OpenStreetMap API aids in geographic coordinate-based location identification. Usability evaluation, conducted through Google Analytics and a Google Form questionnaire, indicates high user satisfaction, particularly regarding information clarity, content completeness, and overall experience. Future enhancements may target optimizing page loading speed and increasing user engagement. In summary, this research contributes to accessible and understandable air quality monitoring, benefiting users seeking reliable AQI data.

Keywords: AQI, API, Open Meteo, Vercel, GitHub, Serverless Architecture.

1. Pendahuluan

Dalam era modern yang dipenuhi inovasi teknologi, pemantauan kualitas udara menjadi semakin penting untuk menjaga kesehatan dan kesejahteraan masyarakat. Kualitas udara yang buruk dapat memiliki dampak serius terhadap kesehatan manusia, terutama bagi kelompok rentan seperti anak-anak, lansia, dan individu dengan masalah pernapasan. Oleh karena itu, pengembangan sistem monitoring kualitas udara menjadi suatu kebutuhan. Penelitian ini mengembangkan solusi yang inovatif dengan mengintegrasikan Open Meteo API dan OpenStreetMap API untuk memberikan informasi kualitas udara berbasis lokasi. Melalui penggunaan data latitude dan longitude, sistem ini memungkinkan pengguna untuk dengan mudah melihat indeks kualitas udara (AQI) pada suatu tempat tertentu [1] [2]. Keunggulan sistem ini tidak hanya terletak pada kemampuannya memberikan informasi secara akurat, tetapi juga pada fleksibilitasnya dalam mengizinkan pengguna untuk mengonfigurasi lokasi dan masa depan. Dengan demikian, pengguna dapat dengan lebih tepat merencanakan aktivitas mereka berdasarkan kualitas udara di suatu wilayah. Selain itu, sistem ini menggunakan arsitektur serverless yang memiliki keunggulan efisiensi biaya, scalability, dan reliability. Efisiensi biaya karena sistem hanya perlu membayar biaya untuk sumber daya yang digunakan. Scalability karena sistem dapat dengan mudah diskala untuk menangani jumlah data yang lebih besar atau beban kerja yang lebih berat. Reliability karena sistem memiliki tingkat keandalan yang tinggi. Terakhir, sistem ini dapat diintegrasikan dengan GitHub untuk memudahkan pengembangan dan pengelolaan sistem. Dimana beberapa kelebihan tersebut merupakan suatu pengembangan dari hasil penelitian sebelumnya. Penggunaan Software Development Life Cycle (SDLC) dalam pengembangan sistem ini memberikan dasar yang kokoh, memastikan bahwa proses pengembangan dilakukan dengan metode yang terstruktur dan terorganisir. Sistem ini juga mengadopsi standar United States Air Quality Index (AQI) untuk kategorisasi hasil pengukuran, memberikan informasi yang mudah dimengerti oleh pengguna

melalui representasi visual menggunakan warna. Dengan melibatkan OpenStreetMap API, sistem ini juga meningkatkan kemampuan identifikasi lokasi berdasarkan koordinat geografis. Inovasi ini memberikan dimensi tambahan pada pengalaman pengguna, menjadikan sistem ini tidak hanya sebagai alat pemantauan kualitas udara, tetapi juga sebagai panduan geografis yang komprehensif. Melalui pendekatan yang holistik, penelitian ini diharapkan dapat memberikan kontribusi positif dalam memahami dan mengatasi permasalahan kualitas udara di berbagai wilayah. Dengan menggunakan teknologi terkini, sistem monitoring ini tidak hanya menjadi alat penting bagi individu, tetapi juga untuk otoritas kesehatan dan lingkungan dalam pengambilan keputusan yang berbasis data [3] [4].

2. Tinjauan Pustaka

Berikut analisis review pada beberapa artikel terkait penelitian yang dilakukan. Artikel "**A Review on Air Quality Indexing System**" dalam Asian Journal of Atmospheric Environment oleh Kanchan, Amit Kumar Gorai, dan Pramila Goyal pada tahun 2015 membahas pentingnya Indeks Kualitas Udara (AQI) atau Air Pollution Index (API) dalam memberikan informasi tentang tingkat keparahan pencemaran udara kepada masyarakat. Meskipun terdapat berbagai metode yang dikembangkan oleh peneliti dan lembaga lingkungan hidup, kekurangan metodologi yang diterima secara universal terjadi. Makalah tersebut menyajikan tujuan utama AQI atau API, yaitu mengidentifikasi wilayah dengan kualitas udara buruk dan memberikan informasi kepada masyarakat tentang tingkat keparahan paparannya. Indeks ini dapat dikategorikan sebagai polutan tunggal atau multipolutan, masing-masing dengan kekuatan dan kelemahan sendiri. Penelitian tersebut bertujuan memberikan tinjauan komprehensif terhadap indeks kualitas udara di seluruh dunia [5]. Sementara itu, makalah "**A Case Study of Web API Evolution**" yang dipresentasikan pada IEEE World Congress on Services 2015 oleh S.M. Sohan, Craig Anslow, dan Frank Maurer, membahas tantangan integrasi aplikasi menggunakan API web. Artikel tersebut menyoroti potensi perubahan dalam API web yang dapat mengganggu aplikasi yang bergantung, mencakup kurangnya dukungan untuk versi API lama, dokumentasi yang tidak memadai, dan komunikasi perubahan yang tidak efektif. Melalui studi kasus evolusi API web, penelitian ini memberikan rekomendasi bagi praktisi dan peneliti mengenai profil perubahan API, strategi versi, serta pendekatan dokumentasi dan komunikasi yang efektif [6]. Selanjutnya, "**Serverless Architecture - A Revolution in Cloud Computing**" oleh R. Arokia Paul Rajan pada Tenth International Conference on Advanced Computing (ICoAC) 2018 memberikan gambaran komprehensif tentang komputasi tanpa server. Artikel tersebut membahas prinsip dasar komputasi tanpa server, membandingkan platform utama seperti AWS Lambda, Google Cloud Functions, Microsoft Azure Functions, dan IBM Cloud Functions, serta mengeksplorasi kasus penggunaan dunia nyata. Artikel ini juga menyoroti tantangan teknis dan organisasi, memberikan wawasan tentang tren masa depan, dan mengakui potensi dampak komputasi tanpa server pada pengembangan perangkat lunak di berbagai industri. Artikel ini menyajikan konten yang terstruktur dengan baik dan bermanfaat bagi para peneliti, pengembang, dan organisasi yang tertarik pada komputasi tanpa server [7]. Secara keseluruhan, ketiga jurnal tersebut memberikan kontribusi yang signifikan dalam pengembangan teknologi atau metode Sistem Monitoring Indeks Kualitas Udara berbasis Open Meteo Air Quality API.

3. Metode Penelitian

Dalam pelaksanaannya, penelitian ini mengaplikasikan pendekatan Software Development Life Cycle (SDLC) sebagai kerangka kerja yang terstruktur untuk memastikan kelancaran pengembangan sistem. Proses SDLC mencangkup serangkaian tahapan, mulai dari perencanaan yang komprehensif, analisis mendalam, desain yang terfokus, implementasi yang cermat, hingga pemeliharaan sistem yang berkelanjutan. Pendekatan ini memberikan fondasi kokoh untuk mengelola setiap langkah pengembangan dengan disiplin dan kerapihan, sehingga diharapkan mampu mencapai tujuan penelitian dengan efektif dan efisien [8] [9].

3.1. Perencanaan

Dalam tahap perencanaan, penelitian ini mengidentifikasi kebutuhan sistem berbasis kualitas udara dan fungsionalitas pengguna dengan cermat. Langkah-langkah ini melibatkan penentuan sumber daya yang diperlukan, termasuk strategi penggunaan Open Meteo API dan OpenStreetMap API untuk memastikan akuisisi data yang tepat dan komprehensif. Proses ini diarahkan pada pemahaman yang mendalam terhadap spesifikasi sistem guna memenuhi kebutuhan pengguna serta memastikan integrasi yang efisien dengan sumber daya eksternal yang relevan.

3.2. Analisis

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Dalam fase analisis, penelitian ini melakukan evaluasi mendalam terhadap kebutuhan pengguna dengan tujuan menentukan fitur utama sistem. Proses ini melibatkan analisis kebutuhan pengguna, khususnya dalam konteks konfigurasi lokasi dan masa depan. Selain itu, tahap ini mencakup penyusunan spesifikasi sistem yang jelas dan merinci serta penentuan data yang dibutuhkan dari API eksternal, seperti Open Meteo dan OpenStreetMap. Hasil dari analisis ini akan menjadi landasan untuk perancangan sistem yang responsif dan sesuai dengan ekspektasi pengguna.

3.3. Desain

Dalam proses desain, penelitian ini mengadopsi pendekatan berbasis pseudocode untuk merinci struktur alur kerja sistem dengan lebih terperinci. Pseudocode digunakan sebagai representasi abstrak dari langkah-langkah dan logika program, memberikan deskripsi secara lebih rinci tentang implementasi setiap tahap, dari penerimaan input hingga penentuan kategori Air Quality Index (AQI). Pseudocode memberikan kejelasan dalam merancang logika sistem, memastikan bahwa setiap langkah diimplementasikan dengan akurat dan efisien sesuai dengan analisis kebutuhan pengguna. Dengan demikian, desain pseudocode ini menjadi dasar untuk implementasi yang berhasil dan sesuai dengan spesifikasi yang telah ditetapkan.

3.4. Implementasi

Dalam tahap implementasi, penelitian ini memanfaatkan HTML, CSS, dan JavaScript untuk mentransformasikan desain menjadi kode sumber yang fungsional. Penerapan ini juga melibatkan penggunaan Bootstrap untuk meningkatkan responsivitas dan estetika tampilan sistem. Selanjutnya, untuk mempublikasikan sistem secara online, dilakukan hosting web menggunakan platform Vercel yang terintegrasi dengan GitHub. Pendekatan ini tidak hanya memastikan ketersediaan sistem secara daring tetapi juga memudahkan pengelolaan kode sumber dan pembaruan melalui repositori GitHub. Dengan implementasi ini, diharapkan sistem dapat diakses dengan mudah dan memberikan pengalaman pengguna yang optimal.

3.5. Pengujian

Dalam fase pengujian, penelitian ini menerapkan pengujian black box, yang berfokus pada fungsionalitas sistem tanpa memperhatikan detail implementasi internal. Pengujian ini bertujuan untuk memverifikasi kemampuan sistem dalam memberikan hasil yang akurat sesuai dengan konfigurasi pengguna. Dengan memanfaatkan metode ini, pengujian dapat dilakukan dengan pendekatan yang lebih holistik, menguji berbagai kasus pengguna tanpa bergantung pada struktur internal sistem. Hasil dari pengujian black box ini menjadi indikator kehandalan dan kesesuaian sistem dalam menghadapi situasi nyata, sehingga memastikan kualitas dan performa yang diinginkan.

3.6. Pemeliharaan

Dalam tahap pemeliharaan (2.7), penelitian ini berfokus pada tindakan menjaga dan memperbarui sistem agar tetap relevan dan berkinerja optimal. Hal ini melibatkan pemantauan dan adaptasi terhadap perubahan pada API eksternal, seperti Open Meteo dan OpenStreetMap, untuk memastikan ketersediaan data yang akurat. Selain itu, respons terhadap umpan balik pengguna menjadi bagian integral dari pemeliharaan ini, dengan tujuan untuk meningkatkan pengalaman pengguna dan mengidentifikasi serta memperbaiki bug yang mungkin terjadi. Dengan pendekatan ini, sistem dapat terus berkembang sesuai dengan kebutuhan pengguna dan lingkungan eksternalnya.

4. Hasil dan Diskusi

Pengembangan sistem ini menggunakan pendekatan Software Development Life Cycle (SDLC) sebagai metodologi inti untuk memastikan bahwa seluruh proses pengembangan berjalan sesuai dengan tahapan yang terstruktur dan terorganisir. Dengan menerapkan SDLC, setiap fase dalam siklus pengembangan, mulai dari perencanaan, analisis, desain, implementasi, hingga pengujian dan pemeliharaan, dikelola dengan cermat untuk memastikan keberhasilan dan kehandalan sistem.

4.1. Perencanaan

Dalam tahap perencanaan penelitian ini menghasilkan identifikasi kebutuhan sistem yang didasarkan pada analisis kebutuhan pengguna, dengan fokus utama pada informasi kualitas udara dan fungsionalitas pengguna. Kebutuhan sistem mencakup kemampuan untuk melihat indeks kualitas udara (AQI) berdasarkan lokasi geografis, memungkinkan konfigurasi latitude dan longitude, serta

memberikan perkiraan cuaca dalam rentang waktu tertentu. Selanjutnya, hasil penentuan sumber daya menetapkan penggunaan Open Meteo API dan OpenStreetMap API sebagai sumber daya utama. Open Meteo API digunakan untuk mengakses data kualitas udara dan perkiraan cuaca, sedangkan OpenStreetMap API digunakan untuk mengidentifikasi nama tempat berdasarkan koordinat geografis. Integrasi kedua API ini menjadi fondasi sistem monitoring indeks kualitas udara untuk menyediakan informasi yang akurat dan relevan kepada pengguna.

4.2. Analisis

Berdasarkan perencanaan yang telah disusun, diperoleh suatu analisis yang menunjukkan bahwa sistem ini membutuhkan data array dari API Air Quality Open Meteo yang dapat diakses melalui penggunaan JavaScript. Data nama lokasi kemudian diperoleh berdasarkan data array Open Meteo, yang selanjutnya diolah menggunakan Open Street Map untuk memperoleh informasi lokasi yang lebih lengkap. Untuk mendukung fungsionalitas sistem, beberapa halaman yang esensial diperlukan, antara lain: index.html sebagai halaman utama, script.js untuk mengelola logika dan interaksi dengan API, serta style.css sebagai file stylesheet untuk penataan tampilan dan gaya halaman web. Dengan pengaturan ini, sistem diharapkan dapat beroperasi dengan baik, menyediakan informasi kualitas udara yang relevan, dan menampilkan hasilnya dengan tata letak yang bersih dan sesuai dengan desain yang diinginkan.

Klasifikasi tingkat AQI menggunakan United States Air Quality Index (AQI) memberikan informasi tentang kualitas udara di suatu daerah berdasarkan rentang nilai yang diperoleh. Rentang nilai ini dapat memberikan gambaran sejauh mana kualitas udara di suatu lokasi, dan masing-masing rentang tersebut memiliki kategori kesehatan yang berbeda. Berikut adalah penjelasan lebih lanjut tentang setiap kategori:

- **Good** (0-50): Rentang ini menunjukkan bahwa kualitas udara dianggap baik, dengan risiko polusi udara yang rendah atau hampir tidak ada. Udara dalam kategori ini dianggap aman untuk pernapasan oleh semua kelompok. Warna umumnya adalah hijau, mencerminkan kondisi yang aman dan baik untuk pernapasan.
- **Moderate** (51-100): Kategori ini menandakan bahwa kualitas udara adalah sedang, dan meskipun secara umum masih aman, beberapa polutan udara mungkin menyebabkan risiko kecil bagi individu yang sangat peka terhadap polusi udara. Warna umumnya adalah kuning atau kuning muda, menandakan tingkat peringatan yang lebih tinggi, meskipun masih dalam kisaran yang dapat diterima.
- **Unhealthy for Sensitive Groups** (101-150): Rentang ini mengindikasikan bahwa individu yang termasuk dalam kelompok sensitif, seperti mereka yang memiliki kondisi pernapasan atau jantung, anak-anak, dan orang dewasa yang lebih tua, mungkin mengalami efek kesehatan jika terpapar dalam waktu yang lama. Warna umumnya adalah oranye, memberikan peringatan terhadap risiko kesehatan bagi kelompok rentan.
- **Unhealthy** (151-200): Kategori ini menunjukkan bahwa kualitas udara menjadi tidak sehat, dan semua orang dapat mulai mengalami efek kesehatan. Individu yang termasuk dalam kelompok sensitif mungkin mengalami efek yang lebih serius. Warna umumnya adalah merah. Ini menunjukkan bahwa semua orang, terlepas dari kondisi kesehatan, mungkin mengalami efek yang lebih serius.
- **Very Unhealthy** (201-300): Rentang ini menggambarkan bahwa kualitas udara sangat tidak sehat, dan populasi umum dapat mulai mengalami efek kesehatan yang serius. Semua orang, terlepas dari kondisi kesehatan, dapat terpengaruh. Warna umumnya adalah ungu. Ini mencerminkan tingkat peringatan yang tinggi dan memerlukan tindakan pencegahan serius untuk melindungi kesehatan masyarakat.
- **Hazardous** (301-500): Kategori ini merupakan tingkatan paling tinggi dalam skala AQI dan menunjukkan kondisi kesehatan yang darurat. Kualitas udara dalam rentang ini dapat menyebabkan efek kesehatan yang serius pada semua orang. Tindakan darurat mungkin diperlukan. Warna umumnya adalah merah tua atau coklat, memberikan peringatan yang sangat serius dan menunjukkan adanya bahaya kesehatan yang mendesak. Tindakan darurat mungkin diperlukan [4].

4.3. Desain

Dalam fase desain sistem, pendekatan pseudocode berhasil diterapkan untuk merinci struktur dan logika alur kerja sistem. Pseudocode memberikan pandangan yang lebih terperinci terkait implementasi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

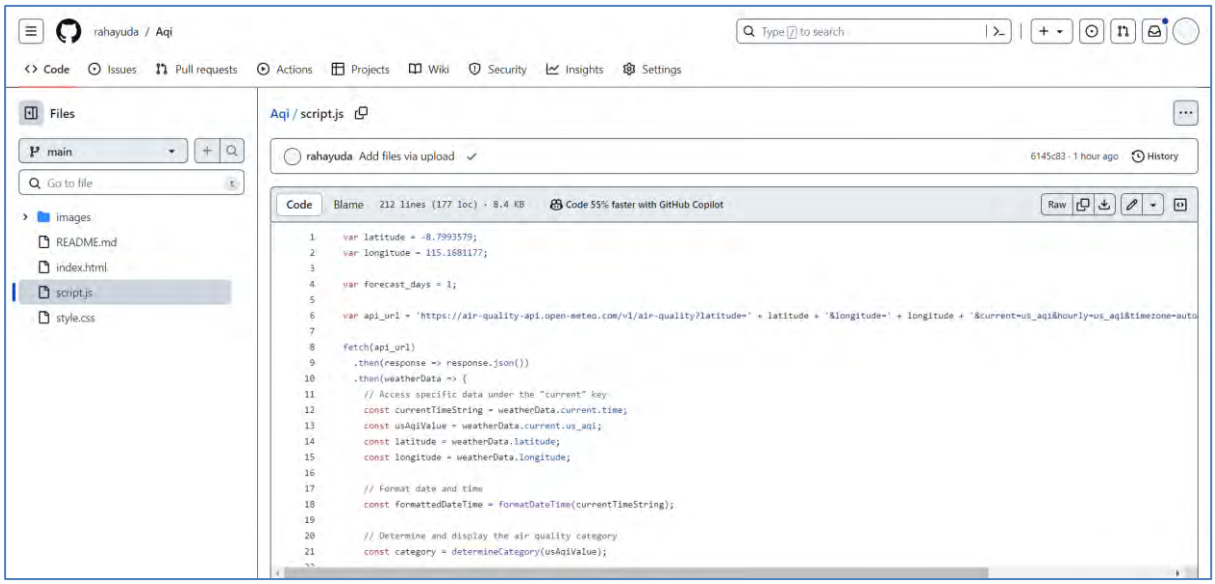
setiap langkah, mulai dari penerimaan input hingga penentuan kategori Air Quality Index (AQI). Hasil desain ini menciptakan landasan yang jelas dan terstruktur untuk implementasi selanjutnya, memastikan bahwa setiap aspek dari sistem diwujudkan dengan akurat dan efisien. Arsitektur serverless, sebagai pendekatan infrastruktur yang dapat diintegrasikan, memberikan landasan yang ideal untuk menerapkan pseudocode tersebut. Dalam arsitektur serverless, sistem dibangun tanpa kebutuhan untuk mengelola infrastruktur server secara langsung. Sebaliknya, fungsi-fungsi kecil yang menyusun alur kerja diimplementasikan sebagai layanan terpisah. Setiap fungsi serverless bertanggung jawab untuk tugas tertentu, seperti penerimaan input, pemrosesan data, dan penentuan kategori AQI. Kelebihan arsitektur serverless terletak pada fleksibilitas dan skalabilitasnya. Fungsi-fungsi serverless dapat diaktifkan secara otomatis ketika diperlukan dan dinonaktifkan saat tidak digunakan, mengoptimalkan penggunaan sumber daya dan mengurangi biaya infrastruktur. Integrasi arsitektur serverless dengan pendekatan pseudocode membantu mencapai konsistensi, keberlanjutan, dan efisiensi dalam pengembangan sistem Monitoring Indeks Kualitas Udara [10].

```
1. Inisialisasi variabel: latitude, longitude, forecast_days, dan
   api_url.
2. Ambil data cuaca menggunakan api_url dan proses data:
   a. Ekstrak informasi yang relevan.
   b. Tampilkan kategori kualitas udara saat ini, nilai, koordinat, dan
      warna latar belakang.
   c. Isi tabel ramalan.
   d. Tangani kesalahan jika ada.
3. Implementasikan fungsi updateData:
   a. Dapatkan dan validasi nilai input baru.
   b. Perbarui variabel dan api_url.
   c. Ulangi langkah 2a hingga 2d.
   d. Tutup 'inputModal'.
   e. Ambil dan tampilkan data lokasi.
4. Implementasikan fungsi fetchLocationData:
   a. Ambil data lokasi.
   b. Tampilkan nama lokasi.
5. Implementasikan berbagai fungsi bantu:
   a. formatDateTime untuk format tanggal dan waktu.
   b. padZero untuk menambahkan nol di depan.
   c. determineCategory untuk menentukan kategori AQI.
   d. determineCaution untuk mendapatkan informasi peringatan.
   e. setBackgroundColor untuk mengatur warna latar belakang.
   f. populateForecastTable untuk mengisi tabel ramalan.
6. Tampilkan data pada halaman web.
7. Implementasikan desain responsif menggunakan Bootstrap.
8. Hosting halaman web di Vercel untuk akses publik:
   https://aqipollution.vercel.app.
9. Lakukan pengujian black box untuk memastikan hasil yang akurat.
10. Jaga dan perbarui sistem untuk menyesuaikan perubahan API
    eksternal, tanggapi umpan balik pengguna, dan perbaiki bug yang
    teridentifikasi.
```

4.4. Implementasi

Sebagai bagian krusial dari siklus pengembangan perangkat lunak, tahap implementasi melibatkan transformasi desain konseptual menjadi suatu produk yang dapat berjalan. Dalam konteks Sistem Monitoring Indeks Kualitas Udara, implementasi dilakukan dengan menggunakan sejumlah teknologi kunci. HTML, CSS, dan JavaScript digunakan untuk menerjemahkan desain sistem ke dalam kode sumber yang dapat dijalankan secara efektif. Web hosting diberdayakan melalui Vercel, platform yang terintegrasi dengan GitHub, memungkinkan penerapan praktis dan manajemen mudah dari kode sumber. Bootstrap, kerangka kerja front-end yang terkenal, dimanfaatkan untuk meningkatkan responsivitas dan estetika antarmuka pengguna, memastikan pengalaman pengguna yang menyenangkan dan mudah diakses. Dengan kombinasi teknologi ini, tahap implementasi menjadi fondasi utama bagi fungsionalitas dan tampilan Sistem Monitoring Indeks Kualitas Udara.

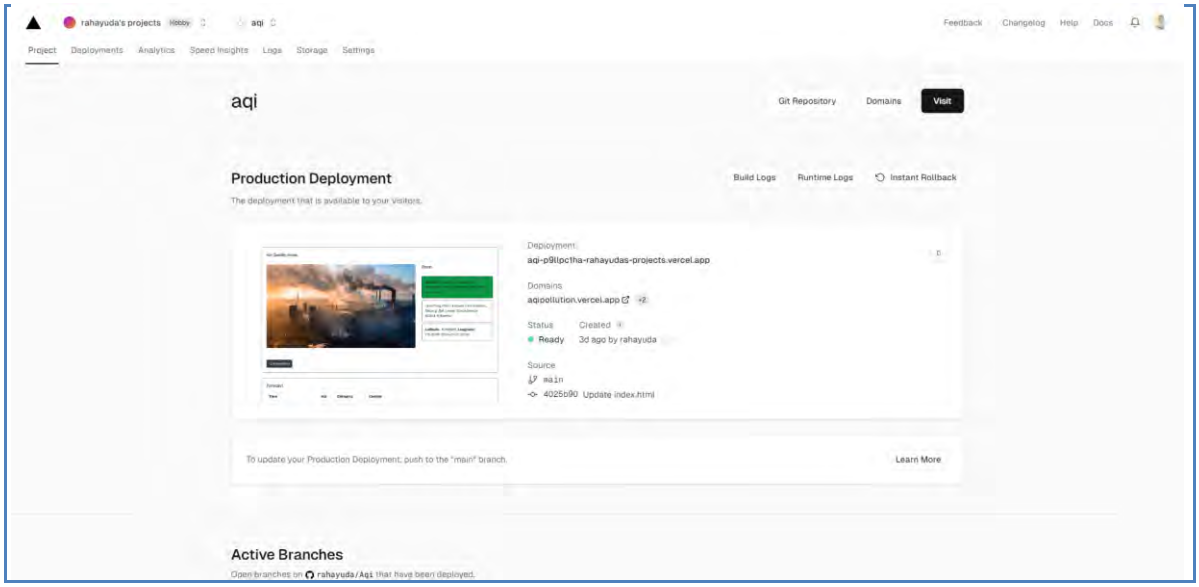
Menerapkan desain ke dalam kode sumber Sistem Monitoring Indeks Kualitas Udara melibatkan integrasi HTML, CSS, dan JavaScript sebagai komponen utama. HTML digunakan untuk menentukan struktur dan elemen-elemen pada halaman web, CSS bertanggung jawab atas penataan dan tampilan estetika, sementara JavaScript memberikan fungsionalitas dinamis dan interaktivitas pada halaman. Dalam repository GitHub, terdapat tangkapan layar (screenshot) yang mencerminkan struktur folder dan berkas serta beberapa cuplikan kode krusial terlihat pada **gambar 1**. **index.html** merupakan berkas utama yang menentukan struktur dan konten dari halaman web. Di dalamnya, elemen-elemen HTML ditempatkan untuk membangun antarmuka pengguna. Struktur yang terorganisir memudahkan pemahaman dan pemeliharaan kode. **script.js** adalah berkas JavaScript yang memberikan logika dan interaktivitas pada halaman. Dengan menggunakan JavaScript, Sistem Monitoring Indeks Kualitas Udara dapat mengambil data dari API, memproses informasi, dan mengupdate tampilan secara dinamis tanpa perlu me-refresh halaman. **style.css** berperan dalam mendefinisikan tata letak dan tampilan estetika elemen-elemen HTML. Melalui CSS, warna, ukuran, dan posisi elemen dapat diatur untuk menciptakan antarmuka pengguna yang menarik dan mudah dibaca.



Gambar 1. Repositori Kode pada Github

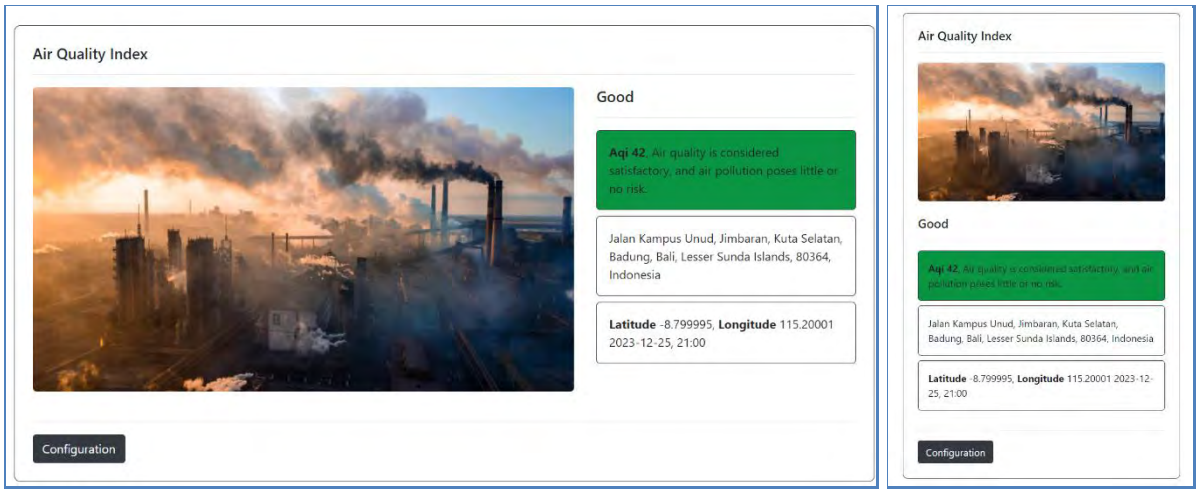
Proses hosting web merupakan langkah krusial dalam menghadirkan sebuah proyek secara daring. Dalam konteks Sistem Monitoring Indeks Kualitas Udara, proyek ini dihosting menggunakan platform Vercel yang terintegrasi secara langsung dengan GitHub. Integrasi ini mempermudah pengelolaan proyek dan penerapan perubahan, membuatnya menjadi pilihan yang efisien bagi pengembang. Vercel adalah platform cloud yang memungkinkan pengguna untuk menyimpan dan mendeploy proyek web secara cepat dan mudah. Dengan integrasi GitHub, setiap kali terjadi perubahan pada repositori GitHub, Vercel secara otomatis memicu proses pembangunan (build) dan merilis (deploy) versi terbaru dari proyek. Hal ini memastikan bahwa proyek selalu tersedia dalam kondisi terkini. Keuntungan utama dari integrasi Vercel dan GitHub adalah efisiensi pengelolaan kode sumber dan implementasi perubahan. Proses deployment yang otomatis menghilangkan kebutuhan untuk melakukan langkah-langkah manual, sehingga pengembang dapat lebih fokus pada pengembangan fitur dan fungsionalitas proyek seperti terlihat pada **gambar 2**. Selain itu, Vercel menyediakan URL unik yang mudah diingat (https://aqipollution.vercel.app), memudahkan pengguna untuk mengakses Sistem Monitoring Indeks Kualitas Udara secara daring dengan cepat.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

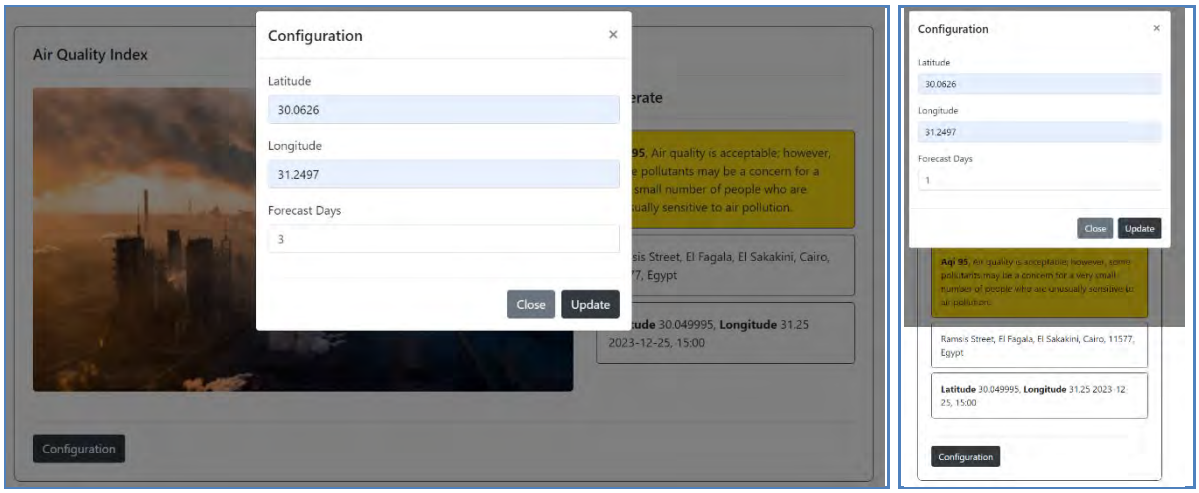


Gambar 2. Vercel Deployment

Dalam pengembangan Sistem Monitoring Indeks Kualitas Udara, responsivitas dan tampilan yang estetis menjadi fokus penting. Untuk mencapai hal ini, proyek ini memanfaatkan Bootstrap, sebuah kerangka kerja CSS yang terkenal, untuk merancang antarmuka yang responsif dan menarik. Bootstrap menyediakan berbagai komponen dan utilitas yang mempermudah pengembangan tata letak yang responsif tanpa harus menulis banyak kode CSS kustom. Tampilan web versi desktop dari Sistem Monitoring Indeks Kualitas Udara mencerminkan desain yang bersih dan terorganisir. Pengguna dapat dengan mudah membaca informasi mengenai kategori indeks kualitas udara, nilai AQI, dan detail lokasi. Tata letak ini memberikan pengalaman pengguna yang intuitif dan informatif. Selain itu, responsivitas proyek ini dijamin melalui penggunaan Bootstrap, yang mengoptimalkan tampilan web saat diakses melalui perangkat mobile. Dengan demikian, pengguna dapat mengakses Sistem Monitoring Indeks Kualitas Udara dengan lancar tanpa mengalami kesulitan tata letak atau keterbacaan, seperti terlihat pada **gambar 3**. Sistem konfigurasi juga dihadirkan melalui modal yang dapat diakses dengan mudah. Modal ini memungkinkan pengguna untuk memasukkan data lokasi baru dan jumlah hari perkiraan cuaca dengan nyaman, terlihat pada **gambar 4**. Tampilan yang bersih dan sederhana memberikan pengalaman pengguna yang lebih baik dalam mengonfigurasi Sistem Monitoring Indeks Kualitas Udara



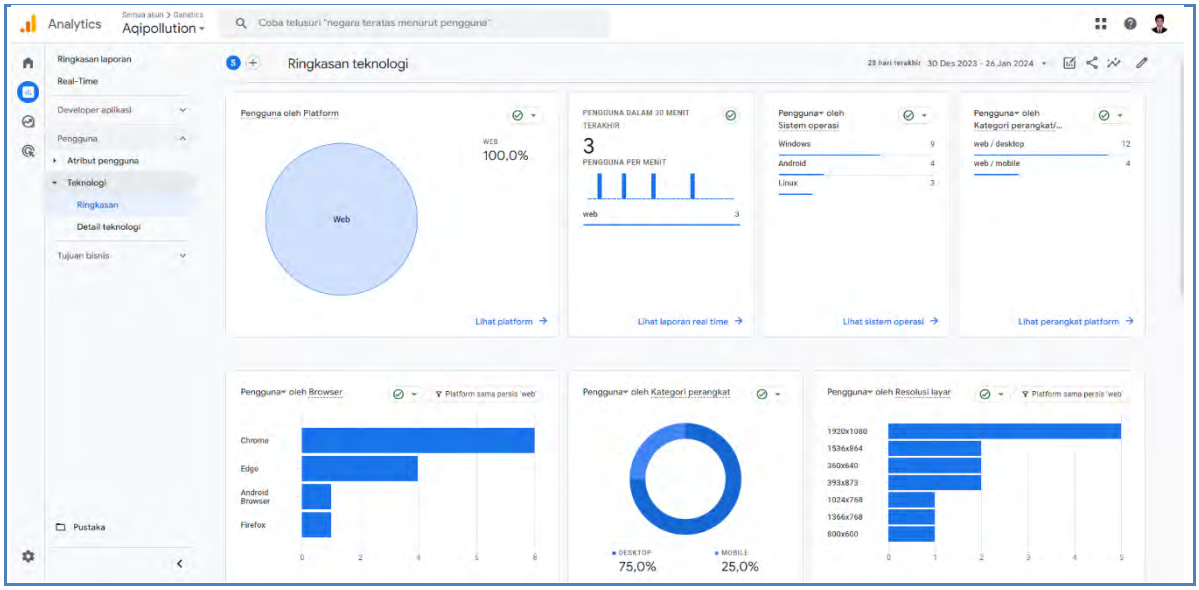
Gambar 3. Tampilan Desktop dan Mobile



Gambar 4. Tampilan Konfigurasi

4.5. Pengujian

Selama periode evaluasi dari tanggal 23 Januari hingga 27 Januari 2024, statistik penggunaan yang tercatat dalam Google Analytics menunjukkan pola yang menarik seperti terlihat pada **gambar 5**. Mayoritas pengguna menggunakan sistem operasi Windows dengan persentase 39.13%, diikuti oleh pengguna Android sebesar 17.39%, dan Linux sebesar 13.04%. Dari segi perangkat, sebagian besar akses dilakukan melalui Web/Desktop dengan persentase 52.17%, sementara penggunaan Web/Mobile mencapai 17.39%. Chrome adalah browser yang paling banyak digunakan dengan persentase 34.78%, diikuti oleh Edge dengan 17.39%, dan Android Browser hanya mencapai 4.35%. Resolusi layar 1920x1080 mendominasi dengan 21.74%, sementara 1536x864 dan 360x640 masing-masing mencapai 8.70%. Data penggunaan juga mencatat Sesi Engagement mencapai 100%, menunjukkan tingkat keterlibatan yang tinggi dari pengguna, dengan Rasio Engagement sebesar 61.76%, menunjukkan sebagian besar pengguna terlibat dalam interaksi. Waktu Engagement Rata-Rata adalah 46 detik, menandakan sebagian besar pengguna menghabiskan waktu singkat di situs, sementara Jumlah Peristiwa mencapai 236, menunjukkan aktivitas yang cukup signifikan dalam interaksi pengguna dengan situs.

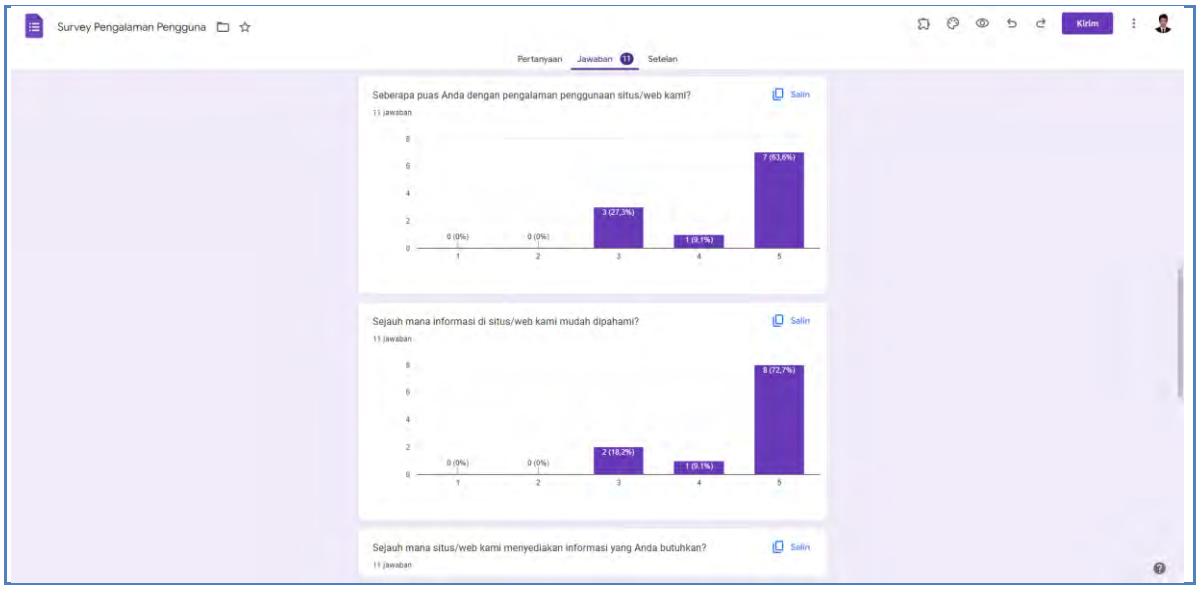


Gambar 5. Google Analytics

Pada tahap pengujian sistem menggunakan metode Black Box, fokus utamanya adalah pada pengujian fungsionalitas tanpa memperhatikan detail implementasi internal. Pengujian ini bertujuan untuk memverifikasi kemampuan Sistem Monitoring Indeks Kualitas Udara dalam memberikan hasil yang

p-ISSN: 2301-5373
e-ISSN: 2654-5101

akurat sesuai dengan konfigurasi pengguna yang diberikan. Hasil pengujian Black Box menunjukkan bahwa semua fitur pada sistem telah berjalan dengan baik dan sesuai dengan harapan. Sistem berhasil memberikan respons yang tepat dan akurat terhadap konfigurasi yang diberikan, termasuk pembaruan lokasi, penggunaan Open Meteo API, dan integrasi OpenStreetMap API. Pengujian ini mencakup evaluasi terhadap kemampuan sistem dalam menampilkan indeks kualitas udara (AQI) berdasarkan lokasi geografis, konfigurasi latitude dan longitude, serta perkiraan cuaca dalam rentang waktu tertentu. Semua aspek ini telah diuji secara menyeluruh, dan hasilnya memastikan bahwa Sistem Monitoring Indeks Kualitas Udara berfungsi dengan baik sesuai dengan kebutuhan pengguna. Dengan demikian, hasil pengujian Black Box memberikan keyakinan bahwa Sistem Monitoring Indeks Kualitas Udara siap untuk digunakan secara luas, memberikan informasi yang akurat dan bermanfaat terkait kualitas udara sesuai dengan lokasi dan preferensi pengguna.



Gambar 6. Kuisioner

Berdasarkan hasil evaluasi menggunakan kuisioner Google Form dengan skala Likert dari 1 hingga 5 seperti terlihat pada **gambar 6**, yang menilai lima aspek berbeda dari pengalaman penggunaan situs/web, diperoleh persentase tingkat kepuasan pengguna sebagai berikut: untuk Kepuasan Umum (P1), persentasenya sekitar 88.89%; untuk Kejelasan Informasi (P2), sekitar 93.33%; untuk Kelengkapan Konten (P3), sekitar 91.11%; untuk Kecepatan Memuat Halaman (P4), sekitar 88.89%; dan untuk Loyalitas (P5), sekitar 86.67%. Hasil ini menunjukkan adanya tingkat kepuasan yang tinggi dari pengguna terhadap berbagai aspek penggunaan situs/web, dengan skor tertinggi terdapat pada Kejelasan Informasi. Hal ini memberikan gambaran yang positif tentang pengalaman pengguna dan memberikan arahan untuk potensi perbaikan di masa depan.

Dari hasil evaluasi tersebut, dapat disimpulkan bahwa Sistem Monitoring Indeks Kualitas Udara telah berhasil dalam memberikan respons yang akurat dan memuaskan kepada pengguna. Namun, perlu dipertimbangkan untuk lebih memperhatikan tingkat kepuasan pada fitur tertentu yang mungkin mendapat penilaian lebih rendah, seperti kecepatan memuat halaman. Langkah-langkah untuk meningkatkan kepuasan pengguna pada area-area tertentu dapat dilakukan berdasarkan temuan ini.

4.6. Pemeliharaan

Pada tahap pemeliharaan, Sistem Monitoring Indeks Kualitas Udara akan terus dikembangkan dan diperbarui guna mengakomodasi perubahan pada API eksternal yang digunakan. Proses ini melibatkan respons terhadap umpan balik dari pengguna, dengan memperbaiki bug yang teridentifikasi dan meningkatkan fungsionalitas sistem sesuai kebutuhan. Selain itu, pemeliharaan sistem juga mencakup pemanfaatan layanan web hosting Vercel. Penggunaan Vercel sebagai platform hosting memastikan akses publik yang stabil dan responsif terhadap pengguna. Dengan demikian, sistem dapat diakses dengan baik oleh pengguna dari berbagai lokasi dan perangkat, menjaga ketersediaan informasi mengenai kualitas udara yang konsisten. Pemeliharaan ini merupakan langkah penting untuk menjaga performa optimal sistem seiring berjalannya waktu. Dengan terus melakukan pembaruan dan

perbaikan, Sistem Monitoring Indeks Kualitas Udara dapat terus memberikan layanan yang handal dan relevan bagi pengguna, serta tetap responsif terhadap perubahan lingkungan atau kebutuhan pengguna yang mungkin muncul.

5. Kesimpulan

Dalam penelitian ini, sebuah Sistem Monitoring Indeks Kualitas Udara berhasil dikembangkan dengan memanfaatkan pendekatan Software Development Life Cycle (SDLC) dan integrasi Open Meteo API serta OpenStreetMap API. Sistem ini memungkinkan pengguna untuk dengan mudah melihat Indeks Kualitas Udara (AQI) berdasarkan lokasi geografis, memberikan perkiraan cuaca dalam rentang waktu tertentu, dan konfigurasi lokasi serta masa depan. Melalui penerapan standar United States Air Quality Index (AQI) dalam representasi visual menggunakan warna, pengguna dapat dengan cepat memahami tingkat risiko kesehatan. Penggunaan SDLC memastikan proses pengembangan yang terstruktur, sementara integrasi OpenStreetMap API meningkatkan kemampuan identifikasi lokasi berdasarkan koordinat geografis. Responsif dan estetis, Sistem ini dihosting secara online menggunakan Vercel dan GitHub, memastikan ketersediaan dan manajemen kode yang efisien. Pengujian Black Box sukses menunjukkan kehandalan sistem, dan melalui pemeliharaan yang berkelanjutan, diharapkan sistem ini dapat terus memberikan informasi kualitas udara yang akurat dan dapat diandalkan kepada pengguna. Hasil evaluasi usability yang positif, seperti yang tergambar dari tingkat kepuasan yang tinggi dalam penggunaan situs/web, menegaskan kelayakan dan efektivitas sistem ini. Dengan pendekatan holistik, penelitian ini memberikan kontribusi positif dalam memahami dan mengatasi permasalahan kualitas udara, dengan memadukan teknologi terkini menjadi alat pemantauan yang accessible dan komprehensif.

Daftar Pustaka

[1] World Health Organization, *Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease*. Geneva: WHO Press, 2018.
[2] Y. Zhang, *Advances in Air Pollution Monitoring and Exposure Assessment*. FL: CRC Press, 2019.
[3] U.S. Environmental Protection Agency (EPA), *Air Quality Index: A Guide to Air Quality and Your Health*. Washington, DC: EPA Publications, 2023.
[4] Meteo France, *CAMS European air quality forecasts, ENSEMBLE data*. opernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store (ADS), 2022.
[5] Kanchan, A. K. Gorai, and P. Goyal, "A Review on Air Quality Indexing System," *Asian J. Atmos. Environ.*, 2015.
[6] S. M. Sohan, C. Anslow, and F. Maurer, "A Case Study of Web API Evolution," *IEEE World Congr. Serv.*, 2015.
[7] R. A. P. Rajan, "Serverless Architecture - A Revolution in Cloud Computing," *Tenth Int. Conf. Adv. Comput.*, 2018.
[8] I. Sommerville, *Software Engineering*. Addison-Wesley, 2023.
[9] Karl Wiegers, *Software Requirements*. Microsoft Press, 2023.
[10] Syamsiah, "Perancangan Flowchart dan Pseudocode Pembelajaran Mengenal Angka dengan Animasi Untuk Anak PAUD Rambutan," *Satuan Tulisan Ris. dan Inov. Teknol.*, 2019.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Perancangan Robot Pembersih Vacuum Pintar Berbasis Mikrocontroller Menggunakan Android

Leonardo Silalahi[a1], Zelvi Gustiana[a2], Zulham[b3]

[a]Teknologi Informasi, Universitas Dharmawangsa
Jl. K.L Yos Sudarso No.224, Medan, Indonesia
[1]leosilalahi966@gmail.com
[2]zelvi@dharmawangsa.ac.id (Corresponding author)

[b] Rekayasa Perangkat Lunak, Universitas Dharmawangsa
Jl. K.L Yos Sudarso No.224, Medan, Indonesia
[3]zulham@dharmawangsa.ac.id

Abstract

Robotics innovation has experienced rapid advancements, bringing significant changes to robots, enabling them to adapt to innovative environments. Integrated control systems with microcontrollers and Android platforms facilitate human tasks. Robots are also beneficial in flexible situations such as clinics, homes, workplaces, and hazardous areas. One practical application is the floor-cleaning device with Android control system. In its implementation, the program utilizes the Arduino Uno IDE integrated with a microcontroller.
This floor-cleaning device effectively removes dust and dirt. The research on "Design of Intelligent Vacuum Robot Based on Microcontroller Using Android" successfully created a cleaning tool that efficiently assists people in floor cleaning. The microcontroller and Android-based vacuum cleaner are highly useful for cleaning dusty and dirty floors. Therefore, the design of a smart vacuum cleaning robot based on microcontroller and Android becomes a beneficial solution for the community
(Justify, Arial 10)

Keywords: *Arduino Uno, Science, Microcontroller, Robotics, Vacuum Cleaner*

1. Pendahuluan

Sains dalam inovasi saat ini mengalami perkembangan pesat, khususnya di sektor robotika [1]. Robot tidak lagi hanya objek yang statis, melainkan entitas yang dapat menjalankan tugas sesuai dengan desain buatan manusia [2]. Konsep robot tidak terbatas pada sudut pandang ilmu material dalam lingkungan inovatif. Inovasi menciptakan perbedaan yang signifikan, membebaskan manusia dari tugas rutin dan jadwal yang monoton [3]. Pada era sekarang, penggunaan sistem kontrol semakin berkembang dengan pesat [4]. Sistem kontrol memberikan bantuan kepada individu untuk mempermudah pekerjaan mereka [5]. Dalam konteks ini, kerangka kontrol yang umum digunakan melibatkan penggabungan mikrokontroler dengan platform Android untuk menjalankan perangkat pendukung lainnya [6]. Kehadiran robot juga memberikan manfaat besar dalam situasi yang lebih fleksibel seperti pengembangan klinik, rumah, tempat kerja, dan lingkungan berbahaya seperti pabrik pengontrol atom dan kimia [7]. Robot manusia menjadi bagian yang terintegrasi dalam kehidupan sehari-hari, memfasilitasi interaksi [8].

Mikrokontroler berfungsi sebagai kerangka kerja komputer utilitarian dalam satu chip, menggabungkan pusat prosesor, memori (termasuk Slam, memori program, atau keduanya), dan perangkat input/output [9]. Perbedaannya dengan chip serbaguna yang umumnya digunakan di PC terletak pada kebutuhan mikrokontroler akan kerangka yang khusus untuk mengelolanya [10]. Kerangka mikrokomputer paling dasar adalah rangkaian elektronik minimal yang diperlukan untuk menjalankan IC mikrokomputer. Mikrokontroler, sering disebut sebagai µC, memiliki peran penting dalam teknologi semikonduktor dengan lebih banyak transistor, tetapi ukurannya kecil dan dapat diproduksi massal untuk menjaga harganya tetap terjangkau dibandingkan dengan mikroprosesor [11]. Robotika, sebagai bidang multi-disiplin, bertujuan untuk mengembangkan robot dengan kegunaan khusus. Disiplin yang terlibat

melibatkan mekanika, elektronika, dan pemrograman komputer. Oleh karena itu, untuk memahami pembuatan robot, penting untuk mengembangkan keterampilan dalam ketiga bidang tersebut [12].

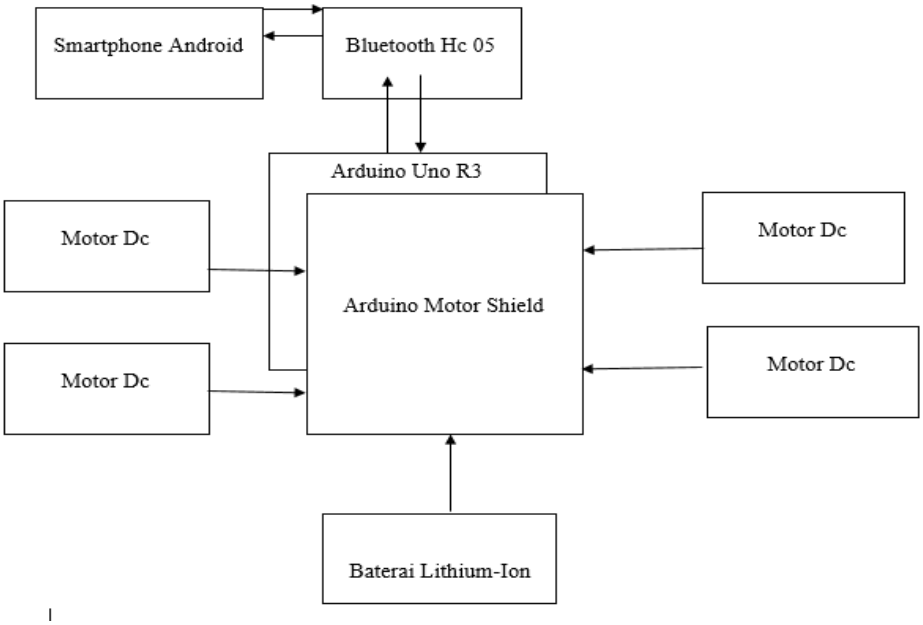
Awalnya, vakum diartikan sebagai ruang yang bebas dari materi, tetapi konsep vakum diperkenalkan dengan asumsi yang keliru. Vakum tidak dapat dihasilkan melalui eksperimen praktis dan tidak dapat didefinisikan secara langsung karena tidak ada respons dari ruang kosong. Absolut vakum bisa dibandingkan dengan lubang hitam yang menyerap semua sinyal materi [13]. Android merupakan sistem operasi bergerak untuk telepon seluler yang berbasis Linux, memberikan platform terbuka bagi para pengembang untuk menciptakan aplikasi. Android menjadi generasi baru dalam platform mobile, memberikan kebebasan kepada pengembang untuk mengembangkan sesuai harapan [14]. Modul Bluetooth HC-05, yang tersedia di pasaran dengan harga yang relatif terjangkau, terdiri dari 6 pin konektor, masing-masing dengan fungsi yang berbeda-beda.

Asal-usul kata "arduino" berasal dari bahasa Italia, di mana "ardu" berarti sulit dan "no" berarti tidak [15]. Arduino merupakan platform open source untuk pembuatan prototipe elektronik, baik dalam perangkat keras (hardware) maupun perangkat lunak (shield), yang mudah digunakan dan fleksibel. Hardware-nya menggunakan prosesor Atmel AVR ATmega 328. Arduino Uno dilengkapi dengan 14 pin input/output digital (termasuk 6 yang dapat digunakan sebagai output PWM), 6 pin input analog, koneksi USB, dan tombol reset [16].

2. Metode Penelitian

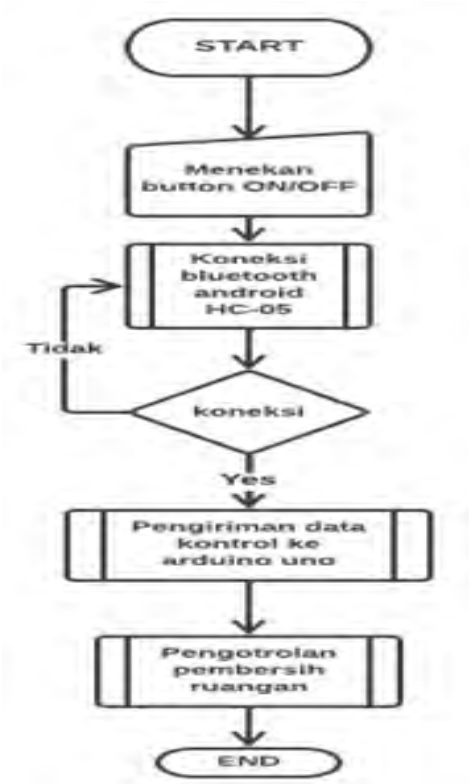
Berdasarkan hasil wawancara dan penelitian dokumenter, diketahui bahwa ibu rumah tangga masih menggunakan sapu untuk membersihkan lantai kotor yang bisa bocor. Oleh karena itu penulis membuat sebuah alat yang membantu anda membersihkan lantai menggunakan sistem android yang sangat efisien dengan remote control nya. Tujuan dari analisis kebutuhan sistem adalah untuk mengurangi masalah yang timbul dalam proses pengolahan data dan pengiriman data serta untuk meningkatkan pelayanan yang lebih baik.

Persyaratan sistem informasi lebih sederhana daripada sistem lain, tetapi harus mencakup komponen sistem seperti perangkat keras, perisai, dan perangkat otak. Persyaratan yang paling penting untuk sistem informasi adalah layanan siklis atau transaksional. Suatu sistem pada dasarnya adalah sekelompok elemen yang terkait erat yang dirancang untuk mencapai tujuan tertentu. Sederhananya, sistem dapat digambarkan sebagai kumpulan atau kumpulan elemen, komponen, atau variabel yang terorganisir, berinteraksi, saling bergantung, dan terintegrasi.



Gambar 1. Diagram Blok Sistem

p-ISSN: 2301-5373
e-ISSN: 2654-5101



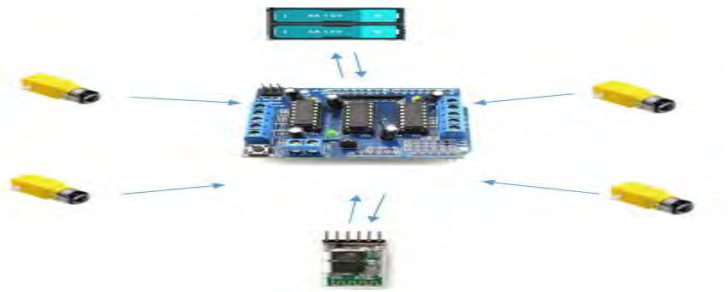
Gambar 2. Diagram Alur Sistem

Diagram ini menggambarkan aliran data sistem menggunakan notasi. DFD sering digunakan untuk menggambarkan sistem yang ada atau baru yang sedang dikembangkan secara logis tanpa memperhatikan lingkungan fisik di mana data mengalir. Diagram aliran data (data flow diagram) adalah representasi grafis dari aliran data dari suatu sumber.

3. Pembahasan dan Hasil

3.1 Perancangan Perangkat Keras

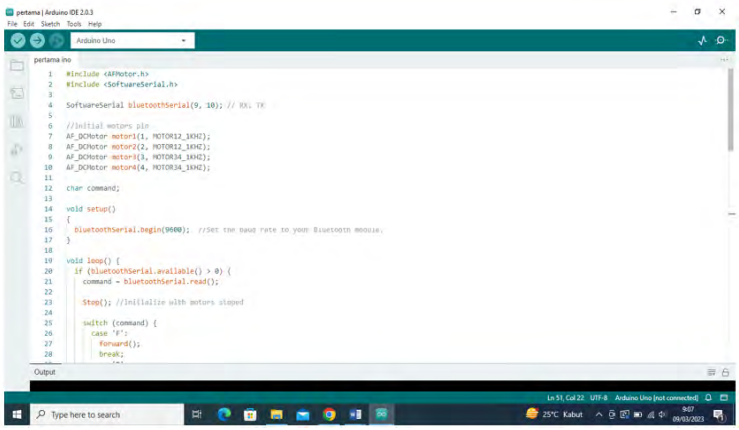
Perangkat keras di komputer Anda adalah bagian fisik dari komputer Anda yang memiliki fungsi berbeda untuk proses yang berbeda. Juga dalam tahap perancangan robot penyedot debu dengan menggunakan android.



Gambar 3. Perancangan Perangkat Keras

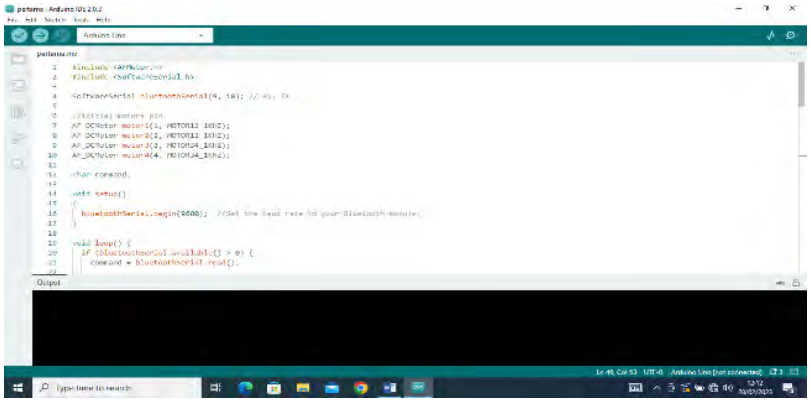
3.2 Perancangan Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini berperan sebagai antarmuka dan pembawa data. Selanjutnya, kode program mikrokontroler merupakan bagian dari perangkat lunak. Berdasarkan konsep perancangan perangkat keras, program yang dirancang akan digunakan dalam sistem vacuum cleaner dengan dengan menggunakan android dengan mikrokontroler berbasis Arduino Uno.



Gambar 4. *Shield* Pemograman Arduino Ide

Untuk membangun sebuah sistem dengan mikrokontroler, diperlukan sebuah aplikasi bernama Arduino IDE yang digunakan untuk membuat perintah atau sering disebut dengan coding program. Intinya, pengkodean yang disusun membantu memberikan perintah ke Bluetooth HC-05 sesuai dengan kondisi yang diperlukan. Dengan kata lain, penyedot debu dengan sistem kontrol yang diberdayakan Android membersihkan lantai yang kotor. Gambar 4.2 di bawah ini adalah software Arduino IDE yang berisi kode untuk diupload ke mikrokontroler untuk sistem robot vacuum.



Gambar 5. S*oftware* Arduino IDE

3.3 Hasil Pengujian

Berdasarkan analisis dan perancangan yang telah dilakukan pada bab sebelumnya, telah terwujud perancangan vakuum cleaner menggunakan adnroid Berbagai tes harus dilakukan untuk mengetahui cara kerja perangkat. Selain itu, perlu juga melakukan pengujian ini untuk mengetahui seperti apa pengkondisian agar alat ini dapat digunakan secara optimal.

Gambar 6. Pengujian *Vakuum Cleaner*

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Aplikasi adalah perangkat lunak yang menggabungkan fitur-fitur tertentu dengan cara yang dapat diakses oleh pengguna. Ada jutaan aplikasi di App Store dan Android App Store yang menyediakan layanan aplikasi. Aplikasi itu sendiri adalah fondasi ekonomi seluler. Sejak kedatangan iPhone pada tahun 2007 dan App Store pada tahun 2008, aplikasi telah menjadi sarana utama bagi pengguna untuk mengakses revolusi smartphone.

Gambar 7 *Bluetooth Rc Controller*

4. Kesimpulan

Hal ini didasarkan pada tahap penelitian perangkat keras, perangkat lunak dan desain implementasi. Berikut kesimpulan yang dapat diambil dari penelitian "Perancangan Robot Cerdas Vakum Berbasis Mikrokontroler Menggunakan Android".

1. pada penelitian ini penulis telah menghasilkan sebuah alat menggunakan mikrokontroler berbasis arduino uno, sehingga dapat memudahkan masyarakat dalam membersihkan lantai yang berdebu
2. vacuum cleaner yang dibuat penulis dapat membersihkan debu atau kotoran yang berada di lantai.
3. Perancangan Robot Pembersih Vacuum Pintar Berbasis Mikrokontroler Menggunakan Android, yang dibuat oleh penulis ini menjadi alat bantu masyarakat untuk membersihkan lantai yang kotor.

References

[1] A. Sigma, "APLIKASI MIKROKONTROLER ARDUINO UNO DALAM RANCANG BANGUN KUNCI PINTU MENGGUNAKAN E-KTP," vol. 7, no. 1, pp. 74–88, 2022.
[2] V. F. Sari, R. Ekawita, and E. Yuliza, "DESAIN BANGUN pH TANAH DIGITAL BERBASIS ARDUINO UNO," vol. 7, no. 1, pp. 36–41, 2021.
[3] P. Trainer, P. Mikrokontroler, T. Mikrokontroler, and A. Uno, "ARDUINO UNO PADA MATA PELAJARAN TEKNIK PEMOGRAMAN MIKROPROSESOR DAN MIKROKONTROLER KELAS XI TEI DI SMKN 1 NGAWI Zainal Mujib Ansori S1 Pendidikan Teknik Elektro , Fakultas Teknik , Universitas Negeri Surabaya Lilik Anifah Jurusan Teknik Elektro , Fakultas Teknik , Universitas Negeri Surabaya I Gusti Putu Asto Buditjahjanto Jurusan Teknik Elektro , Fakultas Teknik , Universitas Negeri Surabaya Nurhayati Jurusan Teknik Elektro , Fakultas Teknik , Universitas Negeri Surabaya," 2021.
[4] R. A. Pratama and I. Permana, "Simulasi Permodelan Menggunakan Sensor Suhu Berbasis Arduino," vol. 10, no. 1, pp. 7–12, 2021.
[5] P. Studi *et al.*, "REVIEW APLIKASI SENSOR PADA SISTEM MONITORING DAN KONTROL BERBASIS MIKROKONTROLER ARDUINO," vol. 8, no. 4, pp. 171–179, 2021.
[6] A. Lestari and O. Candra, "Sistem Otomasi Pensortiran Barang berbasis Arduino Uno," vol. 7, no. 1, pp. 27–36, 2021.
[7] A. Kamolan and L. Sampebatu, "DENGAN INPUT KODE PIN DAN MULTI SENSOR BERBASIS," vol. 6, no. 1, pp. 22–31, 2021.

[8] I. Artikel and A. Info, "RANCANG BANGUN SISTEM KEAMANAN SEPEDA MOTOR BERBASIS ARDUINO UNO MENGGUNAKAN GPS DAN RELAY MELALUI," vol. 1, no. 1, pp. 1–7, 2022.
[9] S. Manurung, I. Parlina, F. Anggraini, and D. Hartama, "Penggunaan Sistem Arduino Menggunakan RFID untuk Keamanan Kendaraan Bermotor," vol. 1, no. 2, pp. 139–148, 2021.
[10] M. Fahreza, "Desain Controling Pengaman Arus Lebih Berbasis Arduino," vol. 2, no. 1, 2021.
[11] M. Muthohir and S. Prayogi, "Prototype Sistem Keamanan Brankas Menggunakan Teknologi RFID Berbasis Arduino Uno RFID Berbasis Arduino Uno," vol. 1, no. 1, pp. 108–117, 2021.
[12] R. Bangun *et al.*, "Jurnal simetrik vol 12, no. 2, desember 2022," vol. 12, no. 2, pp. 606–612, 2022.
[13] F. Nadziroh, "ALAT DETEKSI INTENSITAS CAHAYA BERBASIS ARDUINO UNO SEBAGAI PENANDA PERGANTIAN WAKTU SIANG-MALAM BAGI TUNANETRA ARDUINO UNO-BASED LIGHT INTENSITY DETECTION TOOL AS A DAY-NIGHT ALTERATION MARK FOR THE BLIND
[14] Y. Della, R. Suppa, A. Ali, H. Dani, and U. Andi, "RANCANG BANGUN ALAT PENGUKUR SIFAT FISIS AIR BERBASIS," vol. 5, no. 2, pp. 339–345, 2021.
[15] A. Pratama, S. R. Andani, and A. Wanto, "Penerapan Mikrokontroler Arduino Uno pada Desain Perancangan Sistem Ayunan Bayi Otomatis," vol. 1, no. 3, pp. 108–114, 2021.
[16] H. Fair, B. Mulyati, F. Teknik, and U. Nurtanio, "Rancang Bangun Alat Pengukur Kecepatan Aliran Air Menggunakan Water Flow Sensor Berbasis Ardunino Uno," vol. 2, no. 1, pp. 1–11.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Support Vector Regression untuk Prediksi Harga Rumah Dengan Optimasi Grid Search

Mulyawan[a1], Reza Subagja[b2], Dede Rohman[b3], Deny Indriyana Efendi[b4]

[a]Program Studi Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
[1]mulyawan@gmail.com

[b]Program Studi Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
[2]rezasubagja5@gmail.com
[3]dederohman@gmail.com
[4]dendyindriyaefendi@gmail.com

Abstract

Rumah merupakan kebutuhan pokok dalam kehidupan manusia, berfungsi sebagai tempat perlindungan tidak hanya dari kondisi cuaca eksternal, tetapi juga dari makhluk hidup lainnya. Harga rumah menjadi elemen kunci dalam transaksi properti, baik melalui cara konvensional maupun digital. Penelitian ini memiliki tujuan untuk menerapkan algoritma Support Vector Regression (SVR) dalam konteks prediksi harga rumah berdasarkan karakteristiknya. Dalam upaya mencapai kinerja model yang optimal, dilakukan optimasi parameter menggunakan algoritma Grid Search. Data yang digunakan diperoleh melalui teknik web scraping dari situs properti rumah123.com, dengan fokus pada wilayah Jakarta. Atribut data mencakup lokasi, luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, kapasitas garasi, dan harga rumah. Metode penelitian melibatkan langkah-langkah Obtain, Scrub, Explore, Model, dan Interpret. Dalam pemodelan, dua skenario data dieksplorasi, yakni menggunakan dataset asli dan hasil transformasi logaritma pada variabel target. Hasil penelitian menunjukkan bahwa model terbaik ditemukan pada skenario data hasil transformasi logaritma, dengan nilai metrik RMSE = 0.2774, MAE = 0.2061, MAPE = 0.1453, dan R-Squared = 0.7867. Parameter optimal yang dihasilkan dari metode grid search adalah Cost = 1, epsilon = 0.1, dan gamma = 1.

Keywords: *Prediksi Harga Rumah, Support Vector Regression, Grid Search*

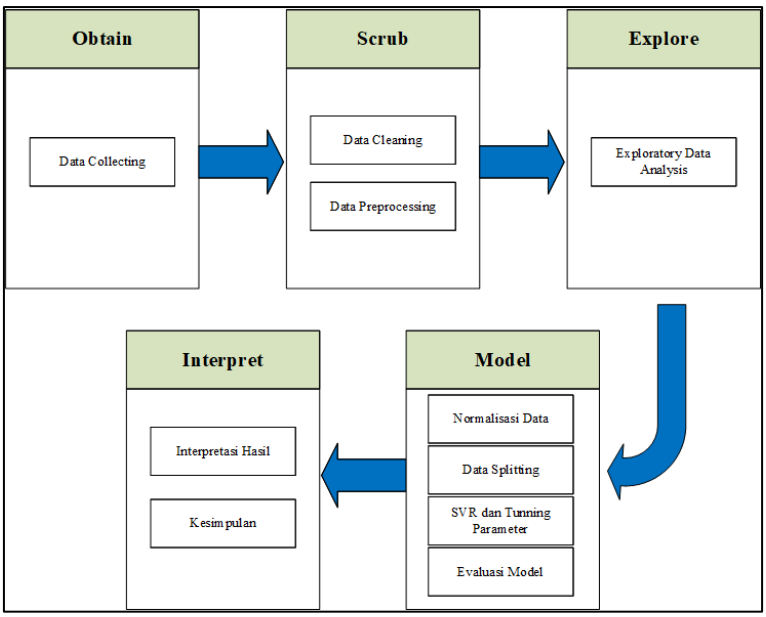
1. Pendahuluan

Rumah merupakan salah satu kebutuhan dasar manusia, bersama dengan kebutuhan sandang dan pangan. Menurut Mulder (2006), seiring dengan pertumbuhan populasi, permintaan akan perumahan akan terus meningkat. Selain itu, perkembangan teknologi informasi telah membawa perubahan besar dalam industri properti. Banyak platform jual beli properti *online* bermunculan, menyediakan layanan bagi pengguna untuk saling bertransaksi secara efisien. Dalam konteks ini, data penjualan rumah yang terakumulasi di platform tersebut menjadi sangat kaya dan berpotensi untuk dieksplorasi. Perkembangan pesat tersebut menghadirkan tantangan dan peluang dalam peningkatan kualitas pengalaman pengguna. Berdasarkan survei peneliti, mayoritas platform jual beli rumah *online* saat ini belum memiliki fitur yang dapat melakukan prediksi harga rumah berdasarkan karakteristiknya. Kehadiran fitur prediksi harga rumah dapat mengatasi ketidakpastian dan meningkatkan transparansi di pasar properti digital. Salah satu teknologi yang mampu menyikapi masalah tersebut adalah *machine learning*. *Machine learning* adalah salah satu cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan sedikit atau tanpa campur tangan manusia secara eksplisit. Konsep utama di balik *machine learning* adalah memberikan kemampuan pada sistem untuk meningkatkan kinerjanya seiring waktu berdasarkan

pengalaman dan data yang telah diterima. Di samping itu, masalah yang kerap dihadapi para pengembang *machine learning* yaitu kesulitan dalam mengoptimasi parameter model. Optimasi parameter dilakukan untuk bisa meningkatkan kinerja dari model yang dihasilkan. Salah satu metode yang umum dilakukan yaitu dengan memanfaatkan algoritma *grid search*. Penelitian terdahulu telah menunjukkan bahwa *machine learning* dapat signifikan meningkatkan akurasi prediksi. Penelitian [1] menyatakan bahwa dalam penentuan harga rumah, terdapat 5 variabel yang dapat berpengaruh. Berdasar survei yang dilakukannya kepada para pengembang (*developer*) rumah, variabel bebas terdiri dari luas lahan, luas bangunan, banyaknya kamar tidur, banyaknya kamar mandi, hingga ketersediaan tempat parkir mobil. Selain variabel tersebut, lokasi rumah juga ikut berperan dalam penentuan harga [2]. Pada penelitian [3] menyatakan bahwa konsep algoritma *support vector regression* (SVR) dapat menghasilkan nilai peramalan yang bagus karena SVR mempunyai kemampuan menyelesaikan *overfitting*. Selain itu, penelitian [4] menunjukan bahwa algoritma *grid search* mampu mengatasi kesulitan dalam menentukan parameter optimal dalam proses pengembangan model *support vector regression* (SVR). Oleh karena itu, dengan mengintegrasikan temuan penelitian sebelumnya, penelitian ini bertujuan menghasilkan model *support vector regression* (SVR) terbaik untuk prediksi harga rumah dengan optimasi algoritma *grid search*. Langkah ini berdasarkan pada hasil penelitian sebelumnya yang telah membuktikan efektivitas SVR dan *grid search* dalam memodelkan hubungan kompleks antara variabel bebas dan variabel target.

2. Metode Penelitian

Penelitian ini dilakukan berdasarkan tahapan metodologi OSEMN, dengan alat analisis data menggunakan google colaboratory versi python 3.10.12. Berikut adalah diagram alir dari tahapan penelititan:



Gambar 1. Diagram Alir Penelitian

Teknik analisis data merupakan tahapan kritis dalam proses penelitian. Tujuan analisis data di sini adalah untuk mengidentifikasi pola umum dari data yang terkumpul melalui proses pengolahan atau penjelasan data tersebut [5]. Berikut teknik analisis data yang akan dilakukan dalam penelitian ini:

2.1. Support Vector Regression

Support Vector Regression (SVR) merupakan teknik analisis data utama yang digunakan dalam penelitian ini. SVR adalah metode regresi yang berdasarkan konsep *Support Vector Machines* (SVM). SVM dirancang dan dikembangkan oleh Boser, Guyon, dan Vapnik, serta pertama kali diperkenalkan dalam *Annual Workshop on Computational Learning Theory* pada tahun 1992 [6].

p-ISSN: 2301-5373
e-ISSN: 2654-5101

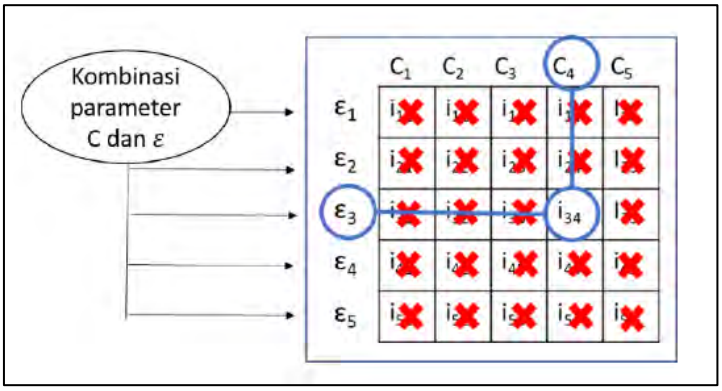
SVM mampumengatasi pemetaan data nonlinier dengan mengubah data uji asli ke dimensi yang lebih tinggi menggunakan fungsi kernel. Tujuan dari SVR adalah untuk mengidentifikasi suatu fungsi yang dapat direpresentasikan sebagai *hyperplane* atau garis pemisah dalam format fungsi regresi. Fungsi ini dimaksudkan untuk sesuai dengan seluruh data input dengan tingkat kesalahan yang minimal, sehingga membuat kesalahan tersebut sekecil mungkin [3]. Misalkan terdapat l data training, (xi,yi), i =1,…,l dengan data input x = {x_(1.)…,x_l} ⊆ ℜN dan y = {y_1,… ,y_l} ⊆ ℜ dan l adalah banyaknya data training. Fungsi regresi dari metode SVR adalah sebagai berikut:

$$f(x)= w\ \varphi(x)+b \qquad (1)\ [3]$$

Keterangan :
W : Vektor pembobot
φ(x) : Fungsi yang memetakan x dalam suatu dimensi
b : Bias

2.2. Algoritma Grid Search

Algoritma *Grid Search* adalah salah satu teknik pencarian parameter yang umum digunakan dalam tahap optimasi parameter model *Support Vector Regression* (SVR). Tujuannya adalah untuk menemukan parameter terbaik dalam dataset pelatihan agar model dapat dengan tepat memprediksi data uji [7]. Pada penerapannya, algoritma ini harus dipandu oleh beberapa metrik kinerja, dan biasanya diukur dengan *cross-validation* pada data latih [8]. *Cross validation* merupakan teknik evaluasi model yang membantu metode *grid search* dalam mengevaluasi kinerja model yang dihasilkan dari setiap kombinasi parameter. Proses kerja *cross validation* adalah dengan cara membagi dataset menjadi beberapa subset, yang disebut lipatan atau *folds*. Kemudian, model dilatih pada beberapa subset dan diuji pada subset yang tidak terlibat pelatihan. Proses ini berulang sampai hasil akhir dari *cross validation* berupa nilai rata-rata performa model yang konsisten. Pada penelitian ini dilakukan *cross validation* dengan jumlah *folds* sebanyak 5. Berikut pada gambar 2 merupakan ilustrasi dari proses metode *grid search*.



Gambar 2. Ilustrasi Algoritma Grid Search [4]

Pada penelitian ini, fungsi kernel yang dioptimasi yaitu kernel *Gaussian Radial Basis Function* (RBF). Hal ini didasarkan penelitian sebelumnya yang membuktikan bahwa kernel ini mampu menghasilkan kinerja model terbaik [9]. Pada SVR dengan kernel RBF, ada parameter lain yang harus dioptimasi yaitu C (*cost*), γ (gamma) dan ε (epsilon) [7]. Parameter *cost* berfungsi mengontrol *trade-off* antara presisi pemodelan data pelatihan dan kompleksitas model. Epsilon berperan menentukan batas kesalahan yang dapat diterima, sedangkan gamma mengontrol seberapa jauh pengaruh satu titik data terhadap titik data yang lain dalam sebuah model SVR. Fungsi kernel RBF dirumuskan dalam persamaan 2 berikut:

$$K(xi,xj) = \exp\left(-\frac{1}{2\sigma^2}\right)(Xi - Xj) \qquad (2)\ [10]$$

Keterangan :
Xi : Data ke – i
Xj : Data ke – j
σ : Standart deviasi
$\frac{1}{2\sigma^2}$: γ (gamma)

2.3. Evaluasi Model

Teknik ini melibatkan penggunaan berbagai metrik evaluasi yang dirancang khusus untuk tipe masalah yang dihadapi. Dalam kasus regresi, ada banyak metrik yang bisa dijadikan alat evaluasi. Pada penelitian ini, proses evaluasi model menggunakan metrik *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *R-Squared*.

Root Mean Square Error (RMSE) adalah metrik evaluasi yang mengukur seberapa besar perbedaan antara nilai prediksi yang dihasilkan oleh model dengan nilai aktual dalam skala yang sama. Untuk menghitung RMSE, langkah pertama dengan mengambil rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. Selanjutnya, nilai RMSE diperoleh dengan mengambil akar kuadrat dari rata-rata kuadrat selisih tersebut. RMSE memberikan gambaran tentang seberapa besar kesalahan prediksi secara keseluruhan, dengan nilai yang lebih kecil menunjukkan tingkat akurasi yang lebih tinggi. Secara matematis, rumus RMSE dijelaskan pada persamaan 3 berikut:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \quad (3)$$

Keterangan :
n : Jumlah sampel dalam pengujian
Yi : Nilai aktual dari observasi ke-i
$\hat{y}i$: Nilai prediksi dari model untuk observasi ke-i

Mean Absolute Error (MAE) adalah metrik yang mengukur kesalahan prediksi suatu model dengan menghitung rata-rata dari nilai absolut selisih antara prediksi dan nilai sebenarnya. Semakin kecil nilai RMSE, semakin baik model dapat melakukan prediksi. Dalam konteks matematis, MAE untuk suatu model dapat dijelaskan sebagai persamaan 4 berikut.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Yi - \hat{Y}i\right| \quad (4)$$

Keterangan :
n : Jumlah sampel dalam pengujian
Yi : Nilai aktual dari observasi ke-i
$\hat{Y}i$: Nilai prediksi dari model untuk observasi ke-i

Mean Absolute Percentage Error (MAPE) merupakan metrik evaluasi yang mengukur rata-rata dari persentase kesalahan absolut antara prediksi dan nilai sebenarnya. MAPE memberikan gambaran persentase kesalahan rata-rata antara prediksi dan nilai sebenarnya, dan seperti MAE, nilai yang lebih rendah menunjukkan kinerja model yang lebih baik. Pada bentuk matematisnya, rumus MAPE dijelaskan pada persamaan 5.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Yi - \hat{Y}i}{Yi}\right| \times 100 \quad (5)$$

Keterangan :

p-ISSN: 2301-5373
e-ISSN: 2654-5101

n	:	Jumlah sampel dalam pengujian
Yi	:	Nilai aktual dari observasi ke-i
$\widehat{Yi}$	:	Nilai prediksi dari model untuk observasi ke-i

R-squared (R^2) sering disebut sebagai koefisien determinasi adalah metrik evaluasi yang memberikan indikasi tentang seberapa besar variabilitas dalam variabel target yang dapat dijelaskan oleh model. Secara umum, *R-squared* menyatakan proporsi variasi dalam variabel target yang dapat dijelaskan oleh variabel independen yang ada dalam model. Metrik ini memiliki nilai rentang dari 0 hingga 1, dimana nilai yang mendekati nilai 1 menunjukan bahwa kinerja model semakin baik. Rumus *R-squared* dapat dinyatakan dalam persamaan 6.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \widehat{yi})^2}{\sum_{i=1}^{n}(yi - \overline{yi})^2} \quad (6)$$

Keterangan	:	
n	:	Jumlah sampel dalam pengujian
Yi	:	Nilai aktual dari observasi ke-i
$\widehat{yi}$	:	Nilai prediksi dari model untuk observasi ke-i
$\overline{yi}$	:	Nilai rata-rata dari variabel target

3. Hasil dan Pembahasan

Berikut merupakan hasil dari setiap tahapan metodologi penelitian:

3.1. Obtain

Data diperoleh melalui proses scrapping dari website rumah123.com. Pemilihan data disesuaikan dengan format data penjualan rumah pada platform tersebut. Lingkup data dibatasi dengan wilayah Jakarta. Dataset yang dihasilkan terdiri dari atribut waktu_scrap, deskripsi, alamat, narahubung, luas_tanah_m2, luas_bangunan_m2, kamar_tidur, kamar_mandi, garasi, dan harga_miliar. Baris data yang berhasil dikumpulkan sebanyak 9.665 record. Data tersebut kemudian disimpan di penyimpanan berbasis cloud milik peneliti dengan format CSV agar mempermudah dalam proses analisis data selanjutnya.

3.2. Scrub

Scrub melibatkan pembersihan dan pengolahan data awal pada dataset. Proses pembersihan memperhatikan faktor *missing values*, penyesuaian format, dan data *outlier*. Langkah-langkah tersebut diterapkan untuk menanggulangi ketidakpastian dan ketidakakuratan yang mungkin terjadi akibat proses *scrapping data*. Proses pembersihan data juga dapat berdampak signifikan pada kinerja sistem *data mining*, sebab penanganan data terkait akan mengalami pengurangan baik dalam hal jumlah maupun kompleksitas [11]. Metode *Inter Quartile Range* (IQR) diterapkan sebagai pendekatan dalam mendeteksi data *outlier*. Teknik ini dilakukan dengan cara mencari selisih antara kuartil ketiga dengan kuartil pertama [12]. Pada tahapan Scrub, dataset yang dihasilkan terdiri dari 7 kolom atribut. Informasi hasil pengolahan dataset dapat dilihat pada gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3990 entries, 5 to 9903
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   lokasi            3990 non-null   int64
 1   luas_tanah_m2     3990 non-null   float64
 2   luas_bangunan_m2  3990 non-null   float64
 3   kamar_tidur       3990 non-null   int64
 4   kamar_mandi       3990 non-null   int64
 5   kapasitas_garasi  3990 non-null   int64
 6   harga_miliar      3990 non-null   float64
dtypes: float64(3), int64(4)
memory usage: 249.4 KB
```

Gambar 3. Dataset Hasil Tahap Scrub

Gambar 3 menjelaskan dataset hasil proses Scrub, memuat informasi total baris data, nama kolom, tipe data, dan informasi kelengkapan data pada tiap atribut. Hasil tersebut menjelaskan bahwa atribut yang dipakai yaitu lokasi, luas_tanah_m2, luas_bangunan_m2, kamar_tidur, kamar_mandi, kapasitas_garasi, dan harga_miliar. Total baris data yang digunakan ke tahap selanjutnya sebanyak 3.990 *record*.

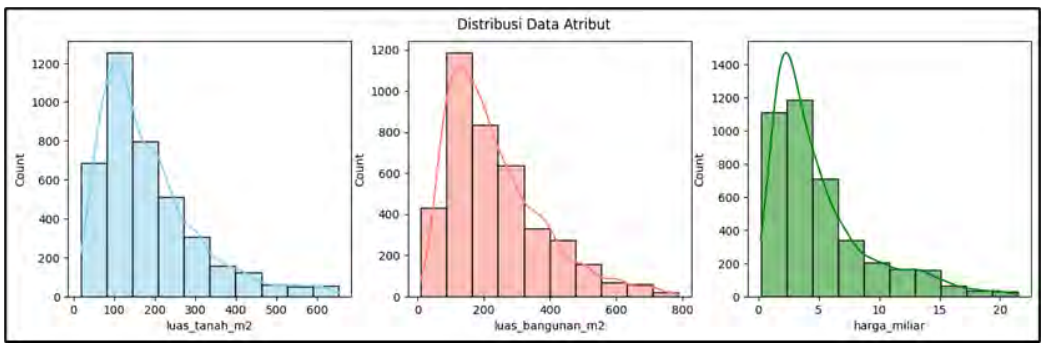
3.3. Explore

Tahapan Analisis eksploratif ini dilakukan untuk memahami pola-pola dan hubungan antarvariabel dalam dataset. Proses ini melibatkan statistika deksriptif, visualisasi distribusi data, dan perhitungan korelasi variabel menggunakan nilai korelasi *Pearson* [12]. Hasil proses Explore ditampilkan pada gambar berikut:

	lokasi	luas_tanah_m2	luas_bangunan_m2	kamar_tidur	kamar_mandi	kapasitas_garasi	harga_miliar
count	3990.000000	3990.000000	3990.000000	3990.000000	3990.000000	3990.000000	3990.000000
mean	2.175689	184.193734	232.573183	3.700251	3.059649	1.541604	5.026924
std	1.293663	120.602909	142.774869	1.103051	1.076167	0.588303	4.032648
min	0.000000	19.000000	10.000000	1.000000	1.000000	1.000000	0.181000
25%	1.000000	98.000000	122.000000	3.000000	2.000000	1.000000	2.180000
50%	2.000000	150.000000	200.000000	4.000000	3.000000	1.000000	3.600000
75%	3.000000	240.000000	300.000000	4.000000	4.000000	2.000000	6.500000
max	4.000000	653.000000	788.000000	8.000000	7.000000	3.000000	21.500000

Gambar 4. Hasil Statistika Deskriptif

Gambar 4 tersebut menunjukan perhitungan statistika deskriptif dari setiap atribut. Statistika deskriptif bertujuan untuk mengetahui nilai-nilai khas atau karakteristik dari sekumpulan data. Nilai *count* merupakan banyaknya data pada tiap atribut kolom. Nilai *mean* adalah nilai rata-rata atribut dan std merupakan nilai simpangan baku. Untuk nilai 25%, 50%, dan 75% mewakili dari nilai kuantil satu sampai tiga dari setiap atribut. baris *min* dan *max* adalah kependekan dari nilai minimal dan maksimal dari setiap kumpulan data.



Gambar 5. Distribusi Data Atribut

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Gambar 5 menunjukan distribusi data atribut luas_tanah_m2, luas_bangunan_m2, dan harga_miliar. Ketiga *chart* tersebut menunjukan bahwa kebanyakan data berkumpul pada nilai yang lebih rendah yang menyebabkan distribusi menjadi asimetris.

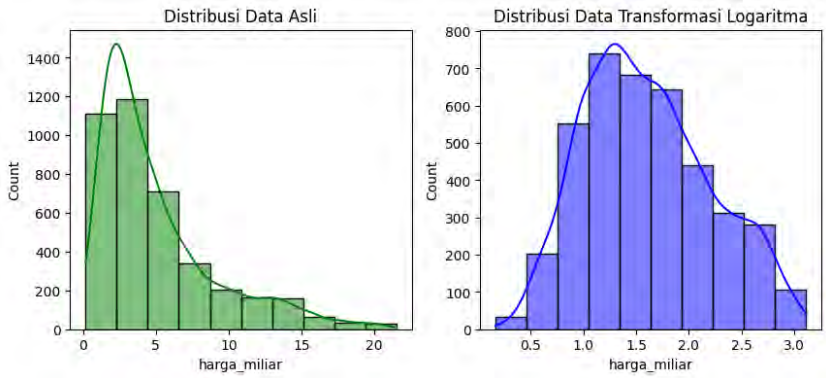
	lokasi	luas_tanah_m2	luas_bangunan_m2	kamar_tidur	kamar_mandi	kapasitas_garasi	harga_miliar
lokasi	1.000000	-0.119756	-0.118734	-0.053385	-0.024096	-0.019982	-0.252905
luas_tanah_m2	-0.119756	1.000000	0.675995	0.469173	0.361070	0.329136	0.744062
luas_bangunan_m2	-0.118734	0.675995	1.000000	0.527637	0.564736	0.351934	0.729006
kamar_tidur	-0.053385	0.469173	0.527637	1.000000	0.627710	0.266849	0.392201
kamar_mandi	-0.024096	0.361070	0.564736	0.627710	1.000000	0.310474	0.411164
kapasitas_garasi	-0.019982	0.329136	0.351934	0.266849	0.310474	1.000000	0.330251
harga_miliar	-0.252905	0.744062	0.729006	0.392201	0.411164	0.330251	1.000000

Gambar 6. Nilai Korelasi Pearson

Gambar 6 tersebut menunjukan nilai korelasi atau hubungan antar atribut dataset. Dari hasil tersebut terlihat bahwa semua variabel kriteria memiliki korelasi positif dengan variabel target, kecuali atribut lokasi.

3.4. Model

Langkah krusial dalam pelaksanaan Model adalah menggunakan data baru untuk menghasilkan perkiraan kasus yang relevan, sehingga dapat diuji keberlakuan model yang telah dibangun dalam memberikan solusi yang sesuai terhadap permasalahan yang dihadapi [13]. Berdasarkan *chart* distribusi data pada variabel luas_tanah_m2, luas_bangunan_m2, dan harga_miliar, terindikasi adanya distribusi data asimetris berjenis *right-skewed*. Distribusi data ini ditandai dengan berkumpulnya sebaran data pada nilai yang mendekati nilai nol. Distribusi data asimetris dapat mempengaruhi tingkat akurasi dalam proses pengembangan model. Dalam mengatasi hal tersebut, transformasi logaritma dapat dilakukan untuk mendekati distribusi normal [14]. Menurut Gujarati dan Porter (2009) dalam paper [15] menyebutkan salah satu model dengan transformasi variabel adalah model semilog, yaitu salah satu variabel muncul dalam bentuk logaritmik, atau dapat disebut model log-lin atau lin-log. Oleh karena itu, dua skenario pemodelan diterapkan, yaitu dengan data asli dan data hasil transformasi logaritma natural pada variabel target. Hal ini dilakukan untuk melihat perbedaan hasil evaluasi model dari kedua skenario tersebut.



Gambar 7. Distribusi Data Target Asli dan Hasil Transformasi

Gambar 7 tersebut menjelaskan perbandingan visualisasi distribusi data asli dengan hasil transformasi pada variabel target harga_miliar. Dari gambar tersebut terlihat bahwa transformasi logaritma natural dapat mengubah distribusi asimetris mendekati distribusi normal.

Pada proses *splitting* data, dilakukan pembagian dengan perbandingan 80% data latih dan 20% data uji [16]. Pembagian data menghasilkan data latih sebanyak 3.192 baris dan 798 data uji. Proses pembagian ini dilakukan pada masing-masing skenario pemodelan.

Proses normalisasi data pada penelitian ini menggunakan teknik *min-max scalling* dengan kisaran nilai antara 0 dan 1. Normalisasi data dilakukan untuk menyamakan skala pada setiap atribut input [17]. Hal ini dilakukan sebagai salah satu upaya dalam mempercepat performa model yang akan dibangun dalam memahami pola-pola dalam data.

```
array([[0.25      , 0.13091483, 0.24967148, 0.42857143, 0.33333333,
        0.        ],
       [0.25      , 0.09621451, 0.13272011, 0.42857143, 0.5       ,
        0.5       ],
       [0.25      , 0.18454259, 0.42049934, 0.28571429, 0.5       ,
        0.5       ],
       ...,
       [0.75      , 0.14037855, 0.18396846, 0.28571429, 0.5       ,
        0.        ],
       [0.25      , 0.28548896, 0.24967148, 0.57142857, 0.33333333,
        0.        ],
       [1.        , 0.15930599, 0.16557162, 0.28571429, 0.33333333,
        0.5       ]])
```

Gambar 8. Data Hasil Normalisasi *Min-Max Scalling*

Gambar 8 menunjukkan hasil perubahan skala pada masing-masing variabel independen menjadi skala yang sama. Proses transformasi ini dilakukan untuk mempermudah proses pemodelan antara variabel independen dengan variabel target di dalam proses Model.

```
Parameter terbaik setelah grid search: {'C': 10, 'epsilon': 1, 'gamma': 1, 'kernel': 'rbf'}
Root Mean Squared Error (RMSE): 2.175646150188057
Mean Absolute Error (MAE): 1.3630525890121044
Mean Absolute Percentage Error (MAPE): 0.3224556491060808
R^2 Score: 0.7028737429175912
```

Gambar 9. Parameter Optimal dan Evaluasi Model Data Asli

Gambar 9 menunjukkan parameter optimal dari proses *grid search* dan hasil evaluasi model menggunakan data asli. Dilihat berdasarkan metrik *R-Squared*, model yang dihasilkan cukup baik dalam memodelkan data input dan data target.

```
Parameter terbaik setelah grid search: {'C': 1, 'epsilon': 0.1, 'gamma': 1, 'kernel': 'rbf'}
Root Mean Squared Error (RMSE): 0.2774275753583595
Mean Absolute Error (MAE): 0.20611210095739088
Mean Absolute Percentage Error (MAPE): 0.1452525638876444
R^2 Score: 0.7866721130857308
```

Gambar 10. Parameter Optimal dan Evaluasi Model Data Hasil Transformasi

Gambar 10 tersebut menunjukkan nilai evaluasi model menggunakan data hasil transformasi logaritma, yang mana hasil tersebut juga merupakan akibat dari parameter optimal metode *grid search*. Berdasarkan nilai *error* dari metrik RMSE, MAE, dan MAPE, kinerja dapat dikategorikan baik karena nilainya yang cenderung kecil.

3.5. Interpret

Pada tahapan metodologi yang telah dilakukan, hasil dapat diinterpretasikan sebagai berikut:

a. Tahapan Obtain menggunakan teknik *scrapping* pada *website* rumah123.com menghasilkan data sebanyak 9.665 *record*. Dataset terdiri 10 atribut dengan kondisi data belum dilakukan pembersihan. Dataset yang terkumpul disimpan peneliti dengan format CSV dengan tujuan mempermudah proses analisis data selanjutnya.

b. Tahapan Scrub dilakukan proses pembersihan dan pengolahan data awal. Kegiatan ini dilakukan dengan memperhatikan faktor *missing values,* tidak konsisten format, dan data *outlier*. Hasil dataset berupa 7 kolom atribut dengan total data sebanyak 3.990 *record*. Atribut data terdiri dari lokasi, luas_tanah_m2, luas_bangunan_m2, kamar_tidur, kamar_mandi, kapasitas_garasi, dan harga_miliar.
c. Tahapan Explore dilakukan untuk mengetahui karakteristik data secara mendalam. Ditemukan indikasi distribusi data asimetris pada variabel luas_tanah_m2, luas_bangunan_m2, dan harga_miliar. Hal ini ditandai dengan kebanyakan data berkumpul pada nilai yang lebih rendah dan mendekati nilai nol. Nilai korelasi *Pearson* variabel bebas terhadap variabel target mayoritas positif terkecuali atribut lokasi.
d. Tahapan Model yang dilakukan menggunakan dua skenario data, data asli dan data hasil transformasi logaritma pada variabel target. Hal ini didasari distribusi data yang asimetris. Pada evaluasi model, data asli menghasilkan metrik RMSE = 2.175646150188057, MAE = 1.3630525890121044, MAPE = 0.3224556491060808, dan *R-Squared* = 0.7028737429175912 dengan parameter optimal *Cost* = 10, epsilon = 1, dan gamma = 1. Data hasil transformasi menghasilkan RMSE = 0.2774275753583595, MAE = 0.20611210095739088, MAPE = 0.1452525638876444, dan *R-Squared* = 0.7866721130857308 dengan parameter optimal *Cost* = 1, epsilon = 0.1, dan gamma = 1.

4. Kesimpulan

Berdasarkan hasil dari seluruh tahapan penelitian, dapat ditarik kesimpulan sebagai berikut:

a. Model terbaik dan parameter optimal dihasilkan dari skenario data hasil transformasi logaritma. Hal tersebut berdasarkan nilai metrik koefisiensi determinasi yang baik dengan nilai metrik *error* yang cenderung kecil dibandingkan model data asli. Model tersebut menghasilkan RMSE = 0.2774275753583595, MAE = 0.20611210095739088, MAPE = 0.1452525638876444, dan *R-Squared* = 0.7866721130857308 dengan parameter optimal *Cost* = 1, epsilon = 0.1, dan gamma = 1.
b. Berdasarkan hasil model terbaik, kualitas data sangat berpengaruh terhadap akurasi model yang dihasilkan. Kesimpulan ini didukung oleh penelitian sebelumnya yang menyatakan bahwa pengolahan data awal sangat mempengaruhi tingkat akurasi dari model yang dihasilkan [1].
c. Distribusi data simetris berperan dalam peningkatan akurasi model. Hal ini terbukti dari perbedaan hasil evaluasi dari kedua skenario pemodelan, distribusi yang lebih simetris menghasilkan nilai metrik yang lebih baik. Hasil penelitian ini juga telah terbukti pada penelitian sebelumnya dengan objek teliti yang berbeda [15].

References

[1] A. Saiful, S. Andryana, eta A. Gunaryati, «Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression», *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, libk. 8, zenb. 1, or. 41–50, 2021, doi: 10.35957/jatisi.v8i1.701.

[2] F. R. Lumbanraja, R. A. Saputra, K. Muludi, A. Hijriani, eta A. Junaidi, «Implementasi Support Vector Machine dalam Memprediksi Harga Rumah pada Perumahan di Kota Bandar Lampung», *J. Pepadun*, libk. 2, zenb. 3, or. 327–335, 2021, doi: 10.23960/pepadun.v2i3.90.

[3] Z. Rais, R. Isnaeni, eta Sudarmin, «Analisis Support Vector Regression (SVR) Dengan Kernel Radial Basis Function (RBF) Untuk Memprediksi Laju Inflasi Di Indonesia», *VARIANSI J. Stat. Its Appl. Teach. Res.*, libk. 4, zenb. 1, or. 30–38, 2022, doi: 10.35580/variansiunm13.

[4] G. H. Saputra, A. H. Wigena, eta B. Sartono, «Penggunaan Support Vector Regression Dalam Pemodelan Indeks Saham Syariah Indonesia Dengan Algoritme Grid Search», *Indones. J. Stat. Its Appl.*, libk. 3, zenb. 2, or. 148–160, 2019, doi: 10.29244/ijsa.v3i2.172.

[5] T. Prasetiya, I. Ali, C. L. Rohmat, eta O. Nurdiawan, «Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor», *INFORMATICS Educ. Prof. J. Informatics*, libk. 5, zenb. 1, or. 93, 2020, doi: 10.51211/itbi.v5i1.1431.

[6] Syafi'i, O. Nurdiawan, eta G. Dwilestari, «Penerapan Machine Learning Untuk Menentukan Kelayakan Kredit Menggunakan Metode Support Vektor Machine», *J. Sist. Inf. dan Manaj.*, libk. 10, zenb. 2, or. 1–6, 2022.

[7] D. I. Purnama, «Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR)», *Jambura J. Math.*, libk. 2, zenb. 2, or. 49–59, 2020, doi: 10.34312/jjom.v2i2.4458.

[8] H. Yasin, A. Prahutama, eta T. W. Utami, «Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search», *Media Stat.*, libk. 7, zenb. 1, or. 29–35, 2014, doi: 10.14710/medstat.7.1.29-35.

[9] R. E. Cahyono, J. P. Sugiono, eta S. Tjandra, «Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen», *J. Teknol. Inf. dan Multimed.*, libk. 1, zenb. 2, or. 106–116, 2019, [Sarean]. Available at: www.siskaperbapo.com

[10] N. D. Maulana, B. D. Setiawan, eta C. Dewi, «Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery)», *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, libk. 3, zenb. 3, or. 2986–2995, 2019.

[11] O. Nurdiawan, A. Irma Purnamasari, eta I. Ali, «Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana», *J. Data Sci. dan Inform.*, libk. 1, zenb. 2, or. 32–35, 2021.

[12] K. R. Putra, S. Umaroh, N. Fitrianti, eta S. Nugraha, «RESULTANT: Data Preparation Techniques to Improve XGBoost Algorithm Performance», *MIND (Multimedia Artif. Intell. Netw. Database) J.*, libk. 8, zenb. 1, or. 42–51, 2023.

[13] F. Firmansyah eta O. Nurdiawan, «Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals», *JATI (Jurnal Mhs. Tek. Inform.*, libk. 7, zenb. 1, or. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.

[14] I. Setiawan, R. Fina Antika Cahyani, eta I. Sadida, «EXPLORING COMPLEX DECISION TREES : UNVEILING DATA PATTERNS AND OPTIMAL PREDICTIVE POWER», *J. Innov. Futur. Technol.*, libk. 5, zenb. 2, or. 112–123, 2023.

[15] D. Wasani eta S. I. Purwanti, «The GRDP Per Capita Gap between Provinces in Indonesia and Modeling with Spatial Regression», *J. Mat. Stat. dan Komputasi*, libk. 19, zenb. 1, or. 65–78, 2022, doi: 10.20956/j.v19i1.20997.

[16] P. Putriyana eta O. Nurdiawan, «Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa SMK Al Huda Kedungwungu Dengan Menggunakan Algoritma Naïve Bayes Classifier», *Exp. Student Exp.*, libk. 1, zenb. 1, or. 1–7, 2023.

[17] A. Toha, P. Purwono, eta W. Gata, «Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV», *Bul. Ilm. Sarj. Tek. Elektro*, libk. 4, zenb. 1, or. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.

Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 12, No 3. Februari 2024 e-ISSN: 2654-5101

Implementasi Docker Container untuk Sistem Monitoring dan Pengontrolan Peralatan Listrik di Laboratorium Cerdas

Sahirul Alam[a1], Sri Lestari[a2], Anni Karimatul Fauziyyah[a3], Dzulfikar[a4]

[a]Departemen Teknik Elektro dan Informatika, Sekolah Vokasi, Universitas Gadjah Mada
Gedung Herman Yohanes, Sekip Unit III, Catur Tunggal, Depok, Sleman, Yogyakarta, Indonesia
[1]sahirul.alam@ugm.ac.id
[2]srilestari59@ugm.ac.id
[3]anni.karimatul.f@ugm.ac.id
[4]dzulfikar.sv@mail.ugm.ac.id

Abstract

Smart laboratory atau laboratorium cerdas adalah laboratorium yang dilengkapi dengan teknologi canggih seperti Internet of Things (IoT), robotika, dan kecerdasan buatan (Artificial Intelligence/AI) untuk meningkatkan efisiensi, akurasi, dan keamanan dalam melakukan penelitian dan pengujian. Dalam sebuah smart laboratory, perangkat IoT dapat digunakan untuk memantau suhu, kelembaban, dan kualitas udara di dalam ruangan, sehingga dapat memastikan kondisi lingkungan yang ideal untuk menjaga kualitas sampel yang diuji. IoT sendiri adalah sebuah konsep yang mengacu pada konektivitas antara berbagai perangkat atau objek yang terhubung ke internet, sehingga memungkinkan perangkat tersebut saling berkomunikasi dan bertukar data. Penelitian ini bertujuan untuk membangun sebuah sistem pemantauan dan pengontrolan peralatan listrik di dalam laboratorium. Sistem yang dibangun menerapkan teknologi IoT sehingga pemantauan dan pengontrolan peralatan listrik akan menjadi lebih mudah dan dapat dilakukan dari mana saja. Selain itu, sistem memanfaatkan teknologi docker container sehingga instalasi dan pengelolaan perangkat lunak dapat dilakukan dengan lebih mudah dan efisien.

Keywords: *Docker Container, Internet of Things, Smart Laboratory, Monitoring System, Energy Efficiency*

1. Pendahuluan

Internet of Things (IoT) adalah sebuah konsep yang mengacu pada konektivitas antara berbagai perangkat atau objek yang terhubung ke internet, sehingga memungkinkan perangakt tersebut untuk saling berkomunikasi dan bertukar data. Secara sederhana, IoT menggambarkan jaringan perangkat yang terhubung secara online, sehingga dapat saling berkomunikasi dan berinteraksi satu sama lain [1]. Perangkat IoT dapat berupa segala jenis perangkat elektronik, mulai dari sensor dan kamera hingga kendaraan dan peralatan rumah tangga, bahkan hewan peliharaan yang dikenakan alat pelacak.

Dalam aplikasi praktis, teknologi IoT dapat digunakan dalam berbagai macam industri, seperti pertanian [2], manufaktur, transportasi, kesehatan, dan rumah tangga. Salah satu manfaat utama dari IoT adalah memungkinkan pengumpulan data yang lebih banyak dan akurat, yang dapat digunakan untuk membantu membuat keputusan yang lebih baik dan meningkatkan efisiensi dalam bisnis atau aktivitas sehari-hari. Namun, keamanan dan privasi menjadi isu penting yang harus diperhatikan dalam pengembangan dan penggunaan teknologi IoT [3].

Smart laboratory atau laboratorium cerdas adalah laboratorium yang dilengkapi dengan teknologi canggih seperti IoT, robotika, dan kecerdasan buatan untuk meningkatkan efisiensi, akurasi, dan keamanan dalam melakukan penelitian dan pengujian. Dalam sebuah smart laboratory, perangkat IoT dapat digunakan untuk memantau suhu, kelembaban, dan kualitas udara di dalam ruangan [4], sehingga dapat memastikan kondisi lingkungan yang ideal untuk menjaga kualitas sampel yang diuji. Selain itu suhu dan kelembaban ruangan juga penting untuk dijaga agar peralatan di laboratorium tidak mengalami kerusakan akibat kelembaban tinggi seperti jamur dan karat.

Penelitian ini bertujuan untuk membangun sistem monitoring dan pengontrolan peralatan listrik di laboratorium cerdas. Peralatan listrik yang dikendalikan dengan IoT yaitu lampu penerangan dan AC. Sedangkan monitoring dilakukan untuk memantau dan mengumpulkan data terkait suhu dan

kelembaban ruang laboratorium. Penerapan IoT ke dalam sistem yang dibangun dengan memanfaatkan docker container. Docker merupakan seperangkat platform yang menggunakan teknologi virtualisasi pada level sistem operasi untuk menjalankan perangkat lunak dengan memuat semua dependensi yang diperlukan di dalam container [5]. Dengan menggunakan docker container, instalasi dan pengelolaan perangkat lunak IoT dapat dilakukan dengan lebih mudah dan efisien.

Beberapa penelitian terkait penerapan IoT untuk mengendalikan perangkat listrik antara lain untuk pengontrolan lampu [6] [7] dan untuk pengontrolan kipas angin dan AC [8]. Teknologi IoT memang telah banyak diimplementasikan untuk pengendalian peralatan listrik. Namun, belum banyak penelitian yang membahas tentang penggunaan docker container untuk aplikasi IoT. Penelitian [9] menggunkan docker container untuk diterapkan pada pengontrolan lampu penerangan di gedung. Perbedaan penelitian tersebut dengan penelitian ini antara lain adalah penggunaan perangkat keras, perangkat lunak, dan juga pemanfaatan. Dari segi perangkat keras, penelitian [9] menggunakan komputer dengan spesifikasi menengah ke atas.

Beberapa penelitian terkait monitoring kualitas udara ruangan dapat ditemui di literatur. Penelitian [10] dan [11] memanfaatkan IoT untuk melakukan monitoring kualitas udara terkait kadar CO2 dan resiko infeksi virus serta melakukan pengontrolan buka tutup jendela. Penelitian [12] memanfaatkan IoT dan CCTV untuk monitoring suhu dan kelembaban ruang pengering umbi bawang merah. Penelitian [13] memanfaatkan IoT untuk monitoring suhu dan kelembaban ruang produksi obat serta melakukan pengontrolan kipas dan dehumidifier. Selain itu, beberapa penelitian juga memanfaat IoT untuk monitoring suhu dan kelembaban udara di ruang sarang burung walet [14], ruang budidaya jamur tiram putih [15], ruang penyimpanan obat [16], hingga ruang server [17] [18]. Perbedaan utama penelitian ini dengan penelitian yang telah disebutkan adalah bahwa penelitian ini mengimplementasikan IoT untuk monitoring dan pengontrolan peralatan listrik di laboratorium, terutama untuk memantau suhu dan kelembaban ruang penyimpanan peralatan laboratorium. Selain itu, penelitian ini membahas implementasi docker container untuk sistem monitoring sebagai kebaruan yang belum pernah dibahas di penelitian yang sudah ada.

2. Metodologi

2.1 Alur Pengembangan Sistem

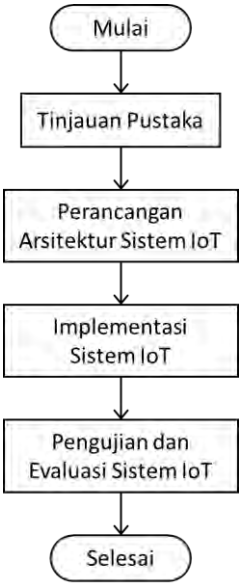
Pada penelitian ini, teknologi IoT akan diimplementasikan untuk mewujudkan konsep laboratorium cerdas (smart laboratory). IoT digunakan untuk memantau dan mengontrol peralatan listrik yang ada di laboratorium seperti lampu, air conditioner, dan LED proyektor. Pada konsep laboratorium cerdas, peralatan listrik tersebut harus dapat dikontrol secara otomatis untuk mendukung suasana laboratorium yang nyaman namun tetap memperhatikan efisiensi penggunaan daya listrik.

Untuk penerangan laboratorium, sistem IoT akan mengatur pengoperasian lampu dengan mempertimbangkan ada tidaknya orang dan intensitas cahaya di dalam ruangan. Oleh karena itu, dua jenis sensor akan digunakan yaitu sensor passive infra-red (PIR) dan light depending resitor (LDR). Selanjutnya, air conditioner akan dioperasikan untuk mengatur suhu dan kelembaban yang kondusif di laboratorium. Dalam hal ini, sistem akan menggunakan sensor DHT-11. Untuk memantau penggunaan daya listrik, KWh meter akan dipasang di sumber tegangan peralatan listrik. Secara sederhana, jalannya penelitian direncakan seperti pada Gambar 1. Tinjauan pustaka dilakukan untuk mengetahui trend impementasi IoT terutama untuk pengontrolan peralatan listrik dan pemantauan kondisi ruangan. Perancangan sistem IoT meliputi pemilihan perangkat keras dan perangkat lunak IoT. Dalam hal ini, peralatan perlu menyesuaikan dengan kondisi ruangan tempat IoT akan diimplementasikan. Setelah sistem IoT diimplementasikan, maka pengujian perlu dilakukan untuk mengetahui fungsionalitas dan kinerja sistem IoT yang sudah dibuat.

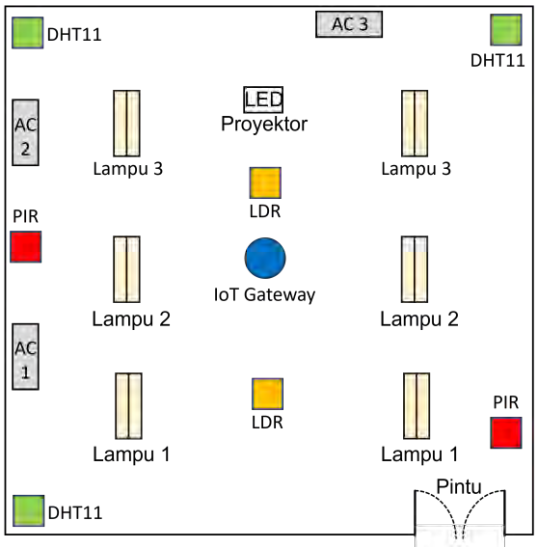
2.2 Perakitan Perangkat Keras

Perancangan arsitektur IoT perlu disesuaikan dengan kondisi ruangan. Dalam hal ini, sistem IoT akan diterapkan di Laboratorium Telekomunikasi Departemen Teknik Elektro dan Informatika, Sekolah Vokasi, Universitas Gadjah Mada. Ilustrasi rencana implementasi sistem IoT di laboratorium adalah seperti ditunjukkan pada Gambar 2. Setiap sensor dan aktuator di Gambar 2 dilengkapi dengan IoT kit. Sedangkan IoT kit sendiri tersusun dari beberapa komponen seperti ditunjukkan pada Gambar 3. Secara lebih spesifik, perangkat keras yang digunakan untuk sistem IoT ditunjukkan pada Tabel 1.

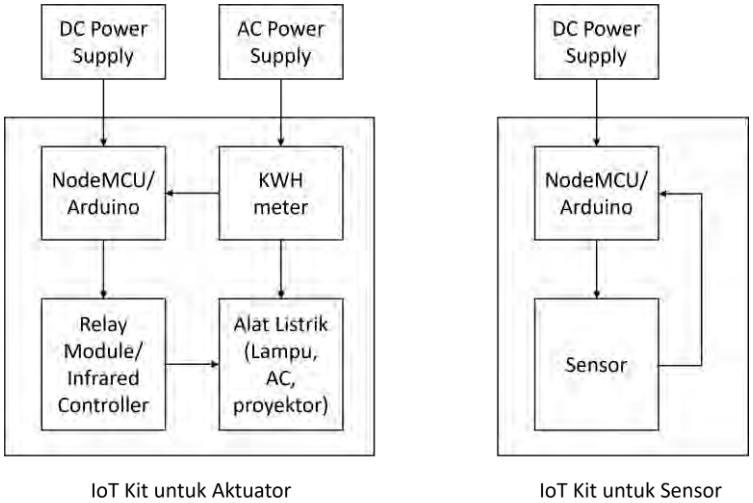
p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 1. Alur pengembangan sistem



Gambar 2. Rencana implementasi sistem IoT di laboratorium

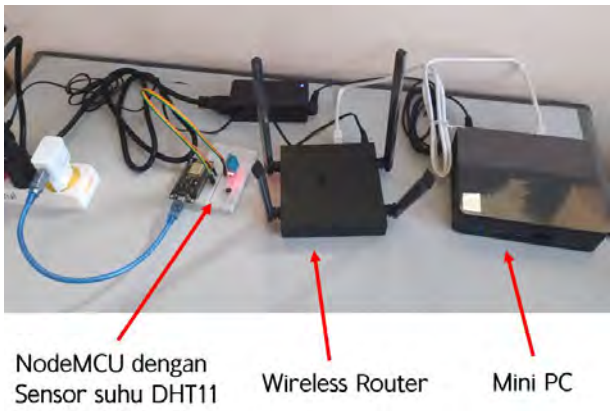


Gambar 3. Komponen penyusun IoT-kit untuk sensor dan aktuator

Tabel 1. Spesifikasi perangkat keras yang digunakan

Perangkat Keras	Spesifikasi/Jenis
NodeMCU board	ESP8266 Lolin
Relay Module	KY-019 1-Channel
Sensor Suhu dan Kelembaban	DHT11
Sensor Gerak	HC-SR501 Human Sensing PIR
IR Transmitter	KY-005
Sensor Tegangan AC	ZMPT101B
Sensor Arus AC	100A SCT-013-000 Non-invasive
IoT Gateway	Wireless Router TP-Link Archer C54
IoT Server	Intel NUC11ATKC4

Pengukuran konsumsi daya listrik menggunakan kWh meter yang terdiri dari sensor tegangan ZMPT101B dan sensor arus 100A SCT-013-000. IoT server yang sudah lebih dulu populer adalah menggunakan Raspberry. Namun, pada penelitian ini, IoT server menggunakan mini PC karena dengan kisaran harga yang hampir sama dengan Raspberry, namun mini PC memiliki skalabilitas yang lebih baik, misalnya dapat ditingkatkan kapasitas memori maupun penyimpanannya. IoT Gateway untuk komunikasi data semua perangkat IoT dan server menggunakan wireless router seperti ditunjukkan pada Gambar 4.



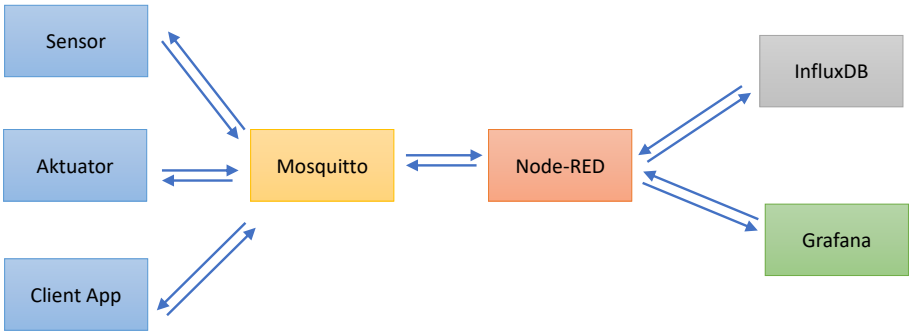
Gambar 4. Perangkat Keras IoT untuk Laboratorium Cerdas

2.3 Instalasi dan Konfigurasi Perangkat Lunak Menggunakan Docker Container

Perangkat lunak dalam sistem IoT digunakan untuk mengatur komunikasi data antara sensor, aktuator, server, dan perangkat pengguna. Selain itu, perangkat lunak juga digunakan untuk mengatur tampilan untuk monitoring kondisi ruangan. Perangkat lunak sistem IoT diinstall di dalam server yang menggunakan sistem operasi Ubuntu Server. Perangkat lunak tersebut diinstall dengan menggunakan docker container. Kelebihan metode instalasi ini yaitu dapat dilakukan dengan cepat dan lebih mudah untuk melakukan manajemen dan konfigurasi perangkat lunak. Secara umum, perangkat lunak yang dijalankan di dalam docker container serta ilustrasi alur komunikasi datanya ditunjukkan pada Gambar 5. Mosquitto adalah perangkat lunak yang digunakan sebagai broker IoT yang berperan mengatur komunikasi data antar perangkat IoT. InfluxDB merupakan perangkat lunak basis data yang digunakan untuk menyimpan data hasil pengukuran dari perangkat IoT. Grafana adalah perangkat lunak front end yang digunakan untuk menampilkan visualisasi data untuk keperluan monitoring kondisi ruangan. Data yang dikirim oleh Mosquitto memiliki format JSON. Oleh karena itu, Node-RED yang berperan sebagai penghubung komunikasi data antar perangkat lunak memiliki kemampuan untuk melakukan interpretasi data dalam format JSON.

Dari sistem operasi, IoTStack dapat dapat diunduh dengan menggunakan command seperti pada Gambar 6, kemudian reboot perlu dilakukan setelah downdload selesai. Setelah itu, instalasi dapat dilanjutkan dengan menggunakan secure socket shell (SSH) dan membuka folder IoTStack. Menu instalasi akan muncul dan perangkat lunak yang dipilih untuk diinstall adalah Mosquitto, Node-RED, InfluxDB, Grafana, dan Portainer-CE. Untuk memulai proses pembuatan container setelah memilih perangkat lunak dapat dilakukan dengan command “Start Stack”. Setelah proses instalasi selesai, pengecekan dapat dilakukan untuk memastikan semua program yang dipilih berhasil diinstall dengan menggunakan command “docker-compose ps”.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 5. Arah komunikasi data dari perangkat keras dan perangkat lunak IoT

```
curl -fsSL https://raw.githubusercontent.com/SensorsIot/IOTstack/master/install.sh | bash
sudo shutdown -r now
```

Gambar 6. Perintah untuk mengunduh IoTStack

Pembuatan database menggunakan InfluxDB dapat dilakukan dengan memasuki influx docker container, kemudian menambahkan database dngan perintah seperti Gambar 7. Data yang dikirimkan dari Mosquitto menggunakan format JSON, sehingga Node-RED perlu menginterpretasi data dalam format JSON tersebut.

```
docker exec -it influxdb influx
CREATE DATABASE sensor_data
quit
```

Gambar 7. Perintah untuk menambahkan database di InfluxDB

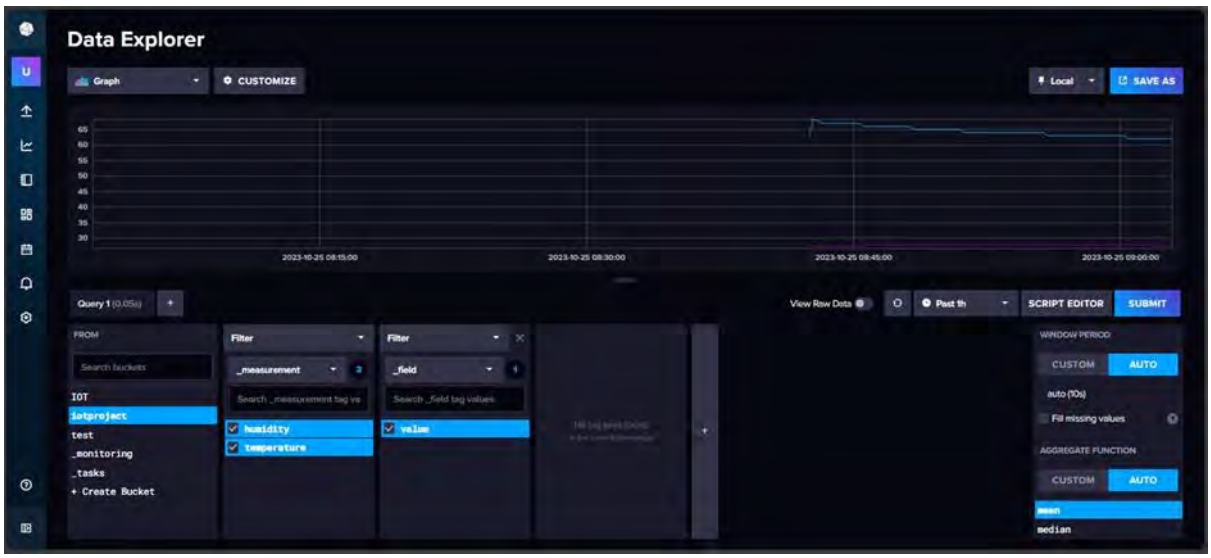
3. Hasil dan Pembahasan

Pengujian sistem yang telah dibangun dilakukan dengan melakukan pengontrolan baik secara manual maupun otomatis. Pengontrolan manual dapat dilakukan melalui perangkat pengguna yang telah terlebih dahulu diinstall aplikasi MQTT client. Pengontrolan yang dapat dilakukan secara manual antara lain ON-OFF peralatan listrik seperti lampu, AC, dan LCD proyektor. Monitoring kondisi ruangan seperti suhu dan kelembaban utara dapat dilakukan dengan mengakses Grafana menggunakan browser. Caranya yaitu dengan mengakses alamat IP server kemudian diikuti port 3000. Langkah selanjutnya adalah proses autentikasi dengan memasukkan username dan password, kemudian dashboard grafana dapat diakses. Ilustrasi tampilan dashboard Grafana untuk monitoring suhu dan kelembaban udara ruangan laboratorium adalah seperti ditunjukkan pada Gambar 8.

Data hasil pengukuran sensor disimpan di dalam basis data. Data tersebut dapat diakses dengan membuka InfluxDB menggunakan browser melalui alamat IP server dengan port 8086. Tampilan InfluxDB ditunjukkan pada Gambar 9. Data yang tersimpan di basis data dapat diunduh dalam format CSV. Contoh log data temperatur dan kelembaban ruangan tanggal 11 Oktober 2023 mulai dari pukul 00:00 hingga 23:59 WIB ditunjukkan pada Tabel 2. Data dari file CSV menggunakan periode sampling 5 detik sehingga ada 5 data pengukuran tiap menit. Namun, data pada Tabel 2 yang diambil dari file CSV menggunakan 1 sampel per jam. Data pada Tabel 2 selanjutnya ditampilkan dalam plot grafik seperti ditunjukkan pada Gambar 10.

Grafik temperatur dan kelembaban udara ruangan di Gambar 10 menunjukkan keterkaitannya dengan okupansi atau ada tidaknya orang di dalam ruangan. Suhu dan kelembaban yang tinggi biasanya terjadi di luar jam kerja. Pada saat jam kerja atau ada kegiatan di dalam laboratorium, suhu dan kelembaban menurun karena AC dinyalakan. Berdasarkan data hasil monitoring tersebut, dapat disimpulkan bahwa pengontrolan kelembaban udara ruangan masih kurang optimal terutama untuk ruang penyimpanan peralatan laboratorium. Oleh karena itu, penelitian selanjutnya dapat diarahkan untuk mengoptimalkan pengontrolan kelembaban udara dengan mengatur pengoperasian AC namun tetap mempertimbangkan efisiensi konsumsi daya.

Gambar 8. Tampilan dashboard Grafana untuk monitoring kondisi udara ruangan



Gambar 9. Tampilan dashboard InfluxDB

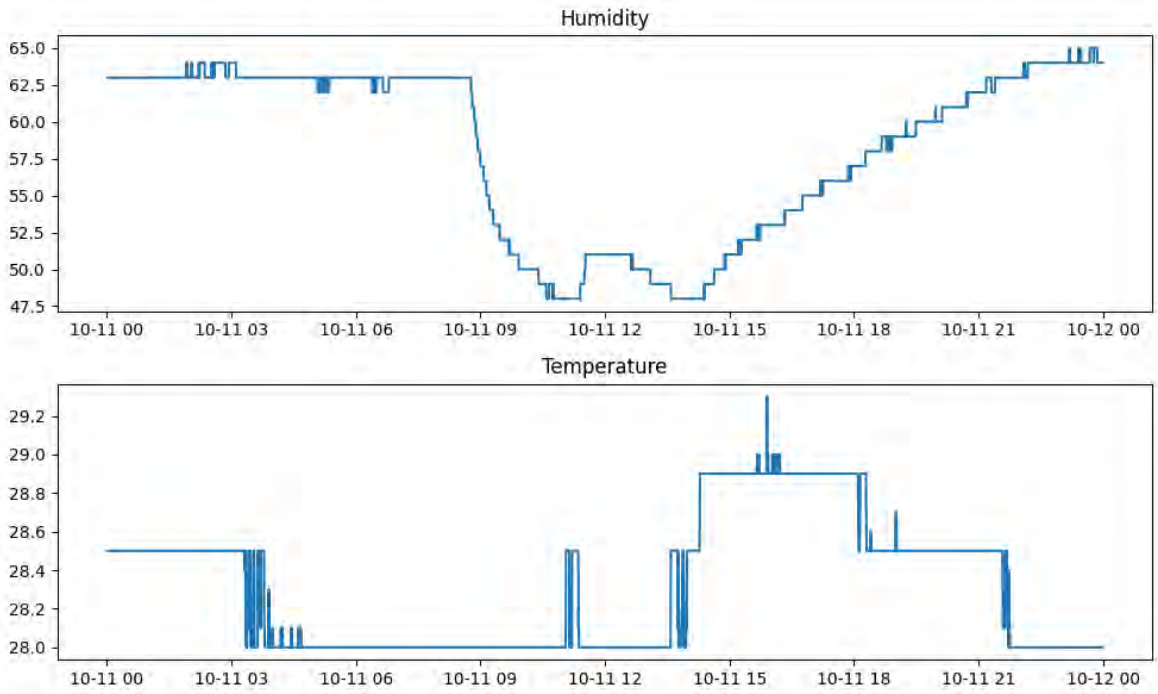
4. Kesimpulan

Penelitian ini mengimplementasikan penggunaan docker container untuk membangun sistem IoT dengan cepat, karena pengguna dapat menggunakan beberapa perangkat lunak seperti basis data dan visualisasi data yang sudah tersedia. Pada penelitian ini, sistem IoT digunakan untuk melakukan pengontrolan peralatan listrik dan monitoring kondisi ruangan di dalam laboratorium. Penggunaan docker container juga mempermudah pengelolaan dan pengaturan konfigurasi perangkat lunak yang digunakan dalam sistem IoT. Dari hasil monitoring, kondisi udara dalam ruang laboratorium berhubungan dengan okupansi ruangan. Biasanya peralatan listrik dioperasikan untuk meningkatkan kenyamanan dalam ruang laboratorium. Namun, kelembaban udara juga perlu dikontrol terutama untuk ruang penyimpanan peralatan laboratorium. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan optimasi pengontrolan kelembaban udara namun dengan tetap memperhatikan efisiensi konsumsi daya.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 2. Log Data Hasil Pengukuran Sensor IoT

No.	Time	Humidity Value	Temperature Value
1	10/11/2023 0:00	63	28.5
2	10/11/2023 1:00	63	28.5
3	10/11/2023 2:00	63	28.5
4	10/11/2023 3:00	64	28.5
5	10/11/2023 4:00	63	28.1
6	10/11/2023 5:00	63	28
7	10/11/2023 6:00	63	28
8	10/11/2023 7:00	63	28
9	10/11/2023 8:00	63	28
10	10/11/2023 9:00	57	28
11	10/11/2023 10:00	50	28
12	10/11/2023 11:00	48	28
13	10/11/2023 12:00	51	28
14	10/11/2023 13:00	50	28
15	10/11/2023 14:00	48	28.5
16	10/11/2023 15:00	51	28.9
17	10/11/2023 16:00	53	28.9
18	10/11/2023 17:00	55	28.9
19	10/11/2023 18:00	57	28.9
20	10/11/2023 19:00	59	28.6
21	10/11/2023 20:00	60	28.5
22	10/11/2023 21:00	62	28.5
23	10/11/2023 22:00	63	28
24	10/11/2023 23:00	64	28



Gambar 10. Plot data temperatur dan kelembaban udara ruangan tanggal 11 Oktober 2023

References

[1] A. F. Isnanto, A. Surriani, S. Lestari and U. Y. Oktiawati, "Prototype of Smart Home and Monitoring Application Based On Internet of Things (IoT) Using Android," *Jurnal Listrik, Instrumentasi, dan Elektronika Terapan (JuLIET),* vol. 1, no. 1, 2020.
[2] K. L. Rivaldo, I. K. A. Mogi, I. P. G. H. Suputra, N. A. Sanjaya, I. D. M. B. A. Darmawan and I. B. G. Dwidasmara, "Sistem Monitoring Tanaman Hidroponik Berbasis Internet of Things Menggunakan Restful API," *Jurnal Elektronik Ilmu Komputer Udayana,* vol. 11, no. 1, 2022.
[3] M. H. Shirvani and M. Masdari, "A survey study on trust-based security in Internet of Things: Challenges and issues," *Internet of Things,* vol. 2023, no. 21, 2023.
[4] H. Yin, "Smart Lab Technologies," in *Handbook of Mobile Teaching and Learning*, Berlin, Springer-Verlag Berlin Heidelberg, 2015, pp. 1-11.
[5] B. S. Kim, S. H. Lee, Y. R. Lee, Y. H. Park and J. Jeong, "Design and Implementation of Cloud Docker Application Architecture Based on Machine Learning in Container Management for Smart Manufacturing," *Applied Science,* vol. 12, no. 13, 2022.
[6] W. Istiana, "Sistem Pengendalian Lampu Menggunakan Raspberry Pi Internet of Things (IoT) Berbasis Mobile," *Jurnal Portal Data,* vol. 2, no. 6, 2022.
[7] S. D. M. Panjaitan, "Prototype Pengendalian Lampu Jarak Jauh Dengan Jaringan Internet Berbasis Internet of Things (IoT) Menggunakan Rasberry Pi 3," *Jurnal Pendidikan Sains dan Komputer,* vol. 2, no. 2, 2022.
[8] F. Z. Rachman, "Smart Home Berbasis IoT," in *Seminar Nasional Inovasi Teknologi Terapan (SNITT)*, Balikpapan, 2017.
[9] Z. Iklima, "Rancang Bangun Prototipe Sistem Kendali Terdistribusi Instalasi Penerangan pada Gedung 3 Lantai berbasis IoTaaS (Internet of Things as a Service) menggunakan Docker Container," *Jurnal Teknologi Elektro,* vol. 10, no. 1, 2019.
[10] A. Zivelonghi and A. Guiseppi, "Smart Healthy Schools: An IoT-enabled concept for multi-room dynamic air quality control," *Internet of Things and Cyber-Physical Systems,* vol. 4, pp. 24-31, 2024.
[11] G. Rescio, A. Manni, A. Caroppo, A. M. Carluccio, P. Siciliano and A. Leone, "Multi-Sensor Platform for Predictive Air Quality Monitoring," *Sensors,* vol. 23, no. 11, pp. 1-17, 2023.
[12] F. Y. Ontowirjo, V. C. Poekoel, P. D. Manembu and R. F. Robot, "Implementasi Internet of Things Pada Sistem Monitoring Suhu dan Kelembaban Pada Ruangan Pengering Berbasis Web," *Jurnal Teknik Elektro dan Komputer,* vol. 7, no. 3, pp. 331-339, 2018.
[13] A. A. M. Khalifa and K. Prawiroredjo, "Model Sistem Pengendalian Suhu dan Kelembaban Ruangan Produksi Obat Berbasis NodeMCU ESP32," *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi dan Komputer,* vol. 6, no. 1, pp. 13-25, 2022.
[14] A. Iskandar, "Implementasi IoT Pada Sistem Monitoring dan Kendali Otomatis Suhu Dan Kelembaban Ruangan Sarang Burung Walet Berbasis Mikrokontroler," *Jurnal Cyber Tech,* vol. 4, no. 8, pp. 1-8, 2022.
[15] I. Y. Syas and F. A. Rakhmadi, "Prototipe Sistem Monitoring serta Kendali Suhu dan Kelembapan Ruangan Budidaya Jamur Tiram Putih Menggunakan Sensor DHT22 Dan Mikrokontroler NodeMCU," *Sunan Kalijaga Journal of Physics,* vol. 1, no. 1, pp. 7-13, 2019.
[16] S. H. W. Sasono, "IoT Smart Health Untuk Monitoring Dan Kontrol Suhu dan Kelembaban Ruang Penyimpan Obat Berbasis Android di Rumah Sakit Umum Pusat Dr. Sardjito," *ReTII,* pp. 53-62, 2020.
[17] E. B. Raharjo, S. Marwanto and A. Romadhona, "Rancangan Sistem Monitoring Suhu Dan Kelembapan Ruang Server Berbasis Internet Of Things," *Teknika ATW,* vol. 6, no. 2, pp. 61-69, 2019.
[18] N. F. Khobariah, P. D. S. Hermawan and R. S. Kusumadiarti, "Sistem Monitoring Suhu dan Kelembapan Ruang Server Berbasis Wemos D1," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika),* vol. 7, no. 1, pp. 32-42, 2022.

Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 12, No 3. Februari 2024 e-ISSN: 2654-5101

Segmentasi Gaya Hidup Mahasiswa Menggunakan Algoritma *K-Means Clustering*

Nur Kholidin[1], Nana Suarna[2], Willy Prihartono[3]

[12] Teknik Informatika, STMIK IKMI Cirebon
Jl.perjuangan No.10B, Karyamulya kec.Kesambi kota Cirebon, Indonesia
[1]Nkholidin657@gmail.com
[2]St_nana@yahoo.com

[3]Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl.Perjuangan No.10B Karyamulya kec.Kesambi kota Cirebon, Indonesia
[3]Willyprihartono@yahoo.com

Abstract

Gaya hidup merupakan salah satu kunci menjaga agar tubuh tetap sehat dan bugar. Mengabaikan pola hidup sehat bisa membuat tubuh lesu dan mudah terserang penyakit. Masalah yang teridentifikasi termasuk sakit maag akibat makan tidak teratur dan efek samping seperti stres, penambahan berat badan, dan penurunan konsentrasi karena seringnya begadang. Penelitian ini berupaya untuk meningkatkan kesadaran mahasiswa untuk hidup sehat dengan memanfaatkan algoritma *K-means clustering* untuk segmentasinya. Data yang digunakan pada proses segmentasi ini berjumlah 214 data dan 16 atribut. Temuan menunjukkan segmentasi optimal pada K = 2 menghasilkan rata-rata jarak pusat *cluster* sebesar 0.815, dengan rata-rata jarak pada *cluster* 0 adalah 0.779 yang berisi 51 mahasiswa dan masuk kedalam kategori gaya hidup sehat. Pada cluster 1 rata-rata jaraknya adalah 0.925 berisi 155 mahasiswa dan masuk kedalam kategori gaya hidup tidak sehat. Kemudian pada evaluasi cluster menggunakan metode *Davies Bouldin Index* memperoleh nilai 0,118. Hasil ini memberikan wawasan berharga untuk inisiatif kesehatan kampus yang ditargetkan dan disesuaikan dengan kebutuhan gaya hidup mahasiswa. Dengan meningkatkan kesadaran kesehatan, penelitian ini mengantisipasi pengaruh positif terhadap kesehatan mahasiswa secara keseluruhan dan prestasi akademik.

Keywords: *segmentasi, gaya hidup mahasiswa, K-means, clustering, davies bouldin index*

1. Introduction

Mahasiswa adalah salah satu kelompok yang sangat rentan akan tekanan akademik[1]. Serta gaya hidup tidak baik atau cenderung tidak sehat seperti kualitas tidur, pola makan, dan aktivitas fisik [2]. Untuk mencapai kesadaran kesehatan, diperlukan kekonsistenan yang mencakup aspek fisik, mental, dan emosional [3]. Pengertian gaya hidup menurut Setiadi (2010:148) dalam jurnal "pengaruh literasi keuangan, gaya hidup pada perilaku keuangan pada generasi milenial" menyatakan bahwa gaya hidup didefinisikan sebagai cara hidup yang didefinisikan oleh bagaimana seseorang menghabiskan waktu mereka, serta pemikiran mereka tentang diri sendiri dan juga dunia sekelilingnya [4]. kemudian penelitian yang dilakukan oleh Vebriyani dengan judul "Pengaruh Media Sosial dan Gaya Hidup Hedonis terhadap Perilaku Konsumtif Mahasiswa (Studi Kasus pada Mahasiswa Fakultas Ekonomi dan Bisnis Islam IAIN Kudus Angkatan 2016)" menjelaskan pengertian gaya hidup ialah cara hidup dan tingkah laku seseorang mengenai bagaimana seseorang tersebut menghabiskan waktunya (aktivitas), apa yang dianggap penting dalam lingkungannya (*interes*) serta bagaimana seseorang memikirkan diri sendiri dan dunia sekelilingnya (*opini*) [5].

Pada penelitian ini hal yang penting dilaukan yaitu pengumpulan dan pengelolaan data kesehatan mahasiswa secara efektif merupakan tantangan besar. Tantangan analitisnya adalah menerapkan metode *K-Means clustering* yang tepat dan representatif untuk mengidentifikasi segmen mahasiswa yang berpengetahuan luas. Teknik clustering merupakan proses

pembentukan kelompok data (*cluster*) dari kumpulan data yang kelompok atau kelasnya tidak diketahui, dan menentukan data mana yang termasuk dalam cluster mana [6]. selain itu teknik *clustering* juga suatu proses untuk menentukan kelas taksonomi atau fikologis untuk analisis topologi data yang ada. Teknik *cluster* mempunyai dua metode dalam pengelompokannya, yaitu hirarki dan non-hirarki. Pengelompokkan hirarki adalah teknik pengelompokan data yang bekerja dengan cara mengelompokkan dua atau lebih bagian data yang memiliki persamaan [7]. Sedangkan Algoritma *K-Means* adalah algoritma analisis data *unsupervised* menggunakan sistem partisi. Algoritma ini mempunyai masukan data dan nilai K yang artinya data atau objek tersebut tersebar ke dalam *cluster*, dan setiap *cluster* mempunyai nilai titik pusat (*centroid*) yang mewakili cluster tersebut [8]. Pendapat lain juga dikemukakan oleh Ahsina, bahwa Algoritma K-means merupakan fungsi yang bertujuan untuk mencapai kesamaan dalam kelompok tinggi dan kesamaan dalam kelompok rendah, pengelompokan menggunakan teknik partisi berbasis titik pusat (*centroid*), dimana setiap centroid kelompok mewakili karakteristik setiap kelompok tersebut [9].

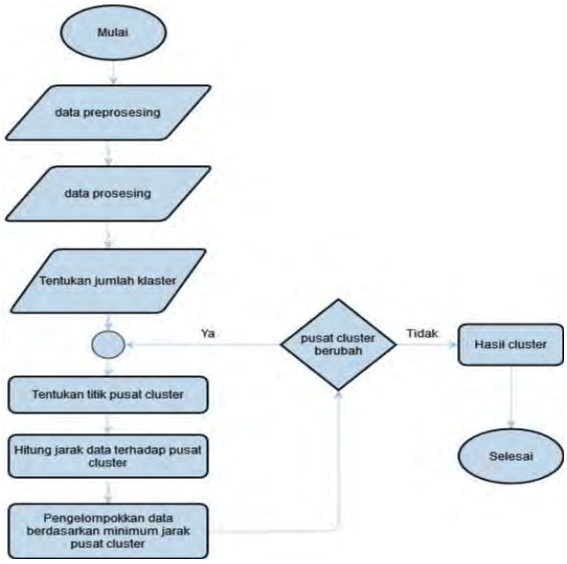
Penelitian terkait juga sudah dilakukan oleh beberapa peneliti sebelumnya. Diantaranya oleh [10] membahas mengenai penerapan algoritma k-means dalam segmentasi mahasiswa saat penerimaan mahasiswa baru UIN Sultan Syarif Kasim Riau berasal dari sekolah unggulan atau bukan. Temuan analisis ini diharapkan bisa memberikan pandangan dan wawasan yang berharga serta menjadi landasan untuk merancang strategi yang lebih dalam penerimaan mahasiswa baru di tahun berikutnya, pengembangan kurikulum, serta langkah-langkah lainnya yang diperlukan untuk meningkatkan kualitas dan kuantitas pengalaman perkuliahan. Penelitian lain oleh [11] menerangkan bahwa *K-Means clustering* dalam Penerapan metode *Crisp-DM* dengan algoritma *K-means clustering* untuk segmentasi mahasiswa berdasarkan kualitas akademik, juga bisa digunakan untuk menggali serta mendapatkan informasi yang berguna mengenai karakteristik dari mahasiswa dan dapat dijadikan sebagai bahan untuk menyusun program-program yang lebih baik. Penelitian oleh [12] yang membahas mengenai penggunaan algoritma k-means untuk segmentasi pelanggan pada aplikasi alfagift. Pada penelitian ini diterapkan strategi-strategi penerapan dengan membagi pelanggan ke beberapa kelompok atau segmen yang berbeda. Beberapa karakteristik yang diperiksa antara lain perilaku serta kebutuhan pelanggan yang berbeda-beda. Hal ini dilakukan sebagai data pendukung dalam menentukan dan mengetahui loyalitas serta preferensi untuk pelanggan sekaligus sebagai sarana media promosi bagai pengembang dimasa yang akan dating. Temuan dari analisis ini diharapkan memberikan wawasan yang baru dan berharga untuk merancang strategi dalam berbagai bidang. Algoritma *K-means clustering* terbukti efektif dalam mengelompokkan objek berdasarkan propertinya, menghasilkan informasi yang berguna untuk perbaikan dan pengembangan lebih lanjut.

Sebanyak 214 data telah berhasil dikumpulkan melalui kuesioner dari mahasiswa Program Studi Teknik Informatika angkatan 2020 di STMIK IKMI Cirebon. Kuesioner ini mencakup 16 atribut yang relevan dengan penelitian yang dilakukan. Data yang terkumpul memberikan gambaran yang komprehensif mengenai berbagai aspek yang diteliti dalam konteks penelitian.

2. Reseach Methods

Metode penelitian yang digunakan dalam penelitian ini adalah algoritma *K-means clustering*. algoritma *K-Means* merupakan metode data *clustering* non hirarki, *clustering* adalah salah satu metode data mining bersifat tanpa pelebelan atau bisa disebut unsupervised [13]. Yang mempartisi menjadi satu kelompok sehingga data yang memiliki karakteristik serupa dikelompokkan ke dalam satu *cluster* yang sama dan data yang karakteristiknya berbeda dikelompokkan ke dalam kelompok lain [14]. Tujuannya untuk mengelompokkan data ke dalam suatu segmen berdasarkan kemiripan karakteristik data. Tahapan untuk menerapkan metode *k-means clustering* bisa dilihat pada gambar 1.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 1. Metode penelitian

Tahapan untuk melakukan segmentasi menggunkan algoritma *k-menas clustering*:

1. Menentukan jumlah *cluster* yang diinginkan sesuai dengan kelompok yang akan dibentuk
2. Tetapkan k pusat *cluster* awal secara random
3. Menghitung jarak data terdekat terhadap pusat *cluster* menggunakan rumus persamaan *euclidean*.

$$d_{Euclidean}\,(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

Dimana d(x,y) yaitu jarak antar variabel data hasil kuesioner dan variabel *centroid*, xi = (x1, x2, …, xi) yaitu variabel data hasil kuesioner, yi = (y1, y2, …, yi) yaitu variabel pada *centroid.*

4. Pengelompokkan data berdasarkan minimum jarak pusat *cluster* terdekat
5. Melakukan perhitungan *centroid* baru

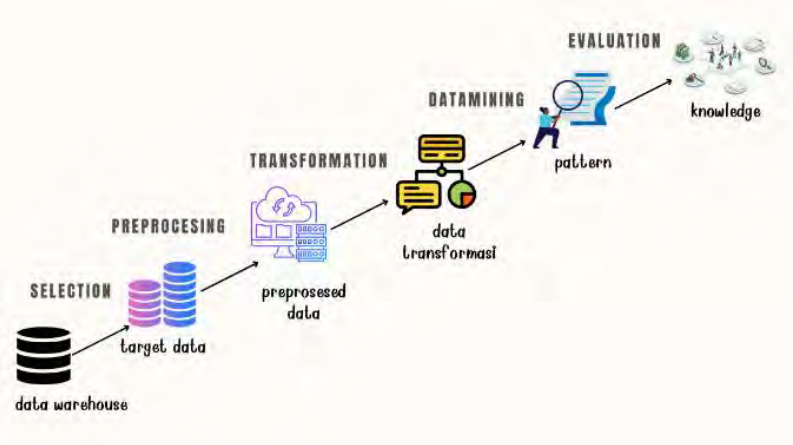
$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

Dimana μ_k= Titik *centroid* dari *cluster* ke-K, N_k= Banyaknya data pada *cluster* ke-K, x_q= Data ke-q pada *cluster* ke-k

6. Mengecek kualitas *cluster* yang dihasilkan oleh algoritma *K-Means* serta melakukan interpretasi hasil

2.1 Teknik Analisis Data

Teknik analisis data pada penelitian menggunakan teknik KDD (*Knowledge Discovery in Databases*). KDD merupakan proses yang melibatkan ekstraksi informasi berguna, yang sebelumnya tidak diketahui, dan berpotensi berharga dari kumpulan data besar. Melibatkan langkah-langkah seperti pada gambar 2.



Gambar 2. Tahapan KDD

1. *Selection*, Proses mengubah data terpilih ke dalam bentuk prosedur.
2. *Data Transformation*, Proses transformasi data terpilih ke dalam bentuk *mining procedure*.
3. *Data Mining*, Proses mengekstraksi pola laten dan menghasilkan data berguna menggunakan berbagai teknik.
4. *Evaluation*, Proses dimana pola-pola yang telah diidentifikasi berdasarkan ukuran yang diberikan berubah seiring waktu..
   a. Jarak *cluster* dengan titik pusat

      Jarak *cluster* dengan titik pusat adalah ukuran jarak antara setiap klaster dengan pusatnya. Pusat *cluster* di *K-means* disebut *centroid*, dan penghitungan jarak ini akan mempengaruhi pembentukan *cluster* yang optimal.
   b. Nilai DBI

      DBI dihitung dengan mempertimbangkan nilai rata-rata jarak antara masing-masing cluster dengan *cluster* lainnya. Metrik ini mengukur kemiripan *cluster* dalam hal bentuk dan ukuran Rumusnya adalah:

$$DBI = \frac{1}{K}\sum_{i=1}^{k} max_{i \neq j}\,(R_{i,j}) \quad (3)$$

      Dari rumus tersebut, k merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang dihasilkan (non-negatif >= 0), maka semakin baik *cluster* hasil pengelompokan *K-means* yang digunakan.
5. *Knowledge Presentation*, Proses paling akhir dari proses KDD, Data-data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis.

2.2. Data

Data adalah kumpulan informasi dasar tentang sesuatu yang diperoleh dari pengamatan dan bisa diolah jadi bentuk menjadi sebuah informasi, database atau solusi dari suatu masalah tertentu[15]. Data yang digunakan dalam penelitian ini adalah data sebanyak 214 responden dan terdiri 16 atribut yang dikumpulkan melalui kuesioner menggunakan google formulir. Data hasil pengisian kuesioner tersebut dalam bentuk dokumen *soft file* format *Microsoft Excel*. Pertanyaan kuesioner pada penelitian ini dapat diakses melalui link berikut ini https://bit.ly/3tqYoqI. Berikut ditampilkan data hasil pengisian kuesioner pada tabel 1.

Tabel 1. Hasil kuesioner gaya hidup mahasiswa

No	Nama	Jenis Kelamin	Usia	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
1	Mela	Perempuan	18-25 tahun	3	3	3	4	4	3	2	2	3	1	2	1	Ya

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2	Salsa	Perempuan	18-25 tahun	3	4	3	5	5	3	5	3	5	3	2	3	Ya
3	Hajaroh	Perempuan	18-25 tahun	3	4	3	4	4	3	3	3	1	1	1	1	Ya
…	….	….	….	..	..	..	..	..	..	..	..	..	..	..	..	…
213	Riska	Perempuan	18-25 tahun	2	2	3	2	5	3	5	2	4	1	2	3	Ya
214	Trian	Laki-laki	18-25 tahun	2	3	3	1	5	3	3	1	4	1	3	2	Ya

2.3. *Data Cleansing*

Data cleansing proses ini berfokus pada identifikasi dan pengecekan masalah yang mungkin ada dalam data mentah sehingga data yang akan diolah nantinya dapat menghasilkan informasi berkualitas. Pada penelitian ini pembersihan data dilakukan mencakup responden yang mengisi kuesioner lebih dari 1 kali, ketidaklengkapan, ketidakbenaran serta ketidakrelevan data. Hasil pembersihan data dapat dilihat pada tabel 2.

Tabel 2. Data Cleansing

No	Nama	Jenis Kelamin	Usia	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
1	Mela	Perempuan	18-25 tahun	3	3	3	4	4	3	2	2	3	1	2	1	Ya
2	Salsa	Perempuan	18-25 tahun	3	4	3	5	5	3	5	3	5	3	2	3	Ya
…	….	….	….	..	..	..	..	..	..	..	..	..	..	..	..	
205	Riska	Perempuan	18-25 tahun	2	2	3	2	5	3	5	2	4	1	2	3	Ya
206	Trian	Laki-laki	18-25 tahun	2	3	3	1	5	3	3	1	4	1	3	2	Ya

2.4. *Data Selection*

Data yang dipakai pada penelitian ini adalah data kuesioner gaya hidup mahasiswa. Yang sebelumnya telah disebarkan kepada responden dan dari 16 atribut yang ada pada kuesioner penelitian tersebut, dilakukan proses *selection* hasil dari data kuesioner yang dilakukan digunakan 14 atribut. Karena 2 atribut memiliki nilai yang tidak relevan untuk dilakukan analisa lebih lanjut. Atribut yang tidak digunakan terdiri dari usia dan P_{13}. Hasil pemilihan atribut dapat dilihat pada tabel 3 berikut.

Tabel 3. *Data Selection*

No	Nama Atribut	Tipe
1	Nama	polynomial
2	Jenis kelamin	polynomial

3	P_1	integer
4	P_2	integer
5	P_3	integer
6	P_4	integer
7	P_5	integer
8	P_6	integer
9	P_7	integer
10	P_8	integer
11	P_9	integer
12	P_{10}	integer
13	P_{11}	integer
14	P_{12}	integer

2.5. *Data Transformation*

Pada tahap ini transformasi data dilakukan pada atribut nama yang dijadikan sebagai ID serta atribut jenis kelamin akan diubah menjadi numerik. Atribut kelamin akan diinisialisasikan menjadi numerik seperti dalam tabel 4 berikut.

Tabel 1. Tranformasi atribut jenis kelamin

Jenis Kelamin	inisialisasi
Laki-laki	0
Perempuan	1

untuk merubah data bertipe nominal menjadi numerik dapat dilakukan menggunakan operator *Nominal to Numeric*. Proses ini dapat dilihat pada gambar 3.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Row No.	Nama	Jenis kelam...	Apakah and...	Apakah and...	apakah and...	Apakah and...	apakah and...	Apakah
1	Abdul	0	3	4	4	4	2	2
2	Ahmad	0	4	3	3	3	3	2
3	Adi	0	3	4	5	3	5	4
4	Adi Pangestu	0	5	3	4	4	2	3
5	ADI SADEWA	0	2	4	4	5	3	3
6	Adi Sudrajat	0	4	4	3	2	4	4
7	Adinda	1	4	4	5	4	3	4
8	Adisty	1	3	1	3	1	4	4
9	Aditia	0	5	5	3	3	4	4
10	Adrian	0	5	5	4	4	5	4
11	Agustin	1	4	5	4	5	4	4
12	AJENG	1	4	4	2	3	5	2
13	Ajeng Galuh ...	1	3	3	3	4	5	5
14	Aji saputra	0	3	4	4	1	3	3

Gambar 1. Data transformation

2.6. *Data Mining*

Pada proses data mining penelitian ini menggunakan *algoritma K-means clustering*. *Algoritma* ini digunakan pada proses segmentasi, mengetahui jumlah *cluster* yang dihasilkan serta menghitung jarak antar *cluster* dengan pusat *cluster* dan mengukur besarnya *Davies Bouldin index* (DBI) menggunakan operator *cluster distance performance.* Rumus yang digunakan untuk perhitungannya adalah

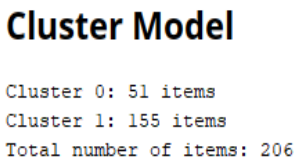
$$d_{Euclidean}(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

Parameter untuk mengukur besarnya *Davies Bouldin index* (DBI) sesuai pada tabel 5.

Tabel 2. Parameter

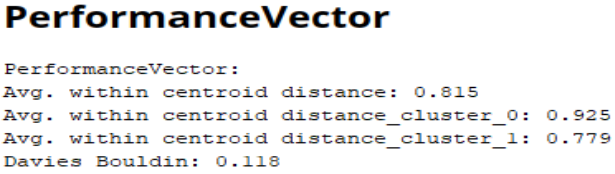
Parameter	Keterangan
K-min	2
Max runs	10
Measure type	NumericalMeasurement
Numerical measure	EuclideanDistance
Clustering algorithm	K-means clustering
Max Optimization Steps	100
Main criterion	Davies bouldin

Setelah pengisian parameter, dilakukan 9 kali percobaan untuk mencari *cluster* terbaik dan paling optimal. Percobaan *cluster* yang dilakukan secara berulang membantu memastikan bahwa hasil segmentasi yang diambil adalah hasil yang paling stabil dan sesuai dengan karakteristik data. Diperoleh hasil *cluster* optimal yaitu 2 *cluster*, dimana *cluster* 0 sebanyak 51 items dan *cluster* 1 sebanyak 155 items sesuai pada gambar 4.



Gambar 2. *Cluster model*

Selanjutnya nilai *Davies Bouldin Index* (DBI) yang dihasilkan dari algoritma *K-means clustering* ini sebesar 0.118 sesuai dengan gambar 5.



Gambar 3. *Performance Vector K=2*

Hasil analisis data menjelaskan bahwa *cluster* 0 dengan jumlah anggota sebanyak 51 mahasiswa masuk kedalam kategori gaya hidup sehat dengan rincian laki-laki sebanyak 19 mahasiswa dan perempuan 32 mahasiswa. Pada cluster 1 dengan jumlah anggota sebanyak 155 mahasiswa masuk kedalam kategori gaya hidup tidak sehat dengan rincian laki-laki 57 mahasiswa dan perempuan 98 mahasiswa.

2.7. *Evaluation*

Metode evaluasi yang digunakan dalam penelitian ini adalah *Davies Bouldin Index* (DBI) dengan rumus

$$\mathrm{DBI} = \frac{1}{K}\sum_{i=1}^{k} max_{i \neq j}\,(R_{i,j}) \quad (3)$$

Dimana *cluster* dengan nilai yang mendekati angka nol merupakan *cluster* terbaik. DBI merupakan salah satu metode yang dikenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 untuk mengevaluasi *cluster,* dinama cluster dengan nilai yang mendekati angka nol merupakan cluster terbaik[16]. Dalam table 6 terdapat hasil evalusai menggunakan DBI dengan jumlah percobaan segmentasi data sebanyak 9 kali.

Tabel 6. Hasil evaluasi menggunakan DBI

Cluster	Hasil nilai DBI
K=2	0.118
K=3	0.170
K=4	0.164
K=5	0.178
K=6	0.162
K=7	0.168
K=8	0.167

p-ISSN: 2301-5373
e-ISSN: 2654-5101

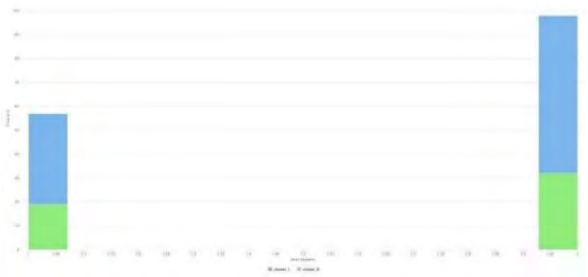
K=9	0.167
K=10	0.150

Dari tabel 6 disimpulkan bahwa K=2 adalah cluster terbaik. *Cluster 0*: Merupakan mahasiswa dengan kategori gaya hidup sehat yang berjumlah 51 mahasiswa. *Cluster 1*: Merupakan mahasiswa dengan kategori gaya hidup tidak sehat yang berjumlah 155 mahasiswa.

3. Result and Discussion

3.1. Result

Pembahasan ini merupakan interpretasi hasil yang berfokus pada K=2 sebagai *cluster* terbaik, kita dapat melihat dengan jelas perbedaan antara kedua kelompok mahasiswa yang dihasilkan. Persentase persebaran mahasiswa berdasarkan jenis kelamin dapat dilihat pada gambar 6.



Gambar 6. Persebaran mahasiswa berdasarkan jenis kelamin

Hasil *Performance* dari percobaan menggunakan nilai parameter sesuai pada tabel 5, percobaan pertama dilakukan pada K=2 dimana *cluster* ini menghasilkan rata-rata jarak pusat *cluster* sebesar 0.815, rata-rata jarak pada *cluster* 0 adalah 0.779, jarak rata-rata pada *cluster* 1 adalah 0.925. Sedangkan hasil evaluasi menggunakan metode DBI memperoleh hasil 0.118 dengan hasil evaluasi ini K=2 menjadi cluster yang paling optimal karena nilainya yang paling mendekati 0 dibanding *cluster* lainya. Hasil *Performance K=2* bisa dilihat pada gambar 7.

PerformanceVector

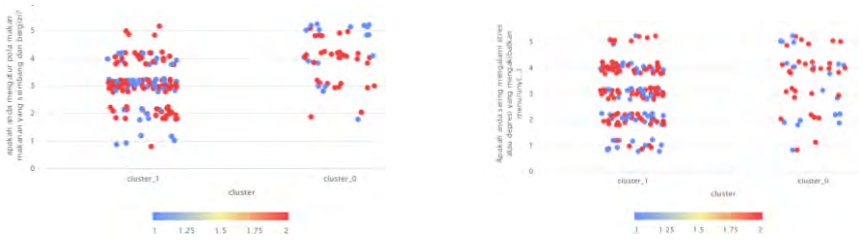
```
PerformanceVector:
Avg. within centroid distance: 0.815
Avg. within centroid distance_cluster_0: 0.779
Avg. within centroid distance_cluster_1: 0.925
Davies Bouldin: 0.118
```

Gambar 7. *Performance K=2*

Hasil dari visualisasi *cluster* berdasarkan seberapa sering berolahraga (P_1) dan melakukan begadang atau tidur saat larut malam (P_2) bisa dilihat pada gambar 8.

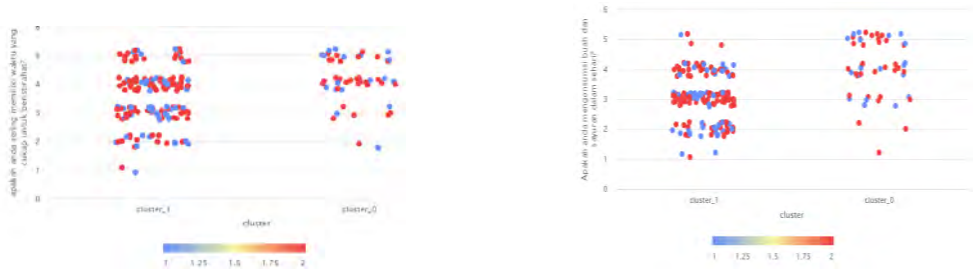
Gambar 8. Visualisasi (P_1) dan (P_2)

Hasil dari visualisasi *cluster* model berdasarkan makan makanan yang seimbang dan bergizi bisa (P_3) dan mengalami stres atau depresi yang mengakibatkan menurunya kondisi fisik (P_4) dilihat pada gambar 9.



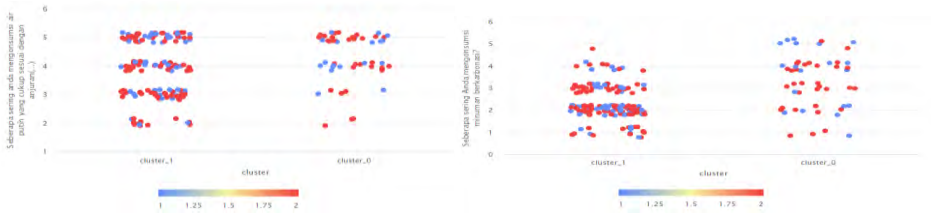
Gambar 9. Visualisasi (P_3) dan (P_4)

Hasil dari visualisasi berdasarkan memiliki waktu yang cukup untuk beristirahat (P_5) dan mengonsumsi buah dan sayuran (P_6) bisa dilihat pada gambar 10.



Gambar 10. Visualisasi (P_5) dan (P_6)

Hasil dari visualisasi berdasarkan mengonsumsi air putih yang cukup (P_7) dan mengkonsumsi minuman berkarbonasi (P_8) pada gambar 11.



Gambar 11. Visualisasi (P_7) dan (P_8)

Hasil dari visualisasi mencari informasi mengenai cara menjaga tubuh agar tetap sehat (P_9) dan mengikuti seminar atau kelas untuk menjaga agar tubuh tetap sehat (P_{10}) pada gambar 12.

Gambar 12. Visualisasi (P_9) dan (P_{10})

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Hasil visualisasi berdasar mengikuti program olahraga atau aktivitas fisik bersama dengan mahasiswa lain (P_{11}) dan berpartisipasi dalam kampanye atau acara yang berhubungan dengan kesehatan (P_{12}) pada gambar 13.

Gambar 13. Visualisasi (P_{11}) dan (P_{12})

Setelah mengetahui hasil visualisasi dari masing-masing atribut pada penelitian ini ditemukan bahwa perbedaan jelas antara *cluster* 0 dan *cluster* 1. Dari visualisasi persebaran *cluster* data gaya hidup mahasiswa, terlihat bahwa sebagian besar mahasiswa cenderung termasuk dalam *cluster* 1. Ini mengindikasikan bahwa mayoritas dari mereka menganut gaya hidup yang tidak sehat. Ada kekurangan dalam aktivitas fisik, kurangnya pola makan yang teratur, kebiasaan tidur larut malam, dan tingginya tingkat depresi serta stres. Situasi ini memprihatinkan karena seharusnya mahasiswa dapat menjaga gaya hidup agar tetap sehat dan bugar, sehingga dapat mencapai prestasi optimal dalam perkuliahan.

3.2. *Discussion*

Berdasarkan persebaran *cluster* yang divisualisasikan dari data penelitian ini terlihat jelas *cluster* yang dominan berada pada *cluster* 1. Ini menunjukkan bahwa mayoritas mahasiswa menerapkan gaya hidup tidak sehat. Berikut adalah beberapa kegiatan inisiasi kampus yang dapat diterapkan untuk mendukung kesehatan mahasiswanya:

1. Promosi Gaya Hidup Sehat: Mengorganisir kampanye dan acara promosi gaya hidup sehat, termasuk olahraga rutin, seminar nutrisi, dan workshop kesehatan mental.
2. Fasilitas Olahraga: Meningkatkan fasilitas olahraga di kampus untuk mendorong partisipasi dalam aktivitas fisik. Lapangan olahraga atau area terbuka dapat menjadi tempat yang menarik bagi mahasiswa.
3. Klinik Kesehatan Mahasiswa: Menyediakan klinik kesehatan khusus mahasiswa di kampus, yang dapat memberikan layanan kesehatan umum, konseling, dan dukungan kesehatan mental.
4. Kantin Sehat: Menyediakan pilihan makanan sehat dan bergizi ini dapat membantu menciptakan lingkungan yang mendukung gaya hidup sehat sekaligus membentuk kebiasaan makan yang baik di kalangan mahasiswa.
5. Edukasi Kesehatan Mental: Membangun kesadaran tentang kesehatan mental melalui seminar, lokakarya dan kampanye informasi. Menyediakan layanan konseling dan dukungan mental untuk mahasiswa yang membutuhkan.
6. Komunitas Kesehatan: Mendorong pembentukan komunitas kesehatan di kampus yang saling mendukung. Kelompok atau klub yang fokus pada kesehatan dapat menjadi tempat dimana mahasiswa dapat berbagi pengalaman dan pengetahuan.

4. *Conclusion*

Hasil penelitian segmentasi gaya hidup mahasiswa menggunakan algoritma *K-means clustering* maka hasilnya dapat disimpulkan sebagai berikut:

1. Dalam penelitian ini, ditemukan bahwa *cluster* terbaik dan paling optimal adalah K=2. Hasil ini mencerminkan kecenderungan bahwa mahasiswa dapat dikelompokkan menjadi dua kategori utama berdasarkan gaya hidup mereka.
2. Rata-rata jarak yang dihasilkan pada K=2, rata-rata jarak dalam pusat cluster sebesar 0.815, menunjukkan sejauh mana mahasiswa dalam suatu kelompok memiliki kesamaan pola gaya hidup sehat. Selain itu, analisis rata-rata jarak dalam pusat *cluster* 0 yaitu 0.779 dan *cluster* 1 yaitu 0.925 memberikan informasi spesifik tentang karakteristik masing-masing *cluster*. Ini membantu mengidentifikasi gaya hidup yang dominan di setiap

klaster. Pada hasil analisis data menjelaskan bahwa *cluster* 0 dengan jumlah anggota sebanyak 51 mahasiswa atau 24.76% masuk kedalam kategori gaya hidup sehat. Pada *cluster* 1 dengan jumlah anggota sebanyak 155 mahasiswa atau 75.24% masuk kedalam kategori gaya hidup tidak.

3. Penentuan *cluster* terbaik menggunakan *Davies Bouldin Index* (DBI) memberikan pemahaman lebih lanjut tentang validitas dan kehomogenan *cluster*. Pada K=2, DBI mendekati nilai 0, yaitu 0.118. Nilai yang mendekati nol menandakan bahwa *cluster* tersebut kompak dan terpisah dengan baik. Oleh karena itu, hasil ini mendukung pemilihan K=2 sebagai konfigurasi optimal dalam segmentasi mahasiswa berdasarkan gaya hidup sehat. Dengan demikian, analisis ini memberikan kerangka yang kuat untuk memahami dan meningkatkan kesadaran kesehatan di kalangan mahasiswa.

References

[1] A. E. Saputri, "HUBUNGAN ANTARA KESEPIAN DENGAN STRES AKADEMIK PADA MAHASISWA RANTAU S1 TAHUN PERTAMA DI UNIVERSITAS ISLAM SULTAN AGUNG SEMARANG," no. 30701900019, 2023.

[2] A. M. Putri, "GAYA HIDUP SEHAT DAN SUBJECTIVE WELL BEING PADA MAHASISWA KEDOKTERAN," vol. 7, no. 2, pp. 635–642, 2023.

[3] A. Ilhamsyah, "Jurnal Masyarakat Sehat Indonesia (JMSI)," *J. Masy. Sehat Indones. 1-4*, vol. 012, pp. 1–4, 2023.

[4] N. S. Azizah, "Pengaruh literasi keuangan, gaya hidup pada perilaku keuangan pada generasi milenial," vol. 01, pp. 92–101, 2020.

[5] T. Vebriyani, "Pengaruh Media Sosial dan Gaya Hidup Hedonis terhadap Perilaku Konsumtif Mahasiswa (Studi Kasus pada Mahasiswa Fakultas Ekonomi dan Bisnis Islam IAIN Kudus Angkatan 2016)," pp. 11–31, 2021.

[6] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," pp. 62–66, 2019.

[7] M. G. Sadewo, A. Eriza, A. P. Windarto, and D. Hartama, "Algoritma K-Means Dalam Mengelompokkan Desa / Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," pp. 754–761, 2019.

[8] A. Aranta, "ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL," vol. 18, no. 2, pp. 152–163, 2021.

[9] N. H. Ahsina, F. Fatimah, and F. Rachmawati, "ANALISIS SEGMENTASI PELANGGAN BANK BERDASARKAN PENGAMBILAN KREDIT DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING," vol. 8, no. 3, 2022.

[10] misra hartati siti monalisa, tengku nurainun, "PENERAPAN ALGORITMA K-MEANS DAN METODE MARKETING MIX DALAM SEGMENTASI MAHASISWA DAN STRATEGI PEMASARAN," 2021.

[11] Y. Suhanda, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. dan Komput.*, vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.

[12] S. A. Perdana, S. F. Florentin, and A. Santoso, "ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT," *Sebatik*, 2022, [Online]. Available: https://jurnal.wicida.ac.id/index.php/sebatik/article/view/1991

[13] L. A. A. R. P. A.A Sagung Prami Apsari Kumala, "MEASURE COMPARISON DISTANCE ON K-MEANS CLUSTERING FOR GROUPING MUSIC ON MOOD," *J. Elektron. Ilmu Komput. Udayana*, vol. 11, no. 4, pp. 663–674, 2023.

[14] M. H. Adiya and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17–24, 2019.

[15] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," vol. 12, 2022, doi: 10.34010/jati.v12i1.

[16] Z. Nabila, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," vol. 2, no. 2, pp. 100–108, 2021.

Jurnal Elektronik Ilmu Komputer Udayana
Volume X, No X. Month Year

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Sistem Monitoring dan Pengendalian Akuarium Berbasis *Internet of Things*

Abdul Yazid[a1], Ridha Febriliana[b1]

[a]Program Studi Teknik Komputer, Fakultas Teknologi dan Informatika, Universitas Dinamika
jl. Raya Kedung Baruk 98, Surabaya, 60298, Indonesia
[1]20410200017@dinamika.ac.id

[b]Sistem Telekomunikasi, Kampus Daerah Purwakarta, Universitas Pendidikan Indonesia, Jl. Veteran No.8, Nagri Kaler, Purwakarta, Jawa Barat, 41115, Indonesia
[2]ridhafebriliana@upi.edu

Abstract

The ESP32 module will be used in this research to build an automatic monitoring and control mechanism for the aquarium. The creation of an automatic feeding device that can increase the productivity of fish farming is the main goal. In addition, this research places focus on IoT-based aquarium monitoring and control. The water level is monitored using a JSN-SR04t ultrasonic sensor, while the water temperature is monitored using a DS18B20 sensor. The Blynk app can be used with these gadgets to monitor and regulate the condition of the aquarium, including regulating the water temperature and water level automatically. This study makes a significant contribution to the advancement of automation technology as well as the improvement of aquarium administration and maintenance. The system was tested to ensure that all its functions were working correctly. These tests include, Measurement of water temperature and water level, automatic fish feeding, control of pumps and lights, monitoring and setting aquarium conditions through the Blynk application. All system functions work correctly, the system can measure water temperature with an accuracy of 96.74%, measure water level with an accuracy of 95.39%, feed fish at the time specified by the user. control pumps and lights according to the settings, monitor and manage aquarium conditions through the Blynk application. The automatic aquarium monitoring and control system was 100% successful in functional testing.

Keywords: *Internet of Things, Monitoring, Controlling, Aquarium,* Automation

Abstrak

Modul ESP32 akan digunakan dalam penelitian ini untuk membangun mekanisme monitoring dan pengendalian otomatis pada akuarium. Pembuatan alat pemberian pakan otomatis yang dapat meningkatkan produktivitas budidaya ikan adalah tujuan utamanya. Selain itu, penelitian ini menempatkan fokus pada pemantauan dan pengendalian akuarium berbasis IoT. Ketinggian air dipantau menggunakan sensor ultrasonic JSN-SR04t, sedangkan suhu air dipantau menggunakan sensor DS18B20. Aplikasi Blynk dapat digunakan dengan gadget ini untuk memantau dan mengatur kondisi akuarium, termasuk mengatur suhu air dan ketinggian air secara otomatis. Studi ini memberikan kontribusi yang signifikan terhadap kemajuan teknologi otomasi serta peningkatan administrasi dan pemeliharaan akuarium. Sistem diuji untuk memastikan bahwa semua fungsinya bekerja dengan benar. Pengujian ini meliputi, Pengukuran suhu air dan ketinggian air, pemberian pakan ikan otomatis, pengendalian pompa dan lampu, pemantauan dan pengaturan kondisi akuarium melalui aplikasi Blynk. Semua fungsi sistem bekerja dengan benar, sistem dapat mengukur suhu air dengan akurasi 96,74%, mengukur ketinggian air dengan akurasi 95,39%, memberikan pakan ikan pada waktu yang ditentukan user. mengontrol pompa dan lampu sesuai dengan pengaturan, memantau dan mengatur kondisi akuarium melalui aplikasi Blynk. Sistem Monitoring dan Pengendalian akuarium otomatis berhasil 100% dalam pengujian fungsional.

Keywords: *Internet of Things, Monitoring, Kontroling, Akuarium, Otomasi*

1. Pendahuluan

Banyak orang memelihara ikan baik dalam akuarium maupun di kolam, dan ikan dapat menjadi sumber penghasilan bagi beberapa orang [1]. Akuarium adalah suatu tempat yang digunakan untuk menampung dan merawat berbagai jenis organisme air, seperti ikan, tanaman air, dan makhluk hidup lainnya. Akuarium ini sebagai media hobi yang populer di kalangan pecinta alam, peternak ikan, dan bahkan dalam dunia akademis [2]. Banyak orang saat ini memiliki minat dalam menjaga ikan hias dalam akuarium sebagai salah satu hobi yang popular [3]. Dalam beberapa tahun terakhir, perkembangan teknologi telah membawa inovasi yang signifikan dalam cara kita merawat dan memonitor akuarium.

Internet of Things (IoT) adalah salah satu konsep teknologi yang telah mengubah cara kita berinteraksi dengan lingkungan sekitar. IoT memungkinkan objek dan perangkat untuk terhubung ke internet dan berkomunikasi satu sama lain, memungkinkan pengguna untuk memantau dan mengendalikan perangkat dari jarak jauh. Penerapan IoT dalam berbagai sektor, termasuk pertanian dan rumah tangga, telah membawa manfaat besar, termasuk efisiensi, pemantauan *real-time*, dan penghematan energi [4]. Penerapan yang sering ditemui dalam *home automation* adalah penggunaan IoT untuk mewujudkan konsep *smart home* [5]. Dalam konsep ini, berbagai perangkat rumah seperti televisi, kulkas, lampu, AC, *smart lock*, dan CCTV dapat diawasi dan dikelola melalui internet dari lokasi yang berjauhan [6].

Dalam penelitian ini terdapat beberapa penelitian terdahulu yang menjadi acuan seperti Smart Akuarium Berbasis IoT Menggunakan Raspberry Pi 3 [1], Penelitian ini fokus pada pengembangan sebuah sistem yang dapat mengotomatisasi pengelolaan akuarium. Tujuan utamanya adalah untuk memonitor dan mengendalikan kondisi akuarium, termasuk pemberian pakan ikan dan pencahayaan. Raspberry Pi 3 digunakan sebagai platform utama, dengan motor servo untuk mengendalikan wadah pakan ikan dan relay untuk mengontrol pencahayaan [7]. Webcam digunakan untuk memantau kondisi di dalam akuarium [8]. Perancangan *Automatic Fish Feeder* Skala Akuarium Berbasis *Internet of Things* (IoT) [9], Penelitian ini bertujuan untuk merancang alat pemberi pakan ikan otomatis dengan menggunakan modul ESP8266 dan sensor ultrasonik HC SR04 [10]. Sistem ini dirancang untuk memantau stok pakan dan memberikan notifikasi kepada pengguna. Fokus utama adalah pengembangan alat pemberi pakan otomatis yang dapat meningkatkan efisiensi dalam budidaya ikan. Rancang Bangun Monitoring Akuarium Dan Pakan Ikan Otomatis Berbasis *Internet of Things* (Iot) [3], Penelitian ini lebih berfokus pada pemantauan dan pengendalian akuarium berbasis IoT. Sensor turbidity digunakan untuk memantau kejernihan air, sensor ultrasonik untuk memonitor stok pakan ikan. Alat ini juga mencakup penggunaan aplikasi Blynk untuk mengontrol dan memantau kondisi akuarium, termasuk pengaturan waktu pemberian pakan ikan secara otomatis. Perancangan dan Implementasi Sistem Pemantauan Suhu dan Ketinggian Air pada Akuarium Ikan Hias berbasis IoT [6], Penelitian ini berfokus pada pengembangan sistem pemantauan suhu dan ketinggian air pada akuarium ikan hias berbasis IoT. NodeMCU ESP8266 digunakan sebagai pengendali utama dengan sensor suhu DS18B20 dan sensor ultrasonik HC-SR04. Sistem ini juga mencakup penggunaan bot Telegram untuk memberikan perintah dan menerima notifikasi terkait kondisi akuarium. Rancang Bangun Alat Monitoring Dan Penanganan Kualitas Air Pada Akuarium Ikan Hias Berbasis *Internet of Things* [11], Penelitian ini berfokus pada pemantauan dan penanganan kualitas air dalam akuarium ikan hias menggunakan teknologi IoT. Sensor pH meter, sensor suhu DS18B20, dan sensor TDS digunakan untuk memantau parameter-parameter air [12]. NodeMCU ESP8266 digunakan sebagai kontroler utama yang terhubung ke internet melalui IoT, dan Telegram digunakan untuk mengirim notifikasi terkait kualitas air. Setiap penelitian memiliki pendekatan yang berbeda untuk memenuhi kebutuhan pemeliharaan akuarium berbasis IoT, baik dalam hal pemberian pakan ikan, pemantauan kondisi air, atau pengendalian peralatan. Dengan berbagai fokus ini, penelitian-penelitian tersebut bertujuan untuk meningkatkan efisiensi dan kualitas pemeliharaan ikan dalam akuarium.

Meskipun banyak penelitian telah dilakukan pada akuarium berbasis IoT, masih ada beberapa kekurangan yang perlu di *addressed*. Penelitian sebelumnya [1, 2, 3, 4, 5] fokus pada pemantauan dan pengendalian parameter air seperti suhu, pH, dan TDS. Namun, penelitian tersebut tidak mempertimbangkan faktor lain seperti pemberian pakan ikan dan kontrol ketinggian air secara otomatis. Perbandingan Penelitian Penulis dengan Penelitian Sebelumnya dijelaskan pada tabel dibawah :

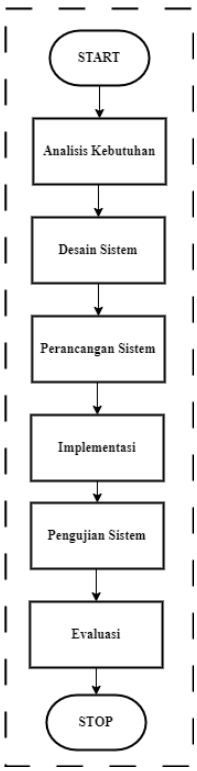
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Table 1. Perbandingan Penelitian

Fitur	Penelitian Penulis	Penelitian Sebelumnya [1]	Penelitian Sebelumnya [2]	Penelitian Sebelumnya [3]	Penelitian Sebelumnya [4]
Pemantauan	Suhu, Ketinggian Air	Suhu, pH, TDS	Suhu, Ketinggian Air	Suhu, Ketinggian Air	Suhu, pH, TDS
Pemberian Pakan Ikan	Otomatis	Manual	Manual	Manual	Otomatis
Kontrol Ketinggian Air	Otomatis	Manual	Manual	Manual	Manual
Pengendalian Peralatan	Lampu, Pompa, Kipas	-	-	-	-
Platform	Esp32	Raspberry Pi 3	ESP8266	NodeMCU ESP8266	NodeMCU ESP8266
Aplikasi	Blynk	-	-	-	Telegram

Hal ini bertujuan untuk memastikan bahwa kondisi di dalam akuarium tetap optimal sesuai dengan kebutuhan ikan yang dipelihara. Pada penelitian ini, penulis akan menggunakan platform Esp32, sensor DS18B20 untuk memantau suhu, sensor Ultrasonic untuk memantau tinggi air, dan motor servo untuk memberikan pakan ikan secara otomatis. Selain itu, akan digunakan Relay 4 channel yang akan mengendalikan perangkat-perangkat seperti pompa pengisian dan pengurasan, lampu, dan kipas. Semua sistem ini akan diintegrasikan melalui aplikasi Blynk untuk memudahkan pemantauan dan pengendalian dari jarak jauh. Metode penelitian yang yang digunakan terdiri dari 7 langkah yaitu, analisis kebutuhan, desain sistem, perancangan sistem, implementasi, pengujian sistem dan evaluasi. Sistem akan diuji untuk memastikan bahwa semua fungsinya bekerja dengan benar. Pengujian ini meliputi, Pengukuran suhu air dan ketinggian air, pemberian pakan ikan otomatis, pengendalian pompa dan lampu, pemantauan dan pengaturan kondisi akuarium melalui aplikasi Blynk.

2. Metode Penelitian



Gambar 1. Rancangan Penelitian

Penelitian yang direpresentasikan dalam Gambar 1 dimulai dengan menganalisis kebutuhan, selanjutnya desain sistem, melanjutkan dengan perancangan sistem, mengujinya untuk mengukur respons sistem, dan akhirnya mengevaluasi hasil uji sistem untuk meninjau ulang dan menganalisisnya.

2.1. Analisis Kebutuhan

Perangkat Keras :

a. Sensor DS18B20
   Sensor suhu digital DS18B20 adalah sensor yang dapat mengukur suhu dengan presisi tinggi dan komunikasi satu kawat (*One-Wire*).
   - Fungsi: Digunakan untuk memantau suhu air dalam akuarium. Suatu perubahan suhu yang drastis dapat mempengaruhi kesehatan ikan, sehingga sensor ini penting untuk menjaga kondisi akuarium.

b. Sensor Ultrasonic
   Sensor ultrasonic JSN-SR04t, digunakan untuk mengukur jarak dari sensor ke objek di depannya dengan mengirimkan gelombang ultrasonik dan mengukur waktu pantulannya.
   - Fungsi : Digunakan untuk memantau tinggi air dalam akuarium. Ini memungkinkan sistem untuk mengukur tinggi air aktual dalam akuarium dan mengambil tindakan berdasarkan tinggi air tersebut.

c. ESP32
   ESP32 adalah sebuah mikrokontroler yang kuat dan memiliki kemampuan konektivitas Wi-Fi dan Bluetooth. Ini digunakan untuk mengendalikan perangkat keras dan mengirimkan data ke server atau aplikasi Blynk.
   - Fungsi : Berperan sebagai otak sistem yang mengumpulkan data dari sensor (DS18B20 dan Ultrasonik), mengontrol peralatan (motor servo, Relay), dan mengelola komunikasi dengan aplikasi Blynk serta pengiriman data ke server.

d. Motor Servo
   Motor servo adalah perangkat yang digunakan untuk menggerakkan mekanisme buka-tutup wadah pakan ikan secara presisi.
   - Fungsi : Mengontrol bukaan dan penutupan wadah makanan ikan. Motor servo memastikan bahwa pemberian pakan terjadi sesuai jadwal yang telah ditentukan.

e. Relay 4 Channel
   Relay 4 channel adalah saklar elektromagnetik yang dapat mengontrol perangkat listrik lainnya. Setiap relay memiliki beberapa saluran atau channel yang dapat digunakan untuk mengaktifkan atau mematikan perangkat listrik.
   - Fungsi : Mengontrol berbagai perangkat seperti pompa pengisian dan pengurasan, lampu, dan kipas dalam akuarium. Relays memungkinkan sistem untuk mengatur dan mengontrol perangkat ini sesuai kebutuhan.

Perangkat Lunak :

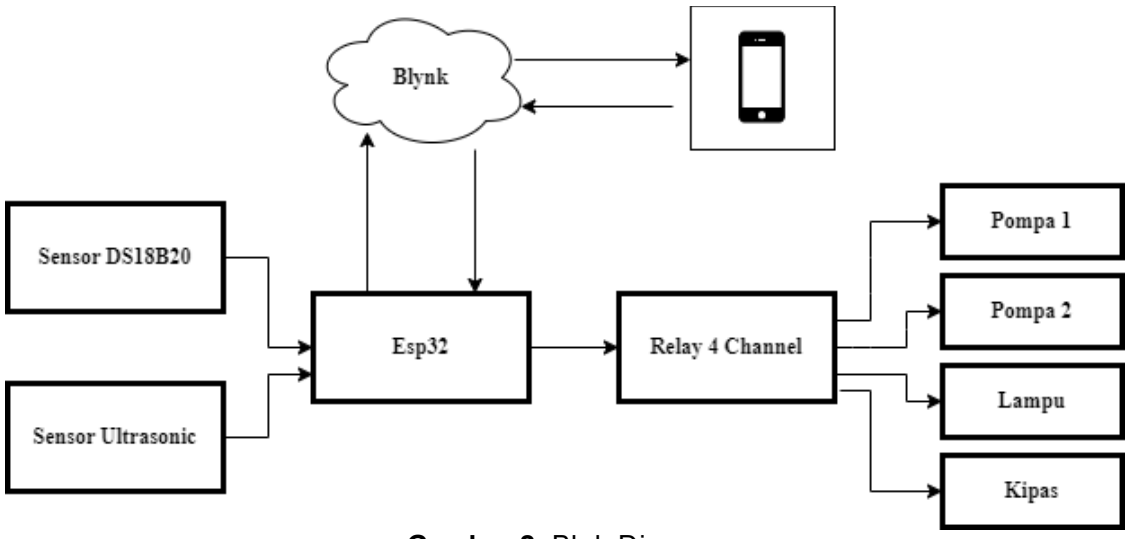
a. Aplikasi Blynk
   Blynk adalah platform IoT yang memungkinkan Anda membuat aplikasi pengendalian berbasis ponsel atau web dengan mudah dan cepat. Aplikasi Blynk digunakan sebagai antarmuka pengguna untuk memantau kondisi akuarium dan mengontrol peralatan seperti lampu, kipas, dan pemberian pakan. Ini memungkinkan pengguna untuk berinteraksi dengan sistem dengan nyaman.

b. Aplikasi Arduino IDE
   Arduino IDE adalah lingkungan pengembangan terpadu yang digunakan untuk menulis dan mengunggah kode program ke mikrokontroler seperti ESP32. Digunakan untuk mengembangkan dan memprogram ESP32. Anda akan menulis kode program dalam Arduino IDE untuk mengontrol semua aspek sistem, termasuk membaca data dari sensor, mengendalikan perangkat keras, dan mengelola komunikasi dengan aplikasi Blynk.

2.2. Desain Sistem

Blok Diagram :

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. Blok Diagram

Berdasarkan blok diagram yang disajikan dalam Gambar 2, sistem ini terdiri dari beberapa komponen utama yang bekerja secara terintegrasi untuk mengelola akuarium dengan cerdas. Cara kerja sistem ini dapat dijelaskan sebagai berikut :

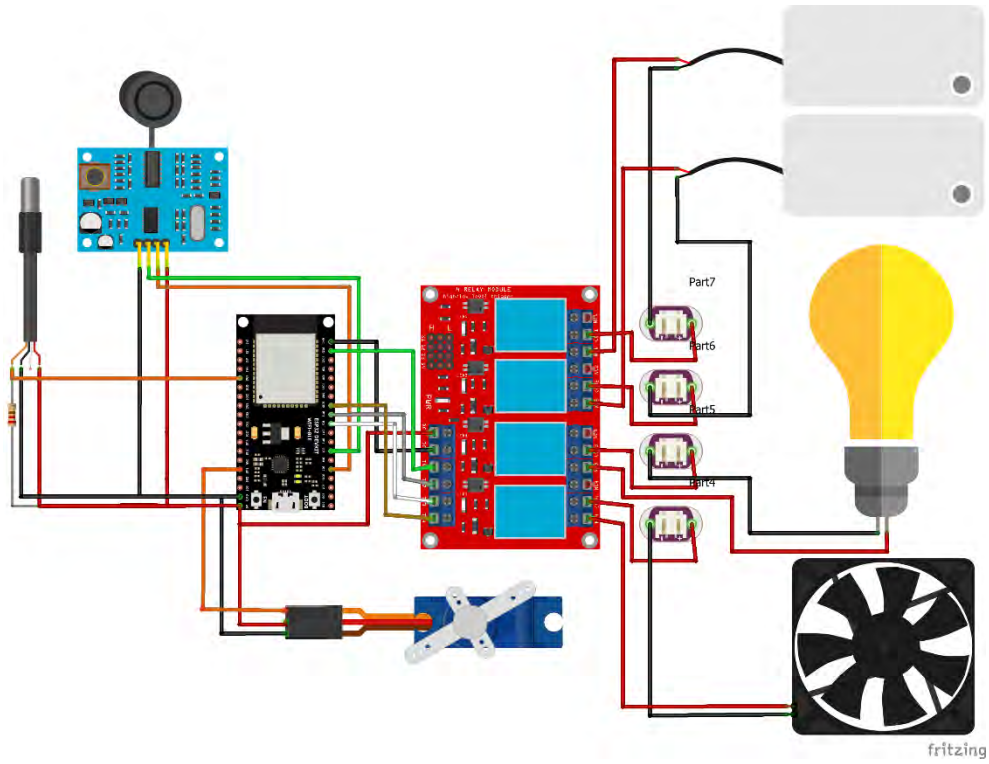
a. Input:
   1. Sensor DS18B20: Sensor ini berfungsi untuk memantau suhu air dalam akuarium. Sensor akan mengukur suhu air dan mengirimkan data suhu ke mikrokontroler ESP32.
   2. Sensor Ultrasonic: Sensor ultrasonik digunakan untuk mengukur tinggi air dalam akuarium. Data yang dihasilkan oleh sensor ini akan digunakan dalam pengendalian sistem.

b. Proses:
   1. Mikrokontroler ESP32 berperan sebagai otak sistem. Ini merupakan modul yang mengendalikan dan mengelola berbagai sensor dan aktuator dalam sistem. Sistem ini memiliki komunikasi dua arah yang aktif antara ESP32 dan aplikasi Blynk di smartphone pengguna. Informasi dan perintah dikirim dari aplikasi Blynk ke ESP32, dan ESP32 mengambil data dari sensor serta mengendalikan peralatan sesuai dengan perintah yang diterima.

c. Output:
   1. Relay Module 4 Channel : Modul relay dengan empat saluran berfungsi sebagai saklar untuk mengendalikan berbagai peralatan dalam akuarium. Setiap saluran relay memiliki tugasnya sendiri :
      - Pompa Pengisian Air : Saluran ini mengendalikan pompa yang digunakan untuk mengisi air ke dalam akuarium. Ini penting jika tinggi air dalam akuarium turun di bawah batas tertentu.
      - Pompa Pengurasan Air : Saluran ini mengontrol pompa yang digunakan untuk menguras air dari akuarium. Hal ini dapat diperlukan jika ada kelebihan air atau jika perlu mengganti air akuarium.
      - Kipas : Saluran ini mengendalikan kipas yang digunakan untuk menjaga suhu air agar tetap dalam rentang yang aman bagi ikan.
      - Lampu : Saluran ini mengontrol lampu dalam akuarium. Lampu ini dapat digunakan untuk memberikan pencahayaan sesuai jadwal atau kondisi tertentu dalam akuarium.

2.3. Perancangan Sistem

Rangkaian Skematik :



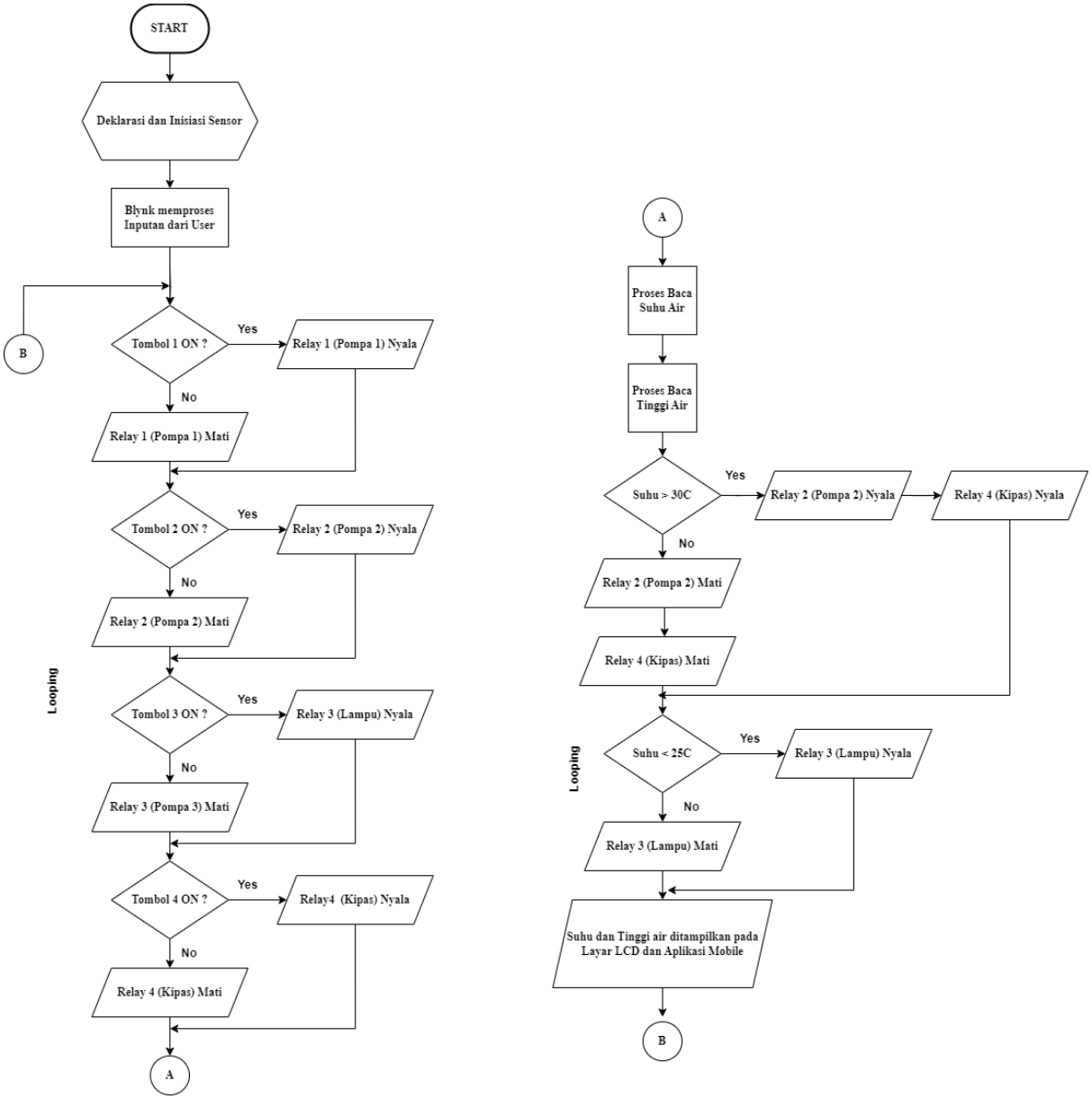
Gambar 3. Rangkaian Skematik

Berdasarkan pada Gambar 3, yang merupakan rangkaian skematik, dirancang untuk menggambarkan manajemen perkabelan yang akan diterapkan pada alat yang akan dibuat. Rangkaian ini bertujuan untuk memberikan pandangan yang jelas tentang bagaimana semua komponen yang terlibat akan terhubung dan berinteraksi dalam sistem yang akan diimplementasikan. Dengan kata lain, gambar ini akan menjadi panduan untuk mengatur kabel-kabel yang menghubungkan sensor-sensor, mikrokontroler, modul relay, dan perangkat keras lainnya dalam satu kesatuan yang terkoordinasi. Ini akan menjadi dasar yang kuat untuk merakit alat sesuai dengan spesifikasi yang telah direncanakan.

Alur Kerja Sistem :

Berdasarkan pada Gambar 4 dibawah, Proses dimulai dengan mengaktifkan program, menentukan variabel yang dibutuhkan, dan mengaktifkan sensor yang dibutuhkan. Program Blynk kemudian akan mengumpulkan data pengguna dan memprosesnya. Selain itu, ada fungsi loop yang terus berjalan, yang merupakan inti dari program. Fungsi ini mengontrol perangkat berdasarkan masukan pengguna dan kondisi saat ini. program ini memantau status empat tombol Pin1, Pin2, Pin3, dan Pin4 yang ada di proyek. Tombol-tombol ini berfungsi untuk mengendalikan berbagai aspek proyek dengan menekan tombol digital. Selanjutnya, program akan memanggil fungsi untuk mengambil data suhu menggunakan sensor suhu, terutama dengan mengakses sensor suhu dengan indeks 0. Ini akan membantu mengukur suhu dengan mengumpulkan data dari sensor yang terhubung. Program juga menggunakan sensor ultrasonik untuk mengukur tinggi air. Ini dilakukan dengan mengirimkan sinyal getaran ke sensor melalui pin TRIGGER_PIN, dan kemudian menghitung berapa lama sensor memberikan sinyal balikan.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Alur Kerja Sistem

2.4. Implementasi

Desain yang telah dibuat sebelumnya kemudian diimplementasikan secara langsung dengan menggabungkan komponen-komponen elektronika secara fisik dalam rangkaian mekanika menjadi satu kesatuan. Proses implementasi dibagi menjadi beberapa langkah :

1. Perakitan Rangkaian Elektronika
   a. Komponen elektronika :
      - Sensor DS18B20
      - Sensor Ultrasonic JSN-SR04t
      - ESP32
      - Motor Servo
      - Relay 4 Channel
      - Breadboard
      - Jumper Wire
   b. Perakitan :
      Komponen elektronika dirakit pada breadboard sesuai dengan skema rangkaian yang telah dirancang.Jumper wire digunakan untuk menghubungkan antar komponen.

2. Pemrograman ESP32
   a. Aplikasi Arduino IDE :
      - Kode program ditulis dalam Arduino IDE untuk mengendalikan semua aspek sistem.
      - Kode program ini memuat fungsi-fungsi seperti, membaca data dari sensor DS18B20 dan Sensor Ultrasonic, mengendalikan Motor Servo untuk membuka dan menutup wadah pakan, mengendalikan Relay 4 Channel untuk mengontrol pompa, lampu, dan kipas, mengirimkan data ke aplikasi Blynk, menerima perintah dari aplikasi Blynk.
      - Kode program diunggah ke ESP32 menggunakan Arduino IDE.
3. Integrasi dengan Aplikasi Blynk
   a. Aplikasi Blynk :
      - Antarmuka pengguna dibuat di aplikasi Blynk untuk menampilkan data sensor dan mengendalikan perangkat.
      - Widget seperti gauge, slider, dan button digunakan untuk menampilkan data dan menerima input dari pengguna.
   b. Koneksi :
      - ESP32 terhubung ke aplikasi Blynk melalui jaringan Wi-Fi.
      - Data sensor dan perintah dari pengguna ditukar antara ESP32 dan aplikasi Blynk.

3. Hasil dan Pembahasan

Hasil dan pembahasan dibagi menjadi dua yaitu dokumentasi alat dan Analisis Parameter berupa pengujian sensor dan aktuator, dalam uji coba ini, peneliti menguji sistem pakan ikan otomatis dengan mengukur akurasi dalam memberikan pakan kepada ikan-ikan di dalam akuarium. Peneliti juga membandingkan pembacaann tinggi air yang dihasilkan oleh sensor Ultrasonic dengan pengukuran manual menggunakan alat pengukur ketinggian air selain itu juga menguji aktuator yaitu Relay.

3.1. Dokumentasi Alat

Gambar 5. Tampilan Blynk Desktop

Gambar 5 menunjukkan tampilan aplikasi Blynk Dekstop yang digunakan untuk mengontrol dan memantau sistem monitoring dan pengendalian akuarium melalaui website. Berikut adalah penjelasan mengenai elemen-elemen pada gambar :

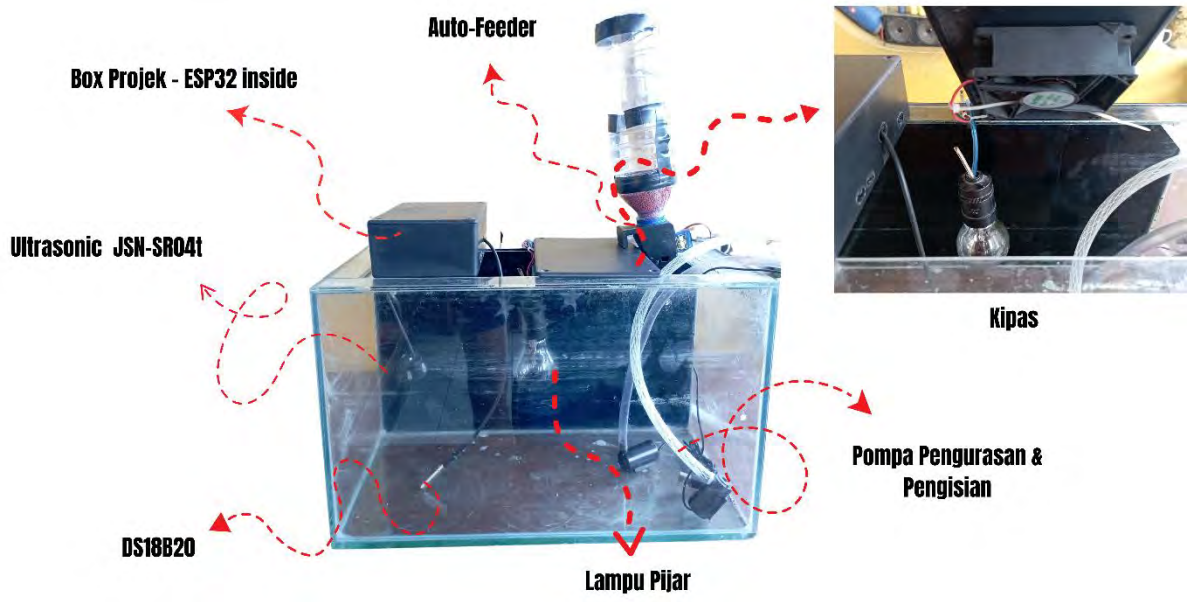
a. Gauge : Menampilkan data sensor secara real-time, seperti suhu air dan tinggi air.
b. Slider : Digunakan untuk mengatur pemberian pakan.
c. Button : Digunakan untuk mengontrol perangkat, seperti pompa, lampu, dan Kipas.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Gambar 6. Tampilan Blynk Mobile

Gambar 6 menunjukkan tampilan aplikasi Blynk versi Mobile yang digunakan untuk mengontrol dan memantau sistem monitoring dan pengendalian akuarium melalui smartphone. Berikut adalah penjelasan mengenai elemen-elemen pada gambar :

a. Gauge : Menampilkan data sensor secara real-time, seperti suhu air dan tinggi air.
b. Slider : Digunakan untuk mengatur pemberian pakan.
c. Button : Digunakan untuk mengontrol perangkat, seperti pompa, lampu, dan Kipas.



Gambar 7. Alat

Gambar 7 menunjukkan bahwa sistem monitoring dan pengendalian akuarium tersusun dari berbagai komponen elektronik yang saling terhubung dan bekerja sama untuk memantau dan mengendalikan kondisi akuarium.. Berikut adalah penjelasan mengenai komponen-komponen pada gambar :

Table 2. Daftar Komponen

Komponen	Fungsi	Spesifikasi
Mikrokontroler ESP32	Otak sistem	- CPU: Xtensa LX106 Dual-core 32-bit RISC - RAM: 400 KB - Flash: 4 MB - WiFi: 802.11 b/g/n - Bluetooth: v4.2 BR/EDR and BLE
Sensor DS18B20	Mengukur suhu air	- Rentang pengukuran: -55°C hingga 125°C - Akurasi: ±0.5°C (dalam rentang 0°C hingga 70°C) - Resolusi: 9 bit
Sensor Ultrasonic JSN-SR04t	Mengukur tinggi air	- Rentang pengukuran: 2 cm hingga 450 cm - Akurasi: ±3 mm - Resolusi: 0.3 mm
Motor Servo	Menggerakkan mekanisme buka-tutup wadah pakan ikan	- Sudut putar: 0° hingga 180° - Tegangan operasi: 4.8 V hingga 6 V
Relay 4 Channel	Mengontrol berbagai perangkat	- Tegangan operasi: 5 V - Arus maksimum per channel: 10 A

Penjelasan Singkat Cara Kerja Rangkaian Elektronik :

a. Sensor DS18B20 dan Sensor Ultrasonic JSN-SR04t: Mengirim data pengukuran ke ESP32.
b. ESP32: Memproses data sensor, mengontrol Motor Servo dan Relay 4 Channel, dan mengirimkan data ke aplikasi Blynk.
c. Motor Servo: Membuka dan menutup wadah pakan ikan sesuai dengan perintah dari ESP32.
d. Relay 4 Channel: Mengontrol pompa pengisian dan pengurasan, lampu, dan kipas dalam akuarium sesuai dengan perintah dari ESP32.

3.2. Analisa Parameter

Pengujian fungsional adalah tahap penting dalam pengembangan sistem yang bertujuan untuk memastikan bahwa setiap fungsi yang ada dalam sistem tersebut beroperasi dengan benar. Proses ini melibatkan uji coba setiap fungsi secara terpisah untuk memverifikasi bahwa sistem dapat melakukan berbagai tugas yang direncanakan. Ini termasuk kemampuan sistem untuk membaca data dari sensor dengan akurasi yang tepat, mengontrol perangkat sesuai dengan instruksi yang diberikan, serta mengirim dan menerima data dari aplikasi Blynk yang terintegrasi. Pengujian juga memastikan bahwa sistem mampu merespons perintah yang diberikan oleh aplikasi Blynk dengan tepat dan tanpa hambatan.

a. Uji Coba Sensor DS18B20
   Untuk mengukur akurasi suhu dalam akuarium, sensor DS18B20 harus diuji dan dibandingkan dengan Termometer Digital. Nilai error menunjukkan perbedaan antara nilai yang diukur oleh sensor dan nilai referensi. Nilai error yang positif menunjukkan bahwa nilai yang diukur lebih tinggi daripada nilai referensi, sedangkan nilai error yang negatif menunjukkan bahwa nilai yang diukur lebih rendah daripada nilai referensi.
   Nilai kesalahan (error) dari sensor DS18B20 diperiksa sebagai berikut :

$$Error = \left|\frac{Thermometer - DS18B20}{Thermometer}\right| \times 100\% \quad (1)$$

Table 3. Uji Coba Sensor DS18B20

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Suhu (C)	Thermometer	Error (%)
27.1	28.5	4.912
27.1	28.5	4.912
27.1	28.5	4.912
27.1	28.5	4.912
27.3	28.6	4.586
27.3	28.6	4.586
27.3	28.6	4.586
27.3	28.6	4.586
27.7	27.2	1.838
27.7	27.2	1.838
27.7	27.2	1.838
27.7	27.2	1.838
27.7	28.4	2.465
28.1	28.4	1.056
28.1	28.4	1.056
Rata-rata		3.26

Dari Tabel 3, Hasil pengujian Sensor DS18B20 dalam mendeteksi suhu di dalam akuarium telah dilaporkan. Pengujian ini dilakukan dengan membandingkan pembacaann suhu yang diberikan oleh Sensor DS18B20 dengan pengukuran yang dilakukan menggunakan termometer. Hasil dari pengujian ini menunjukkan bahwa terdapat kesalahan (error) sebesar 3.26% dalam nilai pembacaann suhu antara Sensor DS18B20 dan termometer.

b. Uji Coba Sensor Ultrasonic

Untuk mengukur akurasi Ketinggian Air dalam akuarium, sensor Ultrasonic harus diuji dan dibandingan dengan Penggaris. Nilai error menunjukkan perbedaan antara nilai yang diukur oleh sensor dan nilai referensi. Nilai error yang positif menunjukkan bahwa nilai yang diukur lebih tinggi daripada nilai referensi, sedangkan nilai error yang negatif menunjukkan bahwa nilai yang diukur lebih rendah daripada nilai referensi.
Nilai kesalahan (error) dari sensor Ultrasonic diperiksa sebagai berikut :

$$Error = \left|\frac{Penggaris - Ultrasonic}{Penggaris}\right| \times 100\% \quad (2)$$

Table 4. Uji Coba Sensor Ultrasonic

Tinggi (CM)	Penggaris	Error (%)
14.3	15.2	5.92
14.3	15.2	5.92
14.3	15.2	5.92
14.3	15.2	5.92
14.5	15.2	4.61
14.5	15.2	4.61
14.5	15.5	6.45
14.5	15.5	6.45
15.0	15.5	3.23
15.0	15.5	3.23
15.0	15.5	3.23
15.1	15.5	2.58
15.1	15.5	2.58
15.1	15.5	2.58
15.1	15.5	2.58
Rata-rata		4.61

Dari Tabel 4, Hasil pengujian Sensor Ultrasonic dalam mendeteksi Tinggi air di dalam akuarium telah dilaporkan. Pengujian ini dilakukan dengan membandingkan pembacaann Tinggi yang diberikan oleh Sensor Ultrasonic dengan pengukuran yang dilakukan menggunakan Penggaris. Hasil dari pengujian ini menunjukkan bahwa terdapat kesalahan (error) sebesar 4.61% dalam nilai pembacaan Tinggi air antara Sensor Ultrasonic dan penggaris.

c. Uji Coba Pakan ikan
Untuk mengukur keberhasilan pakan ikan ,maka dicoba beberapa kali dan menunjukan hasil sebagai berikut :

$$Keberhasilan = \frac{Banyak\ Data\ Berhasil}{Jumlah\ Data} \times 100\% \quad (3)$$

Table 5. Uji Coba Pakan Ikan

Percobaan Ke-	Status		Kondisi Pakan
	On	Off	
1	✓	x	Terbuka
2	x	✓	Tertutup
3	✓	x	Terbuka
4	x	✓	Tertutup
5	✓	x	Terbuka
6	x	✓	Tertutup
7	✓	x	Terbuka
8	x	✓	Tertutup
9	✓	x	Terbuka
10	x	✓	Tertutup
11	✓	x	Terbuka
12	x	✓	Tertutup
13	✓	x	Terbuka
14	x	✓	Tertutup
15	✓	x	Terbuka
Tingkat Keberhasilan			100%

Dalam Tabel 5, terdapat catatan mengenai pengujian pakan dengan 15 percobaan yang dilakukan. Pengujian ini melibatkan penggunaan perintah on-off pada aplikasi Blynk, dan hasilnya menunjukkan bahwa semua percobaan berhasil dengan tingkat keberhasilan sebesar 100%.

d. Uji Coba Relay
Untuk mengukur keberhasilan pakan ikan ,maka dicoba beberapa kali dan menunjukan hasil sebagai berikut :

$$Keberhasilan = \frac{Banyak\ Data\ Berhasil}{Jumlah\ Data} \times 100\% \quad (4)$$

Table 6. Uji Coba Relay

Percobaan Ke-	Status		Kondisi Relay
	On	Off	
1	✓	x	Menyala
2	x	✓	Mati
3	✓	x	Menyala
4	x	✓	Mati
5	✓	x	Menyala
6	x	✓	Mati
7	✓	x	Menyala
8	x	✓	Mati
9	✓	x	Menyala
10	x	✓	Mati
11	✓	x	Menyala
12	x	✓	Mati
13	✓	x	Menyala
14	x	✓	Mati

15	✓	x	Menyala
Tingkat Keberhasilan			100%

Dalam Tabel 6, terdapat catatan mengenai pengujian Relay dengan 15 percobaan yang dilakukan. Pengujian ini melibatkan penggunaan perintah on-off pada aplikasi Blynk, dan hasilnya menunjukkan bahwa semua percobaan berhasil dengan tingkat keberhasilan sebesar 100%.

4. Kesimpulan

Beberapa kesimpulan penting dapat diperoleh dari penelitian ini berdasarkan temuan penelitian yang telah dilakukan. Pertama, berdasarkan uraian alat, penelitian ini berhasil menciptakan sistem pemberian pakan ikan otomatis yang dapat memberi makan ikan di akuarium dengan tingkat keberhasilan hingga 100%. Kedua, ada sejumlah hasil pengujian yang patut diperhatikan dalam analisis parameter. Terdapat error sebesar 3,26% pada data suhu antara termometer dan sensor DS18B20. Namun pengujian kemampuan Sensor Ultrasonik dalam mendeteksi ketinggian air menunjukkan ketidakakuratan sebesar 4,61% jika dibandingkan dengan pengukuran yang dilakukan secara manual menggunakan penggaris. Terakhir, pengujian aktuator Relay menggunakan perintah on-off aplikasi Blynk juga menghasilkan tingkat keberhasilan 100%. Hal ini menunjukkan bahwa sistem dapat dipercaya untuk mengoperasikan aksesori seperti lampu, kipas angin, dan pompa pengisian/pengurasan. Secara keseluruhan, temuan pengujian menunjukkan bahwa alat pemberi makan ikan otomatis ini bekerja dengan baik dan mungkin merupakan alat yang berguna untuk mengawasi dan meningkatkan produktivitas budidaya ikan di akuarium.

Daftar Pustaka

[1] I. Muiz, "Smart Akuarium Berbasis IOT Menggunakan Raspberry Pi 3," *J. Pendidik. Sains dan Komput.*, vol. 2, no. 02, pp. 333–336, 2022, doi: 10.47709/jpsk.v2i02.1742.

[2] Khairunisa, Mardeni, and Y. Irawan, "Smart aquarium design using raspberry Pi and android based," *J. Robot. Control*, vol. 2, no. 5, pp. 368–372, 2021, doi: 10.18196/jrc.25109.

[3] F. Burhani, Z. Zaenurrohman, and P. Purwiyanto, "Rancang Bangun Monitoring Akuarium Dan Pakan Ikan Otomatis Berbasis Internet Of Things (IOT)," *JEECOM J. Electr. Eng. Comput.*, vol. 4, no. 2, pp. 62–68, 2022, doi: 10.33650/jeecom.v4i2.4309.

[4] M. Dhanaraju, P. Chenniappan, K. Ramalingam, S. Pazhanivelan, and R. Kaliaperumal, "Smart Farming: Internet of Things (IoT)-Based Sustainable Agriculture," *Agric.*, vol. 12, no. 10, pp. 1–26, 2022, doi: 10.3390/agriculture12101745.

[5] Y. M, A. E. A, A. S. R, K. S, and P. Dr.V, "Smart Home Automation System Based on Li-Fi Technology," *Int. J. Innov. Res. Eng.*, no. May, pp. 342–345, 2023, doi: 10.59256/ijire.2023040397.

[6] P. Wijaya and T. Wellem, "Perancangan dan Implementasi Sistem Pemantauan Suhu dan Ketinggian Air pada Akuarium Ikan Hias berbasis IoT," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 225, 2022, doi: 10.30865/json.v4i1.4539.

[7] H. Dipak Ghael, L. Solanki, G. Sahu, and A. Professor, "A Review Paper on Raspberry Pi and its Applications," *Int. J. Adv. Eng. Manag. (IJAEM*, vol. 2, no. January 2020, p. 225, 2008, doi: 10.35629/5252-0212225227.

[8] K. Yahya Nashrullah, M. Bhanu Setyawan, and A. C. Fajaryanto, "JURNAL ILMIAH MAHASISWA UNIVERSITAS MUHAMMADIYAH PONOROGO (KOMPUTEK) RANCANG BANGUN IoT SMART FISH FARM DENGAN KENDALI RASPBERRY PI DAN WEBCAM," vol. 0985, pp. 81–91, 2019.

[9] Kennedi Sembiring, Agus Setiawan, Muhammad Alief Tegar Wicaksono, and Abdul Rahman, "Perancangan Automatic Fish Feeder Skala Akuarium Berbasis Internet of Things (IoT) Menggunakan Modul ESP8266," *E-JOINT (Electronica Electr. J. Innov. Technol.*, vol. 3, no. 2, pp. 66–72, 2022, doi: 10.35970/e-joint.v3i2.1671.

[10] N. Ya'acob, N. N. S. N. Dzulkefli, A. L. Yusof, M. Kassim, N. F. Naim, and S. S. M. Aris, "Water Quality Monitoring System for Fisheries using Internet of Things (IoT)," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1176, no. 1, p. 012016, 2021, doi: 10.1088/1757-899x/1176/1/012016.

[11] R. K. Putra Asmara, "Rancang Bangun Alat Monitoring Dan Penanganan Kualitas Ait Pada Akuarium Ikan Hias Berbasis Internet Of Things (IOT)," *J. Tek. Elektro dan Komput. TRIAC*,

vol. 7, no. 2, pp. 69–74, 2020, doi: 10.21107/triac.v7i2.8148.
[12] F. Chuzaini and Dzulkiflih, “IoT Monitoring Kualitas Air dengan Menggunakan Sensor Suhu , pH , dan Total Dissolved Solids (TDS),” *J. Inov. Fis. Indones.*, vol. 11, no. 3, pp. 46–56, 2022.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 3. Februari 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Estimasi Pertumbuhan Penduduk Jawa Barat Menggunakan Metode Regresi Linear Berganda

Opiana[a1], Nana Suarna[a2], Willy Prihartono[b3]

[a]Program Studi Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon Jawa Barat
[1]vanopiana@gmail.com
[2]st_nana@yahoo.com

[b]Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon Jawa Barat
[3]willyprihartono@gmail.com

Abstract

Pertumbuhan penduduk merupakan besaran persentase perubahan jumlah penduduk di suatu wilayah dengan waktu tertentu yang dibandingkan dengan jumlah penduduk pada waktu sebelumnya. Permasalahan yang terjadi saat ini adalah pertumbuhan penduduk yang kian meningkat dari waktu ke waktu, sehingga lembaga sensus sering mengalami kesulitan dalam mengestimasi laju pertumbuhan penduduk tiap tahunnya. Penelitian ini bertujuan untuk mengetahui penerapan data mining dengan metode regresi linear berganda untuk mengestimasi laju pertumbuhan penduduk di Provinsi Jawa Barat. Tahapan yang dilakukan menggunakan proses *KDD (Knowledge Discovery in Database)* yaitu dengan melakukan seleksi data, pra-Proses data, transformasi data, data mining, dan yang terakhir interpretasi dan evaluasi yang akan menghasilkan *output* berupa pengetahuan baru yang dapat memberikan kontribusi yang lebih baik. Hasil prediksi pertumbuhan penduduk di Jawa Barat pada tahun 2023 sebanyak 50.074 juta jiwa atau naik 1.49% dari tahun sebelumnya, pada tahun 2024 sebanyak 50.811 juta jiwa memiliki kenaikan sebesar 1.47% dari tahun sebelumnya. Maka penggunaan metode regresi linear berganda dapat dijadikan referensi dalam estimasi pertumbuhan penduduk di Jawa Barat.

Keywords: Pertumbuhan Penduduk, Data Mining, Regresi Linear Berganda

1. Introduction

Penduduk ialah sekumpulan warga yang tinggal di wilayah tertentu untuk menetap dengan kebutuhan yang berlaku[1]. Salah satu indikator penting untuk mengukur perkembangan suatu wilayah adalah dari pertumbuhan penduduknya. Pertumbuhan penduduk merupakan besaran persentase jumlah penduduk di suatu wilayah dengan waktu tertentu yang dibandingkan dengan jumlah penduduk pada waktu sebelumnya[2]. Permasalahan yang selama ini terjadi adalah pertumbuhan penduduk yang kian meningkat dari waktu ke waktu[3]. Sehingga, lembaga sensus sering mengalami kendala dalam mengestimasi laju pertumbuhan penduduk tiap tahunnya.

Penelitian sebelumnya yang dilakukan Putri Ardiyanti berjudul "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Denpasar Menggunakan Metode Regresi Linear Berganda" dalam mengaplikasian metode regresi linear berganda dalam mengestimasi laju pertumbuhan penduduk berhasil menerapkan metode regresi linear berganda untuk memprediksi laju pertumbuhan penduduk di Kota Denpasar dengan mendapatkan hasil bahwa penduduk Denpasar akan mengalami peningkatan pertumbuhan penduduk sebesar 7,5 % pada tahun 2023 dan 12,25% pada tahun 2024[4].

Penelitian kedua yang dilakukan Purwadi dalam penelitian yang berjudul "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linear Berganda Pada BPS Deli Serdang" berhasil mengestimasi pertumbuhan penduduk pada Kabupaten Deli Serdang periode tahun 2018 telah terjadi penambahan penduduk sebanyak 66.243 jiwa dari tahun sebelumnya, ,dengan menggunakan metode regresi linear berganda ditemukan pola yang berkaitan erat antara atribut jumlah laki-laki dan jumlah perempuan terhadap laju pertumbuhan penduduk[2].

Berdasarkan penelitian yang telah dilakukan oleh peneliti sebelumnya yang telah berhasil menerapkan data mining dengan metode regresi linear berganda di daerah lain. Maka, Penelitian ini akan mengestimasi laju Pertumbuhan penduduk di Provinsi Jawa Barat untuk melihat tingkat akurasi dan validitas estimasi laju pertumbuhan penduduk menggunakan pendekatan data mining dengan metode regresi linear berganda. Tujuan melakukan penelitian ini mengetahui penerapan data mining dengan menggunakan metode regresi linear berganda untuk menghasilkan estimasi laju pertumbuhan penduduk di Provinsi Jawa Barat dengan akurat. Selain itu, penelitian ini memiliki tujuan untuk memberikan kontribusi dalam pengembangan model prediksi yang lebih akurat dalam estimasi laju pertumbuhan penduduk.

Metode yang digunakan dalam penelitian ini adalah metode regresi linear berganda. Regresi linear merupakan salah satu cara prediksi menggunakan garis lurus untuk menggambarkan hubungan diantara dua variabel atau lebih[2]. Penggunaan data mining dengan metode regresi linear berganda dapat membantu dalam mengidentifikasi hubungan antar berbagai faktor yang mempengaruhi pertumbuhan penduduk. Tahapan yang dilakukan menggunakan proses KDD (Knowledge Discovery in Database) yaitu dengan melakukan seleksi data, pra-proses data, transformasi data, data mining, dan yang terakhir interpretasi serta evaluasi yang akan menghasilkan output berupa pengetahuan baru yang dapat memberikan kontribusi lebih baik[5].

2. Reseach Methods

Penelitian ini menggunakan metode kuantitatif. Metode penelitian kuantitatif digunakan untuk mengumpulkan data secara sistematis dan mengukur hubungan antar variabel. Pendekatan kuantitatif dianggap efektif dalam memberikan gambaran yang objektif dan dapat diukur terhadap laju pertumbuhan penduduk. Tahapan yang dilakukan selama penelitian ditampilkan dalam *figure* 1.

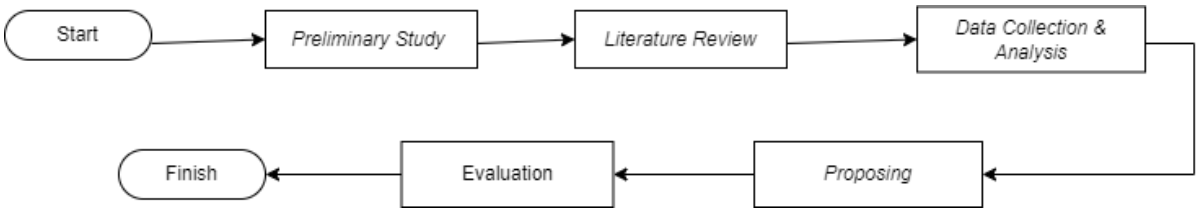


***Figure 1**. Tahapan Metode penelitian*

2.1. Pengumpulan Data

Penelitian ini menggunakan data sekunder, data sekunder merupakan informasi atau data yang telah dikumpulkan, diolah atau dipublikasikan oleh pihak lain atau lembaga tertentu sebelumnya dengan tujuan yang berbeda dan kemudian digunakan kembali dalam penelitian atau analisis terbaru. Adapun data ini diperoleh dari data publik yang bersumber dari Open Data Jabar https://opendata.jabarprov.go.id/id/dataset/jumlah-penduduk-berdasarkan-jenis-kelamin-dan-kabupatenkota-di-jawa-barat.

2.2. Teknik Analisis Data

Tahapan analisis data dilakukan dengan metode Knowledge Discovery in Databases (KDD) melibatkan serangkaian langkah sistematis untuk mengekstrak pengetahuan dari data. Berikut tahapan-tahapan dalam analisis data dengan metode KDD ditampilkan dalam *figure* 2.

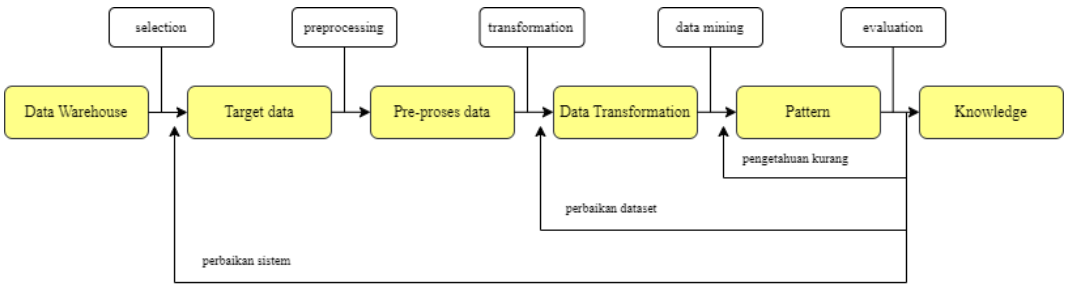


***Figure 2.** Tahapan Proses KDD*

a. Selection
Seleksi data melibatkan pemilihan dataset dan variabel-variabel yang akan menjadi fokus untuk di analisis. Pemilihan data yang relevan bertujuan untuk memastikan bahwa dataset yang

dipilih tidak hanya sesuai kebutuhan penelitian, tapi juga memiliki potensi untuk menghasilkan informasi berharga.

b. Preprocessing
Tahapan pra pemrosesan data melibatkan rangkaian kegiatan yang bertujuan untuk membersihkan dan merapikan data. Proses ini mencakup penanganan nilai yang hilang, deteksi outlier, normalisasi data, dan kegiatan pra pemrosesan data lainnya. Tujuan utama dari tahapan ini adalah untuk mempersiapkan data dengan cermat dan teliti agar memastikan kualitas dan kebersihan data yang memadai untuk analisis lebih lanjut. Dengan mengatasi nilai yang hilang, mendeteksi anomali, dan normalisasi data, tahapan pra pemrosesan ini bertujuan untuk mengoptimalkan integritas data, sehingga hasil analisis yang dihasilkan dapat diandalkan dan memberikan wawasan yang mendalam.

c. Transformation
Pengubahan data menjadi bentuk yang sesuai atau bermanfaat untuk analisis merupakan suatu proses yang melibatkan sejumlah tindakan, seperti pemfilteran variabel, pembentukan variabel baru, atau konversi format data. Proses ini bertujuan untuk mengoptimalkan data agar dapat diandalkan dalam analisis lebih lanjut.

d. Data Mining
Metode regresi linear berganda digunakan untuk menguji pengaruh lebih dari satu variabel independen terhadap variabel dependen. Bentuk persamaan untuk regresi linear berganda didefinisikan dengan rumus berikut[6]:

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n \quad (1)$$

Keterangan:
Y = Variabel tak bebas (nilai variabel yang akan diprediksi)
a = Konstanta
b_1,b_2,⋯,b_n = Nilai koefisien regresi
X_1,X_2,⋯,X_2 = Variabel bebas

Penelitian ini menggunakan 2 variabel bebas, yaitu X1 dan X2, maka bentuk persamaan regresinya adalah persamaan (2):

$$Y = a + b_1X_1 + b_2X_2 \quad (2)$$

Dalam penelitian ini jumlah penduduk didefinisikan sebagai Y atau variabel dependen, penduduk laki-laki dan penduduk di definiskan sebagai X1 dan X2 yaitu sebagai variabel independen. Sehingga apabila dirumuskan maka menjadi persamaan (3):

$$Jumlah\ Penduduk = a + b_1.Penduduk\ laki - laki + b_2.Penduduk\ Perempuan \quad (3)$$

e. Evaluation
Proses evaluasi, sebagai tahap integral dalam KDD, mencakup penilaian mendalam terhadap kualitas hasil dari data mining yang telah dilakukan. Fokus utama evaluasi adalah memastikan bahwa model atau pola yang diidentifikasi melalui data mining dapat dianggap sebagai hasil yang dapat diandalkan dan dapat diimplementasikan secara praktis dalam konteks aplikatif. Selain itu, evaluasi ini juga bertujuan untuk memverifikasi bahwa temuan yang dihasilkan sesuai dengan tujuan penelitian yang telah ditetapkan sebelumnya. Oleh sebab itu, untuk mengevaluasi model regresi linear dalam estimasi laju pertumbuhan penduduk digunakan evaluasi metrik *Root Mean Squared Error (RMSE) dan R-Squared (R2).*

RMSE (Root Mean Square Error adalah akar kuadrat dari Mean Squared Error (MSE) yang dihasilkan. RMSE digunakan karena untuk mengukur seberapa baik model dapat memprediksi nilai respons jumlah penduduk (variabel dependen) berdasarkan Jumlah penduduk laki-laki dan jumlah penduduk perempuan (variabel independen). RMSE mengukur seberapa dekat prediksi model dengan nilai yang sebenarnya[7]. Semakin kecil nilai yang dihasilkan semakin bagus pula hasil yang dihasilkan. Secara sederhana, RMSE merupakan metode untuk menghitung bias dalam model peramalan[8].
Adapun rumus RMSE yang digunakan dirampilkan pada persamaan (4):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - Y'_t)^2}{n}} \quad (4)$$

Keterangan:
n = Jumlah Sampel Dalam Data
Yt = Nilai Aktual
Y't = Nilai Prediksi

3. Result and Discussion

Hasil penelitian yang dilakukan dalam pembahasan ini akan menguraikan proses bagaimana estimasi pertumbuhan penduduk di Jawa Barat dengan menggunakan regresi linear berganda.

3.1. Data

Data yang digunakan dalam penelitian ini adalah dataset sebanyak 540 record dan terdiri 9 atribut, dengan rentang waktu data dari tahun 2013 sampai dengan 2022. Dataset bersumber dari Repository Open Data Jabar dengan situs https://opendata.jabarprov.go.id. Hasil penelusuran didapatkan data jumlah penduduk berdasarkan kabupaten kota dalam bentuk dokumen soft file format Microsoft Excel. Seperti disajikan pada tampilan tabel 1 berikut ini: (Dataset hanya disajikan sebanyak 10 record, dari dataset keseluruhan yang berjumlah sebanyak 540 record).

Table 1. Data Jumlah Penduduk Jawa Barat

	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_kelamin	jumlah_penduduk	satuan	Tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	LAKI-LAKI	1930902	JIWA	2013
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUAN	1826962	JIWA	2013
3	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	LAKI-LAKI	1258939	JIWA	2013
4	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	PEREMPUAN	1171101	JIWA	2013
5	32	JAWA BARAT	3203	KABUPATEN CIANJUR	LAKI-LAKI	1154944	JIWA	2013
…	…	…	…	…	…	…	…	…
536	32	JAWA BARAT	3277	KOTA CIMAHI	PEREMPUAN	281882	JIWA	2022
537	32	JAWA BARAT	3278	KOTA TASIKMALAYA	LAKI-LAKI	379050	JIWA	2022
538	32	JAWA BARAT	3278	KOTA TASIKMALAYA	PEREMPUAN	367660	JIWA	2022
539	32	JAWA BARAT	3279	KOTA BANJAR	LAKI-LAKI	104346	JIWA	2022
540	32	JAWA BARAT	3279	KOTA BANJAR	PEREMPUAN	102884	JIWA	

3.2. Data Selection

Data yang dipilih pada penelitian ini adalah data jumlah penduduk Jawa Barat dari tahun 2013 sampai 2022. Data ini merupakan data publik, dan memiliki 540 record dan 9 Atribut. Hasil selection dari dataset yang tidak digunakan adalah 5 atribut sehingga hanya akan menggunakan 4 atribut yaitu nama kabupaten kota, jenis kelamin, jumlah penduduk dan tahun. Karena 5 atribut yang tidak digunakan memiliki record data yang sama pada masing-masing atributnya. Atribut sebelum di seleksi ditampilkan pada tabel 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Table 2. Sebelum Seleksi Atribut

No	Nama Atribut	Data Type
1	Id	integer
2	kode_provinsi	integer
3	nama_provinsi	object
4	kode_kabupaten_kota	integer
5	nama_kabupaten_kota	object
6	jenis_kelamin	object
7	jumlah_penduduk	integer
8	Satuan	object
9	Tahun	integer

Tabel 2 merupakan tabel yang berisi nama atribut beserta tipe datanya sebelum diseleksi. Kemudian atribut yang digunakan untuk penelitian ini setelah hasil diseleksi akan menggunakan 4 atribut seperti ditampilkan pada tabel 3 berikut.

Table 3. Hasil Setelah Seleksi Atribut

no	Nama Atribut	Data Type
1	Nama_kabupaten_kota	object
2	Jenis_kelamin	object
3	Jumlah_penduduk	integer
4	tahun	integer

3.3. Data Mining

Pada tahapan data mining dilakukan proses perhitungan dengan menggunakan model regresi linear berganda. Model regresi dibuat menggunakan model *linear regression* dari *scikit-learn model.* Model ini dilatih dan diuji menggunakan data training sebesar 70% dan data testing sebesar 30%. Model juga di evaluasi menggunakan evaluasi metrik *root mean squared error (rmse), dan r-squared (r2).*

Tahap pertama dilakukan pendefinisian atribut penduduk laki-laki dan atribut penduduk perempuan sebagai nilai x (variabel independen) dan jumlah penduduk didefinisakan sebagai nilai y (variabel dependen) untuk digunakan dalam analisis atau pemodelan regresi linear berganda. Setelah variabel didefinisikan selanjutnya data dibagi menjadi 2 bagian yaitu *training set* dan *test set*. Dengan menggunakan fungsi *'train_test_split' dari scikit learn.* Setelah dilakukan *split data* selanjutnya proses pembuatan model regresi linear dengan menggunakan *library scikit-learn* dan melatihnya dengan data *training.* Model yang telah dibuat digunakan untuk membuat prediksi pada data baru dan untuk mengevaluasi kinerja model pada *data test.*

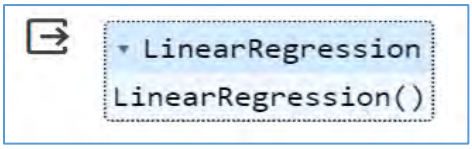


Figure 3. Model Regresi Linear

Pada *Figure* 3 menunjukan bahwa model regresi linear berhasil dibuat. Setelah model berhasil dibuat selanjutnya dilakukan prediksi dengan menggunakan model regresi linear yang telah dilatih pada *data x_test.* Hasil prediksi nantinya digunakan untuk dibandingkan dengan nilai aktual untuk mengevaluasi kinerja model. Hasil prediksi model ditampilkan pada *figure* 4.

```
[1114.08246239 3535.91175116 2451.95432232 3583.90334567 1219.06986152
 2360.97570842 2271.98282404 2480.95066957 1021.08390535 2430.03079491
 1827.03047755 1240.05486343 1139.07313697  531.1205712  1824.02571991
 1253.05065824 2340.99363409 1184.07204924 2218.04410378 1114.07441213
 1573.02749357 1592.05292932 2177.03275109  521.12148377 2509.95909222
 2509.97318018 1712.03996587 1650.04562383  720.11036753 2401.98102341
 2517.00172358 2383.99474144 2544.9387914   726.10981999 1324.05122295
 1199.07168666 1260.06511368 1845.0127344  5326.80365474 2447.96273761
 1851.00011146 3491.97413089  204.14538095 1596.02237581 2467.9468245
 1113.08959763  349.13315491  695.11365525 2371.00498433 1893.00030378
  206.14519843 2500.94683186 1804.02955763 1278.05944592 2210.06294693
  512.1273365  2513.99092824 2611.95984679  376.12968468  429.12585432
 2060.00619574 1797.03120271 2075.01187085 4357.8890643   197.14702603
 1642.05842928 1315.05506311 2323.97807866  337.13425    1733.03905575
  333.13662759 2332.98027619 2640.95619405  437.12512426 1840.0061467
 1674.04142108 3524.92784923 2364.98138108 1187.06473149 2411.96803544
 2056.04681208 1132.07276949 1781.03467539 1165.06673915 2758.96957646
  978.09487339 2527.94134906  713.11000005 1098.08392251 1137.07935719]
```

Figure 4. Hasil Prediksi Model Regresi Linear

Figure 4 merupakan hasil prediksi menggunakan model regresi dari data testing (x_test) yang merupakan data latih dari atribut penduduk laki-laki dan penduduk perempuan. Hasil ini nantinya akan dibandingkan dengan nilai data aktual untuk mengevaluasi kinerja model.

3.4. Evaluasi

Root Mean Squared Error (RMSE) adalah metrik yang digunakan untuk mengukur seberapa baik model regresi dapat memprediksi nilai target jumlah penduduk pada dataset pengujian. RMSE dihitung sebagai akar kuadrat dari Mean Squared Error (MSE). Nilai RMSE yang didapatkan sekitar 0.4685 dalam satuan yang sama dengan variabel dependen menunjukan bahwa sebaran kesalahan relatif kecil. Hasil nilai RMSE adalah 0.4685329901519414. Semakin kecil nilai RMSE, semakin baik kinerja model regresi. Oleh karena itu, nilai 0.4685329901519414 menunjukkan bahwa model regresi memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target pada dataset pengujian. Disimpulkan bahwa model memiliki kemampuan yang baik dalam melakukan prediksi pada data yang tidak terlihat sebelumnya. Ditampilkan pada *figure* 5.

```
Root Mean Squared Error/RMSE = 0.4685329901519414
```

***Figure* 5**. Hasil RMSE

R-squared (R^2) dikenal sebagai koefisien determinasi, adalah metrik evaluasi yang digunakan untuk mengevaluasi seberapa baik model regresi linier cocok dengan data yang diamati. R-squared memberikan informasi tentang seberapa baik variabilitas dalam variabel dependen dapat dijelaskan oleh model regresi. Evaluasi model menggunakan R-squared memiliki rentang Antara 0 hingga 1, dimana 1 menunjukan model yang sempurna yang mampu menjelaskan seluruh variabilitas data dan 0 menunjukan model yang tidak dapat menunjukan variabilitas sama sekali. Hasil yang didapatkan menunjukan nilai sekitar 0.99999 menunjukan bahwa model mampu menjelaskan sebagian besar

p-ISSN: 2301-5373
e-ISSN: 2654-5101

variabilitas dalam data jumlah penduduk berdasarkan variabel independen yang digunakan yaitu penduduk laki-laki dan penduduk perempuan. Model regresi linear sangat cocok dengan data dan dapat menjelaskan sebagian besar variabilitas dalam jumlah penduduk. Ditampilkan pada *figure* 6.

```
R-squared/R2= 0.9999997542871234
```

***Figure* 6**. Hasil R-Squared

Dalam diagram terdapat kemiripan yang sangat tinggi antara nilai aktual dan nilai prediksi, dapat dilihat bahwa sebagian besar titik-titik mendekati garis lurus 45 derajat yang merupakan garis regresi. Garis ini menunjukkan relasi linier antara nilai aktual dan nilai prediksi. Sebaran titik-titik yang rapat menunjukan bahwa model dengan baik memprediksi nilai-nilai dalam data test.ditampillkan pada f*igure* 7.

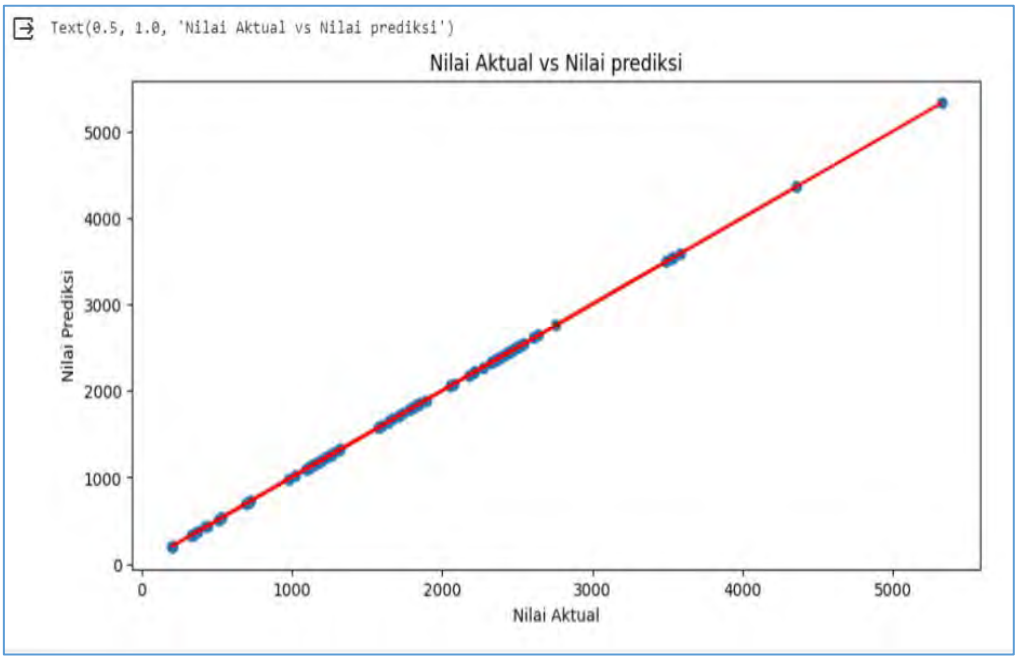


***Figure* 7**. Visualisasi Regresi

Untuk melihat selisih antara nilai aktual dan nilai prediksi dibuat satu dataframe baru untuk memberikan gambaran lebih rinci tentang bagaimana model melakukan prediksi pada data test. Proses ditampilkan pada f*igure* 8.

	nilai aktual	nilai prediksi	selisih
64	1114.0	1114.574542	0.574542
135	3536.0	3535.937879	-0.062121
153	2452.0	2452.191901	0.191901
189	3583.0	3582.820903	-0.179097
253	1218.0	1218.206907	0.206907
...	...	...	...
21	978.0	978.063250	0.063250
234	2528.0	2527.803497	-0.196503
161	714.0	713.665385	-0.334615
16	1098.0	1098.512846	0.512846
118	1137.0	1137.573014	0.573014

90 rows × 3 columns

***Figure* 8**. Hasil Selisish Nilai Aktual Dan Prediksi

Pada Figure 6 dapat dilihat bahwa atribut "nilai aktual" berisi nilai aktual dari variabel dependen (y_test), atribut "nilai prediksi" berisi nilai yang diprediksi oleh model regresi (y_pred) dan atribut "selisih" merupakan perbedaan antara nilai prediksi dan nilai aktual (nilai prediksi - nilai aktual). Dalam gambar terlihat adanya nilai selisih positif yang berarti model memprediksi nilai yang lebih tinggi daripada nilai aktual dan terdapat nilai selisih negatif yang berarti model memprediksi nilai yang lebih rendah daripada nilai aktual. Dari hasil selisih yang didapat terlihat bahwa selisih dari nilai prediksi dan nilai aktual memiliki nilai yang sangat kecil sehingga dapat disimpulkan bahwa model yang dibuat cenderung memberikan prediksi yang baik.

3.5. Prediksi 5 Tahun Berikutnya

Pada hasil menunjukan bahwa pada tahun 2023 jumlah penduduk diperkirakan sekitar 50.074 juta. Tahun 2024 berjumlah sekitar 50.811 juta. Pada tahun 2025 diperkirakan berjumlah 51.548 juta. Kemudian pada tahun 2026 sekitar 52.285 juta dan pada tahun 2027 jumlah penduduk sekitar 52.633 juta jiwa.. Ditampilkan pada *figure* 9.

	Tahun	Penduduk_Laki-laki	Penduduk_Perempuan	Jumlah_Penduduk_Prediksi
0	2023	25334.0	24744.0	50074.0
1	2024	25681.0	25134.0	50811.0
2	2025	26029.0	25524.0	51548.0
3	2026	26376.0	25914.0	52285.0
4	2027	26723.0	25914.0	52633.0

***Figure* 9**. Hasil Prediksi

Untuk melihat bagaimana hasil prediksi jumlah berperilaku terhadap data historis maka dibuat tampilan visualisasi tren data historis. Ditampilkan pada *figure* 10.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

***Figure* 10**. Hasil Visualisasi Data Prediksi

Figure 8 menunjukan bahwa garis biru pada plot (label 'Data Historis') mewakili data historis jumlah penduduk pada tahun tahun sebelumnya. Titik merah pada plot (label 'Prediksi 2023-2027') menunjukkan prediksi jumlah penduduk untuk tahun 2023-2027 berdasarkan hasil dari penerapan model regresi yang telah dibuat. Pada grafik titik merah menunjukan bahwa model regresi memprediksi peningkatan jumlah penduduk dari tahun ke tahun dalam rentang waktu 2023-2027.

4. Conclusion

Berdasarkan hasil penelitian menggunakan regresi linear berganda untuk memprediksi jumlah penduduk di provinsi di Provinsi Jawa Barat dapat disimpulkan bahwa:

Hasil proses prediksi menggunakan metode regresi linear berganda untuk memprediksi jumlah penduduk di Jawa Barat menggunakan model regresi dapat bekerja dengan baik. Model prediksi yang di evaluasi menggunakan berbagai evaluasi metrik menunjukan bahwa model prediksi yang dibuat memiliki akurasi yang baik dalam memprediksi jumlah penduduk berdasarkan penduduk laki-laki dan penduduk perempuan.

Hasil akurasi model untuk prediksi laju pertumbuhan penduduk di Jawa Barat menunjukkan tingkat akurasi yang cukup baik. Hasil evaluasi metrik Root Mean Squared Error (RMSE) menunjukan nilai 0.4685 Dari evaluasi metrik yang telah dilakukan menunjukkan bahwa hasil nilai yang didapatkan semuanya memiliki nilai yang kecil sehingga dapat disimpulkan bahwa model regresi bekerja dengan baik dan memiliki tingkat kesalahan prediksi atau error kecil sehingga dapat memprediksi jumlah penduduk dengan akurat. Selain itu hasil evaluasi menggunakan R-Squared menunjukan nilai 0.9999, hasil yang didapat mendekati nilai 1 yang menunjukan bahwa model regresi dapat membaca variabilitas data sehingga dapat mendukung kesimpulan bahwa model yang digunakan memang memiliki tingkat akurasi yang sangat baik.

Dari hasil penelitian didapatkan hasil prediksi jumlah penduduk untuk 5 (lima) tahun berikutnya yaitu pada tahun 2023 jumlah penduduk diperkirakan sekitar 50.074 juta jiwa atau naik 1.49% dari tahun sebelumnya yang hanya berjumlah 49.339 juta jiwa pada tahun 2022. Pada tahun berikutnya juga diprediksi terjadi kenaikan sebesar 1.47% pada tahun 2024 sehingga jumlah penduduk diperkirakan akan sebanyak 50.811 juta jiwa. Kemudian pada tahun 2025 diprediksi terjadi kenaikan sebesar 1.45% pada tahun ini jumlah penduduk diprediksi sebanyak 51.548 juta jiwa. Tahun 2026 juga diprediksi mengalami peningkatan jumlah penduduk sebesar 1.43% sehingga berjumlah 52.285 juta jiwa dan pada tahun 2027 diperkirakan akan mengalami sedikit kenaikan sekitar 0.66% sehingga pada tahun 2027 jumlah penduduk di Jawa Barat akan sebanyak 52.633 juta jiwa

References

[1] I. Indriani, D. Siregar, and A. P. Windarto, "Penerapan Metode Linear Regression dalam Mengestimasi Jumlah Penduduk," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1112, 2022, doi: 10.30865/jurikom.v9i4.4676.

[2] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 55, 2019, doi: 10.53513/jis.v18i1.104.

[3] E. D. Sri Mulyani *et al.*, "Estimasi Pertumbuhan Penduduk Di Kabupaten Tasikmalaya Menggunakan Metode Regresi Linear Berganda," *Infosys (Information Syst. J.*, vol. 6, no. 1, p. 1, 2021, doi: 10.22303/infosys.6.1.2021.1-11.

[4] A. A. A. P. Ardyanti and A. Abdriando, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Denpasar Menggunakan Metode Regresi Linier Berganda," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 6, no. 1, pp. 37–44, 2023, doi: 10.30813/jbase.v6i1.4317.

[5] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021, doi: 10.47065/jimat.v1i2.104.

[6] I. B. M. Swarbawa, I. G. Arta Wibawa, and I. K. Gede Suhartana, "Prediksi Hasil Panen Padi Di Kabupaten Jembrana Dengan Metode Linear Regression," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 3, p. 671, 2023, doi: 10.24843/jlk.2023.v11.i03.p24.

[7] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.

[8] N. Litha and T. Hasanuddin, "Analisis Performa Metode Moving Average Model untuk Prediksi Jumlah Penderita Covid-19," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 87–95, 2020, [Online]. Available: https://jurnal.yoctobrain.org/index.php/ijodas/article/view/19.

ISSN

E-ISSN