



UNIVERSITAS UDAYANA

Volume 3, Nomor 2, Februari 2025

# JNATIA

Jurnal Nasional Teknologi Informasi dan Aplikasinya

Program Studi Informatika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Udayana

**Susunan Dewan Redaksi**  
**Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)**  
**Volume 3, Nomor 2, Februari 2025**

**Penanggung Jawab**

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

**Ketua Redaktur**

I Gede Surya Rahayuda, M.Kom.

**Editor**

Dr. Drs. I Wayan Santiyasa, M.Si.  
Dr. Dra. Luh Gede Astuti, M.Kom.  
Ir. I Gede Arta Wibawa, S.T., M.Kom.  
I Dewa Made Bayu Atmaja Darmawan, S.Kom., M.Cs.  
I Made Widhi Wirawan, S.Si., M.Si., M.Cs.

**Desain Grafis**

I Gede Yogananda Adi Baskara  
I Gusti Agung Ayu Gita Pradnyaswari Mantara

**Fotografer**

I Kadek Agus Candra Widnyana  
I Komang Dwiprayoga

**Sekretariat**

Ni Ketut Alit Widiastuti, S.Kom.  
Anak Agung Raka Darmawan, S.Kom.  
I Putu Herryawan, S.Kom.

**Reviewer**

Dr. Ir. I Ketut Gede Suhartana, S.Kom., M.Kom., IPM., ASEAN.Eng.  
Ir. I Gede Arta Wibawa, S.T., M.Kom.  
I Made Widiartha, S.Si., M.Kom.  
Gst. Ayu Vida Matrika Giri, S.Kom., M.Cs.  
Ir. I Gusti Agung Gede Arya Kadyanan, S.Kom., M.Kom.  
Dr. Ir. Ngurah Agus Sanjaya ER, S.Kom., M.Kom.  
I Dewa Made Bayu Atmaja Darmawan, S.Kom., M.Cs.  
Luh Arida Ayu Rahning Putri, S.Kom., M.Cs.  
Ir. Agus Muliantara, S.Kom., M.Kom.  
Dr. Dra. Luh Gede Astuti, M.Kom.

Cokorda Rai Adi Pramatha, S.T., M.M., Ph.D.  
Ir. I Gusti Ngurah Anom Cahyadi Putra, S.T., M.Cs.  
Dr. Anak Agung Istri Ngurah Eka Karyawati, S.Si., M.Eng.  
I Gede Santi Astawa, S.T., M.Cs.  
Dr. Made Agung Raharja, S.Si., M.Cs.  
I Komang Ari Mogi, S.Kom., M.Kom.  
Ida Ayu Gde Suwiprabayanti Putra, S.Kom., M.T.  
I Putu Gede Hendra Suputra, S.Kom., M.Kom.  
Dr. Drs. I Wayan Santiyasa, M.Si.  
I Gede Surya Rahayuda, M.Kom.  
I Wayan Supriana, S.Si., M.Cs.

## Daftar Isi

|   |         |
|---|---------|
| Klasifikasi Citra Buah Catur Menggunakan Metode Klasifikasi Random Forest<br>Eugenio Darrin, AAIN Eka Karyawati   | 233-238 |
| Rancangan Sistem Pembuatan Kompos dengan Tempat Sampah Berbasis IOT<br>(Internet of Things)<br>IDG Surya Baaskara, AAIN Eka Karyawati   | 239-246 |
| Komparasi Ekstraksi Fitur BoW dan TF-IDF untuk Klasifikasi SMS Menggunakan<br>Naive Bayes<br>I Komang Dwipayoga, Made Agung Raharja   | 247-254 |
| Pengamanan Data Finansial Menggunakan Enkripsi Homomorfik Pailler<br>Ida Bagus Made Wiguna Tedja Sukmana, Agus Muliantara   | 255-260 |
| Aspect Based Sentiment Analysis Terhadap Ulasan Produk Skincare di E-<br>Commerce Menggunakan CNN-LSTM<br>I Made Treshnanda Mas, AAIN Eka Karyawati   | 261-270 |
| Evaluasi UI Design pada Prototype Aplikasi "Instix" Menggunakan Metode Sistem<br>Usability Scale<br>Ni Putu Dita Ari Pratiwi, AAIN Eka Karyawati  | 271-282 |
| Optimasi Hyperparameter Algoritma Support Vector Machine dalam Klasifikasi<br>Penyakit $\beta$ -Thalassemia<br>I Nyoman Adi Mahendra Putra, Cokorda Pramarta                                  | 283-294 |
| Klasifikasi Berita Berdasarkan Kategori Menggunakan Multinomial Naive Bayes<br>dengan K-Cross Validation dan Seleksi Fitur Chi-Squared<br>Febrian Valentino Agape, Gst. Ayu Vida Matrika Giri | 295-304 |
| Rancang Bangun Website Sebagai Sarana Promosi Usaha Madu Kele-Kele di Desa<br>Bongkasa Pertiwi<br>Ni Made Desni Dwi Arisaputri, Gst. Ayu Vida Matrika Giri                                    | 305-314 |
| Analisis Perbandingan Enkripsi File Teks Berformat .txt dan .docx<br>Menggunakan Algoritma AES<br>Kennardy Andrew Limartha, I Gede Arta Wibawa  | 315-320 |
| Implementasi Algoritma Yolo untuk Deteksi Tanaman Apotik Hidup<br>I Kadek Peri Arta Wijaya, I Gede Santi Astawa   | 321-330 |
| Pemanfaatan Asap Detektor untuk Pembuatan Smart Garasi Berbasis IoT<br>Jamrud Ivan Hartono, I Gede Surya Rahayuda   | 331-336 |
| Penerapan Steganografi untuk Pengamanan Konten Gambar dalam Media Sosial<br>Gede Eka Putra Wijay, I Gusti Agung Gede Arya Kadnyanan   | 337-342 |
| Analisis Keamanan Seed Phrase pada Wallet Cryptocurrency Menggunakan Brute<br>Force<br>I Kadek Adi Sentana, I Gusti Ngurah Anom Cahyadi Putra   | 343-348 |



|  |         |
|--|---------|
| Perancangan UI/UX Aplikasi Kesehatan Mental "MindEase" Berbasis Mobile dengan Metode Design Thinking<br>Pande Komang Bhargo Anantha Yogiswara, I Gusti Ngurah Anom Cahyadi Putra | 349-359 |
| Pengembangan Interface dan Experience Aplikasi Manajemen Kegiatan Notelife Menggunakan Pendekatan Design Thinking<br>Ni Luh Gede Cahaya Putri Mahadewi, I Ketut Gede Suhartana   | 360-371 |
| Perancangan User Interface pada Aplikasi 'LiteraLink' dengan Menggunakan Metode Design Thinking<br>Lusia Elvira Sue Sare1, I Ketut Gede Suhartana                                | 372-383 |
| Pengamanan File Video dengan Enkripsi dan Deskripsi Menggunakan Algoritma Sals0 dan RSA<br>I Putu Herdy Juniawan, I Komang Ari Mogi  | 384-393 |
| Implementasi Algoritma Modified K-Nearest Neighbor untuk Klasifikasi Indeks Kualitas Udara Perkotaan di Berbagai Negara<br>Ni Made Ayu Pranasanthi Dewi, I Made Widiartha        | 394-403 |
| Peningkatan Kualitas Gambar pada Sistem Streaming Video Menggunakan Teknik Kompresi Citra<br>I Made Sastra Wiguna, I Wayan Santiyasa   | 404-413 |
| Sistem Rekomendasi Personalized Music dengan Metode Jenis Gaya<br>Ida Ayu Tri Sabina Putri, I Wayan Supriana   | 414-419 |
| Perancangan Desain UI/UX Aplikasi Somnia untuk Manajemen Pola Tidur Berbasis Mobile<br>Ni Putu Meita Kartika Dewi, Ida Ayu Gde Suwiprabayanti Putra                              | 420-429 |
| Analisis Sentimen Ulasan Aplikasi M-Paspor Menggunakan TF-IDF dan Support Vector Machine<br>Ni Luh Putu Happy Nirmala, Ida Bagus Gede Dwidasmara                                 | 430-437 |
| Analisis Sentimen Ulasan Aplikasi Citilink menggunakan Metode Support Vector Machine dengan TF-IDF<br>David Brave Moarota Zebua, Ida Bagus Gede Dwidasmara                       | 438-445 |
| Implementasi Ekstraksi Fitur VGG-16 dan Pemodelan LSTM untuk Pembangkitan Caption Gambar Otomatis<br>Made Pranajaya Dibyacita, Luh Gede Astuti                                   | 446-457 |
| Implementasi Internet of Things dengan Smart Faucet pada Sistem Irigasi Subak Bali<br>Ida Bagus Rahadi Putra, Ngurah Agus Sanjaya ER   | 458-469 |

## Klasifikasi Citra Buah Catur Menggunakan Metode Klasifikasi Random Forest

Eugenio Darrin<sup>\*1</sup>, AAIN Eka Karyawati<sup>#2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>eugene.darrin16@gmail.com  
<sup>2</sup>eka.karyawati@unud.ac.id

### Abstract

*Chess is known worldwide as an activity that is both considered a sport and a game alike, and yet it is also a way to develop the mind of people so that it might be sharper. In the game of chess there exists multiple types of chess pieces that are designed to be easily differentiated, and yet with the changing of time these differences might be reduced due to artistic reasons. Because of the reasons stated above that this research is conducted, which is to classify the multiple types of chess pieces so that beginner players might be able to differentiate them easily while playing a game of chess. The classification method that is proposed to be used in this research is still in the most basic knowledge and purposed to pique interest towards other researchers in using random forest classification in this topic.*

**Keywords:** Image Processing, Random Forest, Information, GLCM, Chess

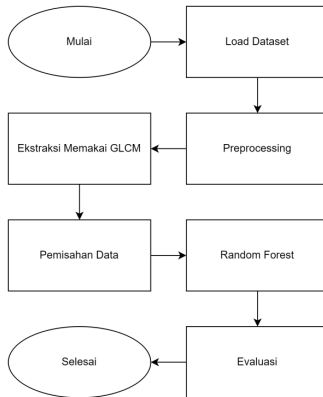
### 1. Pendahuluan

Seiring perkembangannya waktu, berbagai macam kegiatan dalam dunia mengalami perubahan, kegiatan ini dapat terlihat dari kegiatan untuk keberlangsungan hidup sampai dengan kegiatan yang bertujuan untuk hiburan. Salah satu dari kegiatan yang mengalami perkembangan merupakan catur. Permainan catur ini merupakan sebuah permainan yang berasal dari india dengan nama "Chaturanga" yang merupakan kata dari bahasa sanskrit dengan arti "4 pasukan perang". Permainan chaturanga dikembangkan berdasarkan kondisi perang dengan buah catur permainan Chaturanga yang terdiri dari Raja, Menteri, Pasukan Bergajah, Pasukan Berkuda, Kereta Perang (sekarang menjadi Benteng), dan Pasukan Infantri (sekarang menjadi Bidak).[1] Seiring berjalannya waktu, permainan catur juga telah berkembang dan terlebar luaskan menjadi sebuah kegiatan yang dapat dianggap dekat dengan permainan, akan tetapi juga diakui sebagai kegiatan olahraga dalam jangkauan dunia. Perkembangan ini tidak terlepas dari perkembangan teknologi, dimana permainan catur telah dibuat dalam bentuk digital dan menjadi penghubung antara orang diantara dunia. akan tetapi permainan ini juga memiliki sebuah permasalahan dimana dengan perkembangan waktu dapat juga terdapat perkembangan desain yang dapat memberikan kekeliruan antara dua buah catur yang berbeda, hal ini dapat terlihat dari desain buah benteng dan bidak yang walaupun memiliki desain berbeda, terkadang dapat dibuat dengan bentuk yang mirip. Dikarenakan kekeliruan yang dapat terjadi diatas, penelitian ini diajukan dengan tujuan untuk membantu klasifikasi terhadap beberapa buah catur yang ada. Penelitian ini diharapkan akan memberikan solusi untuk memberikan sebuah cara untuk membedakan sebuah citra, atau membantu memberikan informasi kelanjutan dalam mengembangkan sistem klasifikasi dengan nilai-nilai yang lebih tinggi. Kegunaan lanjut dari kegiatan ini dapat membantu dalam mengembangkan sistem permainan catur digital yang akan mendeteksi buah catur berdasarkan bentuknya dalam waktu nyata, atau membantu pembelajaran catur sendiri. Untuk melakukannya maka akan dipakai sebuah algoritma metode klasifikasi random forest dengan ekstraksi GLCM. Metode klasifikasi random forest telah dipakai dalam beberapa penelitian. Salah satu dari penelitian ini adalah pada tahun 2024, telah terdapat sebuah penelitian dengan tujuan untuk meningkatkan performanya dengan judul "Improvement of Random Forest Classifier through Localization of Persian Handwritten OCR".[2] Penelitian

memakai data berkisar 2838 citra biner dari 33 karakter yang dipakai dengan peningkatan akurasi hingga 94%. Dalam materi penelitian ini dapat terlihat bahwa penelitian dilakukan terhadap sebuah sistem penulisan yang biasanya memiliki bentuk yang lebih luas dalam perbedaannya. Sebuah penelitian lain yang dilakukan pada tahun 2022 berjudul "Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal" [3] dilakukan menggunakan data yang didapatkan dari google earth dan memberikan hasil akurasi sebesar 90,5% pada klasifikasi dengan presisi tinggi, akan tetapi presisi tertinggi tetap dimiliki oleh pencampuran GLCM dengan spectrum band untuk mendapatkan hasil akurasi sebesar 92%. Pada tahun 2022 lagi diadakan juga sebuah penelitian memakai GLCM dan random forest untuk mengklasifikasi gambar hasil MRI untuk mengklasifikasi tumor dengan judul "Learning Texture Features from GLCM for Classification of Brain Tumor MRI Images using Random Forest Classifier".[4] Klasifikasi ini dilakukan dengan dataset berukuran 245 gambar dan memiliki tingkat akurasi sebesar 83% sampai dengan 61% dimana sebuah penelitian juga menemukan bahwa tingkat akurasi akan berkurang apabila jarak antara pixel gambar dikurangkan.

## 2. Metode Penelitian

Dalam penelitian ini, data yang digunakan adalah dataset yang didapatkan dari website kaggle yang memberikan data pelatihan dan data uji dalam membuat pohon klasifikasi random forest. Tahap tahap dari penelitian akan selanjutnya dilakukan mengikuti dengan alur yang telah dibuat dalam flowchart dibawah.



Gambar 1. Flowchart Metode Penelitian




Pada flowchart diatas dapat dilihat bahwa tahapan akan melalui beberapa tahapan sebelum dapat dievaluasi kembali dengan teknik klasifikasi metode Random Forest, berikut merupakan beberapa penjelasan dasar tentang alur dari flowchart yang dibuat:

### a. Dataset

Pengumpulan data yang dipakai dalam penelitian merupakan data yang berasal dari website kaggle dengan materi berjudul "Chess Pieces Detection Images Dataset" yang dibuat oleh pengguna Anshul Mehta yang berisikan 5 buah folder citra buah catur yang

terpisahkan menjadi ratu, benteng, kuda, menteri, dan bidak. Pada penelitian kali ini akan dipakai tiga buah dari jenis data yaitu berupa gambar ratu, benteng, dan bidak yang merupakan buah catur yang memiliki bentuk yang kurang lebih mirip. Data ini kemudian akan dilakukan pengolahan kembali untuk membagi data menjadi data pelatihan dan data pengujian dengan perbandingan 80% dan 20%, sehingga data akan berupa 336 gambar 269 data pelatihan dan 67 data uji.

Tabel 1. Contoh Dataset dipakai

| No | Gambar  | Label   |
|----|---|---------|
| 1  |  | Pion    |
| 2  |  | Ratu    |
| 3  |  | Benteng |

b. Preprocessing Data

Setelah didapatkan dataset maka akan dilanjutkan dengan melakukan preprocessing data yang dilakukan dengan melakukan perubahan terhadap label dan deskripsi informasi dari citra yang dipakai.

c. Ekstraksi GLCM

Tahap setelah melakukan preprocessing data adalah ekstraksi fitur dengan memakai GLCM atau *Gray Level Co-occurrence Matrix*, dimana metode tersebut merupakan metode ekstraksi fitur statistik dengan cara menggambarkan hubungan antara dua piksel dalam citra dengan menggunakan matriks *co-occurrence*. GLCM memetakan kemunculan pasangan intensitas piksel tertentu pada jarak dan orientasi tertentu. Orientasi yang digunakan pada penelitian kali ini merupakan orientasi berdasarkan derajat yang merupakan 0°, 45°, 90°, dan 135° masing-masing sementara jarak antara piksel tergantung pada tingkat ketetanggaan yang diinginkan.[5] Pelaksanaan dari tahapan ini akan menghasilkan data berupa CSV yang mengandung 5 nilai dalam GLCM untuk masing-masing derajat, nilai ini berupa *dissimilarity*, *correlation*, *homogeneity*, *contrast*, *ASM*, dan *energy* sehingga akan didapatkan sebuah CSV yang memiliki 24 kolom detail dengan tiap derajat memegang 6 detail pada tabel, dan 1 kolom ekstra yang berisikan label yang didapatkan berdasarkan folder yang dipakai.

Setelah pembuatan tabel akan dilanjutkan dengan memisahkan lagi tabel menjadi tabel lebih kecil yang dapat diolah dengan lebih baik berdasarkan masing-masing derajat.

- d. Pemisahan Data menjadi data latih dan data uji  
Setelah data diekstraksi dengan fitur GLCM akan didapatkan 4 buah data untuk diolah yang berupa data tabel 0°, 45°, 90°, dan 135°. Data pada tabel ini kemudian akan dibagi menjadi 2 data berdasarkan tujuan dan waktu pemakaian data, yaitu sebagai data latih dan data uji, dimana data latih akan dipakai sebagai data pelatihan dari pohon metode *random tree* pada saat pelatihan model, dan data uji merupakan sebuah data yang akan dipakai pada proses pengujian dari hasil latih model pohon RFC.
- e. Random Forest Classification  
RFC atau *random forest classification* merupakan metode klasifikasi machine learning yang dilakukan dengan membuat sebuah pohon pilihan detail untuk menentukan prediksi klasifikasi. Dalam penelitian, metode ini dipakai pada setiap tabel derajat GLCM untuk membuat pohon masing-masing. Akan tetapi pemakaian metode ini memiliki kekurangan dimana metode akan dapat membuat sebuah pohon dengan variasi yang sangat berbeda antara satu sama lain.
- f. Evaluasi  
Evaluasi yang dilakukan berdasarkan hasil dari klasifikasi RFC didasarkan dari tingkatan nilai akurasi yang didapatkan antara percobaan dengan membandingkan label yang telah dibuat sebelumnya dengan label hasil prediksi oleh RFC.

### 3. Hasil dan Diskusi

Pemakaian dataset yang dipakai berupa data citra buah catur jenis ratu, benteng, dan juga bidak/pion berjumlah 336 gambar, dengan dataset pertama dimasukkan ke dalam folder masing masing dengan label sesuai nama foldernya. Dataset yang dibagi menjadi data latih dan data uji telah memiliki label berdasarkan peletakan data, data berupa citra pion sebanyak 82 citra, citra ratu sebanyak 115 data citra, dan 139 data citra benteng. Hal ini dilakukan agar pada saat pelaksanaan tes akurasi dapat dibandingkan dengan data yang sudah memiliki kepastian 100%. Pada tahapan klasifikasi menggunakan *random forest classifier* dilakukan eksperimen terhadap masing masing derajat dari GLCM yang dipakai, yaitu derajat 0°, 45°, 90°, dan 135°, dan memiliki hasil yang berbeda satu sama lain. Hasil dari tahapan ini merupakan akurasi dari klasifikasi label yang diberikan kepada citra buah catur. Hasil dari tingkatan klasifikasi ini dapat dilihat pada tabel 2.

**Tabel 2.** Hasil Akurasi Klasifikasi menggunakan *Random Forest*

| No | Sudut | Akurasi |
|----|-------|---------|
| 1  | 0°    | 48,52%  |
| 2  | 45°   | 47,05%  |
| 3  | 90°   | 45,55%  |
| 4  | 135°  | 52,94%  |

Pada percobaan pada derajat 0°, didapatkan hasil keyakinan klarifikasi sebesar 0.4852 atau 48,52%, selanjutnya pada percobaan dengan data di derajat 45° didapatkan bahwa hasil memiliki persentase keyakinan sebesar 47,05%. Pada klasifikasi di derajat 90° tingkat akurasi hanya sebesar 45,55% dan pada derajat 135° tingkat akurasi tertinggi ditemukan dengan hasil sebesar 52,94%.

### 4. Kesimpulan

Pada penelitian kali ini yang berusaha melakukan awal penelitian ke arah bidang catur memakai klasifikasi *random forest* didapatkan hasil yang masih dapat dianggap sebagai kurang memuaskan, akan tetapi dapat dibuktikan bahwa *random forest* dalam bentuk yang paling sederhana telah dapat mengklasifikasikan buah catur dengan keyakinan akurasi terendah

sebesar 45,55% dan nilai akurasi tertinggi sebesar 52,94%, nilai yang bersifat rendah ini dapat disebabkan oleh beberapa faktor, yaitu dari pemilihan jenis klasifikasi yang dipakai dapat memiliki hasil yang sangat berbeda dari perbedaan gambar yang kecil, dan juga bahwa klasifikasi dapat disebarluaskan lagi dengan detail lebih besar yang mungkin meningkatkan nilai akurasi klasifikasi. Selain dari kemungkinan permasalahan berada dari kekurangpahaman peneliti akan metode klasifikasi yang dipakai, juga dimungkinkan bahwa dataset yang dipakai dapat lebih diperbanyak dan ditambahkan detailnya. Maka dari itu peneliti mengharapkan penelitian ini akan menginspirasi peneliti lain untuk mencoba mengambil materi dan mengembangkannya dengan detail yang lebih banyak atau metode klasifikasi lain untuk dicampurkan.

#### Daftar Pustaka

- [1] (2024, May 6th). History of chess [Online]. Available: [https://en.wikipedia.org/wiki/History\\_of\\_chess](https://en.wikipedia.org/wiki/History_of_chess)
- [2] Zahedi, M., & Eslami, S. (2012). Improvement of Random Forest Classifier through Localization of Persian Handwritten OCR.
- [3] Mohammadpour P, Viegas DX, Viegas C. Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal. *Remote Sensing*. 2022; 14(18):4585. <https://doi.org/10.3390/rs14184585>
- [4] Aggarwal, Ashwani. (2022). Learning Texture Features from GLCM for Classification of Brain Tumor MRI Images using Random Forest Classifier. *WSEAS Transactions On Signal Processing*. 18. 60-63. 10.37394/232014.2022.18.8.
- [5] Neneng, K. Adi, and R. Isnanto, "Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM)", 2016, vol. 1, no. 1, pp.1-10.

Halaman ini sengaja dibiarkan kosong

## Rancangan Sistem Pembuatan Kompos dengan Tempat Sampah Berbasis IOT (Internet of Things)

IDG Surya Baaskara<sup>a1</sup>, AAIN Eka Karyawati<sup>a2</sup>,

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana

Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia

<sup>1</sup>surya.baaskara06@email.com

<sup>2</sup>eka.karyawati@unud.ac.id

### Abstract

*Increased awareness of the importance of household organic waste management has driven the development of innovation in waste treatment. This paper deals with the design and implementation of intelligent garbage cans that can turn household trash, especially food and vegetable residues, into high-quality composite materials. The system uses efficient aerobic composting technology to accelerate the organic decomposition process. Through automatic monitoring and optimum setting of environmental conditions, these garbage cans can produce fertile compost in a short time. In addition to providing practical solutions to household garbage management, the importance of keeping the environment on its own is also emphasized. Through awareness of the negative impact of waste on the environment, individuals can play an active role in promoting sustainable and environmentally friendly practices.*

**Keywords:** *Waste management, compost, environment, environmental awareness, sustainable practices.*

### 1. Pendahuluan

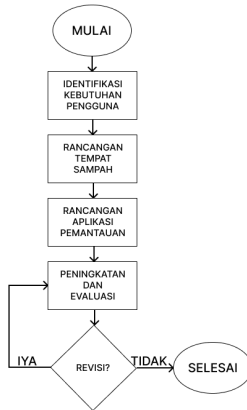
Masalah lingkungan merujuk pada perubahan yang disebabkan oleh aktivitas manusia pada lingkungan alam atau ekosistem bumi yang dinilai buruk oleh manusia. Masalah ini muncul dari interaksi manusia dengan lingkungan alamnya. Perubahan lingkungan dianggap sebagai masalah hanya ketika diinterpretasikan secara negatif dan dilihat sebagai kerusakan yang disebabkan oleh manusia [1]. Menurut Sistem Informasi Pengelola Sampah Nasional (SIPSN) sebanyak 19,296,446.26 ton sampah dihasilkan sepanjang 2023 dan 38,4% dari total sampah tersebut merupakan sampah rumah tangga dan diikuti oleh sampah dari pasar sebanyak 22,9% dan juga sampah perniagaan sebanyak 18,6%. Dari 19,296,446.26 ton sampah itu, di data 41% nya merupakan jenis sampah sisa makanan dan 12,2% nya sampah daun atau kayu, dan dapat dilihat bahwa lebih dari 50% sampah yang ada di Indonesia merupakan sampah organik dan kebanyakan berasal dari rumah tangga [2]. Banyak cara dapat dilakukan untuk mengurangi sampah organik di lingkungan kita, sesuai dengan data yang ditunjukkan bahwa lingkungan kita paling dasar, lingkungan rumah menjadi penghasil sampah nomor 1 dalam hal penyumbang sampah paling besar. Cara yang ditawarkan ada berbagai macam, contoh paling dasar dan bermanfaat adalah pengomposan yang dibuat menggunakan sampah sisa makanan dan jenis sampah organik lainnya. Kompos adalah materi organik yang telah terurai sebagian, menghasilkan bahan yang gelap, remah, dan memiliki aroma tanah. Kompos dibuat melalui proses biologis seperti dekomposisi jaringan tanaman oleh organisme di tanah. Ketika proses peluruhan selesai, kompos menjadi bahan berwarna coklat kehitaman dan berbentuk serbuk yang disebut humus. Bahan organik yang digunakan untuk membuat kompos dapat berupa limbah organik pertanian, limbah dapur seperti sisa makanan, dan bahan organik lainnya yang sesuai dengan kriteria tertentu seperti kelembaban 50-60% dan tidak mengandung lemak tinggi [3]. Pengolahan kompos ini sangat bermanfaat bagi lingkungan, baik dari segi kebersihan dan juga kesuburan tanah. Pulau bali merupakan tempat wisata yang sangat dikenal baik di Indonesia maupun di luar Indonesia sendiri, pulau bali sangat dikenal melalui keindahan alam dan budaya yang dapat



menarik banyak wisatawan, walaupun sangat indah dan dikengkapi dengan budaya yang unik, pulau bali tidak lepas dari masalah sampah. Wisatawan dan orang lokal masih menutup sebelah mata mengenai masalah sampah yang terjadi di pulau bali yang menyebabkan tempat wisata seperti pura ataupun pantai menjadi tercemar dan keindahan yang disajikan berkurang [4]. Kurang optimalnya penanganan sampah menyebabkan berbagai permasalahan lingkungan, dari beberapa permasalahan tentang sampah dapat diketahui bahwa pokok masalah lebih pada minimnya pengetahuan tentang kebersihan lingkungan dan pengelolaan sampah serta fasilitas pendukung yang kurang, seperti tempat sampah [5]. Tempat sampah adalah wadah sementara untuk menyimpan berbagai jenis sampah, umumnya terbuat dari logam atau plastik. Di dalam ruangan, tempat sampah biasanya ditempatkan di dapur untuk membuang sisa makanan seperti kulit buah dan botol. Di perkantoran, digunakan tempat sampah khusus. Beberapa tempat sampah dilengkapi dengan penutup untuk mencegah keluarnya bau. Kebanyakan tempat sampah harus dibuka secara manual, namun kini semakin banyak yang dilengkapi dengan pedal untuk membuka tutupnya dengan mudah [6]. Tempat sampah bisa dimanfaatkan lebih lanjut sebagai tempat pemrosesan sampah organik di lingkungan rumah tangga dengan cara menambahkan beberapa inovasi yang menambah fungsinya seperti sensor yang bisa menjadikan sampah organik yang disimpan menjadi sebuah kompos. Tempat sampah yang terintegrasi dengan sensor ini mempunyai sebuah modul utama yaitu NodeMcu 8226 yang merupakan turunan dari arduino uno. Arduino Uno adalah papan mikrokontroler yang menggunakan ATmega328 sebagai basisnya. Papan ini memiliki 14 pin input/output digital, dengan 6 pin yang bisa berfungsi sebagai output PWM. Selain itu, Arduino Uno juga dilengkapi dengan enam input analog, osilator kristal 16 MHz, port USB, colokan listrik, header ICSP, dan tombol reset. Semua fitur yang diperlukan untuk mendukung mikrokontroler telah ada pada Arduino Uno, sehingga pengguna dapat dengan mudah menghubungkannya ke komputer melalui kabel USB atau memberinya daya melalui adaptor AC-DC atau baterai [7]. Perangkat ini memungkinkan pengguna untuk menjalankan sebuah sensor dan juga memonitoring kelembaban yang ada di dalam tempat sampah yang berguna untuk mempercepat pembuatan kompos, karena komponen paling penting yang harus diperhatikan dalam pembuatan kompos di tempat sampah ini adalah kelembaban dan aliran udara yang ada.

## 2. Metode Penelitian

Metode penelitian yang digunakan dalam artikel ini adalah penelitian dan pengembangan (Research and Development). Penelitian dan pengembangan merupakan proses langkah-langkah untuk mengembangkan produk baru atau memperbaiki produk yang sudah ada. Dalam konteks ini, penelitian dan pengembangan bertujuan untuk menghasilkan tempat sampah yang sudah terintegrasi oleh sebuah sensor dan bisa dijadikan sebuah alat untuk membuat sebuah kompos, tentu ini adalah sebuah pengembangan dari cara lama. Data akan didapatkan melalui sebuah survey dan didapatkan sebuah alur penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

## 2.1 Identifikasi Kebutuhan Pengguna

Pada penelitian kali ini, data identifikasi kebutuhan pengguna dilakukan dengan cara melakukan survei dengan calon pengguna dari sistem ini. Identifikasi dilakukan guna untuk mengetahui seberapa jauh pengguna mengetahui tentang pembuatan kompos dan memperkenalkan sistem baru yang dapat menjadi solusi untuk masalah sampah organik yang ada dalam lingkungan rumah tangga. Identifikasi ini juga diperlukan untuk memenuhi ekspektasi pengguna mengenai sistem ini. Berikut adalah langkah- langkah pengumpulan data yang dilakukan.

- Melakukan survei yang dilakukan menggunakan google form untuk mengetahui seberapa paham pengguna dengan proses pembuatan kompos, serta pengenalan sistem pengolahan kompos yang baru. Bertanya mengenai bentuk dan ekspektasi calon pengguna mengenai sistem ini dan juga saran dari calon pengguna
- Mengumpulkan data yang didapat dari google form dan digunakan untuk membuat dan memenuhi analisis kebutuhan pada penelitian ini.
- Informasi yang didapat digunakan untuk perencanaan pembuatan rancangan sistem pembuatan kompos dengan tempat sampah berbasis IOT ( Internet of Things )

## 2.2 Rancangan Tempat Sampah

Rancangan tempat sampah dibuat berdasarkan preferensi dari calon pengguna, Sensor sudah ditentukan sedari awal. Sensor yang dipakai itu mencakup Sensor Kelembaban (DHT22), Sensor Ultrasonik (HC-SR04), NodeMCU 8226, Motor Servo. Survei dari google form diperlukan untuk mengetahui apa saja kebutuhan tambahan yang diperlukan untuk mengarahkan bentuk tempat sampah yang digunakan serta implementasi saran dari calon pengguna.

## 2.3 Perancangan Aplikasi Pemantauan

Mengembangkan sketsa kasar atau wireframe untuk merancang tampilan antarmuka pengguna berdasarkan kebutuhan dan preferensi pengguna yang telah didapatkan di tahap identifikasi kebutuhan pengguna, setelah didapatkan sebuah sketsa kasar atau wireframe maka akan

dilanjutkan ke tahap pengembangan antarmuka baik itu high-fidelity wireframe sampai mendapatkan sebuah mockup sistem nya.

## 2.4 Peningkatan dan Evaluasi

Setelah mendapatkan sebuah rancangan antarmuka, akan dilakukan evaluasi untuk mendapatkan dan memenuhi ekspektasi dari calon pengguna, baik itu dari sisi tempat sampah maupun aplikasi pemantauan. Evaluasi ini bisa dilakukan baik dari survey ataupun riset internal.

## 3. Hasil dan Diskusi

### 3.1 Identifikasi Pengguna

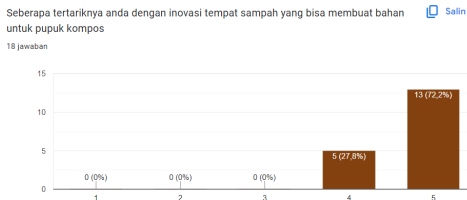
Setelah dilakukanya survey didapatkan hasil- hasil sebagai berikut

- Preferensi pengguna terhadap bentuk tempat sampah sangat penting untuk menentukan seperti apa bentuk tempat sampah pada rancangan ini, dari hasil survey pada Gambar 2 menunjukkan preferensi pengguna menunjukkan bahwa tempat sampah yang lebih baik digunakan merupakan model tempat sampah yang bisa membuka dan menutup, ini juga sejalan dengan cara kerja pembuatan kompos yang cenderung membutuhkan tempat yang agak lembab.



Gambar 2. Data Bentuk Tempat Sampah

- Agar rancangan ini dapat diterima dalam masyarakat maka akan dilakukan survey mengenai antusias pengguna tentang rancangan ini dan didapatkan hasil seperti pada Gambar 3. Gambar 3 menunjukkan bahwa dari ke 18 responden menunjukkan bahwa 13 orang sangat antusias dengan inovasi ini dan 5 orang antusias dalam inovasi ini, ini menunjukkan pandangan masyarakat yang cukup positif dalam inovasi ini kepadanya. Dengan adanya respon yang positif ini diharapkan bahwa rancangan ini dapat diterima dan digunakan dalam masyarakat luas.



Gambar 3. Statistik Antusias Calon Pengguna

- c. Pemahaman mengenai kompos juga sangat penting untuk rancangan ini, karena rancangan ini hanya mempermudah calon pengguna bukan menggantikan kegiatan pembuatan kompos. Gambar 4 menunjukkan bahwa pemahaman orang terhadap pupuk kompos sangat berbeda- beda sehingga kami perlu menambahkan suatu penjelasan dalam sistemnya agar pengetahuan kompos dapat dipergunakan untuk menggunakan alat ini juga.



Gambar 4. Statistik Pemahaman Mengenai Kompos

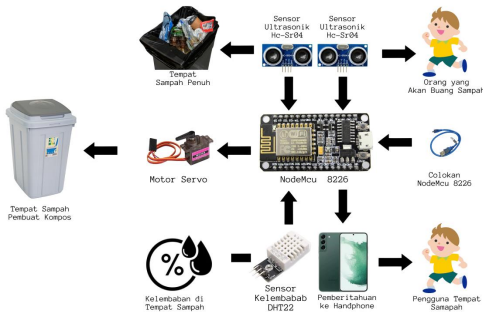
- d. Pada survei mengenai kehadiran sistem yang ditunjukkan pada Gambar 5 ditunjukkan bahwa sistem yang membantu sangat penting bagi pengguna karena semua responden menginginkan sistem pemantauan.



Gambar 5. Rasio Kebutuhan Terhadap Sistem

### 3.2 Rancangan Tempat Sampah

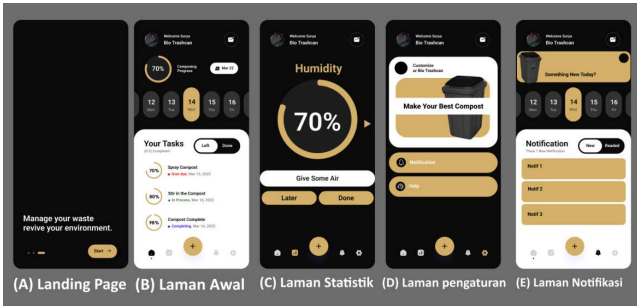
Berdasarkan hasil survey, dapat dibungun sebuah rancangan sistem tempat sampah otomatis seperti yang terlihat pada Gambar 5. Tempat sampah ini sudah dilengkapi dengan sensor ultrasonik yang digunakan untuk 2 hal, yaitu mendeteksi orang yang akan datang dan mendeteksi penuh atau tidaknya tempat sampah, dengan demikian tempat sampah ini bisa terbuka otomatis dan juga bisa mendeteksi tempat sampah itu sudah full atau tidak. Motor servo digunakan sebagai penggerak pembuka dan penutup serta sensor kelembaban digunakan sebagai alat pemantauan agar kompos tidak terlalu basah ataupun kering. Semua komponen di pada Gambar 5 terhubung ke modul NodeMCU 8226 agar bisa berfungsi. Setelahnya monitoring bisa dilakukan melalui handphone karena NodeMCU sudah dilengkapi dengan modul wifi. Cara kerja dari rancangan Gambar 5 adalah dengan memasukan sampah organik ke dalam tempat sampah, nantinya kompos akan terbentuk secara alami dengan sendirinya tetapi kita juga harus memperhatikan kelembaban agar tidak membusuk, campur tangan manusia tidak lepas dari pembuatan pupuk ini.



Gambar 5. Blok Diagram Tempat Sampah

### 3.3 High-fidelity Wireframe

High-fidelity wireframe merupakan langkah selanjutnya dari low-fidelity wireframe dimana dalam tahap ini, rancangan akan diberikan warna dan bentuk untuk mendapatkan sebuah user interface yang menggambarkan sebuah aplikasi. Dilihat pada Gambar 6, ditunjukkan sebuah rancangan high-fidelity wireframe yang memuat laman diantaranya adalah Gambar 6A yang menunjukkan landing page yang digunakan sebagai pengenalan rancangan ini dan juga berisi tujuan dari rancangan ini. Setelah melakukan pengenalan maka akan dilanjutkan ke Gambar 6B yang merupakan laman awal yang menunjukkan persentase pembuatan kompos dari hari ke hari dan memuat tugas harian yang harus dilakukan untuk menambah persentase keberhasilan pembuatan kompos. Gambar 6C menunjukkan laman statistik sensor secara spesifik, pengguna sistem dapat memantau perkembangan dan menjalankan tugas yang diberikan untuk memberi kemajuan pada sistem dan juga pada keberhasilan pembuatan kompos. Setelah semua fitur itu kita akan masuk ke Gambar 6D yang membahas mengenai pengaturan dari rancangan sistem ini, pengaturan disini memuat customer service yang bisa diakses melalui tombol help dan juga pengaturan mengenai notifikasi yang masuk ke perangkat anda. Selain itu ada juga pengaturan mengenai interaksi sensor, tetapi ini mempengaruhi hasil akhir kompos, jadi untuk memaknai pengaturan ini pengguna harus tahu lebih banyak mengenai pembuatan kompos. Terakhir ada sebuah laman notifikasi yang dimuat dalam Gambar 6E, notifikasi ini memuat semua notifikasi yang dikirimkan dari rancangan ini kepada pengguna, pengguna bisa menggunakan ini sebagai pengingat dan melihat notifikasi yang belum dilihat sebelumnya.



Gambar 6. High-fidelity Wireframe

#### 4. Kesimpulan

Dalam penelitian ini, rancangan pembuatan kompos menggunakan tempat sampah berbasis IOT (Internet of Things) sudah berhasil dibuat. Dalam rancangan ini tempat sampah sudah dapat mengukur suhu yang ada dalam tempat sampah sehingga pembuatan kompos dapat terkendali, tempat sampah juga sudah dilengkapi dengan sebuah sistem pembuka dan penutup otomatis sehingga dapat mempermudah pengguna dalam menggunakannya. Penelitian ini juga menghasilkan sebuah rancangan antarmuka yang bisa digunakan untuk mengatur kelembaban yang diinginkan untuk memperingati apabila sudah melebihi atau kurang dari kelembaban yang diinginkan, tentu rancangan aplikasi sudah memuat semua saran yang masuk dalam survei sehingga bisa dipastikan user friendly dan dapat mendorong penggunaan tempat sampah ini lebih jauh lagi. Namun masih banyak pengembangan yang dapat dilakukan dalam sistem ini, pengembangan mencakup konsumen yang lebih luas dan penambahan fitur pada tempat sampah hingga banyak hal yang bisa dilaporkan ke pengguna dan pembuatan kompos menjadi lebih mudah dan presisi. Dengan pembuatan rancangan sistem pembuatan kompos dengan tempat sampah berbasis IOT (Internet of Things) diharapkan dapat mengubah kebiasaan masyarakat rumah tangga dalam kebiasaan membuang sampah organik dan lebih memilih memanfaatkannya ke kompos.

#### Daftar Pustaka

- [1] Geralt "Masalah Lingkungan | Menjelajahi, Memahami Tantangan Planet Kita, Tentang permasalahan lingkungan dan solusinya" Pinter pandai 2023
- [2] K. L. H. Kehutanan "Capaian Kinerja Pengelolaan Sampah adalah Capaian Pengurangan dan Penanganan Sampah Rumah Tangga dan Sampah Sejenis Sampah Rumah Tangga." Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) – Kementerian Lingkungan Hidup dan Kehutanan 2023.
- [3] I. Ridwan, T. Dariati, Nurfaida, C. Wahyuni B Y, A. M. Jaya, A. H. Bahrun "Utilization Of Household Organic Wastes For Composting In Suburb Tamalanrea Jaya City Of Makassar" Jurnal Pengabdian Sriwijaya 2014.
- [4] I. G. A. P. S. I. Svari , I. W. Sutama "Meningkatkan Kesadaran Masyarakat Dalam Mengelola Sampah Rumah Tangga Melalui Lembaga Bank Sampah Jaya Lestari Desa Pemogan. Jurnal Pengabdian, April 2022, Volume 5 Nomor 1
- [5] S. I. P. Yuwana ,M. F. A. S. Adlan "Edukasi Pengelolaan Dan Pemilahan Sampah Organik Dan Anorganik Di Desa Pecalongan Bondowoso"Jurnal Pengabdian Kepada Masyarakat Fordicate (Informatics Engineering Dedication, Vol 1, No. 1 Tahun 2021.
- [6] A. F. Annisa, N. Ahrajabanur, Irmawati. H, N.Amin "Pengadaan Tempat Sampah Untuk

- Lingkungan Sekolah Sehat Mahasiswa KKN-PPL Terpadu angkatan XXI UNM di Pondok Madrasah Al-Wasilah Lemo Desa Kuajang Kecamatan Binuang Kabupaten Polewali Mandar” Jurnal Lepa-lepa Open. Volume 2 Nomor 3, 2022
- [7] S. Samsugi, R. D. Gunawan, A. Thyo, A. T. Prastowo “Penerapan Penjadwalan Pakan Ikan Hias Molly Menggunakan Mikrokontroler Arduino Uno Dan Sensor RTC DS3231” Jurnal Teknologi dan Sistem Tertanam (JTST) Universitas Teknokrat Indonesia, Volume 3 Nomor 2 Tahun 2022

# Komparasi Ekstraksi Fitur BoW dan TF-IDF untuk Klasifikasi SMS Menggunakan Naive Bayes

I Komang Dwiprayoga<sup>1</sup>, Made Agung Raharja<sup>2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>komangyoga835@gmail.com  
<sup>2</sup>made.agung@unud.ac.id

## Abstract

*Short Message Service (SMS) has become one of the most popular communication media. However, the ease and speed of sending SMS is also utilized by irresponsible parties to send spam messages. These spam messages not only annoy users but can also cause financial losses and theft of personal data. The purpose of this research is to compare feature extraction methods that have the best performance such as TF-IDF and Bag of Word tested with Multinomial Naive Bayes machine learning algorithm. For the first research stage, load dataset, data balancing, data preprocessing, feature extraction, modeling with machine learning algorithms, and then testing and comparing confusion matrix models on each feature extraction. The results of this study show that the use of BoW feature extraction has better performance than the TF-IDF feature extraction model with an accuracy value of 94.44%.*

**Keywords:** *Back of Words, TF-IDF, Multinomial Naive Bayes, Sms, Text Classification*

## 1. Pendahuluan

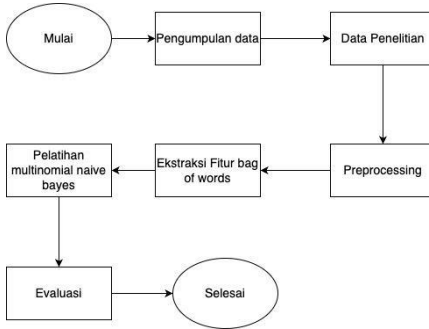
Era ini, data dan informasi merupakan komoditi utama yang dapat diperjualbelikan yang dengan mudah dapat diakses oleh pengguna dan pelanggan. Semuanya itu membawa masyarakat ke dalam suasana yang disebut oleh John Naisbitt, Nana Naisbitt dan Douglas Philips sebagai "Zona Mabuk Teknologi" [1]. Spam, juga disebut sebagai unsolicited commercial email atau unsolicited bulk email telah menyebabkan beberapa masalah komunikasi dalam kehidupan sehari-hari kita. Kerugian yang disebabkan karena spam antara lain spam menempati sumber daya yang besar (termasuk bandwidth jaringan, ruang penyimpanan), contoh kasus spam bisa berupa iklan perjudian maupun pornografi. Spam atau junk email adalah penyalahgunaan dalam pengiriman berita elektronik untuk menampilkan berita, iklan, dan keperluan lainnya yang mengakibatkan ketidaknyamanan bagi para pengguna [2]. Penipuan melalui SMS menjadi ancaman serius karena skema penipuan semakin kompleks dan dapat disesuaikan. Serangan pencurian identitas yang merugikan secara finansial hingga phishing, metode yang menyamar sebagai upaya peretasan dengan menggunakan pesan palsu [3]. Tujuan dilakukan penelitian ini adalah untuk meningkatkan akurasi dari metode yang digunakan sebelumnya oleh peneliti dengan menggunakan metode multinomial naive bayes dengan ekstraksi fitur bag of words. Spam atau penipuan ini dapat diklasifikasikan dengan melihat pola dan kata-kata yang digunakan. Pesan yang masuk ke ponsel dapat diklasifikasikan lagi menjadi tiga jenis pesan, yakni spam/penipuan, promo, dan normal.

## 2. Metode Penelitian

Pada tahap ini diawali dengan mengumpulkan data penelitian yang akan digunakan, selanjutnya tahap load dataset ke python notebook, kemudian dari data mentah tersebut akan dilakukan data balancing atau penyeimbangan data, kemudian dilanjutkan dengan tahap preprocessing data untuk membersihkan data dari kata-kata yang tidak diperlukan. Kemudian dilakukan pemisahan data latih dan data uji sebelum dilakukan tahap ekstraksi fitur. Kemudian tahap selanjutnya tahap perhitungan ekstraksi fitur menggunakan bag of words dan TF-IDF. Kemudian lanjut ke tahap



pelatihan model menggunakan multinomial naive bayes. Tahap terakhir adalah proses evaluasi menggunakan confusion matrix untuk menentukan akurasi dari model yang telah dilatih sebelumnya.



Gambar 1. Alur Penelitian

### 2.1. Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah data sekunder yang diambil langsung di github peneliti yang sebelumnya juga melakukan penelitian terkait. Dataset yang diambil berupa teks atau pesan sms dengan banyak dataset berjumlah 1143 data yang memiliki tiga label, yakni berawal dari sms normal dengan jumlah 569 data yang direpresentasikan dengan index 0, sms penipuan atau spam dengan jumlah 335 data yang direpresentasikan dengan index 1, dan sms promo dengan jumlah 239 data yang direpresentasikan dengan index 2. Gambaran data yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Gambaran Dataset

| id Teks |   | Label |
|---------|---|-------|
| 1       | [PROMO] Beli paket Flash mulai 1GB di MY TELKOMSEL APP dpt EXTRA kuota 2GB 4G LTE dan EXTRA nelpon hingga 100mnt/1hr. Buruan, cek di tsel.me/mytsel1 S&K            | 2     |
| 2       | Yooo sama2, oke nanti aku umumin di grup kelas  | 0     |
| 3       | ""ROXI CELL" Hanya dengan Rp.100rb Anda bisa jadi agen pulsa elektrik ke semua Operator GSM/CDMA, v5=4300, v10=8300, Utk daftar ketik MTRONIK kirim ke 087870870707 | 1     |

Tabel 2. Distribusi Data

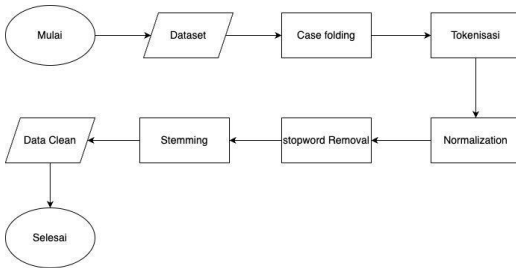
| id Label | Data           |
|----------|----------------|
| 1        | SMS Normal 569 |
| 2        | SMS Spam 335   |
| 3        | SMS Promo 239  |

## 2.2. Data Balancing

Jumlah data di setiap label tidak seimbang, yakni 569 data untuk label normal, 335 data untuk sms spam, dan 239 data untuk sms promo. Oleh karena itu, perlu dilakukannya proses penyeimbangan data di setiap label. Pada penelitian ini, peneliti menggunakan teknik random under sampling untuk menyeimbangkan jumlah data di setiap label. Random under sampling merupakan. Pendekatan undersampling dan oversampling adalah teknik standar yang digunakan dalam menangani data yang tidak seimbang, namun keduanya memiliki keterbatasan masing-masing. Misalnya, undersampling menyebabkan lebih banyak penghapusan sampel data yang pada akhirnya menyebabkan masalah kekurangan data, dengan peningkatan kemungkinan kehilangan data penting. sementara oversampling menyebabkan duplikasi data asli, sehingga menyebabkan overfitting kelas minoritas [4].

## 2.3. Preprocessing

Text preprocessing adalah tahap pertama dalam klasifikasi teks yang mengubah data teks asli yang tidak terstruktur menjadi data yang terstruktur sekaligus juga untuk mengidentifikasi fitur dari teks yang paling signifikan untuk membedakan antara kategori teks [5]. Tahap ini akan menghasilkan data teks yang siap digunakan untuk proses selanjutnya. Adapun tahapan dalam text preprocessing ini ditunjukkan pada Gambar 2.



Gambar 2. Alur Text Preprocessing

## 2.4. Bag of Words

Bag of Words (BoW) adalah sebuah teknik ekstraksi fitur untuk merepresentasikan dokumen teks ke dalam bentuk matriks. Teknik ini bekerja dengan cara mempelajari seluruh kosakata dari dokumen, lalu memodelkan tiap dokumen dengan menghitung jumlah kemunculan tiap katanya [6].

## 2.5. TF-IDF

TF-IDF adalah metode pembobotan kata yang digunakan untuk menentukan pentingnya sebuah kata dalam sebuah dokumen. Ada dua komponen dalam perhitungan nilai TF – IDF, yaitu TF (Term Frequency) dan IDF (Inverse Document Frequency). TF menentukan pentingnya suatu kata relatif terhadap kemunculannya dalam suatu dokumen, sedangkan IDF menentukan suatu kata penting dalam suatu dokumen jika tidak sering muncul di dokumen lain [7]. Rumus dari TF-IDF adalah sebagai berikut:

$$w(d, t) = TF(d, t) \times \log\left(\frac{N}{df(t)}\right) \quad (1)$$

## 2.6. Multinomial Naive Bayes

Multinomial Naive Bayes merupakan implementasi algoritma Naive Bayes yang umumnya digunakan dalam pemrosesan teks dengan mengikuti prinsip distribusi multinomial [3]. Penggunaan model distribusi multinomial menunjukkan bahwa vektor fitur dalam suatu dokumen dibentuk dari frekuensi kemunculan setiap kata pada dokumen tersebut. Adapun algoritma perhitungannya [8]. Algoritma ini akan menghitung probabilitas sebuah dokumen  $d$  terhadap kelas  $C$  yang ditunjukkan pada Persamaan 2 [3].

## 2.7. Evaluasi

Tahap evaluasi dilakukan dengan menguji model dengan data uji yang telah dibagi sebelumnya dengan porsi 20%, yakni 229 data uji dengan menggunakan confusion matrix. Hasil evaluasi akan menghasilkan nilai akurasi, recall, precision, dan f1 score. Adapun rumus yang digunakan untuk melakukan perhitungan dari akurasi, recall, presisi, dan f1 score adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (2)$$

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

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

$$\text{F1 - Score} = \frac{2(\text{Precision}+\text{Recall})}{(\text{Precision}+\text{Recall})} \quad (5)$$

## 3. Hasil dan Diskusi

### 3.1. Data Balancing

Dataset dengan jumlah data 1143 akan di-reduce menjadi 771 data. Data sms spam yang tadinya berjumlah 569 data berkurang menjadi 239 data, begitu juga dengan jumlah data sms spam yang tadinya 335 data menjadi 239 karena mengikuti data minoritas atau dalam penelitian ini yang menjadi data minoritas adalah sms promo sejumlah 239 data. Hasil penyeimbangan data menggunakan teknik random undersampling dapat dilihat pada tabel 2 berikut:

Tabel 3. Hasil Undersampling

| id  | Teks  | Label |
|-----|---|-------|
| 0   | Maaf, Keyword yang anda masukkan ke 234 salah. Info lanjut hub call center: 200   | 0     |
| 1   | Gais aku sudah brgkt. Mau matiin hp   | 0     |
| 2   | "Kan bapaknya ngomong2 ya, trs ditanya siapa itu, aku blg lg rame ma sibuk dikampus nanti yaa"  | 0     |
| 3   | Gimana nuy menghadapi beliau?   | 0     |
| 4   | "Yaudah sekarang mah eta harddisk di laptop maneh keluarin terus dijadiin harddisk external aja.. cari temen yang punya case nya.. entar datarnya copy-in, paling maneh ngerjain di luar dulu aja ga di laptop itu" | 0     |
| 712 | Paket Flash anda 10 MB utk 1 hari akan berakhir pd 08/07/2016. Tarif non paket berlaku setelah tgl tsb. Info hubungi *363#. Cek Kuota *889#.  | 2     |
| 713 | Terima kasih telah menjadi Sahabat IndosatOoredoo. Mau tahu RAHASIA2 TERHEBOH ARTIS-ARTIS TOP + bonus gratis nelpon 60 menit?Tlp *700*1#  | 2     |

| id  | Teks   | Label |
|-----|--|-------|
| 714 | Pembelian Dave's Single Burger Gratis Small Fries & Minum. Periode 5-11 September 2016. Tukarkan SMS ini di Wendy's terdekat! Penawaran terbatas.Promo *606# | 2     |
| 715 | Selamat Paket 50 SMS ke semua operator Anda telah aktif. Cek kuota di *888#  | 2     |
| 716 | Pakai XL tdk perlu repot setting bisa langsung internetan. Yuk aktifkan paket Internet di My.XL.co.id sekarang.  | 2     |

**Tabel 4.** Hasil Distribusi Menggunakan Undersampling

| id | Label      | Data |
|----|------------|------|
| 1  | SMS Normal | 239  |
| 2  | SMS Spam   | 239  |
| 3  | SMS Promo  | 239  |

### 3.2. Preprocessing

Dataset yang berjumlah 771 data dilakukan tahap preprocessing atau pembersihan teks, tahap ini dilakukan untuk meningkatkan performa akurasi dari model yang akan dilatih setelahnya. Hasil dari text preprocessing dapat dilihat dari tabel 2 berikut ini:

**Tabel 5.** Preprocessing

| id | Proses           | Teks                                 |
|----|------------------|--------------------------------------|
| 1  | Data mentah      | Besok kekantor saya tunggu jam 9 ya. |
| 2  | Case Folding     | besok kekantor saya tunggu jam ya    |
| 4  | Normalization    | besok kekantor saya tunggu jam ya    |
| 5  | Stopword Removal | besok kekantor tunggu jam ya         |
| 6  | Stemming         | besok kantor tunggu jam ya           |

### 3.3. Ekstraksi Fitur

Setelah mendapatkan data yang bersih dari tahap preprocessing, maka tahap selanjutnya ke tahap ekstraksi fitur menggunakan Bag of Words dan TF-IDF. Perhitungan ekstraksi fitur menggunakan bantuan dari library scikit-learn dengan modul countvectorizer dan Tfidfvectorizer. Hasil dari perhitungan didapatkan sebanyak 3255 fitur dari 573 data latih. Fitur yang didapat terbilang banyak mengingat jumlah data latih yang digunakan hanya 573 data yang berupa teks. Berikut merupakan hasil dari ekstraksi fitur menggunakan BoW dan TF-IDF:

Ambil data pada dataset yang telah melalui preprocessing:

Dokumen 1 = "transfer tipon nomor ubah nomor rekening terima kasih"

Dokumen 2 = "bayar transfer harap hubungi atas sayabpk ardiansah nomor telepon ubah bayar"

Dokumen 3 = "nomor papa tolong kirim pulsa nomor papa papa kantor polisi telepon papa"

**Tabel 6.** Hasil Pengujian Ekstraksi Fitur BoW

| Doc  | ardiansah | atas | bayar | ..... | tipon | tolong | transfer | ubah |
|------|-----------|------|-------|-------|-------|--------|----------|------|
| Doc1 | 0         | 0    | 0     | ..... | 1     | 0      | 1        | 1    |
| Doc2 | 1         | 1    | 2     | ..... | 0     | 0      | 1        | 2    |

| Doc  | ardiansah | atas | bayar | ..... | tipon | tolong | transfer | ubah |
|------|-----------|------|-------|-------|-------|--------|----------|------|
| Doc3 | 0         | 0    | 0     | ..... | 0     | 1      | 0        | 0    |

Tabel 7. Hasil Pengujian Ekstraksi Fitur TF-IDF

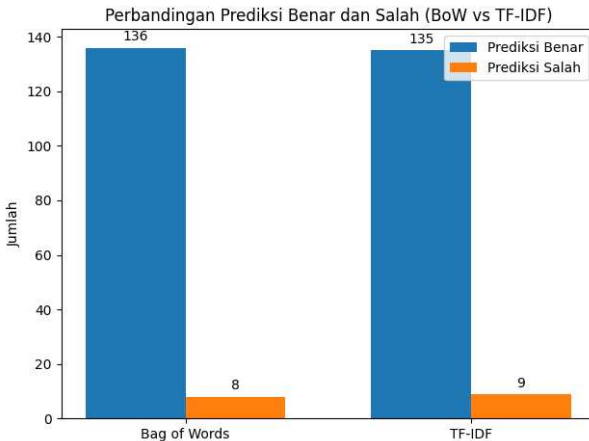
| Doc  | ardiansah | atas     | bayar    | ..... | tipon   | tolong  | transfer | ubah    |
|------|-----------|----------|----------|-------|---------|---------|----------|---------|
| Doc1 | 0.00000   | 0.00000  | 0.00000  | ..... | 0.32493 | 0.00000 | 0.32493  | 0.26303 |
| Doc2 | 0.301647  | 0.301647 | 0.512236 | ..... | 0.00000 | 0.00000 | 0.49580  | 0.49580 |
| Doc3 | 0.00000   | 0.00000  | 0.00000  | ..... | 0.00000 | 0.00000 | 0.00000  | 0.00000 |

### 3.4. Pemodelan

Pemodelan multinomial naive bayes dilakukan dengan bantuan library scikit-learn dari python. Model akan dilatih dengan data latih yang telah melalui proses ekstraksi fitur menggunakan Bag of Words dan TF-IDF. Pemodelan juga menggunakan hyperparameter alpha terbaik yang nanti dicari menggunakan Gridsearch. Setiap ekstraksi fitur akan melalui proses evaluasi menggunakan Gridsearch untuk mencari hyperparameter terbaik dan dengan proses cross validation dengan banyak fold sebanyak 5 kali fold.

### 3.5. Evaluasi

Tahap terakhir adalah tahap pengujian dari model yang telah dilatih dengan menggunakan confusion matrix. Data yang diujikan merupakan data uji yang telah dibagi sebelumnya sebanyak 144 data uji. Setelah melalui proses pengujian, maka didapat nilai akurasi, presisi, recall, dan f1-score terhadap model yang telah dilatih. Hasil pengujian menggunakan confusion matrix dapat dilihat pada gambar 3 dan tabel 8.



Gambar 3. Hasil Prediksi BoW dan TF-IDF

Pada gambar 3, dapat dianalisis, untuk ekstraksi fitur Bag of Words dapat memprediksi 136 data dengan benar dan 8 data yang diprediksi salah, sedangkan pada ekstraksi fitur TF-IDF memprediksi sebanyak 135 data diprediksi benar dan 9 data diprediksi salah. Terdapat perbedaan yang tidak terlalu signifikan dari hasil prediksi kedua ekstraksi fitur.

**Tabel 8.** Hasil Performa Ekstraksi Fitur

| BoW        | TF-IDF    |        |          |        |        |        |        |        |
|------------|-----------|--------|----------|--------|--------|--------|--------|--------|
|            | Precision | Recall | F1-Score | Acc    |        |        |        |        |
| SMS Normal | 97,50%    | 90,70% | 93,98%   | 94,44% | 89,80% | 95,65% | 92,63% | 93,75% |
| SMS Spam   | 90%       | 97,78% | 93,75%   | 94,44% | 89,80% | 95,65% | 92,63% | 93,75% |
| SMS Promo  | 96,30%    | 94,55% | 95,41%   | 94,55% | 94,55% | 94,55% | 94,55% |        |

Berdasarkan tabel 8 di atas, hasil akurasi yang diperoleh menggunakan ekstraksi fitur BoW lebih besar dibandingkan dengan menggunakan TF-IDF, yakni 94,44%. Nilai precision, recall dan F1-Score untuk sms normal didapatkan dengan menggunakan BoW dengan nilai 97,50%, 90,70%, 93,98%, Hasil akurasi untuk precision, recall, dan F1-Score untuk sms spam didapatkan dengan menggunakan BoW dengan nilai 90%, 97,78%, 93,75%. Untuk sms promo didapat nilai precision, recall, dan F1-Score secara berturut-turut 96,30%, 94,55%, dan 95,41%.

#### 4. Kesimpulan

Berisi pernyataan yang menjawab persoalan pada bagian sebelumnya dan karya penelitian yang akan datang [5]. Berdasarkan penelitian yang sudah dilakukan, maka peneliti dapat menarik kesimpulan yang diperoleh, di antara kedua perbandingan ekstraksi fitur, yakni Bag of Words dan TF-IDF yang menghasilkan akurasi dengan performa yang baik adalah ekstraksi fitur Bag of Words dengan nilai akurasi 94,44% sedangkan akurasi yang dihasilkan oleh TF-IDF adalah 93,75%. Perbedaan nilai akurasi antara Bag of Words dan TF-IDF tidak terlalu signifikan, mengingat pada penelitian yang telah dilakukan penulis melakukan penyeimbangan data di setiap label menggunakan teknik random undersampling. Peningkatan performa akurasi dapat dilakukan dengan beberapa cara, seperti menambah data pada setiap label dengan syarat data harus seimbang dan melakukan seleksi fitur.

#### Daftar Pustaka

- [1] L. Mamoto, "Peranan Hukum Pidana dalam Menanggulangi Penipuan Lewat SMS Serta Penegakan Hukumnya," *Lex Crimen*, vol. 5, no. 7, 2016.
- [2] Al Amien, J., Mukhtar, H., & Rucyat, M. A. (2020). *Jurnal Computer Science and Information Technology (CoSciTech)*.
- [3] Aksenta, Almasari, et al. *Literasi Digital: Pengetahuan & Transformasi Terkini Teknologi Digital Era Industri 4.0 dan Society 5.0*. PT. Sonpedia Publishing Indonesia, 2023.
- [4] N. U. Maulidevi and K. Surendro, "SMOTE-LOF for noise identification in imbalanced data classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3413-3423, 2022.
- [5] F. Alzami et al., "Document preprocessing with TF-IDF to improve the polarity classification performance of unstructured sentiment analysis," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 235-242, 2020.
- [6] W. T. H. Putri and R. Hendrowati, "Penggalian Teks Dengan Model Bag of Words Terhadap Data Twitter," *Jurnal Muara Sains, Teknologi, Kedokteran, dan Ilmu Kesehatan*, vol. 2, no. 1, pp. 129-138, 2018.
- [7] B. Harjito, K. N. Aini, and B. Murtiyasa, "Klasifikasi dokumen berkonten serangan jaringan menggunakan multinomial naive bayes," in *Seminar Nasional Teknologi Informasi dan Komunikasi (SEMNASITIK)*, vol. 1, no. 1, pp. 112-118, Oct. 2018.
- [8] N. Rezaeian and G. Novikova, "Persian text classification using naive bayes algorithms

and support vector machine algorithms," Indonesian Journal of Electrical Engineering and Informatics (IJEI), vol. 8, no. 1, pp. 178-188, 2020.

# Pengamanan Data Finansial Menggunakan Enkripsi Homomorfik Pailler

Ida Bagus Made Wiguna Tedja Sukmana<sup>a1</sup>, Agus Muliantara<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>gustedja123@gmail.com  
<sup>2</sup>muliantara@unud.ac.id

## Abstract

*In the current digital era, concerns about data privacy are increasing due to the ease of data creation, storage, access, and dissemination. To protect sensitive data, such as financial information and identities, effective methods are needed to safeguard privacy from unauthorized access. One such method is homomorphic encryption, which allows computations to be performed on encrypted data without the need for decryption. This research implements the Paillier Cryptosystem in financial data security, specifically for customer savings accounts. The encryption and computation processes on savings data are performed using homomorphic encryption, ensuring data confidentiality and customer privacy. The testing results demonstrate that the system can perform homomorphic computations effectively, producing results consistent with plaintext calculations. In conclusion, the use of homomorphic encryption in financial data security can enhance privacy, security, and data reliability while providing more control to data owners.*

**Keywords:** *Cryptography, Encryption, Homomorphic Encryption, Data Security, Paillier Cryptosystem*

## 1. Pendahuluan

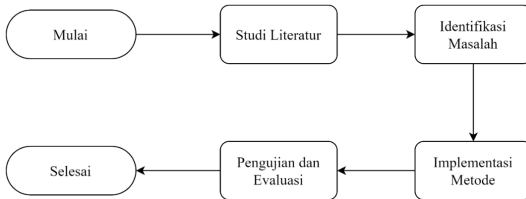
Di era digital ini, data-data dengan sangat mudah dibuat, disimpan, diakses, dan disebarluaskan. Dengan kemudahan tersebut, kekhawatiran akan privasi data kita semakin meningkat. Privasi diperlukan di berbagai jenis data, contohnya informasi keuangan, identitas, dan alamat. Jika terjadi kebocoran data yang meliputi informasi finansial, korban dapat mengalami kerugian harta termasuk pembobolan akun rekening. Maka dari itu diperlukan metode untuk menjaga privasi data tersebut dari pihak tidak berwenang. Terdapat beragam metodologi untuk menjaga privasi data[1] seperti Anonimisasi yang melibatkan penghapusan atau perubahan komponen-komponen yang dapat diidentifikasi dalam *dataset*, sehingga melindungi identitas individu sambil memungkinkan analisis dan penggunaan yang sah; Privasi diferensial yang memperkenalkan kebisingan atau ketidakpastian ke dalam respons kueri, menjaga privasi kontribusi individu dalam *dataset* sambil mempertahankan akurasi statistik; Enkripsi yang mengamankan data dengan mengubahnya menjadi format yang tidak dapat dibaca tanpa kunci dekripsi yang sesuai. Metodologi ini memastikan kerahasiaan, bahkan dalam kasus akses yang tidak sah. Enkripsi mengubah informasi yang dapat dipahami manusia atau yang disebut *plaintext* dengan algoritma tertentu menjadi bentuk yang tidak dapat dipahami atau yang disebut *ciphertext*. Enkripsi biasanya dilakukan untuk kepentingan militer, namun dengan perkembangan teknologi informasi enkripsi juga dilakukan kepada transmisi data melalui internet untuk menjaga privasi data pengguna. Enkripsi dapat dilakukan dengan banyak cara, salah satunya adalah enkripsi homomorfik. Enkripsi homomorfik memiliki cara unik untuk melindungi data. Enkripsi homomorfik memungkinkan data untuk dihitung saat dalam keadaan terenkripsi, tanpa perlu didekripsi terlebih dahulu. Dengan kata lain, data dapat dilakukan operasi matematika dan komputasi dalam bentuk *ciphertext* tanpa harus mengembalikannya ke bentuk *plaintext*. Hal ini sangat berguna karena memungkinkan analisis dan pengolahan data yang aman, tanpa mengorbankan keamanan privasi.



Kajian sebelumnya telah mengimplementasikan enkripsi homomorfik ke dalam sistem *electronic voting* menggunakan skema *Paillier Cryptosystem* dalam jumlah vote kandidat dan dalam penjumlahan vote setiap kandidat[2]. Di sistem tersebut, pemilih perlu melakukan login terlebih dahulu melalui *face-recognition* untuk mendapatkan akses melakukan pemilihan suara. Suara dari pemilih kemudian dienkripsi dengan *paillier* yang kemudian dimasukkan ke *database mysql*. Kandidat yang dipilih, *ciphertext* mereka akan ditambah nilai satu sehingga *ciphertext* berubah. Penelitian ini akan mengimplementasikan *Paillier Cryptosystem* untuk keamanan data finansial seperti tabungan. *Paillier cryptosystem* dipilih dalam penelitian ini karena merupakan jenis enkripsi homomorfik yang memungkinkan operasi penambahan dan pengurangan dilakukan langsung pada data yang terenkripsi tanpa perlu melakukan dekripsi terlebih dahulu. Hal ini memungkinkan proses komputasi dilakukan tanpa mengungkapkan informasi sensitif dalam bentuk plaintext, yang sangat penting untuk menjaga privasi data finansial nasabah. Selain itu, *Paillier Cryptosystem* juga menawarkan tingkat keamanan yang tinggi, terutama dalam konteks keamanan kriptografi modern. Kombinasi dari kemampuan homomorfik dan keamanan yang tinggi membuat *Paillier Cryptosystem* menjadi pilihan yang tepat untuk aplikasi keamanan data finansial seperti tabungan. Oleh karena itu, penelitian ini memilih enkripsi homomorfik *Paillier Cryptosystem* sebagai solusi untuk menjaga privasi finansial nasabah dari pihak yang tidak berwenang. Dalam implementasinya di sistem, hal pertama yang akan dilakukan adalah generasi kunci privat dan publik untuk setiap nasabah. Setelah itu, tabungan nasabah dapat dilakukan penambahan atau pengurangan saldo dalam bentuk *ciphertext* dengan jumlah saldo yang terenkripsi. Hal ini membantu menjaga privasi finansial nasabah dari pihak yang tidak berwenang.

## 2. Metode Penelitian

Penelitian akan dilakukan sesuai dengan flowchart pada gambar 1.



Gambar 1. Flowchart Penelitian

### 2.1. Identifikasi Masalah

Pada tahap ini, peneliti memahami permasalahan yang ingin dipecahkan. Permasalahan yang dibawakan dalam penelitian ini adalah rentannya informasi finansial nasabah, seperti saldo tabungan dan transaksi, sangat sensitif dan rentan terhadap serangan siber. Tanpa enkripsi yang tepat, data ini dapat diretas atau dicuri, mengakibatkan kerugian finansial dan pencurian identitas bagi nasabah. Selain itu, nasabah memiliki hak untuk menjaga privasi informasi keuangan mereka. Namun, dalam sistem konvensional, pihak bank atau pihak lain yang terlibat dalam pemrosesan transaksi dapat mengakses data tabungan nasabah. Maka dari itu, enkripsi homomorfik memungkinkan pemrosesan data yang aman tanpa mengorbankan privasi nasabah.

### 2.2. Enkripsi Homomorfik

Enkripsi homomorfik adalah metode enkripsi yang memungkinkan terjadinya proses komputasi antara *ciphertext* layaknya *plaintext* tanpa perlu di-dekripsi terlebih dahulu. Melakukan komputasi pada data yang terenkripsi berarti bahwa jika seorang pengguna memiliki fungsi  $f$  dan ingin

mendapatkan  $f(m_1, \dots, m_n)$  untuk beberapa input  $m_1, \dots, m_n$ , maka dimungkinkan untuk melakukan komputasi pada enkripsi dari input ini,  $c_1, \dots, c_n$ , dan mendapatkan hasil yang didekripsi menjadi  $f(m_1, \dots, m_n)$  [3]. Nilai fitur ini sangat bermanfaat dan berharga untuk privasi data. Ketika kita mengenkripsi data kita, bahkan jika penyedia layanan atau pengolah data lainnya dapat mengaksesnya, mereka tidak dapat memahaminya. Ini berarti risiko kebocoran data pribadi yang sensitif secara signifikan berkurang, sambil juga memberikan lebih banyak kontrol kepada pemilik data [4].

### 2.3. Paillier Cryptosystem

Paillier cryptosystem adalah enkripsi homomorfik aditif yang diajukan oleh Paillier pada tahun 1999[5]. Sistem ini memiliki kunci publik untuk melakukan enkripsi dan kunci privat untuk melakukan dekripsi. Enkripsi homomorfik aditif berarti *ciphertext* dari enkripsi ini dapat melakukan operasi penambahan pada data yang terenkripsi lainnya dan kemudian mendapatkan hasilnya dalam teks biasa setelah dekripsi.

Adapun langkah-langkah generasi kunci sebagai berikut:

- a. Pilih dua bilangan prima besar  $p$  dan  $q$  secara acak dan independen. Pastikan bahwa

$$\gcd(pq, (p-1)(q-1)) = 1 \quad (1)$$

$\gcd$  adalah *greatest common divisor* atau faktor persekutuan terbesar. Jika  $\gcd$  tidak sama dengan 1, ulangi lagi.

- b. Hitung nilai

$$n = pq \quad (2)$$

dan

$$\lambda = \text{lcm}(p-1, q-1) \quad (3)$$

$\text{lcm}$  adalah least common multiple atau Kelipatan persekutuan terkecil

- c. Pilih sebuah bilangan bulat acak  $g$  dimana  $g$  adalah bilangan bulat antara 1 dan  $n^2$

$$g \in \mathbb{Z}_{n^2}^* \quad (4)$$

- d. Hitung invers modular dari

$$\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n \quad (5)$$

Dimana fungsi  $L$  didefinisikan dengan

$$L(u) = \frac{u-1}{n} \quad (6)$$

multiplicative inverse hanya ada jika dan hanya jika nilai generator valid telah didapat dari tahap sebelumnya.

- e. Kunci publik adalah pasangan  $(n, g)$ . Kunci ini digunakan untuk proses enkripsi.

- f. Kunci privat adalah  $(\lambda, \mu)$ . Kunci ini digunakan untuk proses dekripsi.

Enkripsi pesan  $m$  dapat dilakukan dengan cara berikut:

- a.  $m \in \mathbb{Z}_n$  (7)

Jadi enkripsi dapat bekerja selama  $0 \leq m < n$

- b. Pilih nilai acak  $r$  dalam range  $0 < r < n$
- c. Hitung *ciphertext* dengan persamaan:

$$c = g^m \cdot r^n \text{ mod } n^2 \tag{8}$$

Deenkripsi pesan  $c$  dapat dilakukan dengan cara berikut:

- a.  $c \in \mathbb{Z}_{n^2}$  (9)

Jadi enkripsi dapat bekerja selama  $0 \leq c < n^2$

- b. Hitung *plaintext* dengan persamaan:

$$m = L(c^A \text{ mod } n^2) \cdot \mu \text{ mod } n \tag{10}$$

Adapun properti homomorfik dari paillier cryptosystem sebagai berikut:

- a. Produk perkalian dari dua buah *ciphertext* akan ter-deenkripsi menjadi hasil penjumlahan *plaintext* terkait.

$$D(E(m_1, r_1) \cdot E(m_2, r_2) \text{ mod } n^2) = m_1 + m_2 \text{ mod } n \tag{11}$$

- b. Produk dari perkalian *ciphertext* dengan *plaintext* yang ditingkatkan ke  $g$  akan ter-deenkripsi menjadi jumlah dari teks terkait.

$$D(E(m_1, r_1) \cdot g^{m_2} \text{ mod } n^2) = m_1 + m_2 \text{ mod } n \tag{12}$$

#### 2.4. Rancangan Pengujian

Pengujian yang akan dilakukan adalah pengamanan data saldo nasabah dan penambahan serta pengurangan saldo tersebut dalam bentuk *ciphertext*. Pengujian ini dilakukan untuk mengetahui apakah sistem dapat melakukan melakukan komputasi homomorfik. Waktu eksekusi juga dihitung saat melakukan komputasi dan key generation dengan perulangan sebanyak 50 kali. Pengujian kali ini tidak serta menguji aplikasi dikarenakan aplikasi untuk sistem ini masih dalam tahap perancangan.

### 3. Hasil dan Diskusi

#### 3.1. Pengujian Sistem

- a. Enkripsi

```

Masukan nama: Nama
Masukan saldo awal: 1000000
Ciphertext: 59587455257205647795523678891918458259851447046181281049161512243151159774681399768875331871843130639904312324963519390
2249021033107783984678897276951289486828391605499251827198915873678139229983761209421395842094464019861501439159684397789139812308
45182128450789544978999501749232991731851745124445967487424892457725181483858918864587886643982226924741576818561616238595284062
28382978164553209683558231791339683762897461173725123212118396535081416590488979535113866434346889645254537749356728317
68989892194429913676781011111596392286804680901582121597742668003728762423284680931314096427961200590890740954984040016235768076329
7516442861325138414355551817831926099587700437219888811972236731529886943607666424854700944874279358348431316763483753564861911481
728896676125120967732198510987718778094845201927163836274378806388236694120357690966589943790911118337660809165178669825171917398
22339706795691327963725484013520166858597808616792097286663220037027793231124842722426873569733893311295131439914434044
11915862168293868772386884785231385927549280739388651726566799123912156873692029792853993117375157529845259064573380025820921404
986914989549540894966200474217496408601178043475622994316649483694836523686816493015434001740073805955236893922268043751712327821
18980187929523390596268572198629787549444628291506082461103634851689552486145335182339619715666080456338231609689342228920213149
78477204935487820722028178860150706564999249869786933514355879774163140726958703400408391513210711398704519540002312055072
40695142280172874833931764167835494227352910867459928183073309347159166498459896310246481404815873051268101627819708485380571687
8197004858055716874666892427855479669647949404047301128086226040045682852743581624816381018900886252734246271493682077300754693
83091264620058527
    
```

Gambar 2. Output Enkripsi

Proses dimulai dengan pembuatan objek nasabah1 dan diberikan nama dan saldo awal. Saldo awal tersebut akan diubah menjadi *ciphertext* sebelum disimpan. Pada gambar 2, saldo awal yang dimasukkan adalah 1.000.000 yang kemudian dijadikan *ciphertext*. Rata-rata waktu eksekusi yang diperlukan untuk membuat objek nasabah tersebut dalam 50 kali percobaan adalah sebanyak 4,270 detik.

b. Komputasi

```
Tabungan terenkripsi: Nana
Jumlah deposit: 50000
Tabungan Nana (setelah dekripsi): 1050000
Tabungan terenkripsi: Nana
Jumlah pengambilan: 23700
Tabungan Nana (setelah dekripsi): 1026300
```

Gambar 3. Output Komputasi

Penambahan dan pengurangan dilakukan kepada tabungan nasabah1. Jumlah deposit saldo dan penarikan diubah menjadi *ciphertext* sebelum dihitung bersama dengan tabungan yang sudah ada di objek. Hasil komputasi yang telah didekripsi sesuai dengan jumlah perhitungan *plaintext*. Setelah dilakukan percobaan sebanyak 50 kali, didapatkan total waktu waktu 7,230 detik, sehingga rata-rata waktu yang diperlukan adalah 0.347 detik

c. Keamanan

Dari segi keamanan, sistem kriptografi Paillier adalah sebuah sistem kriptografi kunci publik dengan keamanan semantik. Untuk menghitung dan menilai kelebihan redundansi  $n$  kali pada  $\mathbb{Z}_{n^2}$  dalam kriptosistem paillier sangat sulit, sehingga kriptosistem Paillier memiliki tingkat keamanan yang tinggi. Selain itu, dalam proses enkripsi, karena sifat acak dari nilai  $r$ , bahkan jika *plaintext* yang sama didekripsi setiap kali, akan dihasilkan *ciphertext* yang berbeda. Oleh karena itu, sulit untuk melakukan serangan dengan menggunakan *plaintext*, dan keamanan algoritma tersebut meningkat secara proporsional[6]. Sebagai contoh, jika kita menggunakan kunci 2048-bit, itu berarti modulus  $n$  dalam algoritma Paillier memiliki panjang 2048-bit. Kunci 2048-bit dianggap setara dengan kekuatan keamanan 112-bit, itu berarti akan membutuhkan sekitar  $2^{112}$  operasi dasar untuk memecahkan kunci dengan serangan brute force.

#### 4. Kesimpulan

Berdasarkan hasil percobaan, enkripsi saldo dapat dilakukan yang kemudian disimpan sehingga jumlah saldo nasabah tidak dapat dipahami jumlahnya. Hal ini membantu melindungi privasi dan keamanan data dari akses yang tidak sah. Dengan menggunakan teknik enkripsi, bahkan jika data tersebut diakses oleh pihak yang tidak berwenang, mereka tidak akan dapat membaca atau memanipulasi informasi tersebut. Proses komputasi juga dilakukan pada data yang terenkripsi, yang kemudian didekripsi untuk menghasilkan hasil yang sesuai dengan perhitungan aslinya. Untuk penelitian yang akan datang terdapat hal yang harus dikembangkan, yaitu penyimpanan data pada sebuah database seperti mysql.

#### Daftar Pustaka

- [1] P. Mwiinga, "Privacy-Preserving Technologies: Balancing Security and User Privacy in the Digital Age", Dec. 2023, doi: 10.5281/zenodo.10406538.
- [2] A. Rajak and R. Agustia, "Purwarupa Sistem E-Voting Menggunakan Enkripsi Homomorphic Di Komisi Pemilihan Umum Kota Bandung," Jurnal Penelitian Mahasiswa Teknik Dan Ilmu Komputer (JUPITER), vol. 1, pp. 1–10, 2021, doi: 10.34010/jupiter.v1i1.5403.
- [3] F. Armknecht, C. Boyd, C. Carr, K. Gjøsteen, A. Jäschke, C. A. Reuter, and M. Strand, "A Guide to Fully Homomorphic Encryption," IACR ePrint Archive, 2015. [Online]. Available:

- <https://eprint.iacr.org/2015/1192>.
- [4] B. Li, D. Li, and M. Zhu, "Application analysis of data encryption technology," *Applied and Computational Engineering*, vol. 50, pp. 199–205, 2024, doi: 10.54254/2755-2721/50/20241502.
  - [5] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1999, doi: 10.1007/3-540-48910-X\_16.
  - [6] E. M., & Geng, Y. Homomorphic Encryption Technology for Cloud Computing. *Procedia Computer Science*, 154, 73–83, 2019. <https://doi.org/10.1016/j.procs.2019.06.012>

# Aspect Based Sentiment Analysis Terhadap Ulasan Produk Skincare di E-Commerce Menggunakan CNN-LSTM

I Made Treshnanda Mas<sup>a1</sup>, AAIN Eka Karyawati<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Udayana

Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia

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

<sup>2</sup>eka.karyawati@unud.ac.id

## Abstract

*Skincare product is any type of formulation intended to be topically applied with the intention of improving skin appearance, texture, or health. The burgeoning skincare product market in Indonesia, particularly in E-Commerce, underscores a shift towards online purchasing habits. With platforms like Shopee, Tokopedia, and Lazada dominating, user reviews play a pivotal role when making decision. However the sheer volume of reviews necessitates efficient processing methods. Aspect Based Sentiment Analysis (ABSA) emerges as a solution, delving deeper into sentiment analysis by identifying specific aspects within user feedback. Previous studies have shown Convolutional Neural Network based algorithms to have high performance in similar contexts. In this study, the author used CNN mixed with Long Short Term Memory (LSTM) alongside Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) as feature selection techniques to classify aspects into 5 classes and sentiment into 2 classes. The result of aspect classification produced accuracy, precision, recall, and F1-Score of 87.3%, 87%, 87%, and 85.9%. Meanwhile sentiment classification produced accuracy, precision, recall, and F1-Score of 87%, 87%, 87%, and 87%.*

**Keywords:** Skincare Product, E-Commerce, Aspect Based Sentiment Analysis, CNN, LSTM, TF-IDF, BoW

## 1. Pendahuluan

Produk skincare adalah formulasi apa pun yang dimaksudkan untuk dioleskan pada kulit dengan tujuan memperbaiki penampilan, tekstur, atau kesehatannya. Produk skincare tersedia dalam berbagai bentuk seperti krim, lotion, gel, masker, pembersih, dan pengelupas kulit. Produk skincare sering kali mengandung kombinasi bahan aktif, pelembab, emolien, dan senyawa lain yang dirancang untuk mengatasi permasalahan kulit tertentu. Penggunaan produk skincare di Indonesia semakin meningkat seiring tahun. Menurut laporan yang dirilis oleh [1] terhadap Produk FMCG (Fast Moving Consumer Goods) di E-Commerce Indonesia pada tahun 2023, produk kecantikan dan perawatan menempati posisi paling atas dengan persentase pembeliannya sebesar 49% dengan produk skincare sebagai sebagian besar dari produk kecantikan dan perawatan dengan persentasenya sebesar 39.4%. Menurut [2] E-Commerce merujuk pada proses distribusi, penjualan, promosi, serta pembelian barang atau layanan menggunakan teknologi elektronik seperti internet, televisi, jaringan komputer, dan lainnya. Ini juga mencakup transaksi keuangan online, pengelolaan stok secara otomatis, pertukaran informasi digital, dan sistem pengumpulan data secara otomatis. Pada tahun 2023, jumlah penduduk Indonesia yang merupakan pengguna E-Commerce adalah sebanyak 196 juta, angka tersebut meningkat dari tahun 2022 dengan jumlah pengguna E-Commerce adalah 178.94 juta [3]. Sebagian besar masyarakat Indonesia memanfaatkan platform E-Commerce seperti Shopee, Tokopedia, dan Lazada untuk berbelanja secara online. [4] melakukan analisis terhadap data SimilarWeb dengan hasilnya adalah Shopee merupakan platform E-Commerce dengan peringkat teratas paling sering dikunjungi, Tokopedia mendapati posisi kedua sedangkan Lazada menempati posisi

ketiga. Pengguna E-Commerce dapat memberikan suatu ulasan terhadap produk yang mereka beli setelah mereka menerima atau menggunakan produk tersebut, ulasan yang diberikan oleh pengguna berupa teks dan nilai rating dengan rentan 1 sampai 5. Ulasan-ulasan yang diberikan oleh pengguna terhadap suatu produk dapat digunakan oleh pengguna lainnya sebagai bahan pertimbangan ketika mereka ingin membeli suatu produk. Tetapi ulasan-ulasan ini cenderung berada dalam jumlah yang banyak dan setiap ulasan dapat membahas topik atau aspek yang berbeda-beda, sehingga jika menggunakan metode manual yaitu dengan membaca ulasan satu per satu akan memerlukan waktu yang cukup panjang. Oleh karena itu dibutuhkan suatu pendekatan untuk memproses ulasan-ulasan dari pengguna terhadap suatu produk agar informasi yang terdapat dalam ulasan-ulasan tersebut dapat dimanfaatkan dengan baik oleh pengguna. Aspect Based Sentiment Analysis (ABSA) merupakan area khusus dalam Text Sentiment Analysis yang bertujuan untuk menentukan polaritas sentimen pengguna terhadap istilah aspek tertentu dalam sebuah kalimat [5]. Tidak seperti Sentiment Analysis pada umumnya, informasi yang dapat diperoleh dari ABSA lebih banyak karena ulasan-ulasan tidak hanya dibagi menjadi dua kategori, yaitu "Positif" dan "Negatif", tetapi juga pada aspek apa kedua kelas tersebut. Sebagai contoh, kalimat "Harga produk ini sangat mahal" pada Sentiment Analysis biasa akan diklasifikasikan menjadi "Negatif" tetapi pada ABSA akan diklasifikasikan menjadi "Aspek Harga" dan "Negatif". Oleh karena itu, penggunaan ABSA diharapkan dapat memudahkan pengguna untuk mendapatkan informasi dari ulasan-ulasan terhadap suatu produk skincare. Penelitian yang dilakukan [6] menggunakan algoritma Convolutional Neural Network (CNN) untuk melakukan ABSA pada ulasan aplikasi Peduli Lindungi, hasil pengujian yang dilakukan menghasilkan performa yang bagus dengan nilai f1 score untuk klasifikasi aspek adalah 92.23% dan untuk klasifikasi sentimen adalah 95.13%. Dan penelitian yang dilakukan [7] penggunaan CNN-LSTM dalam analisis sentimen produk halar di Twitter menghasilkan nilai akurasi tertinggi sebesar 93.78% ketika dibandingkan dengan Long Short Term Memory (LSTM), CNN, dan CNN+Bi-RNN+Bi-LSTM. Sebab itu, ABSA dengan metode deep learning menggunakan algoritma CNN-LSTM akan dilakukan pada penelitian ini.

## 2. Metode Penelitian

### 2.1. Alur Penelitian

Gambar 1. Menunjukkan alur penelitian yang akan dilakukan pada penelitian ini.



**Gambar 1.** Alur Penelitian ABSA menggunakan CNN-LSTM

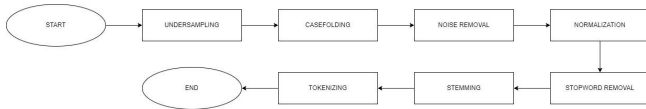
Pada penelitian ini pengumpulan data pertama dilakukan dengan scraping ulasan-ulasan pengguna terhadap produk skincare di platform E-Commerce. Selanjutnya setelah data terkumpul preprocessing terhadap data dilakukan dengan tujuan membersihkan data ulasan-ulasan pengguna tersebut. Pelabelan sentimen dilakukan dengan cara memberi kelas "Negatif" dan "Positif" kepada ulasan dengan menggunakan nilai rating yang diberi pengguna. Pelabelan topik atau aspek dilakukan menggunakan algoritma Latent Dirichlet Allocation (LDA) dengan data ulasan-ulasan pengguna yang sudah dilakukan preprocessing. Pada penelitian ini, fitur diekstraksi dengan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW). Setelah melakukan ekstraksi fitur, langkah selanjutnya adalah melatih model berbasis CNN untuk klasifikasi aspek dan klasifikasi sentimen. Kemudian kedua model klasifikasi yang telah dilatih selanjutnya akan dievaluasi performanya.

## 2.2. Data Collecting

Data ulasan-ulasan pengguna terhadap produk skincare di platform E-Commerce yang digunakan dalam penelitian ini dikumpulkan melalui penerapan teknik scraping. Scraping dilakukan dengan cara mengirim request atau permintaan kepada Application Programming Interface (API) E-Commerce, response atau respons yang didapat dari pengiriman permintaan tersebut berupa file dengan ekstensi json. Respons tersebut kemudian diproses untuk mengambil data teks dan rating saja dan disimpan dalam format csv. Data ulasan-ulasan pengguna diambil dari 1400 produk dengan kategori skincare dan secara keseluruhan terdapat 252 ribu ulasan yang diperoleh.

## 2.3. Data Preprocessing

Adapun tahapan-tahapan preprocessing yang akan dilakukan pada penelitian ini terlampir pada Gambar 2.



Gambar 2. Data Preprocessing

Undersampling adalah langkah pertama yang dilakukan, di mana jumlah data ulasan-ulasan dikurangi dari 252 ribu menjadi 12 ribu. Selanjutnya casefolding dilakukan untuk mengubah semua kata kapital menjadi kata kecil agar semua kata direpresentasikan dalam bentuk yang seragam yaitu dengan kata kecil. Sehabis itu dilakukan noise removal yang bertujuan untuk menghapuskan elemen-elemen selain kata seperti tanda baca, angka, dan elemen html. Lalu normalization dilakukan untuk mengubah kata-kata yang ditulis dengan salah, dengan singkat, dan dengan bentuk tidak formal menjadi bentuknya yang seharusnya. Lalu dilakukan stopwords removal dengan tujuan untuk menghapus kata-kata yang tidak mengandung makna atau arti. Setelah itu stemming akan dilakukan untuk mengubah kata-kata yang memiliki imbuhan menjadi bentuk dasarnya. Hasil dari proses stemming selanjutnya akan dilakukan proses tokenizing untuk memecah data yang berbentuk kalimat menjadi terpisah kata per kata.

## 2.4. Sentiment Labeling

Setelah melakukan preprocessing, selanjutnya ulasan-ulasan pengguna terhadap produk skincare akan diberi label sentimen berdasarkan nilai rating dari ulasan tersebut. Untuk ulasan yang memiliki nilai rating 1 sampai 3 maka akan diberikan label "Negatif", sedangkan untuk ulasan yang memiliki nilai rating 4 sampai 5 maka akan diberikan label "Positif". Secara keseluruhan, terdapat 6121 ulasan pengguna terhadap produk skincare dengan sentimen positif dan 6121 ulasan pengguna terhadap produk skincare dengan sentimen negatif.

## 2.5. Topic Labeling

Pelabelan topik dalam penelitian ini dilakukan dengan menggunakan algoritma LDA. LDA merupakan model probabilitas yang terkenal sebagai salah satu algoritma yang populer untuk pemodelan topik [8]. LDA mencoba menganalisis dokumen dengan melihat ke belakang untuk menemukan beragam topik yang mungkin ada. Pendekatan ini memungkinkan LDA untuk menyajikan dokumen dengan kumpulan topik yang diperkirakan berdasarkan probabilitasnya. Dalam penelitian ini, LDA digunakan untuk mengelompokkan ulasan-ulasan pengguna terhadap produk skincare ke dalam 5 topik yaitu "Delivery", "Skin Compatibility and Usage", "Ordering Experience", "Packaging Quality", "Product Experience" dengan jumlahnya masing-masing adalah 3873, 2601, 2309, 1761, dan 1596.



## 2.6. Feature Extraction

Dalam penelitian ini, dilakukan dua jenis ekstraksi fitur yaitu TF-IDF dan BoW. TF-IDF adalah metode ekstraksi fitur dengan algoritma yang bekerja dengan cara menghitung nilai bobot bagi setiap kata dalam sebuah kalimat, untuk menetapkan nilai bobot untuk setiap kata dalam dokumen, pendekatan ini akan menghitung seberapa sering kata tersebut muncul pada dokumen tersebut [9]. Dalam TF-IDF, semakin sering suatu kata itu muncul dalam dokumen maka semakin rendah nilai pembobotannya.

$$w_{ij} = tf_{ij} \cdot \log \left( \frac{D}{df_i} \right) \quad (1)$$

Sedangkan BoW adalah sebuah metode representasi yang mengubah data teks menjadi vektor dengan panjang yang telah ditentukan, dengan cara melakukan perhitungan frekuensi kata yang berulang. Dalam representasi ini, kalimat atau dokumen dianggap sebagai himpunan kata tanpa memperhatikan urutan dan direpresentasikan dalam vektor dengan ukuran tetap yang mencatat jumlah kemunculan kata, tanpa memperhatikan struktur tata bahasa atau urutan kata.

## 2.7. Convolutional Neural Network

Model yang akan digunakan dalam penelitian ini adalah CNN. CNN merupakan sebuah teknologi kecerdasan buatan yang berbasis pada jaringan saraf berlapis-lapis yang mampu mengenali, mengidentifikasi, dan mengklasifikasikan objek, serta mampu mendeteksi dan memisahkan objek dalam gambar. Meskipun CNN umumnya digunakan dalam Computer Vision seperti deteksi objek dan klasifikasi gambar tetapi CNN telah terbukti efektif dan memiliki performa bagus dalam klasifikasi teks [10]. CNN umumnya terdiri dari 4 layer, yaitu Convolution Layer, Embedding Layer, Pooling Layer, dan Fully Connected Layer. Convolution Layer adalah layer utama di mana proses konvolusi terjadi untuk input dan filter, ketika input memasuki lapisan ini, operasi konvolusi yang melibatkan filter diterapkan ke jendela kata untuk menghasilkan fitur baru, filter diterapkan berulang kali ke setiap jendela kata dalam kalimat untuk menghasilkan peta fitur [10]. Embedding Layer adalah layer yang berfungsi untuk memetakan input menjadi vektor berdimensi rendah, hasil pemetaan di Embedding Layer selanjutnya diberikan kepada Convolution Layer. Pooling Layer adalah layer yang memiliki fungsi untuk mengurangi jumlah parameter secara bertahap, kompleksitas komputasi model, dan kontrol overfitting [11]. Fully Connected Layer adalah layer dalam CNN yang membentuk neuron satu dimensi, layer ini terdiri dari neuron-neuron yang saling terhubung dengan neuron-neuron pada layer sebelum dan seterusnya [12].

## 2.8. Confusion Matrix

Confusion Matrix akan digunakan sebagai metode evaluasi untuk model klasifikasi aspek dan klasifikasi sentimen. Confusion Matrix merupakan sebuah tabel yang digunakan untuk menilai performa suatu model klasifikasi, dalam Confusion Matrix, setiap baris mewakili kelas sebenarnya dari data, sedangkan setiap kolom mewakili kelas prediksi oleh model. Terdapat 4 perhitungan yang dilakukan dalam Confusion Matrix yaitu akurasi, presisi, recall, dan F1-score.

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

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

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

$$\text{Recall} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

### 3. Hasil dan Diskusi

#### 3.1. Data Collecting

Adapun contoh respons dari request data ulasan-ulasan pengguna terhadap produk dengan kategori skincare kepada API E-Commerce sebagai berikut.

Tabel 1. Hasil Data Collecting

| index | Review Message  | Rating |
|-------|---|--------|
| 0     | Sangat rekomen, pemakaian 1 bln terbukti ampuh  | 5      |
| 1     | maap baru kasih ulasan, thanks bgt pesanan semuanya sesuai, good seller !   | 4      |
| 2     | maaf di tipe kulit ku yg berminyak kurang cocok krna bikin tmbh berminyak   | 3      |
| 3     | ku kira cocok buat pria ternyata oh ternyata putihnya pucet ga cocok banget malah timbul jerawat,kapak dah barang kebuang sia sia | 2      |
| 4     | tidak terjadi apa2 setelah pemakaian selama 3 bulan   | 1      |

#### 3.2. Data Preprocessing

Setelah berhasil mengumpulkan data, selanjutnya jumlah baris yang akan digunakan adalah 12 ribu baris, jumlah baris tersebut ditentukan berdasarkan dari jumlah ulasan yang memiliki nilai rating 3 dan dibawahnya. Selanjutnya akan dilakukan Case Folding untuk mengubah semua kata kapital menjadi kata dasar, Noise Removal, Normalization, Stopword Removal, Stemming, dan Tokenizing. Berikut adalah hasil setelah

Tabel 2. Hasil Data Preprocessing

| Tahapan          | Text  | Rating |
|------------------|---|--------|
| Casefolding      | sangat rekomen, pemakaian 1 bln terbukti ampuh  | 5      |
| Noise Removal    | sangat rekomen pemakaian bln terbukti ampuh     | 5      |
| Normalization    | sangat recommend pemakaian bulan terbukti ampuh | 5      |
| Stopword Removal | recommend pemakaian terbukti ampuh              | 5      |
| Stemming         | recommend pakai bukti ampuh                     | 5      |
| Tokenizing       | [recommend, pakai, bukti, ampuh]                | 5      |

#### 3.3. Sentiment Labeling

Setiap baris pada dataset yang telah dilakukan preprocessing kemudian akan diberikan label sentimen berdasarkan dari nilai rating yang dimiliki setiap baris. Untuk baris dengan nilai rating 4 sampai 5 maka akan diberikan sentimen label "Positif", sedangkan untuk baris dengan nilai rating 1 sampai 3 akan diberikan label sentimen "Negatif". Setelah pemberian label sentimen, terdapat 6121 baris dalam dataset dengan label sentimen "Positif", dan 6121 baris dalam dataset dengan label sentimen "Negatif".

Tabel 3. Hasil Sentiment Labeling

| Text                             | Rating | Sentimen |
|----------------------------------|--------|----------|
| [recommend, pakai, bukti, ampuh] | 5      | Positif  |
| ['enggak', 'cocok', 'muka']      | 3      | Negatif  |

### 3.4. Topic Labeling

Pelabelan topik atau aspek dengan LDA dilakukan dengan menggunakan hasil dari ekstraksi fitur BoW. Adapun 5 topik dengan 10 kata kunci masing masing topik sebagai berikut.

**Tabel 4.** Pengelompokan Topik dengan Kata Kuncinya

| Topic                        | Kata Kunci  |
|------------------------------|---|
| Delivery                     | [kirim, barang, lama, cepat, aman, pesan, packing, banget, produk, proses]    |
| Product Experience           | [enggak, pakai, beli, kalo, kecewa, sih, tidak, beda, komplain, kayak]        |
| Ordering Experience          | [terima, kasih, barang, sesuai, harga, beli, tokopedia, produk, tidak, paket] |
| Skin Compatibility and Usage | [cocok, kulit, pakai, moga, coba, enggak, jerawat, muka, banget, wajah]       |
| Packaging Quality            | [enggak, tidak, kemas, isi, buka, tanggal, beli, botol, tutup, pas]           |

Kemudian setiap baris pada dataset akan diberi label topik atau aspeknya berdasarkan dari hasil pengelompokan topik berdasarkan kata-kata yang dilakukan dengan menggunakan model LDA. Distribusi jumlah baris setiap topik dapat dilihat pada tabel sebagai berikut.

**Tabel 5.** Distribusi Jumlah Topik

| Topic                        | Jumlah |
|------------------------------|--------|
| Delivery                     | 3873   |
| Product Experience           | 1596   |
| Ordering Experience          | 2309   |
| Skin Compatibility and Usage | 2601   |
| Packaging Quality            | 1761   |

### 3.5. Data Preparation

Sebelum diberikan kepada model untuk pelatihan, dataset yang telah dilakukan preprocessing menjalani tahap pemisahan untuk memastikan keberhasilan pelatihan, validasi, dan evaluasi model yang efektif. Proses ini dimulai dengan pembagian dataset menjadi tiga subset utama yaitu pelatihan, validasi, dan evaluasi. Pembagian dataset dilakukan dengan proporsi yang telah ditetapkan sebelumnya, di mana 80% dari data digunakan untuk pelatihan model, 10% dialokasikan untuk validasi selama proses pelatihan, dan 10% tersisa untuk mengukur performa model setelah pelatihan selesai.

### 3.6. TF-IDF Feature Extraction

Teknik ekstraksi fitur TF-IDF dilakukan untuk mengubah data tekstual menjadi representasi dalam bentuk numerikal untuk diberikan sebagai input dalam model CNN. Proses ini dimulai dengan tokenisasi dan pengkodean urutan dari data teks input, teks disegmentasi menjadi token atau kata individual dan mengkodekannya ke dalam urutan panjang tetap menggunakan padding untuk memastikan keseragaman dalam dimensi input. Setelah itu, representasi TF-IDF dibuat untuk data teks. Adapun hasil dari ekstraksi fitur TF-IDF pada gambar berikut.

```
print(tfidf_train.shape)
print(tfidf_train)

(9712, 6116)
(0, 5052) 0.6650708789751087
(0, 5058) 0.31338233187547343
(0, 362) 0.44377947657805895
(0, 3130) 0.41388246414479885
(0, 2039) 0.302048542537632
(1, 3190) 0.8665414472226423
(1, 2688) 0.4991051194340615
(2, 1144) 0.11683185976136362
(2, 3218) 0.21011689495370203
```

Gambar 3. Hasil Ekstraksi TF-IDF

### 3.7. BoW Feature Extraction

Selain TF-IDF, ekstraksi fitur dengan menggunakan teknik BoW juga dilakukan. Mirip dengan proses ekstraksi fitur TF-IDF, dalam ekstraksi fitur BoW pertama dilakukan tokenisasi dan mengkodekan urutan teks. Setelah itu representasi BoW dihasilkan untuk data teks dengan mengubah urutan teks yang telah diberi token menjadi vektor dengan menghitung frekuensi setiap istilah dalam kosakata. Hasil ekstraksi fitur ini kemudian akan digunakan sebagai input dalam model CNN. Adapun hasil ekstraksi fitur BoW dapat terlihat pada gambar sebagai berikut

```
print(bow_train.shape)
print(bow_train)

(9712, 6116)
(0, 2039) 1
(0, 3130) 1
(0, 362) 1
(0, 5058) 1
(0, 5052) 1
(1, 2688) 1
(1, 3190) 1
(2, 4344) 1
```

Gambar 4. Hasil Ekstraksi BoW

### 3.8. Pelatihan Model CNN-LSTM Klasifikasi Aspek

Model CNN yang akan dilatih untuk klasifikasi aspek pada penelitian ini digabungkan dengan LSTM untuk secara efektif menangkap pola spasial dan temporal dalam data teks. Model ini menggabungkan representasi TF-IDF dan BoW dari data teks input sebagai fitur tambahan. Representasi ini digabungkan dengan output dari lapisan LSTM dan dimasukkan ke dalam lapisan padat yang terhubung sepenuhnya untuk klasifikasi. Arsitektur model CNN-LSTM yang digunakan untuk pelatihan terdiri dari tiga komponen utama: Embedding Layer yang diikuti oleh dua Convolution Layer, LSTM Layer, dan Dense Layer berikutnya untuk klasifikasi. Berikut adalah parameter yang digunakan dalam pelatihan model CNN-LSTM untuk klasifikasi aspek.

Tabel 6. Parameter Pelatihan Model Klasifikasi Aspek

| Parameter | Nilai |
|-----------|-------|
| Epochs    | 20    |

| Parameter     | Nilai   |
|---------------|---|
| Batch Size    | 128   |
| Conv1D (1)    | Filters=16, kernel_size=4. Activation = relu  |
| Conv1D (2)    | Filters=4, kernel_size=2. Activation = relu   |
| LSTM          | Units=128. Dropout=0.5, recurrent_dropout=0.2 |
| Dense (1)     | Units=128, dropout=0.5, activation=relu       |
| Learning Rate | 0.001   |

### 3.9. Pelatihan Model CNN Klasifikasi Sentimen

Model CNN untuk klasifikasi sentimen juga menggabungkan LSTM untuk secara efektif menangkap pola spasial dan temporal dalam data teks. Model ini juga menggabungkan representasi TF-IDF dan BoW dari data teks input sebagai fitur tambahan. Representasi ini digabungkan dengan output dari lapisan LSTM dan dimasukkan ke dalam lapisan padat yang terhubung sepenuhnya untuk klasifikasi. Arsitektur model CNN-LSTM yang digunakan untuk analisis sentimen terdiri dari Embedding Layer yang diikuti oleh 1 Convolution Layer, LSTM Layer, dan Dense Layer berikutnya untuk klasifikasi. Berikut merupakan nilai parameter yang digunakan dalam pelatihan model CNN-LSTM untuk klasifikasi sentimen.

Tabel 7. Parameter Pelatihan Model Klasifikasi Sentimen

| Parameter     | Nilai   |
|---------------|---|
| Epochs        | 30  |
| Batch Size    | 128   |
| Conv1D (1)    | Filters=128, kernel_size=5. Activation = relu |
| LSTM          | Units=128. Dropout=0.2, recurrent_dropout=0.2 |
| Dense (1)     | Units=128, dropout=0.2, activation=relu       |
| Learning Rate | 0.0001  |

### 3.10. Evaluasi Model CNN-LSTM Klasifikasi Aspek

Setelah model CNN-LSTM untuk klasifikasi aspek selesai dilatih, langkah selanjutnya adalah mengukur dan mengevaluasi performa model tersebut menggunakan data evaluasi yang sebelumnya telah dipisah dari keseluruhan dataset. Berikut adalah hasil evaluasi dari model CNN-LSTM untuk klasifikasi aspek.

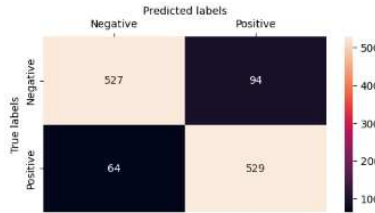


Gambar 5. Confusion Matrix Model Klasifikasi Aspek

Berdasarkan dari Confusion Matrix pada gambar 5. Model klasifikasi aspek memiliki nilai akurasi sebesar 87.3%, presisi sebesar 87%, recall sebesar 87%, dan F1-Score sebesar 85,9%.

### 3.11. Evaluasi Model CNN-LSTM Klasifikasi Sentimen

Pengukuran dan evaluasi performa model CNN-LSTM untuk klasifikasi sentimen dilakukan setelah pelatihan selesai. Evaluasi performa model akan dilakukan menggunakan data evaluasi yang telah dipisah dari keseluruhan dataset. Hasil evaluasi dari model CNN-LSTM untuk klasifikasi sentimen dapat dilihat sebagai berikut.



Gambar 6. Confusion Matrix Model Klasifikasi Sentimen

Berdasarkan Confusion Matrix gambar 6. Yang didapat dengan menggunakan data evaluasi, didapatkan bahwa model CNN-LSTM untuk klasifikasi sentimen memiliki nilai akurasi sebesar 87%, nilai presisi sebesar 87%, nilai recall sebesar 87% dan nilai F1-Score sebesar 87%.

## 4. Kesimpulan

Dapat disimpulkan bahwa penggunaan model CNN-LSTM dalam Aspect Based Sentiment Analysis terhadap ulasan produk skincare di E-Commerce memberikan hasil yang cukup bagus. Dengan menggunakan teknik ekstraksi fitur TF-IDF dan BoW, model CNN-LSTM untuk klasifikasi aspek mendapatkan nilai akurasi sebesar 87.3%, presisi sebesar 87%, recall sebesar 87%, dan F1-Score sebesar 85,9%. Sedangkan, model CNN-LSTM untuk klasifikasi sentimen mendapatkan nilai akurasi sebesar 87%, nilai presisi sebesar 87%, nilai recall sebesar 87% dan nilai F1-Score sebesar 87%.

## Daftar Pustaka

- [1] Kompas, "Kompas Market Insight: Indonesian FMCG E-commerce Report 2023," Kompas. <https://kompas.co.id/article/kompas-market-insight-indonesia-fmcg-e-commerce-report-2023/> (accessed May 9, 2024).
- [2] H. Harmayani, D. Marpaung, A. Hamzah, N. Mulyani, and J. Hutahaean, E-Commerce: Suatu Pengantar Bisnis Digital. Yayasan Kita Menulis, 2020.
- [3] S. Denny, "Masih Digandrungi, pengguna e-commerce Indonesia capai 196 Juta di 2023," liputan6, <https://www.liputan6.com/bisnis/read/5447108/masih-digandrungi-pengguna-e-commerce-indonesia-capai-196-juta-di-2023?page=2> (accessed May 9, 2024).
- [4] A. Ahdia, "5 e-commerce Dengan Pengunjung terbanyak sepanjang 2023." Databoks: Pusat Data Ekonomi dan Bisnis Indonesia, <https://databoks.katadata.co.id/datapublish/2024/01/10/5-e-commerce-dengan-pengunjung-terbanyak-sepanjang-2023> (accessed May 9, 2024).
- [5] K. Ahmed, M. I. Nadeem, Z. Zheng, D. Li, I. Ullah, M. Assam, Y. Y. Ghadi, and H. G. Mohamed, "Breaking down linguistic complexities: A structured approach to aspect-based sentiment analysis," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 8, pp. 101651, 2023. [Online]. Available: <https://doi.org/10.1016/j.iksuci.2023.101651>.

- [6] P. Aritonang, M. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using Convolutional Neural Network," *Ultima InfoSys: J. Ilmu Sist. Inf.*, vol. 13, pp. 54-61, 2022. [Online]. Available: <https://doi.org/10.31937/si.v13i1.2684>.
- [7] A. Feizollah, S. Ainin, N. Anuar, N. Abdullah, and H. Hanif, "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms," *IEEE Access*, vol. 7, pp. 1-1, 2019. [Online]. Available: <https://doi.org/10.1109/ACCESS.2019.2923275>.
- [8] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Front. Artif. Intell.*, vol. 3, pp. 42, 2020. [Online]. Available: <https://doi.org/10.3389/fraci.2020.00042>
- [9] S. A. Pratomo, S. Al Faraby, and M. D. Purbolaksono, "Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF dan Lexicon Pada Ulasan Film Menggunakan Metode KNN," 2021.
- [10] Y. Kim, "Convolutional Neural Networks for Sentence Classification," 2014. [Online]. Available: <https://arxiv.org/abs/1408.5882>.
- [11] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," 2015. [Online]. Available: <https://arxiv.org/abs/1511.08458>.
- [12] M. Bangsa, S. Priyanta, and Y. Suyanto, "Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, pp. 123, Apr. 2020. [Online]. Available: <https://doi.org/10.22146/ijccs.51646>.

# Evaluasi UI Design pada Prototype Aplikasi “Instix” Menggunakan Metode Sistem Usability Scale

Ni Putu Dita Ari Pratiwi<sup>a1</sup>, AAIN Eka Karyawati<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>pratiwi.2208561003@student.unud.ac.id  
<sup>2</sup>eka.karyawati@unud.ac.id

## Abstract

*The success of an application is measured by conducting evaluations to ensure their product is easy to use, understand, and provides a positive experience for users. The Instix application is designed to help business owners centrally manage their products across multiple platforms. This research aims to evaluate the UI design of the Instix application using the System Usability Scale (SUS) method. This research involved 20 participants who were asked to complete certain tasks in the Instix application. Participants then fill out a questionnaire to assess several aspects. The research results show that the evaluation score using the Instix application's Usability Scale (SUS) System method is 87 with the best imaginable category, which shows that this application meets UI and UX quality standards, which can provide an optimal experience in using the application.*

**Keywords:** *UI Evaluation, System Usability Scale, mobile, Marketplace Integration, Product Management*

## 1. Pendahuluan

Internet adalah elemen krusial dalam mendukung perkembangan teknologi, terutama dalam mengembangkan bisnis dan menyebarkan informasi kepada masyarakat yang aktif menggunakan internet. Pertumbuhan *e-commerce*, khususnya di Indonesia, diprediksi mencapai tingkat pertumbuhan tertinggi pada tahun 2024, dengan proyeksi pertumbuhan sebesar 30,5%, yang jauh melebihi rerata global sebesar 10,4% [1]. Hal ini juga sejalan dengan pertumbuhan UMKM digital yang mencapai 21,8 juta pada tahun 2023[2]. Membantu UMKM meningkatkan penjualan adalah salah satu tujuan dari pengembangan *e-commerce*. Namun, perlu diingat bahwa kesuksesan dalam implementasi *e-commerce* tidak selalu mudah. Integrasi perusahaan dan rantai pasokan, serta strategi pemasaran yang tepat, merupakan faktor penting dalam keberhasilan *e-commerce*. Pengembangan *e-commerce* tidak hanya tentang teknologi atau aplikasinya saja, tetapi juga melibatkan berbagai faktor lain seperti strategi marketing yang harus dipertimbangkan agar dapat berhasil dengan baik[3]. Para pemilik *marketplace* dihadapkan pada tantangan dalam mengelola produk mereka di berbagai platform secara terpusat, yang seringkali memakan waktu dan tidak efisien. Solusi yang ditawarkan adalah dengan menggunakan aplikasi Instix, yang memungkinkan para pemilik usaha atau UMKM yang memiliki banyak cabang *marketplace* di berbagai *e-commerce* untuk terintegrasi dan melakukan manajemen barang hanya melalui satu aplikasi. Dengan berbagai fitur yang mendukung, termasuk kemampuan posting produk di berbagai platform *e-commerce* dan proses pengecekan transaksi yang terintegrasi, Instix menjadi solusi efektif untuk memudahkan pengelolaan bisnis online. Dalam pengembangan aplikasi Instix, metode pengujian yang digunakan adalah *System Usability Scale* (SUS), yang telah terbukti efektif dalam mengevaluasi tingkat kebergunaan aplikasi salah satunya dalam jurnal "Evaluasi tingkat kebergunaan aplikasi Administrasi Penduduk menggunakan teknik *System Usability Scale*" [4]. "Perancangan User Interface Aplikasi Rumah Sehat Retno Berbasis Smartphone Dan Evaluasi *System Usability Scale* (Sus)"[5]. Diharapkan penelitian ini dapat memberikan masukan yang berharga untuk memenuhi standar desain UI dan UX, sehingga aplikasi Instix menjadi lebih mudah digunakan



dan memberikan pengalaman yang positif bagi pengguna.

## 2. Metode Penelitian

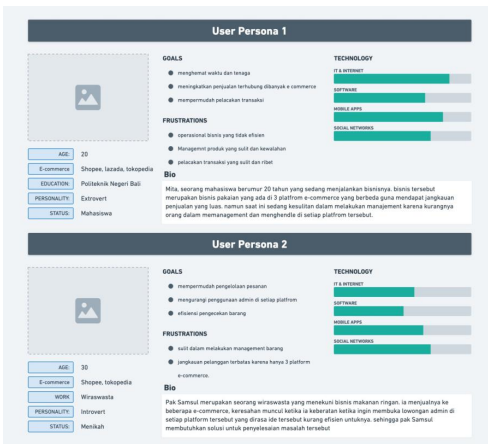
Pada penelitian ini menggunakan metode *System Usability Scale* (SUS) sebagai alat ukur kegunaan aplikasi. Adapun beberapa tahapan yang dilakukan dalam penelitian ini secara sistematis alur penelitian yang digunakan pada perancangan design UI aplikasi Instix dimulai dengan melakukan studi literatur, tahapan analisis kebutuhan pengguna (*user research*) dengan membuat user persona yang dilanjutkan user journey map, untuk membantu memvisualisasikan skenario dan pandangan pengguna dalam menggunakan produk. tahap perancangan desain antarmuka aplikasi dimulai dengan pembuatan *wireframe* dilanjutkan dengan pembuatan desain yang lebih rinci (*high fidelity*). dilanjutkan dengan evaluasi kegunaan aplikasi, yang menggunakan *simple random sampling* sebagai pengambilan sampel responden sebanyak 20 responden yang merupakan jumlah yang efisien dalam pengujian usability [6]. Kuesioner yang digunakan adalah versi terjemahan dalam bahasa Indonesia dari *System Usability Scale* (SUS) yang dikembangkan oleh John Brooke. Kuesioner ini terdiri dari 10 pertanyaan dengan 5 pilihan jawaban, mulai dari "sangat tidak setuju" hingga "sangat setuju". Dalam banyak penelitian, rata-rata skor SUS adalah 68. Oleh karena itu, jika skor SUS di atas 68, dianggap di atas rata-rata, sedangkan jika di bawah 68, dianggap di bawah rata-rata [7].

## 3. Hasil dan Diskusi

Hasil dan diskusi pada penelitian ini dibuat berdasarkan alur penelitian yang telah dirancang sebelumnya.

### 3.1. User persona

Pada Gambar 1 merupakan dua bentuk user persona dengan rentang umur dan latar belakang pengguna yang berbeda sebagai respirasi dari *user* yang akan menggunakan aplikasi, dengan menjabarkan terkait tujuan dan apa yang menjadi kendala yang dialami pengguna.

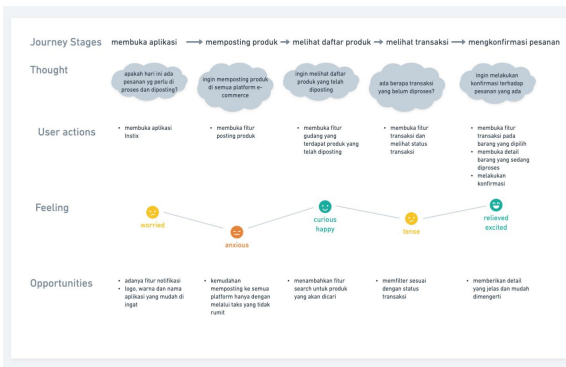


Gambar 1. User Persona

User persona pada Gambar 1 akan membantu tim untuk memahami siapa pengguna, apa yang mereka butuhkan, dan bagaimana produk dapat memenuhi kebutuhan mereka.

### 3.2 User Journey Map

User Journey Map pada Gambar 2 adalah Gambaran visual interaksi pengguna dengan produk atau layanan dari awal hingga akhir. Dalam membantu untuk memahami pengguna diperlukannya pemetaan menggunakan user journey map seperti pada Gambar Berikut terdapat *journey stages* yang merupakan tahapan-tahapan yang akan dilalui pengguna, *thought* hal yang dipikirkan atau yang menjadi keinginan pengguna, *user actions* merupakan aksi yang dilakukan pengguna didalam aplikasi, *feeling* sebagai bentuk perasaan yang dialami user sehingga kita dapat mengetahui tingkat kenyamanan pengguna, dan *opportunities* merupakan hal yang dapat dikembangkan atau ide yang dapat menjadi solusi dari permasalahan.

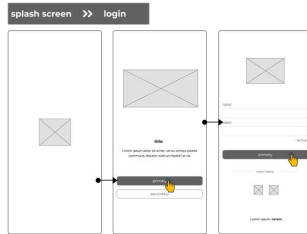


Gambar 2. User Journey Map

### 3.3 Wireframe dan Wireflow

#### a. Alur Ketika Pengguna Pertama Kali Menggunakan Aplikasi

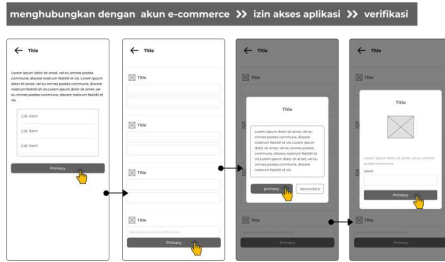
Ketika pengguna membuka aplikasi, mereka akan disambut dengan splash screen pada halaman pertama. Setelahnya, mereka akan diarahkan untuk melakukan login dengan menekan tombol masuk dan mengisi data seperti email dan *password*.



Gambar 3. Alur Masuk Aplikasi

**b. Alur Ketika Pengguna Menghubungkan Akun Dengan E-Commerce**

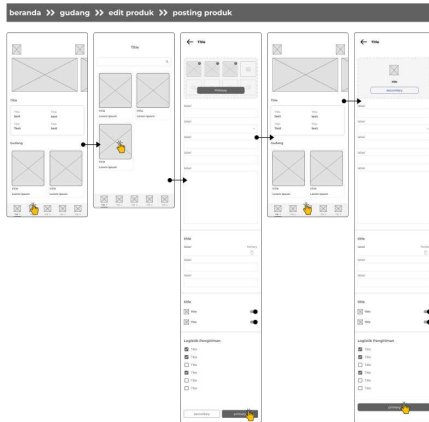
Ketika pengguna telah melewati alur masuk aplikasi, selanjutnya yaitu menghubungkan akun dengan *e-commerce*. Pengguna akan diarahkan untuk membaca arahan dan menekan *button* untuk melanjutkan ke tahap berikutnya yaitu mengisi data akun *e-commerce* yang ingin ditautkan kemudian akan muncul *pop up* untuk melakukan konfirmasi izin akses aplikasi dan *pop up* untuk verifikasi kode OTP, seperti pada Gambar 4.



Gambar 4. Alur Menghubungkan dengan Akun *E-commerce*

**c. Alur Untuk Melihat Produk Yang Kita Pasarkan, Melakukan Edit Dan Posting**

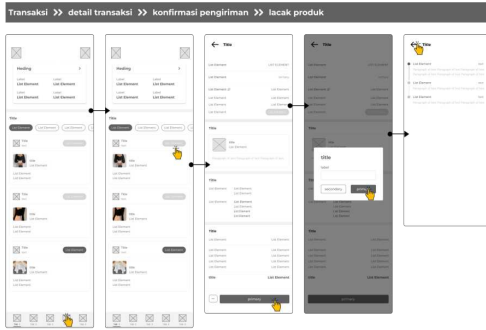
Setelah menghubungkan akun pengguna akan diarahkan ke halaman beranda seperti Gambar 5 yang menunjukkan tampilan beranda dari aplikasi yang berisikan promosi dan beberapa fitur, dilanjutkan dengan halaman gudang yang berisi produk yang telah kita posting. Ketika mengklik produk tersebut maka akan muncul tampilan edit produk yang telah diposting tersebut. Sedangkan pada menu navigasi posting akan menampilkan halaman posting, dengan berbagai *input* data yang diminta untuk diposting.



Gambar 5. Produk Yang Kita Pasarkan, Melakukan Edit Dan Posting

#### d. Alur Mengecek Transaksi

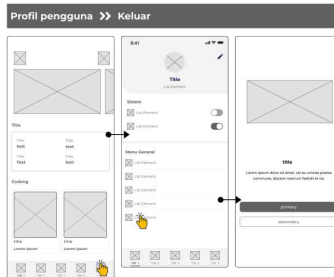
Pengguna juga dapat melakukan pengecekan transaksi. Gambar 6 menunjukkan alur untuk melakukan pengecekan transaksi. pada halaman transaksi pengguna akan dapat melihat kalkulasi pemasukan, jumlah produk terjual, serta total transaksi. Disana juga terdapat filter berdasarkan kategori daftar transaksi. ketika pengguna menekan card produk maka akan menampilkan terkait detail dari transaksi tersebut. pengguna dapat menekan *button* konfirmasi pengiriman untuk transaksi dengan label pesanan yang sedang di proses, pengguna akan mengisi *pop up* konfirmasi dengan memasukkan nomor resi dan menekan *button* simpan. Selanjutnya pengguna akan melihat detail transaksi, dan dapat menekan *button* lacak yang akan menunjukkan status dari pemesanan.



Gambar 6. Alur Mengecek Transaksi

#### e. Profile Pengguna

Bagian profil pengguna akan menampilkan alur interaksi pengguna saat mengakses dan memperbarui informasi profil mereka, serta navigasi kembali ke bagian lain dalam aplikasi setelah selesai. Sementara itu, pengguna juga dapat keluar dari akun mereka dengan menekan *button* keluar seperti pada Gambar 7.

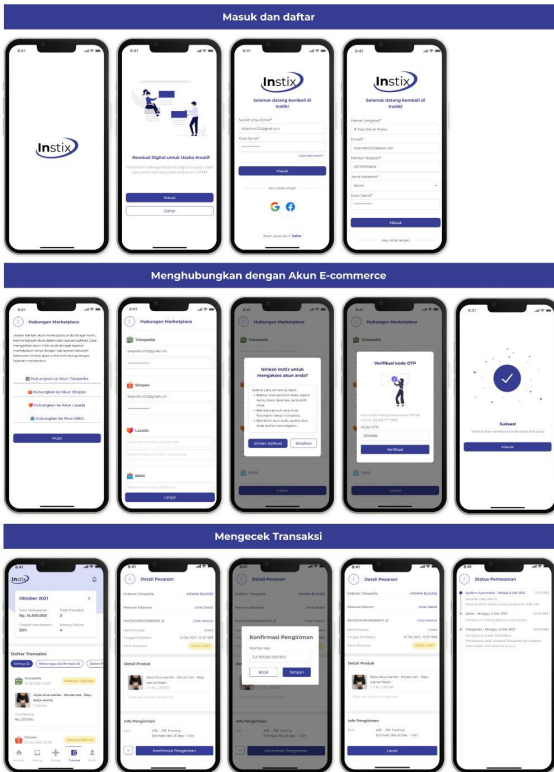


Gambar 7. Profile Pengguna

### 3.4 High Fidelity

Adapun untuk hasil desain High Fidelity dalam Gambar 8 dan Gambar 9 telah dikelompokkan berdasarkan alur dari wireflow serta pengelompokan fitur yang telah disusun sebelumnya. Warna biru mendominasi dengan kode warna #373D93, yang melambangkan rasa kepercayaan dan kenyamanan bagi pengguna. Pembuatan desain didasarkan pada desain sistem yang sudah dirancang sebelumnya, untuk memudahkan proses desain secara keseluruhan. Saat pengguna membuka aplikasi akan terlihat *splash screen* yang terdapat logo aplikasi sebagai *brand identity*. logo aplikasi Instix di design *simple* untuk memudahkan pengguna mengingat aplikasi. Alur pertama penggunaan aplikasi pengguna akan memilih daftar

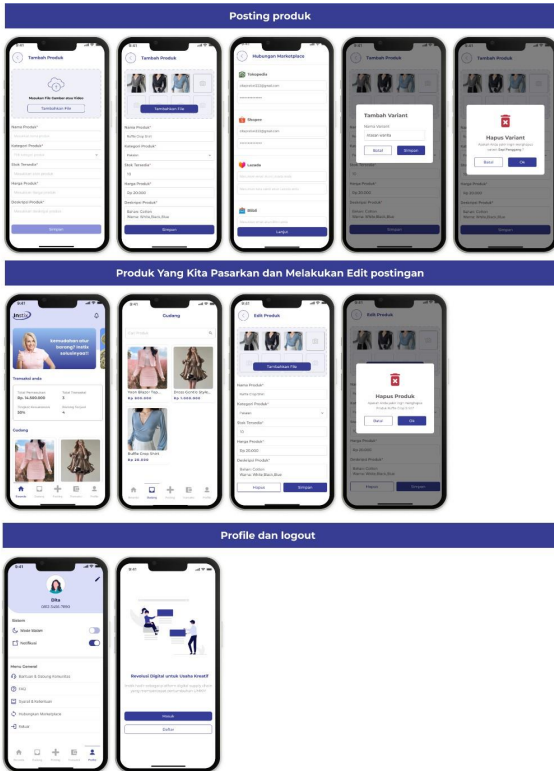
atau masuk bagi yang telah memiliki akun dan hanya menginputkan email serta *password*, berbeda dengan daftar yang memerlukan data lainnya seperti nama, jenis kelamin, dan nomor telepon. Pada alur pengguna untuk menghubungkan akun dengan e-commerce, pengguna akan menekan *button* mulai dilanjutkan dengan input akun e-commerce yang ingin dihubungkan, mengonfirmasi izin akses, verifikasi OTP, serta meningkatkan interaksi pengguna dengan memunculkan screen sukses ketika telah berhasil menghubungkan akun dengan e-commerce.



**Gambar 8.** High Fidelity tampilan masuk, daftar, hubungkan akun, dan transaksi setelah

menghubungkan akun tersebut pengguna akan dapat menjelajahi berbagai fitur yang nantinya akan membantu pengguna seperti di Gambar 8 pada bagian pengecekan transaksi,

sedangkan pada Gambar 9 yaitu melakukan posting produk, memmanagement produk yang telah dipasarkan, baik melakukan edit maupun menghapusnya serta pengguna juga dapat mengatur profile atau akun mereka baik dengan mengubah foto profile, mengubah mode aplikasi, maupun keluar dari aplikasi.



**Gambar9.** High Fidelity tampilan posting produk, fitur gudang, dan profile pengguna

pada bagian high fidelity ini sudah ditambahkan dengan ilustrasi yang senada untuk menambah visualisasi di bagian pilihan ketika ingin masuk atau daftar. Ilustrasi juga ditambahkan ketika melakukan verifikasi kode OTP dan ilustrasi sukses terhubung dengan e-commerce pilihan, penggunaan ilustrasi yang dipilih berguna untuk menambah kesan dalam menyampaikan informasi secara visual, serta membangun brand identity dengan pemilihan ilustrasi yang

sejenis. selain itu pada desain aplikasi ini juga telah menggunakan pop up di beberapa tugas yang harus dilaksanakan oleh user sebagai bentuk untuk menambah konversi yang mendorong *user* untuk melakukan tindakan tertentu seperti pada Gambar 8 dan Gambar 9 ketika pengguna harus menyetujui izin akses, melakukan verifikasi kode OTP, melakukan penambahan varian produk, serta pop up persetujuan untuk menghapus produk.

### 3.4 Evaluasi

Pada tahap pengujian menggunakan metode *System Usability Scale*, dilakukan terhadap 20 responden yang mengisi kuesioner secara valid. Berikut adalah tabel data untuk setiap jawaban pertanyaan.

**Tabel 1.** Hasil penilaian responden terhadap aplikasi instix

| No Reponden Usia |     |                       | Skor Asli |    |    |    |    |    |    |    |    |     |
|------------------|-----|-----------------------|-----------|----|----|----|----|----|----|----|----|-----|
|                  |     |                       | Q1        | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
| 1                | R1  | Remaja (Usia 17 - 25) | 5         | 2  | 5  | 1  | 4  | 2  | 4  | 2  | 5  | 2   |
| 2                | R2  | Remaja (Usia 17 - 25) | 4         | 1  | 4  | 3  | 4  | 3  | 3  | 1  | 3  | 1   |
| 3                | R3  | Remaja (Usia 17 - 25) | 4         | 1  | 5  | 1  | 4  | 2  | 4  | 2  | 3  | 1   |
| 4                | R4  | Remaja (Usia 17 - 25) | 5         | 1  | 5  | 1  | 5  | 1  | 5  | 1  | 5  | 1   |
| 5                | R5  | Remaja (Usia 17 - 25) | 4         | 2  | 4  | 1  | 5  | 2  | 4  | 1  | 4  | 2   |
| 6                | R6  | Dewasa (Usia 26 - 64) | 5         | 2  | 5  | 2  | 5  | 2  | 5  | 2  | 4  | 2   |
| 7                | R7  | Remaja (Usia 17 - 25) | 4         | 2  | 4  | 1  | 4  | 2  | 3  | 1  | 5  | 1   |
| 8                | R8  | Dewasa (Usia 26 - 64) | 4         | 2  | 5  | 2  | 5  | 2  | 4  | 2  | 4  | 2   |
| 9                | R9  | Remaja (Usia 17 - 25) | 5         | 1  | 5  | 1  | 4  | 1  | 5  | 1  | 5  | 2   |
| 10               | R10 | Remaja (Usia 17 - 25) | 5         | 1  | 5  | 2  | 5  | 1  | 5  | 1  | 5  | 2   |
| 11               | R11 | Remaja (Usia 17 - 25) | 4         | 1  | 5  | 1  | 4  | 1  | 5  | 1  | 4  | 2   |
| 12               | R12 | Remaja (Usia 17 - 25) | 5         | 1  | 4  | 1  | 5  | 1  | 3  | 2  | 3  | 1   |
| 13               | R13 | Remaja (Usia 17 - 25) | 5         | 1  | 4  | 2  | 4  | 2  | 5  | 2  | 4  | 1   |
| 14               | R14 | Remaja (Usia 17 - 25) | 5         | 1  | 5  | 1  | 5  | 1  | 5  | 1  | 5  | 1   |
| 15               | R15 | Remaja (Usia 17 - 25) | 5         | 2  | 4  | 2  | 5  | 1  | 4  | 1  | 4  | 1   |
| 16               | R16 | Remaja (Usia 17 - 25) | 4         | 2  | 4  | 1  | 4  | 1  | 4  | 2  | 4  | 1   |
| 17               | R17 | Remaja (Usia 17 - 25) | 4         | 1  | 5  | 1  | 4  | 1  | 4  | 2  | 1  | 2   |
| 18               | R18 | Remaja (Usia 17 - 25) | 4         | 1  | 4  | 1  | 5  | 2  | 4  | 1  | 4  | 1   |
| 19               | R19 | Dewasa (Usia 26 - 64) | 5         | 2  | 4  | 1  | 5  | 1  | 4  | 1  | 5  | 1   |
| 20               | R20 | Dewasa (Usia 26 - 64) | 5         | 1  | 5  | 1  | 4  | 2  | 4  | 1  | 5  | 1   |

Setelah data terkumpul, langkah berikutnya adalah melakukan perhitungan skor SUS sesuai dengan aturan yang telah ditetapkan [8]. Berikut adalah aturan yang digunakan untuk menghitung skor pada kuesioner. Setiap pertanyaan dengan nomor ganjil, skornya dikurangi 1 (satu), sedangkan untuk pertanyaan dengan nomor genap, skor akhir diperoleh dengan mengurangkan nilai 5 (lima) dari skor pertanyaan tersebut. Dengan demikian, dapat dituliskan persamaan berikut:

$$N_{ganjil} = n \times (x - 1) \tag{1}$$



$$N_{\text{genap}} = n \times (5 - x) \quad (2)$$

$$\Sigma qx = N_1 + N_2 + N_3 + N_4 + N_5 \quad (3)$$

$$\Sigma x = \Sigma q \times 2,5 \quad (4)$$

$$\bar{x} = \Sigma x/n \quad (5)$$

keterangan:

- $N_{\text{ganjil}}$  : Jumlah skor untuk setiap jawaban pada nomor ganjil  
 $N_{\text{genap}}$  : Jumlah skor untuk setiap jawaban pada nomor genap  
 $n$  : Jumlah responden per jawaban pertanyaan  
 $x$  : Skor jawaban (lihat tabel 1)  $\Sigma qx$  : Jumlah skor tiap pertanyaan  
 $\Sigma x$  : Total skor SUS  
 $\Sigma q$  : Total skor pertanyaan  
 $\bar{x}$  : rata - rata skor SUS

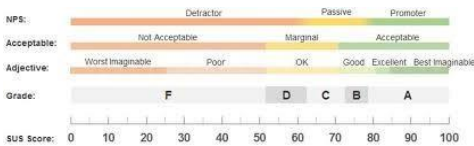
Berdasarkan evaluasi yang dilakukan oleh responden seperti yang tercatat dalam Tabel 1, akan dijalankan proses perhitungan SUS dengan menggunakan formula yang telah ditetapkan sebelumnya. Hasil dari proses ini akan menghasilkan total penilaian, yang kemudian akan dicantumkan dalam Tabel 2.

**Tabel 2.** Hasil rata rata skor SUS

| No | Rep. | Usia                  | Skor Hasil Hitung |    |    |    |    |    |    |    |    |     | Jum. | Nilai<br>(J. x 2.5) |
|----|------|-----------------------|-------------------|----|----|----|----|----|----|----|----|-----|------|---------------------|
|    |      |                       | Q1                | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |      |                     |
| 1  | R1   | Remaja (Usia 17 - 25) | 4                 | 3  | 4  | 4  | 3  | 3  | 3  | 3  | 4  | 3   | 34   | 85                  |
| 2  | R2   | Remaja (Usia 17 - 25) | 3                 | 4  | 3  | 2  | 3  | 2  | 2  | 4  | 2  | 4   | 29   | 73                  |
| 3  | R3   | Remaja (Usia 17 - 25) | 3                 | 4  | 4  | 4  | 3  | 3  | 3  | 3  | 2  | 4   | 33   | 83                  |
| 4  | R4   | Remaja (Usia 17 - 25) | 4                 | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4   | 40   | 100                 |
| 5  | R5   | Remaja (Usia 17 - 25) | 3                 | 3  | 3  | 4  | 4  | 3  | 3  | 4  | 3  | 3   | 33   | 83                  |
| 6  | R6   | Dewasa (Usia 26 - 64) | 4                 | 3  | 4  | 3  | 4  | 3  | 4  | 3  | 3  | 3   | 34   | 85                  |
| 7  | R7   | Remaja (Usia 17 - 25) | 3                 | 3  | 3  | 4  | 3  | 3  | 2  | 4  | 4  | 4   | 33   | 83                  |
| 8  | R8   | Dewasa (Usia 26 - 64) | 3                 | 3  | 4  | 3  | 4  | 3  | 3  | 3  | 3  | 3   | 32   | 80                  |
| 9  | R9   | Remaja (Usia 17 - 25) | 4                 | 4  | 4  | 4  | 3  | 4  | 4  | 4  | 4  | 3   | 38   | 95                  |
| 10 | R10  | Remaja (Usia 17 - 25) | 4                 | 4  | 4  | 3  | 4  | 4  | 4  | 4  | 4  | 3   | 38   | 95                  |
| 11 | R11  | Remaja (Usia 17 - 25) | 3                 | 4  | 4  | 4  | 3  | 4  | 4  | 4  | 3  | 3   | 36   | 90                  |
| 12 | R12  | Remaja (Usia 17 - 25) | 4                 | 4  | 3  | 4  | 4  | 4  | 2  | 3  | 2  | 4   | 34   | 85                  |
| 13 | R13  | Remaja (Usia 17 - 25) | 4                 | 4  | 3  | 3  | 3  | 3  | 4  | 3  | 3  | 4   | 34   | 85                  |
| 14 | R14  | Remaja (Usia 17 - 25) | 4                 | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4   | 40   | 100                 |
| 15 | R15  | Remaja (Usia 17 - 25) | 4                 | 3  | 3  | 3  | 4  | 4  | 3  | 4  | 3  | 4   | 35   | 88                  |
| 16 | R16  | Remaja (Usia 17 - 25) | 3                 | 3  | 3  | 4  | 3  | 4  | 3  | 3  | 3  | 4   | 33   | 83                  |
| 17 | R17  | Remaja (Usia 17 - 25) | 3                 | 4  | 4  | 4  | 3  | 4  | 3  | 3  | 0  | 3   | 31   | 78                  |
| 18 | R18  | Remaja (Usia 17 - 25) | 3                 | 4  | 3  | 4  | 4  | 3  | 3  | 4  | 3  | 4   | 35   | 88                  |

| No                                  | Rep. | Usia                  | Skor Hasil Hitung |    |    |    |    |    |    |    | Jum.      | Nilai<br>(J. x 2.5) |    |     |
|-------------------------------------|------|-----------------------|-------------------|----|----|----|----|----|----|----|-----------|---------------------|----|-----|
|                                     |      |                       | Q1                | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 |           |                     | Q9 | Q10 |
| 19                                  | R19  | Dewasa (Usia 26 - 64) | 4                 | 3  | 3  | 4  | 4  | 4  | 3  | 4  | 4         | 4                   | 37 | 93  |
| 20                                  | R20  | Dewasa (Usia 26 - 64) | 4                 | 4  | 4  | 4  | 3  | 3  | 3  | 4  | 4         | 4                   | 37 | 93  |
| <b>Skor Rata-rata (Hasil Akhir)</b> |      |                       |                   |    |    |    |    |    |    |    | <b>87</b> |                     |    |     |

kualitas antarmuka pengguna dari aplikasi yang diuji, menggunakan beberapa kriteria antara lain NPS (Net Promoter Score), Acceptability Scale, Adjective Ratings, dan Grade seperti yang ditunjukkan dalam Gambar 10.[9]



**Gambar 10.** Kategori Pengukuran Metode SUS

tabel 2 menampilkan hasil skor rata-rata sebesar 87, dengan detail sebagai berikut:

- a. Kategori NPS, menunjukkan skala promotor, sehingga pengguna kemungkinan besar merekomendasikan produk tersebut kepada orang lain.
- a. Untuk Kategori Acceptability, tingkatnya diterima, yang menunjukkan bahwa pengguna merasa aplikasi tersebut dapat diterima dan memenuhi kebutuhan mereka.
- b. Pada kategori Adjective, peringkatnya best imaginable, menunjukkan bahwa aplikasi memiliki kekuatan yang signifikan, solusi efektif dan sudah memenuhi kebutuhan pengguna dengan baik.
- c. Secara keseluruhan, aplikasi mendapat Grade A, menandakan bahwa pengguna merasa aplikasi mudah digunakan dan memenuhi kebutuhan mereka dengan sangat baik.

#### 4. Kesimpulan

Berdasarkan evaluasi menggunakan metode Sistem Usability Scale (SUS), aplikasi Instix memperoleh skor tinggi sebesar 87 dengan kategori "best imaginable", menunjukkan bahwa desain UI/UXnya telah memenuhi standar kualitas yang tinggi. Hasil tersebut menunjukkan bahwa Instix berhasil memberikan pengalaman yang optimal bagi pengguna, dengan tingkat kepuasan yang tinggi terhadap kemudahan penggunaan, penerimaan, dan kepuasan pengguna. Secara keseluruhan, Instix mendapat Grade A, menegaskan bahwa aplikasi ini tidak hanya mudah digunakan, tetapi juga mampu memenuhi kebutuhan pengguna dengan baik, memperkuat posisinya sebagai solusi yang efektif dan terpercaya dalam pengelolaan produk bisnis di berbagai platform.

#### Daftar Pustaka

- [1] E. F. Santika, "ECDB: Proyeksi Pertumbuhan E-Commerce Indonesia Tertinggi Sedunia pada 2024," *Katadata.co.id*, Apr. 29, 2024. <https://databoks.katadata.co.id/datapublish/2024/04/29/ecdb-proyeksi-pertumbuhan-e-commerce-indonesia-tertinggi-sedunia-pada-2024> (accessed May 06, 2024).
- [2] M. Ibrahim, "Sebanyak 21,8 Juta UMKM Sudah Go Digital, Ini Buktinya | Infobanknews," *Infobanknews*, Mar. 07, 2023. <https://infobanknews.com/sebanyak-218-juta-umkm-sudah-go-digital-ini-buktinya/> (accessed May 06, 2024).

- [3] R. Varela, A. F. Araújo, G. G. Vieira, V. K. Manupati, and K. Manoj, "Integrated Framework based on Critical Success Factors for E-Commerce," *Journal of information systems engineering & management*, vol. 2, no. 1, Jan. 2017, doi: <https://doi.org/10.20897/jisem.201704>.
- [4] Ependi, U., Putra, A., & Panjaitan, F, "Evaluasi Tingkat Kebergunaan Aplikasi Administrasi Penduduk Menggunakan Teknik System Usability Scale". *Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 63-76, 2019.
- [5] Pradea Wulandari, "Perancangan User Interface Aplikasi Rumah Sehat Retno Berbasis Smartphone Dan Evaluasi System Usability Scale (SUS)," *Jurnal Ilmu Kesehatan Bhakti Setya Medika/Jurnal ilmu kesehatan Bhakti Setya Medika*, vol. 6, no. 1, pp. 21–29, Sep. 2021, doi: <https://doi.org/10.56727/bsm.v6i1.50>.
- [6] J. Nielsen, "How Many Test Users in a Usability Study?" *Nielsen Norman Group*, Jun. 04, 2012. <https://www.nngroup.com/articles/how-many-test-users/> (accessed May 06, 2024).
- [7] A. P. Sukma, R. Yusuf, and R. H. Dai, "Analisis Pengukuran Usability Sistem Informasi Manajemen Baznas (SIMBA) Menggunakan Metode System Usability Scale (SUS)," *Diffusion Journal of System Information Technology*, vol. 3, no. 2, pp. 224-231, Jul. 2023.
- [8] A. P. Kurniawan, R. Roedavan, A. Putra, dan S. Apriliani, "Pengukuran Antarmuka Pengguna 3D Menggunakan Metode System Usability Scale pada Aplikasi Simulasi Videography Virtual Reality," *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, vol. 5, no. 2, pp. 77-87, 2023. [Online]. Available: <https://doi.org/10.35746/jtim.v5i2.340>
- [9] Kesuma, D. P. "Penggunaan Metode System Usability Scale untuk Mengukur Aspek Usability pada Media Pembelajaran Daring di Universitas XYZ," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, pp. 1615-1626, Sept. 2021.

# Optimasi Hyperparameter Algoritma Support Vector Machine dalam Klasifikasi Penyakit $\beta$ -Thalassemia

I Nyoman Adi Mahendra Putra<sup>a1</sup>, Cokorda Pramatha<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>henpra@gmail.com  
<sup>2</sup>cokorda@unud.ac.id

## Abstract

*Beta-thalassemia, a type of thalassemia disease caused by genetic variations, forces sufferers to receive regular blood transfusions for survival. Therefore, classification of this disease is important to reduce the number of births of beta-thalassemia patients in the future. 5066 beta-thalassemia carrier patient data from the Punjab Thalassemia Prevention and Program (PTPP) case study was taken as the source in this study which was accessed through the github.com website. Preprocessing is done for class alignment to avoid data imbalance, feature selection to streamline model performance, and normalization of feature values to a certain scale. In this research, the main focus is on applying the performance of the Support Vector Machine (SVM) algorithm to obtain classification results. Before entering the final model, hyperparameter tuning is required to obtain suitable parameter values to be entered into the model. hyperparameter tuning that will be carried out include "C" (regularization parameter), "kernel" (kernel type), "gamma" (kernel parameter for non-linear kernels), and "degree" (polynomial degree for polynomial kernels), carried out before the model is evaluated. The accuracy results were evaluated using confusion matrix, resulting in precision of 99.81%, recall of 99.62%, f1-score of 99.71%, and accuracy of 99.71% after hyperparameter tuning where the best parameters are "C": 1, "gamma": 100, "kernel": "rbf" with an average test score of 0.993494149.*

**Keywords:** Beta-thalassemia, Classification, Support Vector Machine, Hyperparameter Tuning, Confusion Matrix

## 1. Pendahuluan

Thalassemia adalah sebuah kelainan genetik yang disebabkan oleh kurangnya bahkan tidak ada pembentukan satu atau lebih rantai globin yang membentuk hemoglobin (Hb) [1]. Data Kementerian Kesehatan RI menunjukkan bahwa sekitar 2500 bayi Indonesia terlahir dengan thalassemia mayor atau pada tingkat thalassemia yang berbahaya setiap tahunnya [2]. Beta-thalassemia merupakan salah satu jenis penyakit thalassemia yang disebabkan oleh pengaruh sejumlah variasi genetik pada berbagai tahap produksi globin beta, seperti proses transkripsi, translasi, atau stabilitas produk beta-globin [3]. Dampak dari beta-thalassemia adalah terbentuknya hemoglobin yang cacat, yang lebih mungkin untuk mengalami kerusakan dan terurai. Kerusakan pada hemoglobin menyebabkan individu yang menderita beta-thalassemia harus secara rutin menerima transfusi darah untuk keberlangsungan hidupnya [4]. Oleh karena itu, pentingnya mengklasifikasikan jenis penyakit beta-thalassemia untuk mengurangi jumlah kelahiran penderita beta-thalassemia kedepannya. Klasifikasi merupakan proses penting dalam analisis data yang bertujuan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan fitur-fitur yang dimilikinya. Dalam klasifikasi data terdapat banyak algoritma yang sering digunakan untuk mendapatkan hasil klasifikasi yang baik salah satunya algoritma Support Vector Machine (SVM). SVM memiliki keunggulan dalam kemampuannya untuk menentukan superclass yang berbeda sehingga dapat memaksimalkan margin antara dua kelas yang berbeda [5]. Terdapat banyak penelitian yang sudah menerapkan algoritma Support Vector Machine (SVM) ke dalam berbagai kasus terutama penyakit. Pada penelitian sebelumnya algoritma Support Vector Machine (SVM) diterapkan untuk mengklasifikasikan kategori obat dengan hasil

akurasi terbaik didapatkan melalui penggunaan kernel "linear" dan "polinomial" sebesar 95,0% [6]. Adapun penelitian lain yang juga menggunakan algoritma Support Vector Machine (SVM) dalam medis yaitu memprediksi penyakit stroke. Permasalahan dataset yang digunakan tidak seimbang sehingga penulis menggunakan teknik oversampling untuk menyamakan distribusi kelas. Pada penelitian tersebut, dua hyperparameter dari Support Vector Machine (SVM) yakni "C" dan "Gamma" dilakukan tuning untuk mendapatkan hasil akurasi yang baik. Dalam uji parameter yang dilakukan, didapatkan hasil terbaiknya yaitu akurasi sebesar 96% dengan hyperparameter tuning dan 77% sebelum hyperparameter tuning [7]. Dalam upaya mengklasifikasikan penyakit beta-thalassemia, penting untuk menggunakan algoritma yang efektif guna memastikan akurasi yang optimal. Penelitian ini bertujuan untuk memanfaatkan performa algoritma Support Vector Machine (SVM) untuk mendapatkan model terbaik dalam mengklasifikasikan penyakit beta-thalassemia. Dengan melakukan hyperparameter tuning yang diterapkan pada penggunaan algoritma Support Vector Machine (SVM), diharapkan mampu mengoptimalkan hasil akurasi akhir yang baik. Oleh karena itu, diharapkan bahwa hasil dari penelitian ini akan memberikan manfaat bagi masyarakat luas serta menjadi referensi penelitian lebih lanjut kedepannya.

## 2. Metode Penelitian

### 2.1. Pengumpulan Data

Data yang diambil dalam penelitian ini berasal dari sumber sekunder, yakni pasien yang merupakan pembawa beta-thalassemia, yang diperoleh dari studi kasus Punjab Thalassemia Prevention and Programme (PTPP) melalui situs github.com. Data disimpan dalam bentuk file ".csv" dengan 5066 pasien yang telah melakukan screening untuk beta-thalassemia pada tahun 2019. Dari jumlah tersebut, 2594 pasien terbukti sebagai pembawa beta-thalassemia, sementara 2472 normal. Fitur yang terdapat dalam dataset dapat dilihat pada Tabel 1.

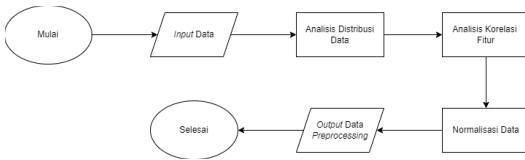
Tabel 1. Fitur Dataset

| No | Fitur | Tipe    | Keterangan  |
|----|-------|---------|---|
| 1  | Age   | Numerik | Umur pasien, yang terdiri dari angka 0 untuk pasien dewasa dan angka 1 untuk pasien anak-anak                     |
| 2  | Sex   | Numerik | Jenis kelamin pasien, yang terdiri dari angka 0 untuk pasien laki-laki dan angka 1 untuk pasien perempuan         |
| 3  | WBC   | Numerik | White blood count (WBC) yang merupakan jumlah sel darah putih pasien  |
| 4  | RBC   | Numerik | Red blood cell (RBC) yang merupakan jumlah sel darah merah pasien yang normalnya sekitar                          |
| 5  | HGB   | Numerik | Jumlah hemoglobin (HGB) pasien  |
| 6  | HCT   | Numerik | Jumlah hematokrit (HCT) pasien  |
| 7  | MCH   | Numerik | Mean corpuscular hemoglobin (MCH) yang merupakan rata-rata sel darah merah pasien                                 |
| 8  | MCV   | Numerik | Mean corpuscular volume (MCV) yang merupakan rata-rata volume sel darah pasien                                    |
| 9  | MCHC  | Numerik | Mean corpuscular hemoglobin concentration (MCHC) yang merupakan rata-rata konsentrasi hemoglobin sel darah pasien |
| 10 | RDW   | Numerik | Red cell distribution width (RDW) yang merupakan lebarnya distribusi sel darah merah pasien                       |
| 11 | PLT   | Numerik | Platelet (PLT) yang merupakan kadar trombosit dari pasien   |

| No | Fitur | Tipe    | Keterangan  |
|----|-------|---------|---|
| 12 | Class | Numerik | Terdiri dari dua buah kelas yaitu 0 untuk pasien normal dan 1 untuk pasien dengan pembawa penyakit beta-thalassemia |

## 2.2. Preprocessing Data

Tahapan preprocessing data merupakan langkah utama yang dilakukan untuk melakukan pengolahan data agar dapat memaksimalkan proses tahap selanjutnya [8]. Adapun tahapan dari preprocessing data yang dilakukan pada penelitian ini ditunjukkan oleh Gambar 1.

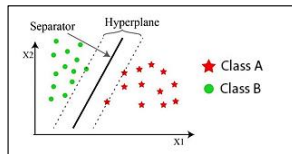


Gambar 1. Alur Tahapan Preprocessing Data

Dalam preprocessing data, tahapan pertama yang dilakukan adalah menganalisis distribusi data. Tahapan tersebut merupakan proses menganalisa seberapa seimbang kelas pada dataset yang digunakan untuk meminimalisir terjadinya kesetimpangan data antar kelas [9]. Tahapan selanjutnya adalah menganalisis korelasi fitur yang bisa dilakukan dengan mengamati korelasi setiap fitur melalui matrix correlation. Tahapan analisis korelasi fitur dapat menentukan fitur yang layak untuk dilakukan proses selanjutnya. Hal tersebut dapat meminimalisir kinerja dari model yang dibangun terhadap dataset yang cukup besar. Langkah terakhir pada preprocessing data yakni melakukan normalisasi data, dimana nilai data disamakan pada skala tertentu untuk dapat menghasilkan nilai akhir yang lebih optimal saat membangun pemodelan.

## 2.3. Membangun Model

Penelitian ini menggunakan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi. SVM bertujuan untuk menemukan hyperplane atau sebuah bidang pemisah dengan memaksimalkan margin antara kelas yang berbeda dalam data [10]. SVM memiliki keunggulan dalam menciptakan hyperlane yang tidak linear, yang dapat dimanfaatkan untuk meningkatkan jarak margin antara kelas [11]. SVM bekerja dengan memetakan data ke dalam ruang fitur berdimensi tinggi untuk mengategorikan titik-titik data. Meskipun data tidak dapat dipisahkan secara linear, pembatas atau pemisah antara kategori dapat ditemukan, seperti yang ditunjukkan dalam Gambar 2 [7].



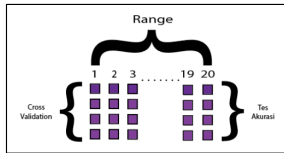
Gambar 2. Ilustrasi algoritma Support Vector Machine (SVM)

Pembangunan model dilakukan dengan menggunakan data yang sudah melalui tahap preprocessing. Data yang digunakan untuk melakukan pemodelan akan dibagi menjadi 80%

untuk data latih dan 20% untuk data uji. Pemodelan pada penelitian ini menggunakan algoritma Support Vector Machine (SVM) dengan mengimplementasikan bahasa pemrograman Python, serta untuk mendukung penerapan algoritma SVM tersebut, penelitian ini memanfaatkan library dari sklearn.

#### 2.4. Hyperparameter Tuning

Hyperparameter tuning adalah tahapan penyesuaian parameter untuk mendapatkan nilai terbaik. Tahapan tersebut diperlukan dalam penelitian ini untuk memaksimalkan akurasi model yang sudah dibangun sebelumnya. Pada tahap hyperparameter tuning, penelitian ini memanfaatkan library dari GridSearchCV. GridSearchCV merupakan sebuah bagian dalam modul scikit-learn yang memungkinkan validasi untuk beberapa model secara efisien dengan memetakan berbagai nilai hyperparameter ke dalam suatu grid dan mengevaluasi kinerja model menggunakan teknik cross-validation [12]. Cross-validation adalah sebuah teknik pengembangan dari validasi split di mana validasi dilakukan dengan membagi data menjadi beberapa subset, kemudian dilakukan iterasi pengujian model dengan menggunakan setiap subset sebagai data uji secara bergantian seperti pada Gambar 3 [13].



Gambar 3. Cross-validation GridSearchCV

Dalam algoritma Support Vector Machine SVM, parameter-parameter yang disesuaikan (tuned) meliputi "C" (parameter regularisasi), "kernel" (jenis kernel), "gamma" (parameter kernel untuk kernel non-linear), dan "degree" (derajat polinomial untuk kernel polinomial). Pada penelitian ini, peneliti menggunakan parameter "kernel" untuk menjadi acuan pencarian parameter terbaik. Hal tersebut dikarenakan didalam library sklearn, setiap "kernel" memiliki parameter yang mempengaruhi untuk pembuatan model yang mempengaruhi hasil klasifikasi model. Parameter default atau bawaan dari algoritma SVM yaitu "C":1, "kernel":'rbf', "gamma":'scale'. Penggunaan GridSearchCV dalam penelitian ini bertujuan untuk menemukan kombinasi nilai-nilai parameter yang menghasilkan model SVM dengan performa optimal untuk tugas klasifikasi yang diberikan [14].

#### 2.5. Analisis Hasil Pemodelan

Evaluasi dilakukan dengan membandingkan hasil tuning hyperparameter dengan hasil tanpa melakukan hyperparameter tuning ke dalam model yang sudah dibuat. Pada tahap evaluasi, penelitian ini menggunakan confusion matrix untuk memperoleh beberapa hasil klasifikasi yakni berupa nilai akurasi, presisi, recall, dan f1-score. Confusion matrix akan menggunakan beberapa nilai untuk mendapatkan hasil klasifikasi yaitu true negative (TN) dan true positive (TP) untuk menyatakan bahwa model berhasil melakukan klasifikasi, sementara false negative (FN) dan false positive (FP) untuk menyatakan bahwa model salah dalam melakukan klasifikasi. Untuk menghitung hasil klasifikasi yang akan diperoleh, dapat menggunakan rumus berikut [15].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

### 3. Hasil dan Pembahasan

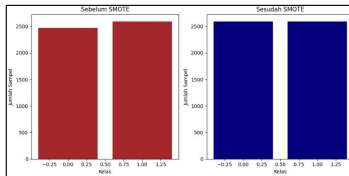
#### 3.1. Preprocessing Data

##### a. Analisis Distribusi Data

Tahapan pertama dalam preprocessing data yakni melakukan analisis distribusi data untuk mengetahui seberapa seimbang kelas dalam dataset yang akan digunakan. Dalam dataset yang digunakan, dari total 5066 data, terdapat selisih 122 data antara dua kelas. Nilai tersebut memiliki perbedaan yang tidak terlalu signifikan, namun pada penelitian ini akan tetap menggunakan metode Synthetic Minority Oversampling Technique (SMOTE) untuk mendapatkan hasil akurasi model yang optimal. Penggunaan metode SMOTE dapat dilihat pada Gambar 4 dengan memanfaatkan library dari imblearn. Sementara hasil perbandingan distribusi data sebelum dan sesudah oversampling dapat dilihat pada Gambar 5.

```
[1]: from imblearn.over_sampling import SMOTE
sm = SMOTE(random_state=0)
#sampling smote
x_sampling,y_sampling = sm.fit_resample(x,y)
```

Gambar 4. Implementasi Metode SMOTE pada Oversampling

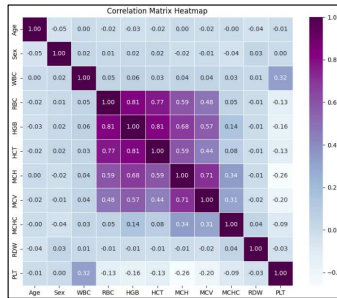


Gambar 5. Hasil Perbandingan Distribusi Data Sebelum dan Sesudah Oversampling

##### b. Analisis Korelasi Fitur

Menganalisis korelasi fitur pada dataset yang sudah melakukan oversampling digunakan untuk melihat seberapa erat hubungan antara setiap pasangan fitur. Analisis korelasi fitur yang dilakukan pada penelitian ini menggunakan correlation matrix heatmap dengan memanfaatkan library dari seaborn, pandas, dan pyplot. Correlation matrix heatmap sebagai analisis korelasi fitur dapat dilihat pada Gambar 6.





Gambar 6. Correlation Matrix Heatmap Terhadap Fitur Dataset

Terlihat bahwa beberapa fitur tidak terlalu berkorelasi dengan fitur lainnya seperti "Age", "Sex", "WBC", "RDW", dan "PLT" maka daripada itu penelitian ini memutuskan untuk membuang fitur tersebut dikarenakan untuk meminimalisir kinerja mesin saat digunakan untuk dataset yang besar. Implementasi pembuangan fitur pada dataset dapat dilihat pada Gambar 7.

```
[26] x = df_scaled.drop(['Class', 'Age', 'Sex', 'WBC', 'RDW', 'PLT'],axis=1)
     y = df_scaled['Class']
```

Gambar 7. Implementasi Pembuangan Fitur pada Dataset

c. Normalisasi Data

Tahapan normalisasi data dilakukan untuk menyelaraskan nilai-nilai dari fitur, dengan menggunakan dataset yang sudah melalui tahap seleksi fitur. Dalam penelitian ini, tahapan normalisasi menggunakan fungsi "minmax" seperti yang ditunjukkan pada Gambar 8.

```
[305] def minmax(df_input):
      list_fitur = df_input.columns[1:-1]
      for fitur in list_fitur:
          max = df_input[fitur].max()
          min = df_input[fitur].min()
          df_input[fitur] = (df_input[fitur]-min)/(max-min)
      return df_input
```

Gambar 8. Fungsi "minmax" pada Normalisasi Data

3.2. Model dan Hyperparameter Tuning

Pada penelitian ini, pembangunan model Support Vector Machine (SVM) dibangun dengan memanfaatkan library dari sklearn yang ditunjukkan pada Gambar 9. Tahapan pembuatan model, dilakukan dengan memanfaatkan dataset yang sudah melalui proses oversampling. Dataset yang digunakan dibagi menjadi 80% dataset uji dan 20% dataset latih.

```
(18) from sklearn.model_selection import train_test_split
    x_train,x_test,y_train,y_test=train_test_split(x_sampling,y_sampling, test_size=0.20, random_state=0)

    from sklearn.svm import SVC
    # Proses Algoritma
    svm = SVC()
    svm.fit(x_train,y_train)
```

**Gambar 9.** Implementasi Model Algoritma Support Vector Machine (SVM)

Tahapan hyperparameter tuning pada penelitian ini akan menyesuaikan hyperparameter dari model yang dibangun yaitu Support Vector Machine (SVM) berdasarkan parameter kernel. Penelitian ini memanfaatkan library sklearn yaitu GridSearchCV untuk menentukan hyperparameter terbaik, dimana nilai parameter berdasarkan kernel yang diuji ditunjukkan pada Tabel 2. Implementasi tahapan tuning ditunjukkan pada Gambar 10.

**Tabel 2.** Nilai Parameter untuk Hyperparameter Tuning Berdasarkan Kernel

| No | Kernel   | Parameter dan Nilai Uji   |
|----|----------|---|
| 1  | 'linear' | 'C': [0.001, 0.01, 0.1, 1, 10, 100, 1000]   |
| 2  | 'rbf'    | 'C': [0.001, 0.01, 0.1, 1, 10, 100, 1000]<br>'gamma': [0.001, 0.01, 0.1, 1, 10, 100]                        |
| 3  | 'poly'   | 'C': [0.001, 0.01, 0.1, 1, 10, 100, 1000] 'gamma': [0.001, 0.01, 0.1, 1, 10, 100]<br>'degree': [2, 3, 4, 5] |

```
from sklearn.model_selection import GridSearchCV

# hyperparameter kernel yang akan diuji
parameters = {
    # masukan setiap nilai parameter kernel yang ingin diuji
}

svm_model = svm.SVC()
svm_grid = GridSearchCV(estimator=svm_model, param_grid=parameters, cv=3, verbose=2, n_jobs=-1)
svm_grid.fit(x_train, y_train)
```

**Gambar 10.** Implementasi GridSearchCV untuk Hyperparameter Tuning

Dalam proses hyperparameter tuning pada setiap nilai parameter uji yang ditunjukkan pada Tabel 2, didapatkan hasil lima kombinasi parameter terbaik setiap kernel ditunjukkan pada Tabel 3 untuk kernel "linear", Tabel 4 untuk kernel "rbf", dan Tabel 5 untuk kernel "poly".

**Tabel 3.** Hasil Parameter Terbaik menggunakan Kernel "linear"

| mean_fit_time | mean_score_time | params                          | mean_test_score |
|---------------|-----------------|---------------------------------|-----------------|
| 0.172570388   | 0.032081842     | {'C': 100, 'kernel': 'linear'}  | 0.946747214     |
| 0.783475955   | 0.046331565     | {'C': 1000, 'kernel': 'linear'} | 0.946747214     |
| 0.098690748   | 0.032089551     | {'C': 10, 'kernel': 'linear'}   | 0.945783126     |
| 0.081038952   | 0.037542741     | {'C': 1, 'kernel': 'linear'}    | 0.940963732     |
| 0.105776548   | 0.04589661      | {'C': 0.1, 'kernel': 'linear'}  | 0.92722792      |

**Tabel 4.** Hasil Parameter Terbaik menggunakan Kernel "rbf"

| mean_fit_time | mean_score_time | params                                     | mean_test_score |
|---------------|-----------------|--|-----------------|
| 0.343053341   | 0.209666332     | {'C': 1, 'gamma': 100, 'kernel': 'rbf'}    | 0.993494149     |
| 0.050074657   | 0.021288713     | {'C': 100, 'gamma': 10, 'kernel': 'rbf'}   | 0.993493453     |
| 0.052993377   | 0.02069219      | {'C': 1000, 'gamma': 10, 'kernel': 'rbf'}  | 0.99180717      |
| 0.043960174   | 0.025547822     | {'C': 10, 'gamma': 10, 'kernel': 'rbf'}    | 0.991806996     |
| 0.141822259   | 0.08481582      | {'C': 1000, 'gamma': 100, 'kernel': 'rbf'} | 0.991566496     |

**Tabel 5.** Hasil Parameter Terbaik menggunakan Kernel "poly"

| mean_fit_time | mean_score_time | params  | mean_test_score |
|---------------|-----------------|---|-----------------|
| 108.3636426   | 0.011723359     | {'C': 1000, 'degree': 5, 'gamma': 10, 'kernel': 'poly'} | 0.99060206      |
| 108.7903892   | 0.016993364     | {'C': 10, 'degree': 5, 'gamma': 10, 'kernel': 'poly'}   | 0.99060206      |
| 108.7040025   | 0.017054876     | {'C': 100, 'degree': 5, 'gamma': 10, 'kernel': 'poly'}  | 0.99060206      |
| 84.76238942   | 0.011989037     | {'C': 1, 'degree': 5, 'gamma': 10, 'kernel': 'poly'}    | 0.990361735     |
| 4.198003848   | 0.016676346     | {'C': 1000, 'degree': 5, 'gamma': 1, 'kernel': 'poly'}  | 0.989878994     |

Pada hyperparameter tuning, didapatkan parameter terbaik yaitu "'C': 1, 'gamma': 100, 'kernel': 'rbf'" dengan skor uji rata-rata 0.993494149. Parameter terbaik tersebut akan dimasukkan kedalam model untuk mencari hasil klasifikasi setelah mengalami proses hyperparameter tuning. Implementasi ke dalam pemodelan, ditunjukkan pada Gambar 11.

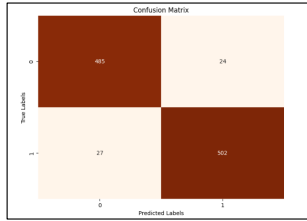
```
# Masukan hyperparameter terbaik
best_modelsvm = SVC(C=1, gamma = 100, kernel='rbf')
best_modelsvm.fit(x_train, y_train)

y_pred = best_modelsvm.predict(x_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1_score = f1_score(y_test, y_pred, average='macro')
```

**Gambar 11.** Implementasi Pemodelan Support Vector Machine dan Hasil Klasifikasi

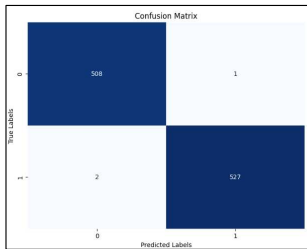
### 3.3. Evaluasi Model

Hasil klasifikasi model sebelum hyperparameter tuning dan sesudah melalui tahap hyperparameter tuning, akan dievaluasi dengan menggunakan confusion matrix. Untuk hasil klasifikasi tanpa melakukan hyperparameter tuning ditunjukkan pada Gambar 12.



**Gambar 12.** Hasil Pemodelan Tanpa Menggunakan Hyperparameter Tuning

Gambar 12 menunjukkan bahwa masih terdapat prediksi yang salah terhadap kelas pada dataset yang digunakan. Terdiri dari 24 data yang masih salah dalam prediksi kelas yang seharusnya bukan merupakan pembawa beta-thalassemia namun diprediksi pembawa beta-thalassemia, serta 27 data masih salah dalam prediksi kelas yang seharusnya merupakan pembawa beta-thalassemia namun diprediksi bukan pembawa thalassemia. Untuk hasil klasifikasi disertai penggunaan hyperparameter tuning ditunjukkan pada Gambar 13.



**Gambar 13.** Hasil Pemodelan Disertai Penggunaan Hyperparameter Tuning

Gambar 13 menunjukkan bahwa masih terdapat prediksi yang salah terhadap kelas pada dataset yang digunakan, namun sudah berkurang jauh dari prediksi yang ditunjukkan pada Gambar 12. Terdiri dari 1 data yang masih salah dalam prediksi kelas yang seharusnya bukan merupakan pembawa beta-thalassemia namun diprediksi pembawa beta-thalassemia, serta 2 data masih salah dalam prediksi kelas yang seharusnya merupakan pembawa beta-thalassemia namun diprediksi bukan pembawa thalassemia. Peningkatan prediksi tersebut membuat perbedaan hasil klasifikasi yang ditunjukkan pada Tabel 6.

**Tabel 6.** Perbandingan Hasil Akhir Pemodelan

|                                      | Akurasi | Presisi | Recall | F1-score |
|--------------------------------------|---------|---------|--------|----------|
| <b>Sebelum Hyperparameter Tuning</b> | 95,08%  | 95,43%  | 94,89% | 95,16%   |

---

|  |        |        |        |        |
|--|--------|--------|--------|--------|
| <b>Setelah<br/>Hyperparameter<br/>Tuning</b> | 99,71% | 99,81% | 99,62% | 99,71% |
|--|--------|--------|--------|--------|

---

Berdasarkan Tabel 6, hasil akurasi, presisi, recall, f1-score yang diperoleh sebelum melakukan hyperparameter tuning secara berturut-turut yaitu 95,08%, 95,43%, 94,89%, dan 95,16%. Sementara hasil akurasi, presisi, recall, f1-score yang diperoleh setelah melakukan hyperparameter tuning secara berturut-turut yaitu 99,71%, 99,81%, 99,62%, dan 99,71%. Hal tersebut secara keseluruhan membuktikan bahwa hyperparameter tuning berpengaruh dalam pemodelan menggunakan algoritma Support Vector Machine (SVM) yang mengalami peningkatan hasil akurasi, presisi, recall, f1-score yang diperoleh.

#### 4. Kesimpulan

Penerapan algoritma Support Vector Machine (SVM) pada klasifikasi penyakit beta-thalassemia dapat disimpulkan bahwa optimasi hyperparameter dapat meningkatkan hasil klasifikasi yang diperoleh. Hyperparameter tuning yang dilakukan memanfaatkan GridSearchCV untuk menentukan parameter terbaik sesuai dengan nilai yang ingin diujikan. Hasil terbaik yang didapatkan setelah melakukan hyperparameter tuning yaitu nilai akurasi sebesar 99,71%, nilai presisi sebesar 99,81%, nilai recall sebesar 99,62%, dan nilai f1-score sebesar 99,71% dengan parameter terbaiknya yaitu "C": 1, 'gamma': 100, 'kernel': 'rbf' dengan skor uji rata-rata 0.993494149. Untuk hasil klasifikasi yang lebih baik, perlu diadakannya tahap hyperparameter tuning lebih lanjut untuk memaksimalkan kinerja dari model yang dibangun serta pengujian dengan algoritma lain untuk membandingkan performa berdasarkan hasil akhir klasifikasi.

#### Daftar Pustaka

- [1] D. Setiawan, H. Setiawan, and A. Nurmalasari, "Edukasi Penyakit Talasemia Pada Mahasiswa Stikes Muhammadiyah Ciamis," *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, vol. 6, no. 3, pp. 1098-1102, 2022.
- [2] R. Roslaeni, S. Ratunanda, and A. L. Susanti, "Skринing Talasemia Bagi Mahasiswa Fakultas Kedokteran Universitas Jenderal Achmad Yani," *Jurnal Abdimas Kartika Wijayakusuma*, vol. 5, no. 1, pp.168-173, 2024. doi: 10.26874/jakw.v5i1.369.
- [3] M. N. Praramdana, M. A. Rusydi, and M. Rizky, "Sebuah Tinjauan Pustaka: Penatalaksanaan Beta Talasemia," *Jurnal Medika Utama*, vol. 4, no. 2, pp. 3257-3264, 2023.
- [4] W. Chauhan, S. Shoab, R. Fatma, Z. Zaka-ur-Rab, and M. Afzal, "Beta-Thalassemia and the Advent of New Interventions Beyond Transfusion and Iron Chelation," *British Journal of Clinical Pharmacology*, vol. 88, no. 8, pp. 3610-3626, 2022. doi: 10.1111/bcp.15343.
- [5] D. S. Rahayu, Nursafika, J. Afifah, and S. Intan, "Classification of Diabetes Mellitus Using C4.5 Algorithm, Support Vector Machine (SVM) and Linear Regression," *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, vol. 1, no. 1, pp. 56-63, 2023.
- [6] Mardewi, N. Yarkuran, Sofyan, and F. Aziz, "Klasifikasi Kategori Obat Menggunakan Algoritma Support Vector Machine," *Journal Pharmacy and Application of Computer Sciences*, vol. 1, no. 1, pp. 27-32, 2023.
- [7] Yennimar, A. Rasid, and S. Kenedy, "Implementation of Support Vector Machine Algorithm with Hyper-Tuning Randomized Search In Stroke Prediction," *Journal of Information Systems and Computer Science Prima*, vol. 6, no. 2, pp. 61-65, 2023.
- [8] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naive Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820-826, 2021. doi: 10.29207/resti.v5i4.3146.
- [9] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 89-93, 2020. doi: 10.14710/jtsiskom.8.2.2020.89-93.

- [10] H. S. Wafa, A. Id Hadiana, and F. R. Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics And Digital Expert (INDEX)*, vol. 4, no. 1, pp. 40-45, 2022.
- [11] P. W. A. Wibawa and C. Pramarta, "Systematic Literature Review: Machine Learning Methods in Emotion Classification in Textual Data," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 3, pp. 425–433, 2023, doi: 10.32736/sisfokom.v12i3.1787.
- [12] Z. M. E. Darmawan and A. F. Dianta, "Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM," *Teknologi: Jurnal Ilmiah Sistem Informasi*, vol. 13, no. 1, pp. 8–15, 2023, doi: 10.26594/teknologi.v13i1.3098.
- [13] S. T. Kusuma and T. B. Sasongko, "Optimasi K-Nearest Neighbor dengan Grid Search CV pada Prediksi Kanker Paru-Paru," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 4, pp. 2162-2171, 2023.
- [14] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search," *Journal of Applied Informatics and Computing (JAIC)*, vol. 7, no. 1, pp. 10-15, 2023.
- [15] R. Yunita and M. Kamayani, "Perbandingan Algoritma SVM Dan Naive Bayes Pada Analisis Sentimen Kebijakan Penghapusan Kewajiban Skripsi," *Indonesian Journal of Computer Science*, vol. 12, no.5, pp. 2879-2890, 2023.

Halaman ini sengaja dibiarkan kosong

# Klasifikasi Berita Berdasarkan Kategori Menggunakan Multinomial Naïve Bayes dengan K-Cross Validation dan Seleksi Fitur Chi-Squared

Febrian Valentino Agape<sup>a1</sup>, Gst. Ayu Vida Matrika Giri<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>febrian.valentino1402@gmail.com  
<sup>2</sup>vida@unud.ac.id

## Abstract

*Classifying news articles based on categories is an important challenge in text analysis and natural language processing. Most categorization of online news articles is often done manually, making it a complex and time-consuming process. To address this issue, the development of an automatic system capable of classifying news articles into various categories such as technology, sports, and entertainment is needed. The system is built using an approach to classify news articles into several appropriate categories using the Naïve Bayes method with TF-IDF weighting and feature selection using Chi-Squared. The Naïve Bayes model training uses the reduced feature results of 10,000 features from 54,091 features. Evaluation results show that the Naïve Bayes approach is able to produce a news classification model with good accuracy, with accuracy, precision, recall, and f1-score values of 96%.*

**Keywords:** News Classification, Multinomial Naïve Bayes, Feature Weighting, TF-IDF (Term Frequency-Inverse Document Frequency), Text Analysis

## 1. Pendahuluan

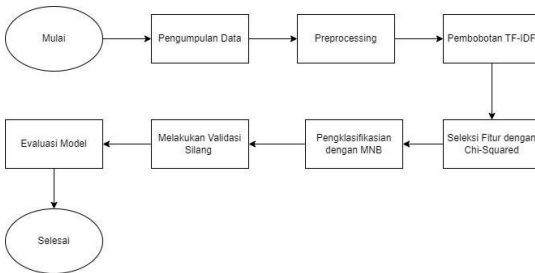
Di tengah banjir informasi yang kita hadapi hari ini, menimbulkan beragam topik berita yang bermunculan bersamaan. Berita mengandung elemen-elemen fakta dan pendapat yang penting untuk dikomunikasikan ke publik, namun tidak setiap elemen tersebut layak untuk dijadikan berita oleh media[1]. Seiring berjalannya waktu, media penyampaian berita telah berubah dari surat kabar, majalah, radio, dan televisi menjadi dominan di platform digital. Biasanya, konten berita di situs berita internet dikategorikan ke dalam berbagai segmen seperti politik, olahraga, ekonomi, hiburan, teknologi, kesehatan, dan sebagainya. Permasalahan timbul yaitu, Melimpahnya dokumen digital di internet dapat menjadi tantangan bagi masyarakat dalam mengakses informasi jika tidak ada sistem pengelolaan yang memadai. Cara umum untuk mengatur konten berita adalah dengan mengklasifikasikan setiap artikel berdasarkan kategori tertentu. Kategori ini bisa ditentukan berdasarkan situasi sosial yang berlaku atau sesuai dengan standar yang telah ditetapkan [2]. Kemampuan untuk mengklasifikasikan berita berdasarkan kategori menjadi sangat penting. Klasifikasi ini tidak hanya membantu dalam menyaring informasi yang relevan tetapi juga memudahkan pengguna dalam menemukan konten yang mereka cari. Teknik penambahan teks merupakan salah satu metode yang efektif untuk klasifikasi dokumen. Proses ini melibatkan penentuan kategori yang sesuai untuk dokumen berdasarkan isi teksnya [3]. Tujuan dari klasifikasi adalah untuk mengembangkan sebuah model atau fungsi yang dapat membedakan antara berbagai konsep atau kategori data, sehingga memungkinkan prediksi kategori untuk objek yang diberikan. Ini mengasumsikan adanya beragam kategori yang dapat diterapkan pada objek tersebut. Klasifikasi telah banyak dilakukan oleh para peneliti dengan menerapkan berbagai metode, salah satunya adalah Multinomial Naïve Bayes [4]. Terdapat cukup banyak penelitian terdahulu yang telah melakukan klasifikasi teks dengan metode ini. Metode Multinomial Naïve Bayes telah menunjukkan efektivitasnya dalam memberikan hasil yang memadai untuk klasifikasi teks [5]. Sebagai ilustrasi, penelitian yang dilakukan oleh Wayan Firdaus Mahmudy dan Agus



Wahyu Widodo menggunakan Naive Bayes Classifier yang belum dimodifikasi. Mereka mengevaluasi berbagai rasio pembagian data latih dan uji, yaitu 5:95, 10:90, 15:85, 20:80, 25:75, dan 30:70. Hasilnya, akurasi klasifikasi yang diperoleh secara bertahap adalah 54%, 65%, 65%, 69%, 71%, dan 76% [2]. Selain itu, dari hasil penelitian Bobby Suryo Prakoso dkk penelitian klasifikasi berita menggunakan metode Multinomial Naive Bayes Dengan Seleksi Fitur Dan Boosting dimana menghasilkan tingkat akurasi, recall, dan presisi sebesar 73,2%. Tidak ada perbedaan antara model yang lebih rinci dan model Naive Bayes Classifier, yang berarti keduanya memiliki performa evaluasi yang sama. Ini menandakan bahwa kedua model tersebut memiliki kemampuan prediksi yang sebanding [6]. Dalam penelitian ini melakukan klasifikasi teks diperlakukan ekstraksi fitur untuk mengubah format tekstual yang tidak terstruktur menjadi terstruktur sehingga dapat diproses oleh model machine learning seperti Multinomial Naive Bayes untuk mengklasifikasikan ke kelas yang telah ditentukan [7]. Pada penelitian ini data dibagi menjadi 80% data training dan 20% testing menggunakan Teknik ekstraksi fitur yaitu TF-IDF serta dilakukan optimisasi model dengan cara melakukan seleksi fitur untuk mencegah masalah overfitting dengan menggunakan metode Chi-squared yang membedakan dengan penelitian penelitian sebelumnya karena dilihat dari hasil akurasi menghasilkan persentase lebih tinggi karena adanya metode optimalisasi tersebut.

## 2. Metode Penelitian

Bagian ini akan menggambarkan secara umum tahapan yang akan dilakukan oleh peneliti dalam penelitian yang bertujuan agar mempermudah peneliti dalam melakukan percobaan serta pengkajian data. Alur metodologi penelitian dapat dilihat pada gambar 1.



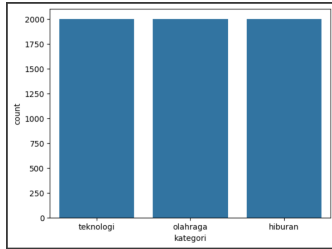
Gambar 1. Alur Metode Penelitian

Penelitian akan dimulai dengan mengumpulkan data teks berupa dataset yang bersumber dari repository Github iwanlaudin0101 yang berisi kumpulan berita yang sudah dilabeli kategorinya. Setelah itu, akan dilakukan tahap preprocessing untuk mempersiapkan data tersebut. Selanjutnya, data akan diolah dengan menggunakan metode pembobotan kata menggunakan Term Frequency Inverse Document Frequency (TF-IDF). Setelah mendapatkan hasil pembobotan kata, langkah selanjutnya adalah melakukan proses seleksi fitur untuk menghindari masalah proses klasifikasi seperti overfitting selanjutnya dilanjutkan dengan proses pengklasifikasian dengan model Multinomial Naive Bayes. Pada tahap akhir penelitian, dilakukan pengujian dan evaluasi terhadap kinerja metode yang digunakan.

### 2.1. Tahapan Pengumpulan Data

Pada proses pengumpulan data, data yang digunakan dalam penelitian ini merupakan data sekunder yang berasal dari repository GitHub dengan nama iwanlaudin0101. Dataset ini terdiri dari tiga variabel utama, yaitu sumber, kategori, dan berita. Namun, untuk keperluan penelitian ini, peneliti hanya akan menggunakan dua variabel, yaitu kategori dan berita. Dataset ini terdiri

dari lima kategori berita yang berbeda, namun pada penelitian ini akan berfokus pada tiga kategori utama, yaitu teknologi, olahraga, dan hiburan. Oleh karena itu, dua kategori lainnya, yaitu showbiz dan tajuk utama, akan dihilangkan dari dataset. Sebagai hasilnya, penelitian ini menggunakan data set yang terdiri dari 6000 record dimana distribusi tiap kategori dapat dilihat pada gambar 2.

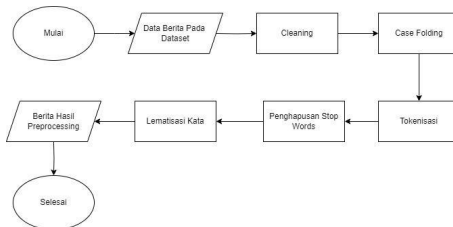


Gambar 2. Distribusi Data Per Kategori

Masing-masing kategori memiliki 2000 record dimana distribusi antara kategori teknologi, olahraga, dan hiburan adalah seimbang, dengan masing-masing kategori memiliki jumlah yang sama. Data ini akan dibagi menjadi data training dan data testing dengan perbandingan 80% : 20%. Oleh karena itu, sebanyak 4800 record akan digunakan sebagai data training, sedangkan 1200 record sisanya akan digunakan sebagai data testing.

## 2.2. Text Preprocessing

Text Preprocessing merupakan tahap awal dalam membangun sebuah model machine learning dalam text mining. Pada langkah ini, dilakukan pra-pemrosesan data teks yang telah dikumpulkan sebelumnya. Preprocessing teks adalah suatu proses untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur, atau dengan kata lain, mengubah teks menjadi indeks kata sesuai kebutuhan [8]. Tujuan dari proses ini adalah untuk mempersiapkan teks agar siap digunakan dan diolah. Pra-pemrosesan melibatkan serangkaian langkah, meliputi cleaning, case folding, tokenization, stopwords removal, dan Lemetazing. Dari tahapan preprocessing text ditujukan untuk mengurangi informasi yang tidak relevan atau tidak dibutuhkan dalam data tersebut dengan menghilangkan kata atau teks yang tidak perlu. Semua langkah ini bertujuan untuk mempermudah proses pembobotan. Ilustrasi alur dari tahap pra-pemrosesan dapat dilihat pada Gambar 3

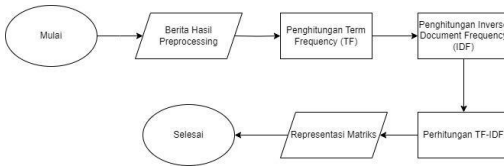


Gambar 3. Alur Text Preprocessing

Gambar 3 menunjukkan bahwa proses pra-pemrosesan meliputi beberapa tahapan. Pada tahap awal, pembersihan teks dilakukan untuk mengeliminasi karakter non-alfabetis seperti simbol, emotikon, dan angka. Selanjutnya, proses case folding diaplikasikan untuk menyeragamkan semua teks menjadi huruf kecil. Proses berikutnya adalah tokenisasi, yang membagi teks menjadi kata-kata individu berdasarkan spasi. Kemudian, kata-kata yang tidak memberikan informasi signifikan, atau stop words, dihilangkan. Tahap akhir adalah lemmatisasi, di mana kata-kata dikembalikan ke bentuk dasar mereka, memfasilitasi analisis teks yang lebih konsisten dan umum.

### 2.3. Ekstraksi Fitur TF-IDF

Setelah proses pra-pemrosesan selesai, langkah selanjutnya adalah tahap pembobotan atau ekstraksi fitur. Dalam analisis teks, konversi kata menjadi format numerik adalah esensial karena komputer hanya mampu mengolah data numerik. Teknik TF-IDF (Term Frequency-Inverse Document Frequency) merupakan salah satu metode ekstraksi fitur yang sering digunakan. TF-IDF berfungsi untuk mengevaluasi seberapa signifikan sebuah kata dalam dokumen atau keseluruhan korpus. Metode ini memanfaatkan kata-kata yang sudah diproses pada tahap pra-pemrosesan sebagai masukan, dan proses ekstraksi fitur TF-IDF ini diilustrasikan pada gambar 4.



Gambar 4. Alur Ekstraksi Fitur TF-IDF

Berdasarkan gambar 4, proses ekstraksi fitur menggunakan metode TF-IDF melibatkan dua tahapan, yaitu term frequency (TF) dan Inverse Document Frequency (IDF). Dengan menggunakan TF-IDF, nilai penting atau bobot dari setiap kata dalam dokumen dapat ditentukan berdasarkan seberapa sering kata tersebut muncul dalam dokumen tersebut dan di seluruh korpus secara keseluruhan [9].

#### 2.3.1 Term Frequency (TF)

Tahap awal ini bertujuan untuk menghitung frekuensi kemunculan setiap kata (term) dalam setiap dokumen. Frekuensi ini diukur dengan membagi jumlah kemunculan kata tersebut dengan jumlah total kata dalam dokumen tersebut. Persamaan matematis untuk menghitung TF ini dapat dirumuskan sebagai berikut (1) [10]:

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata (t) dalam dokumen (d)}}{\text{Total jumlah kata dalam dokumen (d)}} \quad (1)$$

Dimana  $TF(d,t)$  adalah term frequency,  $t$  adalah Kata yang sedang dievaluasi,  $d$  adalah Dokumen tempat kata  $t$  muncul.

#### 2.3.2 Inverse Document Frequency (IDF)

Tahap kedua ini bertujuan untuk mengukur seberapa penting atau unik sebuah kata terhadap seluruh korpus atau kumpulan dokumen. Ini membantu mengurangi bobot kata yang muncul secara umum di seluruh dokumen. Penulisan matematis dari penghitungan TF ini diberikan dalam persamaan (2)[10].

$$IDF(t, D) = \log\left(\frac{\text{Total jumlah dokumen dalam korpus (D)}}{\text{Jumlah dokumen yang mengandung kata (t)}}\right) \quad (2)$$

Kemudian setelah mendapat kan sail IDF dengan persamaan (2) selanjutnya adalah mendapat kan jumlah keseluruhan TF-IDF dengan persamaan yang terdapat dalam persamaan (3) [10].

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

### 2.3.3. Seleksi Fitur Chi-Squared

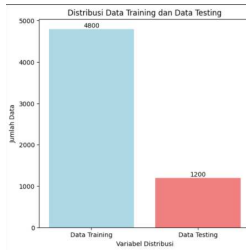
Metode seleksi fitur Chi-Squared memungkinkan kita untuk menentukan fitur mana yang memiliki keterkaitan paling kuat dengan variabel target dalam dataset. Proses ini melibatkan perhitungan seberapa besar ketergantungan antara fitur dan kelas yang ingin diprediksi. Khususnya dalam analisis sentimen atau klasifikasi teks, teknik ini efektif untuk menyortir fitur yang memiliki pengaruh signifikan terhadap prediksi klasifikasi. Model matematis untuk seleksi fitur Chi-Squared dapat diringkas seperti persamaan (4)[11].

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (4)$$

Dimana  $\chi^2$  adalah nilai statistik Chi-Squared, O adalah frekuensi observasi, dan E adalah frekuensi yang diharapkan. Nilai ( $\chi^2$ ) yang tinggi menunjukkan bahwa fitur tersebut memiliki ketergantungan yang signifikan dengan kelas target dan oleh karena itu penting untuk model klasifikasi.

### 2.3.4. Pemisahan Data

Langkah berikutnya adalah memisahkan data menjadi data latih dan data uji. Dalam penelitian ini, digunakan 80% data sebagai data latih dan 20% data sebagai data uji dari total 6000 record.



Gambar 5. Proporsi Data Latih dan Data Uji

Berdasarkan ilustrasi pada Gambar 5, terdapat 4800 data yang ditetapkan sebagai data pelatihan, sementara 1200 data lainnya diidentifikasi sebagai data uji. Proses pembagian antara data pelatihan dan data uji dilaksanakan secara acak dengan tujuan mempertahankan proporsi yang seimbang antara kelas-kelas yang terlibat.

### 2.3.5. Klasifikasi Multinomial Naive Bayes

Metode Naive Bayes Multinomial merupakan algoritma yang bergantung pada prinsip teorema Bayes dan umumnya dipakai dalam penugasan klasifikasi teks. Algoritma ini mengoperasikan asumsi bahwa setiap fitur dalam dokumen, seperti kata-kata, adalah independen satu sama lain. Hal ini memungkinkan perhitungan probabilitas kelas berdasarkan frekuensi kata menjadi lebih

sederhana. Penghitung probabilitas sebuah dokumen  $d$  terhadap kelas  $C$  yang ditunjukkan pada persamaan (5)[12].

$$P(C) = \frac{N_C}{N} \quad (5)$$

Dimana  $N_C$  adalah jumlah kelas  $C$  pada seluruh dokumen dan  $N$  adalah jumlah seluruh dokumen. Untuk probabilitas dari kata ke- $n$  ditentukan dengan menggunakan persamaan (6)[12]:

$$P(X_n|C) = \frac{N_{x_n,c} + \alpha}{N(C) + V} \quad (6)$$

Dimana  $N_{x_n,c}$  mencerminkan jumlah kemunculan term  $X_n$  dalam seluruh data pelatihan pada kelas  $C$  dan  $N(C)$  menyatakan total kemunculan term dalam seluruh data pelatihan pada kelas  $C$ , dan  $\alpha$  merupakan parameter laplace smoothing,  $V$  adalah jumlah total kata pada data untuk melatih model. Sementara rumus Multinomial yang digunakan dalam pembobotan TF-IDF adalah sebagai berikut:

$$P(X_n|C) = \frac{\sum_{d \in C} \text{tf}(X_n, d \in C) + \alpha}{\sum_{d \in C} N_{d \in C} + V} \quad (7)$$

Dimana  $\sum_{d \in C} \text{tf}(X_n, d \in C)$  adalah jumlah pembobotan kata  $X_n$  dari seluruh data training pada kelas  $C$  dan  $\sum_{d \in C} N_{d \in C}$  adalah jumlah bobot seluruh term pada data training pada kelas  $C$ .

### 2.3.6. Pengujian dan Evaluasi

Evaluasi dilakukan untuk menilai kinerja model yang telah dikembangkan. Kinerja ini bisa diukur menggunakan tabel confusion matrix, yang merupakan alat visual yang sering digunakan untuk menampilkan performa model klasifikasi pada sejumlah data uji yang memiliki nilai sebenarnya yang diketahui. Kinerja model dapat dihitung dengan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score dengan rumus-rumus yang sesuai seperti di bawah ini [13].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

## 3. Hasil dan Diskusi

### 3.1. Hasil Preprocessing

Hasil dari seluruh proses preprocessing pada dataset disajikan dalam bentuk tabel yang dapat ditemukan pada Gambar di bawah ini. Berikut adalah contoh beberapa kalimat yang diambil dari dataset.

### Hasil Preprocessing



```
df.head()
```

|   | kategori  | berita  | Hasil  |
|---|-----------|---|--|
| 0 | teknologi | Uber pada hari Jumat mengatakan akan menguak d... | uber jumat menguak data perjalanan paris publi...  |
| 1 | teknologi | Menyusul jejak NES Classic Edition , SNES Clas... | menyusul jejak ne classic edition snes classic...  |
| 2 | teknologi | MDI Ventures , perusahaan modal ventura yang d... | mdi venture perusahaan modal ventura didukung ...  |
| 3 | teknologi | Mazda masih menutup rapat informasi soal sport... | mazda menutup rapat informasi sportscaar anyar ... |
| 4 | teknologi | Sampai di akhir tahun 2017 ini , frasa " print... | tahun frasa printer mencetak video terdengar m...  |

Gambar 6. Hasil Preprocessing

Gambar 6 menunjukkan contoh kalimat yang belum dilakukan preprocessing dan hasil akhirnya contohnya pada berita berindex 1 dari kategori berita yang berisi kalimat "Menyusul jejak NES Classic Edition," setelah dilakukan preprocessing menjadi "menyusul jejak ne classic edition snes classic"

### 3.2. Hasil Pembobotan TF-IDF

Untuk mengonversi teks berita menjadi bentuk yang dapat diproses oleh model klasifikasi, peneliti menggunakan teknik ekstraksi fitur TF-IDF. Teknik ini memungkinkan peneliti untuk mengukur pentingnya setiap kata dalam dokumen berdasarkan frekuensi kemunculannya dalam dokumen tersebut dan dalam keseluruhan korpus.

Hasil Setelah diberi Bobot Nilai dan diubah ke bentuk matriks

```
[ ] X_tf_idf = tf_idf.transform(x).toarray()
X_tf_idf

array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Gambar 7. Hasil Pembobotan TF-IDF

### 3.3. Hasil Seleksi Fitur dengan Chi-Squared

Untuk meningkatkan kinerja model klasifikasi, dilakukan seleksi fitur dengan metode chi-squared untuk memilih fitur-fitur yang paling relevan dari dataset yang sudah dilakukan preprocessing. Peneliti melakukan evaluasi terhadap berbagai nilai k, yang mewakili jumlah fitur yang akan dipilih dari dataset. Hasil evaluasi menunjukkan bahwa menggunakan nilai k=10000 dari jumlah asli fitur data menghasilkan akurasi yang lebih tinggi dibandingkan dengan menggunakan nilai k=5000, k = 1000, k = 2000.

```

from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2

# 5000 dengan highest chi-squared statistics are selected
chi2_features = SelectKBest(chi2, k=10000)
X_kbest_features = chi2_features.fit_transform(X, y)

# Reduksi features
print("Angka Original Feature :", X.shape[1])
print("Angka Reduksi Feature :", X_kbest_features.shape[1])

Angka Original Feature : 54091
Angka Reduksi Feature : 10000
    
```

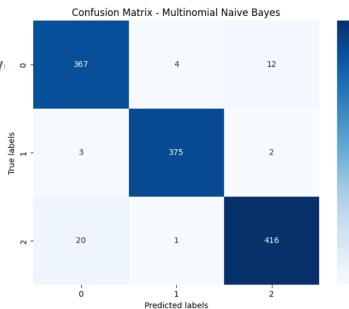
**Gambar 8.** Hasil Seleksi Fitur

Berdasarkan gambar 8 menunjukkan bahwa peneliti mempertahankan jumlah fitur sebanyak nilai k yaitu k=10000 dari jumlah asli fitur data yaitu sebanyak 54091. Ini menunjukkan bahwa mempertahankan lebih banyak fitur memberikan informasi tambahan yang diperlukan bagi model untuk membedakan antara kelas-kelas target dengan lebih baik. Seleksi fitur menggunakan nilai k=10000 telah meningkatkan kinerja model dengan meningkatkan akurasi klasifikasi. Dengan mempertahankan lebih banyak fitur, dapat dipastikan bahwa model memiliki akses ke informasi yang lebih banyak dan relevan untuk melakukan klasifikasi dengan lebih baik.

**3.4. Evaluasi Model Multinomial Naïve Bayes**

Model Multinomial Naive Bayes dibangun dengan menggunakan bantuan library sklearn dalam bahasa pemrograman Python. Setelah itu model akan dilatih dengan data training yang telah dilakukan ekstraksi fitur TF-IDF dan seleksi fitur dengan Chi-squared untuk kemudian dilakukan evaluasi menggunakan data testing. Pelatihan Model dan Prediksi Model Multinomial Naïve Bayes dilatih menggunakan data latih yang telah diseleksi fitur-fiturnya. Hasil akurasi validasi silang menunjukkan bahwa model memiliki kinerja yang stabil dan konsisten dalam mengklasifikasikan data seperti dibuktikan pada gambar 9 dan 10.

Algorithm: Multinomial Naive Bayes  
 Confusion Matrix:  
 [[367 4 12]  
 [ 3 375 2]  
 [ 20 1 416]]  
 Accuracy: 0.965  
 Precision: 0.965  
 Recall Score: 0.965  
 F1 Score: 0.965



**Gambar 9.** Hasil Akurasi Validasi silang

**Gambar 10.** Confusion Matrix Validasi silang

Berdasarkan Gambar 9 waktu yang diperlukan untuk melatih model adalah 0.337 detik, sedangkan waktu yang dibutuhkan untuk melakukan prediksi terhadap data uji adalah 0.057 detik. Dilakukan validasi silang menggunakan metode K-Fold Cross Validation dengan 5 lipatan. Hasil validasi silang menunjukkan akurasi rata-rata sebesar 97.562% dengan deviasi standar sebesar 0.252%. Selanjutnya evaluasi dilakukan menggunakan Confusion Matrix untuk mengukur akurasi yang diperoleh dari metode yang digunakan. Pada pemodelan menggunakan Multinomial Naïve Bayes, diperoleh akurasi yang tinggi pada tahap pelatihan dan pengujian. Informasi evaluasi yang lebih rinci disajikan dalam Tabel 1.

Tabel 1. Hasil Evaluasi

|           | Precision | Recall | F-1 Score | Akurasi |
|-----------|-----------|--------|-----------|---------|
| Teknologi | 95.19%    | 95.19% | 95.19%    | 95.19%  |
| Olahraga  | 98.68%    | 98.68% | 98.68%    | 98.68%  |
| Hiburan   | 95.82%    | 95.82% | 95.82%    | 95,9%   |
| Average   | 96.57%    | 96.57% | 96.57%    | 96.57%  |

#### 4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, dapat disimpulkan bahwa metode Multinomial Naïve Bayes berhasil mengklasifikasikan berita berdasarkan kategori dengan tingkat akurasi tinggi. Dataset teks berita yang dikumpulkan berasal dari repository GitHub dan diolah dengan beberapa tahap yaitu tahap preprocessing, pembobotan TF-IDF, dan seleksi fitur chi-squared. Model klasifikasi yang dikembangkan menunjukkan akurasi yang tinggi dengan nilai sebesar 97.562% dalam validasi silang. Hasil evaluasi menunjukkan bahwa pendekatan ini efektif dalam memproses data teks dan menghasilkan prediksi yang akurat, menunjukkan potensi aplikatifnya dalam pengelolaan berita secara otomatis. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem yang dapat membantu dalam pengelolaan dan analisis berita secara efisien.

#### Daftar Pustaka

- [1] D. N. Chandra, G. Indrawan, dan I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," *Jurnal Ilmiah Teknologi dan Informasi Asia (JITIKA)*, vol. 10, hlm. 11, 2016.
- [2] W. F. Mahmudy dan A. W. Widodo, "Klasifikasi Artikel Berita Secara Otomatis Menggunakan Metode Naive Bayes Classifier Yang Dimodifikasi," *TEKNO*, vol. 21, Mar 2014.
- [3] A. F. Hidayatullah dkk., "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," 2016.
- [4] S. Kumar, A. Sharma, B. K. Reddy, S. Sachan, V. Jain, dan J. Singh, "An intelligent model based on integrated inverse document frequency and multinomial Naive Bayes for current affairs news categorisation.," *International Journal of System Assurance Engineering and Management*, vol. 13, hlm. 1–15, Nov 2021.
- [5] A. Sabrani, I. W. Gede Putu Wirarama Wedashwara, dan F. Bimantoro, "Metode Multinomial Naive Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia (Multinomial Naive Bayes Method for Classification of Online Article About Earthquake in Indonesia)." [Daring]. Tersedia pada: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [6] S. K. Dirjen dkk., "Terakreditasi SINTA Peringkat 2 Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting," *masa berlaku mulai*, vol. 1, no. 3, hlm. 227–232, 2017.
- [7] irwan Budiman, R. F. M, dan D. T. Nugrahadi, "Studi Ekstraksi Fitur Berbasis Vektor Word2vec Pada Pembentukan Fitur Berdimensi Rendah," *Jurnal Komputasi*, vol. 8, 2020.
- [8] E. Indrayuni, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *Jurnal Khatulistiwa Informatika*, vol. 7, no. 1, hlm. 29–36, 2019.



- [9] P. M. Prihatini, "Implementasi Ekstraksi Fitur Pada Pengolahan Dokumen Berbahasa Indonesia," *Jurnal Manajemen Teknologi dan Informatika*, vol. 6, no. 3, 2016.
- [10] N. Komang dkk., "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *MERPATI*, vol. 6, no. 2, 2018.
- [11] S. Goswami, "Using the Chi-Squared test for feature selection with implementation," Nov 2020.
- [12] B. Harjito, K. N. Aini, dan B. Murtiyasa, "Klasifikasi Dokumen berkonten Serangan jaringan menggunakan Multinomial Naive Bayes," *Seminar Nasional Teknologi Informasi dan Komunikasi (SEMNASITIK)*, vol. 1, no. 1, hlm. 112–118, 2018.
- [13] E. Mas'udah, E. Wahyuni, dan A. Anjani, "Analisis sentimen: Pemindahan ibu kota Indonesia pada twitter," *Jurnal Informatika dan Sistem Informasi*, vol. 1, no. 2, hlm. 397–401, 2020.

# Rancang Bangun Website Sebagai Sarana Promosi Usaha Madu Kele-Kele di Desa Bongkasa Pertiwi

Ni Made Desni Dwi Arisaputri<sup>a1</sup>, Gst. Ayu Vida Matrika Giri<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>dwiarisa644@gmail.com  
<sup>2</sup>vida@unud.ac.id

## Abstract

*Kele-Kele Honey from Bongkasa Pertiwi Village is a superior product with large market potential but is still little known by the wider community. This research aims to develop a website as an effective promotional tool to increase awareness and sales of Kele-Kele honey. By implementing a Rapid Application Development (RAD) approach, website development is carried out quickly and iteratively, enabling active involvement of business owners in the development process and feature adjustments based on user feedback. The research results show that the website developed has succeeded in increasing public awareness about Kele-Kele honey and supporting the growth of local businesses. The practical implication of this research suggests that employing Rapid Application Development (RAD) in web development could be an effective strategy to enhance the sales and visibility of local products such as Kele-Kele honey. To further promote the growth of small and medium-sized enterprises, future studies could focus on developing new features and conducting a more comprehensive evaluation of the effectiveness of online promotions.*

**Keywords:** Kele-Kele Honey, Website, RAD, Business Promotion, Web Development

## 1. Pendahuluan

Pemasaran digital atau Digital marketing memiliki dampak yang signifikan dan positif dalam meningkatkan kinerja penjualan. Memanfaatkan platform digital yang mudah digunakan dapat meningkatkan upaya pemasaran dengan memungkinkan koneksi dengan audiens yang beragam. Sebagai pelengkap, toko fisik tetap memegang peran penting dalam menjangkau konsumen secara langsung [1]. Saat ini, Instagram adalah platform yang banyak digunakan, terutama di kalangan pebisnis yang memanfaatkannya untuk mempromosikan perusahaan mereka. Platform ini memungkinkan penjual untuk menampilkan informasi yang luas dan menarik tentang produk atau layanan mereka. Jangkauan media sosial dan teknologi situs web memungkinkan penjual untuk terhubung dengan audiens yang lebih luas, yang berpotensi meningkatkan penjualan dan pendapatan mereka.[2]. Kelompok budidaya madu "kele-kele" di Desa Bongkasa Pertiwi diprakarsai oleh para petani madu yang memiliki hobi. Program Kampung Mandiri (Program CSR Aqua Mambal) memberikan dukungan kepada Desa Wisata Bongkasa Pertiwi, yang bertujuan untuk membantu masyarakat yang rentan. Program ini memilih lebah "kele-kele" yang merupakan lebah asli daerah tersebut, karena kemampuannya beradaptasi dengan lingkungan setempat, tidak memiliki perilaku menyengat, mudah dalam pengelolaan, dan tidak memerlukan alat atau teknologi khusus. Selain itu, investasi awal dan biaya perawatan yang lebih rendah dibandingkan dengan spesies lebah lainnya, menyebabkan produksi mereka meluas. Kelompok Sarining Trigona Pertiwi masih menggunakan pemasaran konvensional dengan sedikit pemanfaatan media sosial. Namun, pemanfaatan tersebut belum maksimal. Terdapat kendala dalam meningkatkan pemasaran online karena kurangnya pengetahuan dan keterampilan. Perlu ditingkatkan strategi pemasaran online dan pengetahuan anggota dalam hal tersebut. [3]. Penelitian sebelumnya telah mencoba membangun sebuah website untuk kelompok budidaya madu kele-kele, namun sayangnya, website tersebut menghadapi masalah aksesibilitas yang menyebabkan hanya satu tangkapan gambar yang dapat ditampilkan. Tidak

hanya itu, juga tidak ada tautan ke website tersebut di platform media sosial lainnya, menyulitkan untuk memperoleh informasi lebih lanjut tentang konten website yang sudah dibuat. Karena kendala tersebut, saya bertekad untuk merancang ulang website ini dengan tujuan memperluas promosi usaha madu kele-kele secara lebih efektif. Selain itu, dari penelitian sebelumnya, saya menyadari bahwa tampilan website yang disajikan hanya sebatas satu tampilan gambar pada bagian awal, menyulitkan untuk memahami isi konten secara keseluruhan. Oleh karena itu, keinginan saya untuk menciptakan website dengan desain yang lebih menarik tidak hanya berakar dari urgensi untuk mengatasi masalah aksesibilitas, tetapi juga karena saya percaya bahwa tampilan visual yang menarik dapat lebih efektif menarik perhatian calon pembeli madu kele-kele.

## 2. Metode Penelitian

### 2.1 Metode Rapid Application Development (RAD)

Rapid Application Development (RAD) adalah metodologi pengembangan sistem yang dibangun berdasarkan prinsip-prinsip Waterfall Model untuk membuat aplikasi secara bertahap. Pendekatan ini menekankan pada kemajuan yang konsisten dan bertahap, sehingga memungkinkan pengembangan aplikasi dalam waktu yang relatif singkat[4]. Di sisi lain, perlu diakui bahwa pengembang perangkat lunak sering dihadapkan pada tugas yang menuntut untuk merinci kebutuhan secara teknis, yang sering kali menjadi titik yang kurang dipahami oleh pelanggan. Pelanggan cenderung lebih memahami kebutuhan mereka dalam konteks fungsionalitas yang diinginkan daripada aspek-aspek teknis yang terlibat dalam implementasinya. Hal ini dapat menyebabkan kesenjangan antara harapan pelanggan dan pemahaman pengembang tentang apa yang sebenarnya diperlukan untuk membangun solusi perangkat lunak yang sesuai. Dalam konteks ini, Model RAD (Rapid Application Development) menawarkan pendekatan yang efektif untuk mengatasi ketidakpahaman pelanggan terhadap aspek teknis. Dengan menggunakan RAD, pengembang dapat mengadakan pertemuan atau sesi kolaborasi dengan pelanggan untuk secara lebih terperinci memperjelas kebutuhan teknis. Proses RAD yang iteratif dan responsif memungkinkan pelanggan untuk secara langsung terlibat dalam pengembangan perangkat lunak, yang pada gilirannya membantu dalam menyatukan persepsi mereka tentang apa yang dimaksudkan dengan spesifikasi kebutuhan yang diinginkan. Berikut merupakan detail dari RAD, yaitu:

- a. **Pemodelan Bisnis**  
Langkah dalam pemodelan ini bertujuan untuk mengidentifikasi informasi yang terlibat dalam sebuah proses khususnya proses bisnis beserta orang-orang yang terlibat, bagaimana fungsinya, dan apa saja yang perlu dibuat. Selama tahap proses pemodelan bisnis ini, observasi dilakukan untuk mengumpulkan lebih banyak wawasan tentang entri kele-kele.
- b. **Pemodelan Data**  
Fase pemodelan data menghasilkan kumpulan objek data yang diperlukan untuk mendukung bisnis madu.
- c. **Pemodelan Proses**  
Pemodelan proses, yang melibatkan penerapan fungsi bisnis berdasarkan data, sangat penting untuk pengembangan ini. Saat ini, ini adalah tahap yang sangat efektif dan mudah dipahami.
- d. **Pembuatan website**  
Tahap selanjutnya dalam pembangunan sistem adalah pengembangan aplikasi, di mana program diimplementasikan sesuai dengan pemodelan proses dan data. Pada tahap ini, pemodelan data dan pemodelan proses dieksekusi dengan menggunakan bahasa komputer, sehingga menghasilkan sistem berbasis web. Sistem ini menggunakan bahasa pemrograman HTML, CSS, dan JavaScript.
- e. **Pengujian dan Pergantian**  
Langkah selanjutnya adalah menguji komponen yang telah dirakit. Setelah sistem diuji, Anda dapat melanjutkan dengan mengembangkan komponen berikutnya. Pada tahap ini, sistem yang sedang dikembangkan dievaluasi, dan modifikasi yang diperlukan pada sistem yang telah selesai dapat dilakukan.

- f. Maintenance  
Pemeliharaan melibatkan proses pemeliharaan fasilitas yang ada dan melakukan perbaikan, modifikasi, atau penggantian yang diperlukan untuk mempertahankan kondisinya saat ini. Hasil terbaik dari upaya beberapa tim akan diterapkan selama pemeliharaan.

## 2.2 Pengumpulan Data

Para penulis penelitian ini menggunakan metode berikut untuk mengumpulkan data:

- a. Observasi  
Penulis melakukan observasi dengan cara datang ke lokasi langsung, dan melihat proses pemanenan yang dilakukan pemilik sampai dengan proses pengemasan. Serta melakukan pengamatan terhadap akun sosial media terkait promosi dilakukan.
- b. Studi Pustaka  
Studi literatur dilakukan dengan menggunakan referensi yang ditemukan secara online atau dari sumber lain.

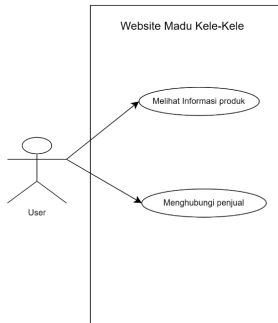
## 2.3 Teknik Pengujian

Pengujian website dilakukan dengan perhitungan usability, usability salah satu metode yang berfungsi untuk mengevaluasi suatu perangkat software, website, atau produk yang dibuat. Dimana dengan menyebarkan responden ke masyarakat sekitar lalu melakukan penilaian terhadap website yang dibuat. Dari nilai atau skor yang di peroleh dari perhitungan usability merupakan hasil dari presentasi untuk website yang dibuat.

## 3. Hasil dan Diskusi

### 3.1 Use Case Diagram

Diagram kasus atau use case diagram penggunaan, salah satu jenis diagram UML (Unified Modeling Language), menggambarkan interaksi antara aktor dan sistem. Aktor dan Use Case (UC) adalah dua elemen utama dari Use Case Diagram. Aktor mencakup apa pun yang secara langsung berkomunikasi dengan sistem, yang dapat berupa sistem komputer lain atau seseorang, yang diidentifikasi berdasarkan perannya, bukan namanya. Aktor diwakili oleh gambar tongkat dengan peran atau sistem yang terdaftar sebagai kata benda di bagian bawah [5]

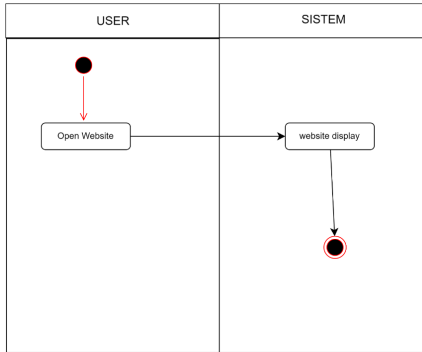


Gambar 1. Use Case Diagram

### 3.2 Activity Diagram

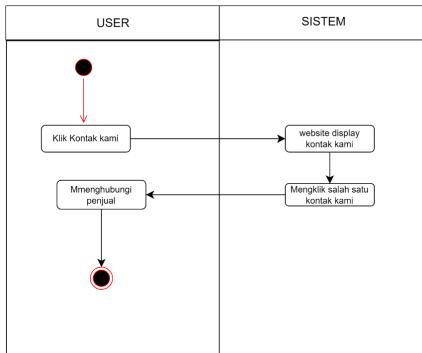
Activity Diagram adalah jenis diagram yang dapat mensimulasikan proses sistem. Diagram ini menjelaskan urutan proses dalam sistem secara vertikal. Diagram aktivitas merepresentasikan pembuatan kasus penggunaan dengan aliran aktivitas [6]. Activity Diagram yang dibangun adalah membuka situs website dan menghubungi penjual.

#### 3.2.1. Membuka Situs Website



Gambar 2. Membuka Situs Website

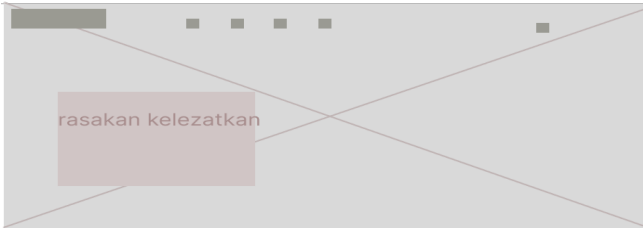
#### 3.2.2 Menghubungi Konta Penjual



Gambar 3. Menghubungi Kontak Penjual

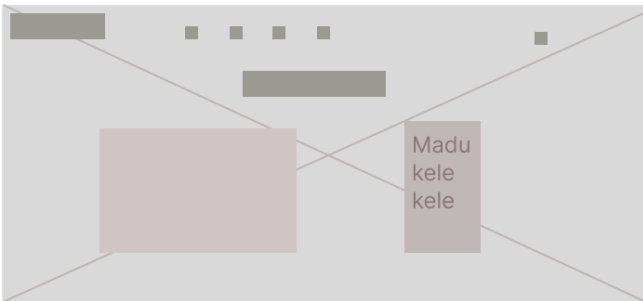
### 3.3 Desain WireFrame

Desain WireFrame di buat dengan Figma untuk merancang desain website sebelum websitenya dibuat, WireFrame digunakan sebagai bayangan dari gambar websitenya. Berikut merupakan gambar desain tampilan awal di website nantinya. Paling atas pojok kiri ada nama websitenya, dan di Tengah-tengah ada 4 bar tentang apa saja yang ada di website. Pada bagian utama di kiri ada tulisan yang mendeskripsikan madu, berikut desain MockUp home:



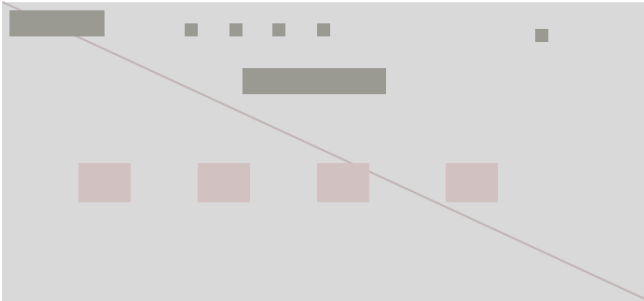
**Gambar 4.** Desain MockUp Home

Selanjutnya merupakan desain dari MockUp Tentang kami, ada perbedaan di bagian utama yaitu di Tengah ada tulisan tentang kami, sebelah kiri nanti akan gambar madu dan sebelah kanan deskripsi tentang Madu, berikut desain MockUp nya:



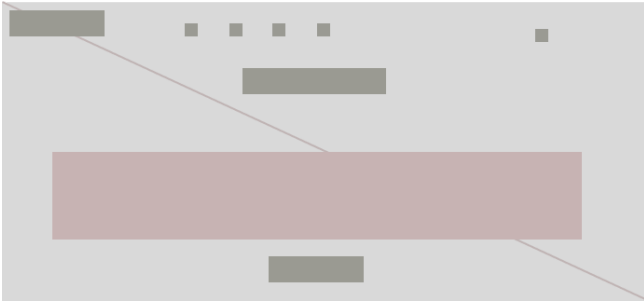
**Gambar 5.** Desain MockUp Tentang Kami

Selanjutnya desain tentang Menu, dimana di bagian utama yang membedakan dengan yang lainnya akan menampilkan produk-produk yang di jual, berikut desain MockUp nya:



**Gambar 6.** Desain MockUp Menu

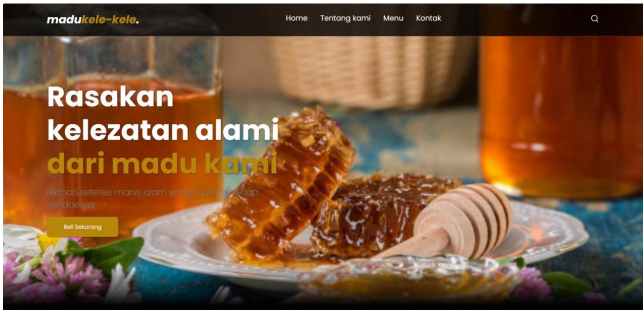
Selanjutnya bagian akhir yaitu Kontak kami, pada bagian ini akan berisikan maps lokasi tempat masu, dan kontak person yang bisa di hubungi. Di bawah ini merupakan desain MockUp kontak kami



**Gambar 7.** Desain MockUp Kontak

### 3.2. Desain Website

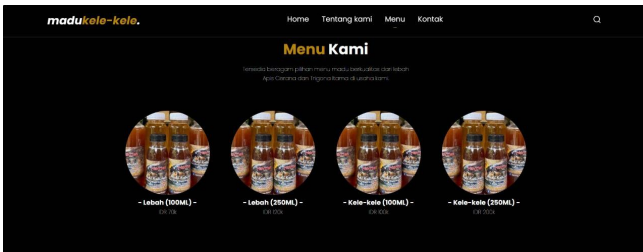
Desain Website yang dibuat berpacu pada desain wireframe, akan tetapi ada beberapa perubahan dan tambahan dari desain wireframe yang sudah dibuat sebelumnya. Desain website menggunakan compiler Vscode, untuk Bahasa pembrogramannya ada 3 yaitu html,css, dan JavaScript. Berikut desain website yang sudah di buat. Berikut dibawah ini ialah desain dari bagian Home, bedanya dari yang sebelumnya sudah terlihat di bagian ini. Pada website sebelumnya latar bagian awalnya putih dasarnya, sedangkan yang saya bangun berupa gambar dari madu, untuk bagian lain tidak bisa saya bandingkan hanya tercantum satu gambar pada penelitian sebelumnya di laporan, sehingga untuk detail lebih lengkap tidak ada. Berikut gambar dari desain websitenya.



Gambar 8. Desain website Home

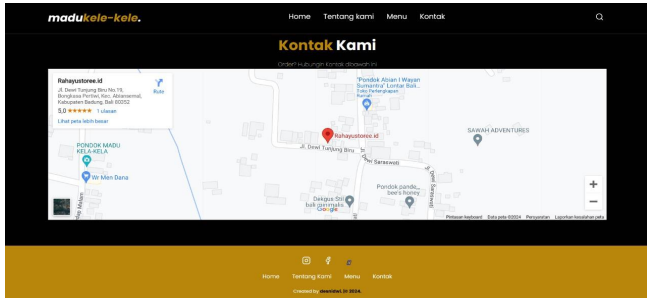


Gambar 9. Desain website Tentang Kami



Gambar 10. Desain website Menu





Gambar 11. Desain website Kontak

### 3.3 Pengujian Usability

Berikut hasil respnden sebanyak 8 orang, dimana dari 10 pertanyaan untuk soal nomor 3 dan 6 perhitungan skornya terbalik, yaitu jika jawabanya sangat tidak setuju memiliki skor 5 sedangkan untuk soal bukan nomor 3 dan 6 jika jawabanya sangat setuju memiliki skor 5. Total skor 400 memiliki akurasi 100%, berikut merupakan perhitunganya:

Tabel 1. Perhitungan Usability

| SKOR HASIL PERHITUNGAN |    |    |    |    |    |    |    |    |     | JUMLAH     |
|------------------------|----|----|----|----|----|----|----|----|-----|------------|
| Q1                     | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |            |
| 5                      | 5  | 5  | 3  | 4  | 5  | 4  | 5  | 4  | 4   | 44         |
| 5                      | 5  | 5  | 5  | 5  | 5  | 4  | 5  | 4  | 5   | 48         |
| 4                      | 4  | 2  | 5  | 5  | 5  | 4  | 4  | 4  | 4   | 41         |
| 4                      | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4   | 40         |
| 4                      | 4  | 4  | 3  | 4  | 5  | 4  | 4  | 4  | 4   | 40         |
| 4                      | 4  | 4  | 4  | 4  | 5  | 4  | 4  | 3  | 5   | 41         |
| 5                      | 5  | 5  | 5  | 5  | 5  | 5  | 5  | 5  | 5   | 50         |
| <b>SKOR</b>            |    |    |    |    |    |    |    |    |     | <b>304</b> |

Dari hasil perhitungan skor kita dapatkan hasil Usability mempunyai nilai 76% dari 100% dengan skala Baik

### 4. Kesimpulan

Dari uraian yang telah dijelaskan, penulis dapat menyimpulkan hal tersebut:

- Perancangan dan Pembangunan sebuah website pada usaha madu kele-kele yang berada di Desa Bongkasa Pertiwi, Abiansemal, Badung Dari website yang sudah dibuat semoga bisa menjadikan promosi usaha masu dan orang-orang ingin melihat menu apa saja yang ada dengan muda bisa di lihat pada website.
- Hasil uji kegunaan menunjukkan bahwa situs web ini mencapai peringkat 76% dari 100%, yang menunjukkan bahwa situs web ini cocok untuk digunakan.

#### Daftar Pustaka

- [1] Hendrawan, Andi."Pengaruh marketing digital terhadap kinerja penjualan produk umkm asli gauri di kecamatan bantarsari cilacaP." *Jurnal Administrasi Dan Kesekretarisan* 4.1 (2019): 50-61.
- [2] Hasdiana and Nurjamiyah. "Pemanfaatan Website Dan Sosial Media Sebagai Media Promosi Bina Lestari Adat Dan Budaya." *Wahana Inovasi: Jurnal Penelitian dan Pengabdian Masyarakat UISU*, vol. 11, no. 1, pp. 167-171, 2022.
- [3] K. S. A. S. G. N. A. K. I Gusti Ayu Widari Upadani, "Penguatan Kapasitas Kelompok Dalam Mengembangkan Usaha Madu Melalui Perbaikan Usaha Dan Digital Marketing," *Jurnal Widya Laksana*, vol. 12, no. 2, pp. 330-340, 2023.
- [4] G. B. A. L. Oky Irnawati, "Metode Rapid Application Development (RAD) pada Perancangan Website Inventory PT. Sarana Abadi Makmur Bersama (S.A.M.B) Jakarta," *Evolusi: Jurnal Sains Dan Manajemen*, vol. 6, no. 2, 2018.
- [5] Kurniawan T, "Pemodelan use case (UML): evaluasi terhadap beberapa kesalahan dalam praktik" *J. Teknol. Inf. dan Ilmu Komput.*, (2018), 77, 5(1)
- [6] "Activity Diagram," Digoding, Accessed: Mei. 9, 2024. [Online]. Available: <https://www.digoding.com/blog/apa-itu-activity-diagram/>

Halaman ini sengaja dibiarkan kosong

# Analisis Perbandingan Enkripsi File Teks Berformat .txt dan .docx Menggunakan Algoritma AES

Kennardy Andrew Limartha<sup>a1</sup>, I Gede Arta Wibawa<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>andrewkennardy@gmail.com  
<sup>2</sup>gede.arta@unud.ac.id

## Abstract

*In this research, the Advanced Encryption Standard (AES) algorithm is used to encrypt text files. Data stored in various formats has the potential to cause the size of the data on the data storage device to become large and does not necessarily guarantee the security of the data. One method of data security is to encrypt data. This research utilizes data in text file format with data types in the form of ".txt" and ".docx" which refer to the file size for testing. Based on the test results, it shows that the AES encryption method provides good results for text file formats. The time required for the encryption process using the AES algorithm is influenced by each different file size, the larger the file size used, the longer the computing time required and vice versa.*

**Keywords:** Advanced Encryption Standard (AES), Encrypt, Data Security, Text Files, .TXT, .DOCX

## 1. Pendahuluan

Dunia sudah memasuki era digitalisasi yang semakin berkembang pesat. Komputer merupakan hal yang sangat berpengaruh bagi sarana distribusi data dan informasi. Pastinya di dalam setiap komputer terdapat file atau dokumen yang berisi data-data rahasia. Data yang disimpan dalam berbagai macam format berpotensi menimbulkan data yang memiliki ukuran besar pada perangkat penyimpanan data dan keamanannya belum tentu terjamin. Untuk itu diperlukan suatu pengamanan data yang baik sehingga data dan informasi dapat disimpan dengan keamanan yang terjamin [1]. Keamanan ini diperlukan agar data tersebut hanya dapat dibuka dan diakses oleh si penerima atau orang-orang yang memiliki kewenangan atas data tersebut sehingga pihak lain yang tidak memiliki hak akses tidak dapat mengetahui makna atau isi dari data tersebut [2]. Kriptografi merupakan bentuk solusi yang dapat ditawarkan di dalam keamanan komputer. Teknik yang digunakan dalam kriptografi adalah menyamarkan atau mengubah informasi sehingga ketika dikirim menjadi sesuatu yang tidak bermakna. Misalkan seseorang mengirimkan data yang berisikan informasi rahasia berupa "keuangan", maka pada proses pengiriman kata akan diganti menjadi sesuatu yang tidak dapat dibaca yaitu "#%! @!&". Kata yang tidak bermakna tersebut dinamakan dengan *ciphertext* sedangkan kata atau pesan yang aslinya disebut *plaintext* [2]. Dalam kriptografi ini terdapat berbagai metode untuk mengatasi keamanan data. Salah satunya metode yang kita perlukan adalah enkripsi. Enkripsi merupakan proses mengubah data menjadi format yang tidak dapat dibaca atau sulit dimengerti tanpa memiliki kunci enkripsi yang sesuai, sehingga melindungi kerahasiaan dan integritas data selama transmisi. Enkripsi terjadi pada saat proses pengiriman berlangsung sedangkan dekripsi terjadi saat proses penyampaian pesan ke tujuan yang dituju dengan cara merubah data rahasia pada proses enkripsi kembali lagi menjadi data aslinya. Enkripsi juga sering diartikan sebagai kode atau (*chipper*). Pengkodean dilakukan dengan algoritma tertentu untuk mengkodekan semua aliran data bit dari suatu pesan asli (*plaintext*) menjadi pesan rahasia (*chipertext*). Proses enkripsi menjadikan suatu informasi akan lebih sulit untuk diketahui oleh orang yang tidak berhak. Keamanan tersebut sangatlah diperlukan untuk menghindari terjadinya penyadapan maupun pembajakan terhadap file dengan informasi yang penting, maka dari itu dibutuhkan sebuah algoritma yang dapat memproteksi file penting yaitu Algoritma *Advanced Standard Encryption* (AES). AES adalah algoritma simetris 2

yang menggunakan kunci yang sama untuk proses enkripsi dan dekripsi dengan panjang kunci yang bervariasi, yakni: AES 128-bit, AES 192-bit, AES 256-bit [3]. AES merupakan *chipper* yang berorientasi pada bit, sehingga memungkinkan dalam pengimplementasian algoritma yang efisien ke dalam software dan hardware. One Time Pad (OTP) yang dipilih untuk mengkombinasikan masing-masing karakter pada *plaintext* dengan satu karakter pada satu kunci dienkripsi dengan satu algoritma kemudian diteruskan dengan algoritma yang lainnya [4]. Algoritma AES ini memiliki kelebihan dari segi jenis kunci simetri yang digunakan. Dengan panjang kunci paling sedikit yaitu 128 setidaknya terdapat  $2^{128}$  kemungkinan kunci. Jika komputer tercepat dapat mencoba 1 juta kunci tiap detik maka akan dibutuhkan waktu  $5,45,4 \times 10^{24}$  tahun untuk mencoba seluruh kunci [5]. Tujuan dilakukannya penelitian ini yaitu untuk menganalisa perbandingan hasil terhadap keamanan proses enkripsi dengan menggunakan algoritma AES pada file text berupa txt dan docx. Kedua file ini seringkali dikirimkan ataupun dipublikasikan melalui jaringan publik sehingga sangat rentan terhadap pencurian karena dapat dibaca dan dimengerti oleh semua pihak manusia [6]. Maka dari itu, dilakukannya analisis ini untuk mengetahui ukuran file yang diperoleh setelah dilakukannya proses enkripsi melalui perbandingan besar atau kecilnya ukuran file dengan waktu yang dibutuhkan dalam proses komputasi. Sehingga diketahui manakah dari kedua file ini yang lebih efektif menggunakan algoritma AES.

## 2. Metode Penelitian

### 2.1 Data Penelitian

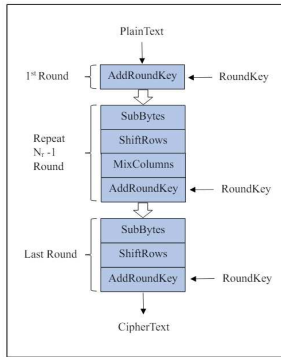
Data yang digunakan untuk melakukan enkripsi dengan algoritma AES adalah dengan menggunakan format file teks. Akan digunakan 2 tipe data yang berbeda pada file teks untuk melakukan perbandingan kinerja algoritma AES dalam melakukan enkripsi pada 2 buah format file. Tipe data yang akan digunakan yakni, ".txt", dan untuk setiap data akan disamakan ukurannya dengan ukuran yang akan diuji yakni berukuran 100 KB, 200 KB, 500 KB dan 1000 KB. Evaluasi perbandingan didasarkan pada kecepatan atau waktu komputasi dan ukuran setelah dilakukannya proses enkripsi pada data tersebut.

### 2.2 Algoritma AES

Advanced Encryption Standard (AES) memiliki nama lain yaitu Algoritma Rijndael yang merupakan suatu algoritma enkripsi tipe simetrik *block cipher* dengan sistem permutasi dan substitusi. Ada tiga jenis algoritma AES, yaitu AES-128, AES-192, dan AES-256. Pengelompokan ini berdasarkan panjang kunci yang digunakan pada algoritma AES. Selain itu ada beberapa hal lain yang membedakan antar jenis algoritma AES, yaitu *round* yang digunakan. AES-128 menggunakan 10 *round*, AES-192 menggunakan 12 *round*, dan AES-256 menggunakan 14 *round* [7].

### 2.3 Proses Enkripsi Algoritma AES

Dalam enkripsi terjadi proses perubahan pesan asli (*plaintext*) menjadi pesan bersandi (*ciphertext*). Secara garis besar proses enkripsi AES terdiri dari 4 jenis transformasi byte, yaitu *SubBytes*, *ShiftRows*, *MixColumns* dan *AddRoundKey*. Proses enkripsi dimulai dari input yang telah dimasukkan ke dalam *array state* akan mengalami perubahan byte *AddRoundKey*. Setelah itu, *array state* akan mengalami perubahan *SubBytes*, *ShiftRows*, *MixColumns*, dan *AddRoundKey* secara berulang-ulang sebanyak putaran. Proses ini dalam algoritma AES disebut sebagai *round function*. *Round* yang terakhir agak berbeda dengan *round round* sebelumnya dimana pada *round* terakhir, *array state* tidak mengalami perubahan *MixColumns* [8]. Dibawah ini merupakan gambar diagram proses enkripsi seperti yang ditunjukkan pada Gambar 1.

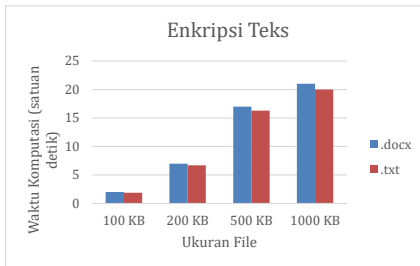


**Gambar 1.** Diagram Proses Enkripsi AES

- a. *AddRoundKey*, melakukan XOR antara pesan asli dengan *cipherkey*
- b. *Repeat N<sub>r</sub> - 1 Round*, memiliki proses yang akan dilakukan setiap putaran yaitu:
  - *SubBytes*: Mensubstitusi byte dengan menggunakan tabel substitusi (S-box).
  - *ShiftRows*: Pergeseran baris-baris *array state* secara *wrapping* (perubahan permutasi).
  - *MixColumns*: Mengalikan data di kolom-kolom *array state* (perubahan pengacakan).
  - *AddRoundKey*: Melakukan XOR antara *state* sekarang dengan *round key* (perubahan penambahan kunci).
- c. *Last Round*, proses round terakhir sama dengan *round* sebelumnya namun tanpa perubahan *MixColumns* yang meliputi:
  - *SubBytes* (perubahan substitusi).
  - *ShiftRows* (perubahan permutasi).
  - *AddRoundKey* (perubahan pengacakan).

### 3. Hasil dan Diskusi

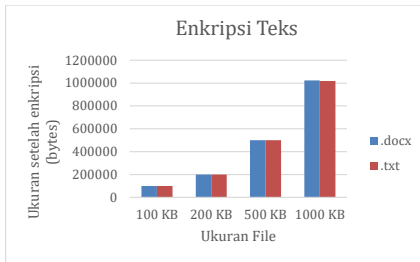
#### 3.1 Perbandingan Waktu Komputasi Tipe Data “.txt” dan “.docx”



**Gambar 2.** Perbandingan Waktu Komputasi Pada File Teks

Berdasarkan pada grafik diatas didapatkan hubungan antara ukuran file dengan waktu komputasi yang dibutuhkan dalam satuan detik. Semakin besar ukuran file yang digunakan maka semakin besar pula waktu yang dibutuhkan untuk komputasi. Grafik diatas menyajikan ukuran file yang sama terhadap kedua tipe data namun terdapat perbedaan yang tergolong tidak terlalu signifikan. Pada ukuran file 1000 KB dengan tipe data ".txt" dibutuhkan waktu komputasi 20 detik. Untuk tipe data ".docx" dibutuhkan waktu 21 detik. Sehingga dapat disimpulkan bahwa tipe data ".txt" membutuhkan waktu komputasi yang lebih cepat dibandingkan dengan tipe data ".docx" dalam proses enkripsinya. Hal tersebut disebabkan oleh struktur tipe data ".txt" yang lebih sederhana. Meskipun struktur yang dimiliki oleh tipe data ".docx" lebih kompleks perbedaan yang terlihat tidak terlalu besar. Kemungkinan disebabkan oleh ukuran file yang sama dan efisiensi dari algoritma AES.

### 3.2 Perbandingan Ukuran File Enkripsi Tipe Data ".txt" dan ".docx"



Gambar 3. Perbandingan Ukuran Enkripsi File Teks

Berdasarkan grafik pada gambar 3 menunjukkan bahwa ukuran file setelah dienkripsi antara kedua tipe data hampir tidak berbeda. Pada ukuran file 100 KB dengan tipe data ".txt" ukuran file setelah di enkripsi menjadi 100062 bytes. Untuk tipe data ".docx" ukuran file setelah dienkripsi menjadi 100093 bytes. Pada ukuran file 200 KB dengan tipe data ".txt" ukuran file setelah di enkripsi menjadi 200087 bytes. Untuk tipe data ".docx" ukuran file setelah dienkripsi menjadi 20121 bytes. Pada ukuran file 500 KB dengan tipe data ".txt" ukuran file setelah di enkripsi menjadi 500173 bytes. Untuk tipe data ".docx" ukuran file setelah dienkripsi menjadi 500385 bytes. Pada ukuran file 1000 KB dengan tipe data ".txt" ukuran file setelah di enkripsi menjadi 1018 KB. Untuk tipe data ".docx" ukuran file setelah dienkripsi menjadi 1023 KB. Dari data yang diperoleh menunjukkan bahwa ukuran tipe data ".txt" dan ".docx" terenkripsi menjadi lebih besar dibandingkan dengan ukuran aslinya sebelum dienkripsi. Hasil dari enkripsi menunjukkan bahwa ukuran tipe data ".txt" menjadi lebih kecil dibandingkan dengan ukuran tipe data ".docx".

## 4. Kesimpulan

Berdasarkan hasil pembahasan yang telah dilakukan di atas dapat disimpulkan bahwa:

- Algoritma AES merupakan salah satu cara yang terbaik dan efisien untuk meningkatkan keamanan data dalam hal enkripsi.
- Ukuran file menjadi salah satu faktor yang berpengaruh dalam waktu proses enkripsi dan waktu enkripsi menjadi salah satu tolak ukur dari cepat atau lambatnya suatu proses enkripsi terhadap file.
- Pada ukuran file 1000 KB dengan tipe data ".txt" dibutuhkan waktu komputasi 20 detik dan ukuran file setelah di enkripsi menjadi 1018 KB. Untuk tipe data ".docx" dibutuhkan waktu 21 detik dan ukuran file setelah dienkripsi menjadi 1023 KB.
- Analisis perbandingan file teks dengan tipe data ".txt" dan ".docx" menggunakan algoritma

Advanced Encryption Standard (AES) untuk mendapatkan hasil perbandingan berupa waktu komputasi dan ukuran file setelah dienkripsi. Hasil keduanya tidak terlalu jauh perbedaannya, untuk waktu komputasi tipe data ".txt" lebih cepat dibandingkan ".docx" karena tipe data ".txt" memiliki struktur yang lebih sederhana.

#### Daftar Pustaka

- [1] Chandra, R, V, H., Ari, K., dan Mahendra, D. Analisis Performa Proses Ekripsi dan Dekripsi Menggunakan Algoritma AES-128 Pada Berbagai Format File. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(1): 481-486.
- [2] Indrayani, L, A., dan I Made, S. 2019. Implementasi Kriptografi dengan Modifikasi Algoritma Advanced Encryption Standard (AES) untuk Pengamanan File Document. *Journal of Informatics and Computer Science*, 01(01): 42-47.
- [3] Handoyo, J., dan Yulieo, M, S. 2020. Keamanan Dokumen Menggunakan Algoritma Advanced Encryption Standard (AES). *Jurnal Sistem Informasi dan Teknologi*, 3(2): 144-150.
- [4] Abdullah, I, N., Dewi, K., dan Mohammad, A. 2018. Aplikasi Enkripsi File Dokumen Menggunakan Metode Algoritma AES (Advanced Encryption Standard) dan OTP (One Time Pad) Berbasis Web Pada PT. MNC Sky Vision. *Jurnal Telematika MKOM*, 10(1):11-16
- [5] Lana, F, N. 2020. *Pengamanan File MP3 Menggunakan Algoritma Advanced Encryption Standard (AES)*. Skripsi. Universitas Airlangga
- [6] Sancaka, T, M, P., dan Veronica, L. 2020. Penerapan Metode Playfair Cipher Dalam Aplikasi Enkripsi Dekripsi File Teks. *Jurnal Ilmiah Elektronika dan Komputer*, 15(02): 260-270.
- [7] Kromodimoeljo, S. 2009. *Teori dan Aplikasi Kriptografi*. SPK IT Consulting
- [8] Tulloh, A, R. 2016. *Kriptografi Advanced Encryption Standard (AES) Untuk Penyandian File Dokumen*, Bandung :Universitas Islam.



Halaman ini sengaja dibiarkan kosong

# Implementasi Algoritma Yolo untuk Deteksi Tanaman Apotik Hidup

I Kadek Peri Arta Wijaya<sup>a1</sup>, I Gede Santi Astawa<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>periarta2004@gmail.com  
<sup>2</sup>santi.astawa@unud.ac.id

## Abstract

*Living apothecary is a plant that is used to provide natural treatment or medicine. Therefore, to distinguish these plants, we carry out detection to distinguish live pharmaceutical plants. In this research we used (You Only Look Once) YOLOv8, Google Colab, Roboflow. Where we get the database from Google then we process the image using Roboflow then we create (You Only Look Once) YOLOv8, and we test it on Google Colab. The results of this research confirm that the YOLOv8 (You Only Look Once) algorithm in identifying and modeling 13 plant objects with a total of 1204 images obtained a precision percentage value of 59%, recall of 58%, and a MAP value of 60%, so the average value the accuracy is above 59%. In addition, it provides important contributions in the field of image recognition and visual data processing obtained through the application of the YOLOv8 algorithm to the problem of identifying objects in images.*

**Keywords:** *You Only Look Once (YOLO), Google Colab, Living Pharmacy, Roboflow*

## 1. Pendahuluan

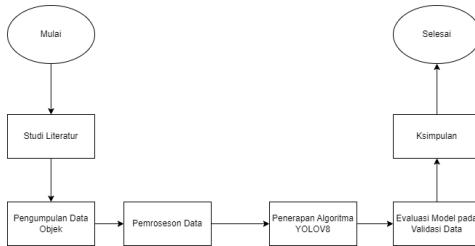
Perkembangan dunia semakin pesat, canggih dan dunia Kesehatan sangat maju dan modern. Salah satu obat-obatan yang mudah ditemui khususnya di Indonesia dan masih banyak digunakan oleh Masyarakat Indonesia adalah Apotik Hidup. Untuk mengetahui tanaman Apotik Hidup kita dapat melakukan penelitian yang terkait. Dari hasil penelitian didapatkan bahwa pemanfaatan Nguyen Widrow telah berhasil meningkatkan akurasi backpropagation sebanyak 25%. Akurasi klasifikasi dari aplikasi yang telah dikembangkan rata-rata diatas 83%. Hal ini mengindikasikan aplikasi yang telah dibangun dapat menjadi salah satu solusi untuk hal klasifikasi jenis tanaman [1]. Penelitian ini mengembangkan metode untuk melakukan deteksi jenis rambu lalu lintas menggunakan metode YOLO V5 yang merupakan salah satu pengembangan dari metode CNN. Nilai identifikasi model yang diperoleh diantaranya precision sebesar 0,923, recall sebesar 0,826, mAP50 sebesar 0,965, dan mAP50-95 sebesar 0,924 [2]. Hasil pengujian relief yang dilakukan dengan menggunakan metode Minimum Square Difference mencapai akurasi sebesar 84,31% sehingga dapat dikatakan cukup akurat untuk sistem deteksi relief Candi Borobudur. Sedangkan dengan menggunakan metode Normalized Cosine Coefficient persentase keberhasilan sistem hanya 23,52%. Hal ini menyebabkan metode Normalized Cosine Coefficient tidak dianjurkan untuk digunakan dalam sistem ini [3]. Hasil penelitian ini menegaskan bahwa algoritma YOLOv5 efektif dalam mengidentifikasi dan memodelkan variasi warna pada 1266 citra bunga lantana camara melalui putaran pelatihan sebanyak 50 epoch dan membutuhkan waktu selama 0.483 jam [4]. *You Only Look Once (YOLO)* merupakan salah satu model Deep Learning yang dapat digunakan untuk pengenalan objek. Penelitian ini bertujuan untuk pengenalan objek pada citra buah menggunakan YOLO [5]. Berdasarkan pengujian yang telah dilakukan dengan data training, beserta data validasi, serta data test menggunakan algoritma YOLOv5 dapat disimpulkan bahwa metode pendeteksian ini dapat mengenali objek secara konsisten dengan tingkat akurasi yang tinggi [6]. Hasil dari penelitian ini dapat mengenali jenis bunga angrek pada gambar. Hasil pengujian dari dataset peneliti menunjukkan akurasi sebesar 30.7% untuk color space grayscale, 37% untuk color space RGB, 34.6% untuk color space HSI, 41% untuk color space YIQ, dan 40.2% untuk color space oRGB dalam mengenali jenis bunga angrek [7].

Penerapan deteksi 10 buah ini dilakukan pada google colaboratory melalui citra dengan dua kali pengujian. Akurasi dalam deteksi 10 buah ini bisa mencapai lebih dari 90% pada pengujian pertama dari setiap buah dan rata - rata 70% pada pengujian kedua untuk citra diluar data test [8].

## 2. Metode Penelitian

Penelitian ini bersifat eksperimental dengan objek penelitian yaitu tanaman Apotik Hidup (Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat) diperoleh dengan cara mengambil foto secara langsung dari Google. Citra yang telah didapatkan selanjutnya akan diproses melalui beberapa tahapan seperti anotasi, pra- pemrosesan, augmentasi (pelebaran dataset), serta mengeksport dataset ke dalam format *YOLOv8*. Proses tersebut dilakukan secara otomatis dengan memanfaatkan *platform roboflow*. Setelah dataset ter-ekspor ke dalam format *YOLOv8* langkah selanjutnya masuk ke *platform* Google Collab memanggil model *YOLOv8* dengan cara menjalankan code cloud, kemudian mengupload show donload code. Kemudian, dilanjutkan ke beberapa proses yaitu pelatihan, pendeteksian dan evaluasi dataset. *Output* dari seluruh proses ini adalah pemodelan dataset dengan nilai *confusion metric* seperti akurasi, recall, precision, serta f1 score.

Berikut merupakan tahapan dalam metodologi penelitian yang dijelaskan melalui diagram alur penelitian dan dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

### a. Studi Literatur

Pada tahap ini mempelajari berbagai referensi atau rujukan yang mendukung penelitian bersumber dari internet, jurnal, tugas akhir, penelitian dan modul -modul yang berkaitan dengan penerapan algoritma *YOLOv8* untuk sistem deteksi pada dataset baru citra tanaman apotik hidup serta penelitian - penelitian sejenis.

### b. Pengumpulan Gambar

Tahap ini merupakan persiapan yang harus dilakukan sebelum melakukan penelitian. Tahap yang dimaksud adalah pengumpulan gambar tanaman apotik hidup. Data yang tersaji merupakan data primer yang diambil dari mesin Google. Gambar yang diambil terdiri dari 13 jenis tanaman yaitu Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat.

### c. Pemrosesan Data

Setelah data mentah dikumpulkan, langkah selanjutnya adalah tahap pemrosesan data. Data yang sudah ada dimasukkan ke dalam *roboflow* untuk menjalani beberapa proses seperti:

- Anotasi gambar atau pemberian nama gambar berdasarkan 13 jenis tanaman yaitu

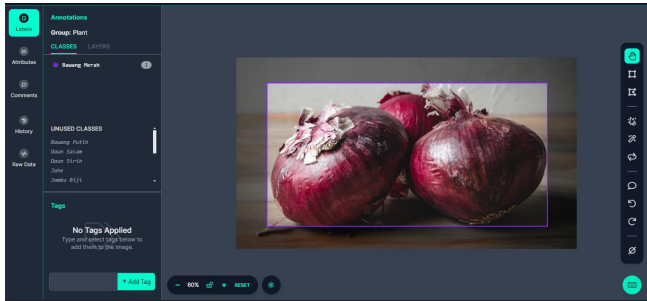
Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat.

- Pra pemrosesan data dilakukan dengan mengekstrak bagian - bagian yang tidak penting seperti latar belakang yang tidak relevan, objek orientasi otomatis sekaligus mengubah ukuran menjadi 640 x 640 piksel.
  - Ekspor data ke dalam format *YOLOv8*
- d. Penerapan Algoritma *YOLOv8*
- Setelah data melewati proses augmentasi, tahap selanjutnya adalah melakukan proses penerapan algoritma *YOLOv8* pada data tindakan ini merupakan proses training data yang ada untuk mengupload file Lantana zip dan data yam1 ke dalam codingan Google Collab lalu melakukan proses runing di dalam coding setelah itu menginput nilai epoch berapa kali data akan dilatih. Setelah hasil nilai keluar lanjut ke proses pendeteksian dengan cara memasukan gambar yang ingin dideteksi ke codingan dalam Google Collab setelah itu model akan segera membaca dan menampilkan hasil deteksi. Setelah objek benar maka tahap selanjutnya melihat dan mencocokkan hasilnya. Jika objek yang ada sudah terdeteksi sama dengan jenis warna pada lantana camara untuk tiap klasifikasi atau kelasnya, maka dapat dikatakan proses training nya berhasil
- e. Evaluasi Model pada Validasi Data
- Setelah data melewati proses augmentasi, tahap selanjutnya adalah melakukan proses penerapan algoritma *YOLOv8* pada data tindakan ini merupakan proses training data yang ada untuk mengupload show donlowad code ke dalam codingan Google Collab lalu melakukan proses runing di dalam coding, setelah itu menginput nilai epoch berapa kali data akan dilatih. Setelah hasil nilai keluar lanjut ke proses pendeteksian dengan cara memasukan gambar yang ingin dideteksi ke codingan dalam Google Collab setelah itu model akan segera membaca dan menampilkan hasil deteksi. Setelah objek benar maka tahap selanjutnya melihat dan mencocokkan hasilnya. Jika objek yang ada sudah terdeteksi sama dengan jenis warna pada lantana camara untuk tiap klasifikasi atau kelasnya, maka dapat dikatakan proses trainingnya berhasil.
- f. Kesimpulan
- Tahap ini model melakukan prediksi pada validasi data, dan hasil prediksi bisa dibandingkan dengan label sebenarnya. Evaluasi ini memberikan wawasan tentang seberapa baik model dapat melakukan prediksi pada data yang belum pernah dilihat sebelumnya hasil ini dapat dilihat dari hasil best epoch dengan nilai rata-rata accuracy, precision, recall dan MAP yang dihasilkan oleh Google Collab.

### 3. Hasil dan Diskusi

#### 3.1. Implementasi *YOLOv8*

Tahap Implementasi YOLO dimulai dengan mempersiapkan dataset 13 jenis tanaman yaitu Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat. Dengan total gambar sebanyak 1204 gambar. Gambar yang dikumpulkan kemudian dimasukan ke dalam *robottow* dilakukan beberapa perlakuan seperti anotasi gambar atau pemberian nama kelas pada objek, hal ini dapat dilihat pada Gambar 2

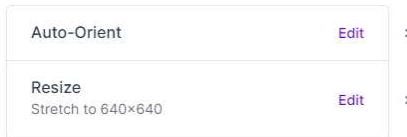


Gambar 2. Contoh Anotasi Objek

Setelah melalui tahap anotasi lanjut ke tahap preprocessing di mana tahapan ini dilakukan dengan cara mengekstrak bagian - bagian yang tidak penting seperti latar belakang yang tidak relevan, objek orientasi diatur otomatis sekaligus mengubah ukuran menjadi 640 x 640 piksel. Seperti yang diperlihatkan pada Gambar 3.

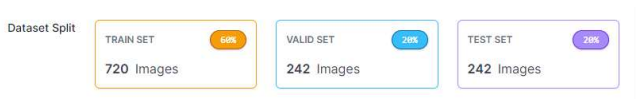
### Preprocessing

Decrease training time and increase performance by applying image transformations to all images in this dataset.



Gambar 3. Tampilan Gambar *Preprocessing*

Contoh tampilan dataset yang dibagi dengan menggunakan *platform roboflow* dapat dilihat pada Gambar 4. Dataset dibagi dengan proporsi 60% untuk Train set, 20% untuk Valid set, dan sisanya 20% untuk Test set. Proporsi pembagian ini merujuk pada penelitian terdahulu yang berhasil dilakukan oleh Keraf, dkk tahun 2024[1]



Gambar 4. Tampilan Dataset Split

Setelah dataset disiapkan langkah selanjutnya masuk ke *platform* Google Collab, dan lanjut melakukan cloning, dimana proses ini menggambarkan sebagian source code yang digunakan untuk memanggil script algoritma *YOLOv8* dalam laman pekerjaan di Google Collab. Hal ini

terjadi karena sifat Google Collab yang bekerja secara runtime (sesaat) yang tidak memerlukan ruang penyimpanan dan spesifikasi laptop yang canggih langkah ini dapat dilihat pada Gambar 5.

```
# Pip install method (recommended)

!pip install ultralytics==8.0.196

from IPython import display
display.clear_output()

import ultralytics
ultralytics.checks()
```

Gambar 5. Tampilan source code untuk mengunduh *YOLOv8*

Setelah mengunduh *YOLOv8* ke dalam Google Collab, selanjutnya mengunduh Show Donload Code yang didapatkan di *Platform roboflow*. Hal ini perlu dilakukan agar memungkinkan Google Collab untuk mengakses ke data atau berkas yang diperlukan untuk melatih ataupun menguji model. Hal ini dapat dilihat pada Gambar 6 dimana Show Donload Code.

```
!mkdir {HOME}/datasets
%cd {HOME}/datasets

!pip install roboflow

from roboflow import RoboFlow
rf = RoboFlow(api_key="gukztetzd5ZKzn4rTy2N")
project = rf.workspace("belajar-sljic").project("live-pharmacy")
version = project.version(2)
dataset = version.download("yolov8")
```

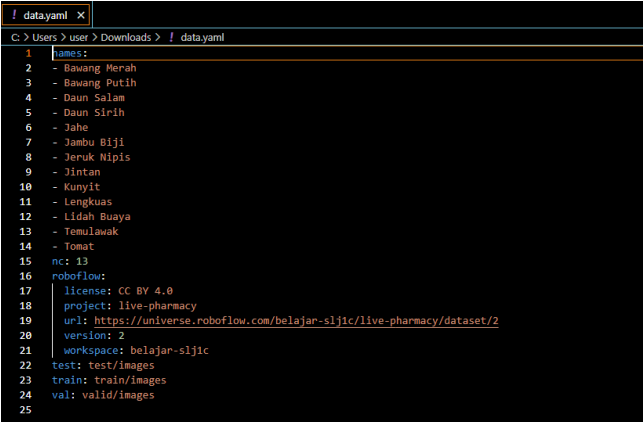
Gambar 6. Tampilan *Show Donload Code*

Langkah selanjutnya masuk ke tahap pelatihan data yang dapat dilihat pada Gambar 7, dimana tahap `!yolo task=detect mode=train model=YOLOv8s.pt data={dataset.location}/data.yaml epochs=25 imgsz=800 plots=True`, bagian ini menggunakan program *YOLOv8* untuk melaksanakan training data. `model=YOLOv8s.pt`, menentukan model awal yang akan digunakan untuk pelatihan. Dalam hal ini, `YOLOv8s.pt` adalah model *YOLOv8* versi ringan yang sudah disediakan. `data={dataset.location}/data.yaml`, menentukan lokasi file konfigurasi data pelatihan Anda. Ganti `{dataset.location}` dengan nama folder aktual dataset Anda. File `.yaml` ini berisi informasi penting tentang dataset, seperti lokasi gambar, label objek, dan lainnya. `epochs=25`, menentukan jumlah iterasi pelatihan. Model akan dilatih selama 25 iterasi. `imgsz=800`, menentukan ukuran gambar yang akan digunakan selama pelatihan. Dalam hal ini, gambar akan diubah ukurannya menjadi 800x800 piksel. `plots=True`, menampilkan grafik pelatihan selama proses berlangsung.

```
%cd {HOME}

!yolo task=detect mode=train model=yolov8s.pt data={dataset.location}/data.yaml epochs=25 imgsz=800 plots=True
```

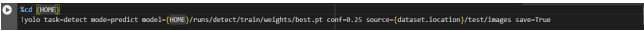
Gambar 7. Tampilan *Source Code Untuk Training Data*



```
1 names:
2   - Bawang Merah
3   - Bawang Putih
4   - Daun Salam
5   - Daun Sirih
6   - Jahe
7   - Jambu Biji
8   - Jeruk Nipis
9   - Jintan
10  - Kunyit
11  - Lengkuas
12  - Lidah Buaya
13  - Temulawak
14  - Tomat
15 nc: 13
16 roboflow:
17   license: CC BY 4.0
18   project: live-pharmacy
19   url: https://universe.roboflow.com/belajar-sljic/live-pharmacy/dataset/2
20   version: 2
21   workspace: belajar-sljic
22 test: test/images
23 train: train/images
24 val: valid/images
25
```

Gambar 8. Tampilan Data Yaml

Setelah melakukan training data, langkah berikutnya melakukan pendeteksian objek seperti terlihat pada Gambar 9. Proses ini diperlukan untuk menguji apakah data train berhasil dikenali oleh model atau tidak. Proses ini dilakukan dengan menjalankan perintah python task=detect. Untuk %cd (HOME), Ini akan mengubah direktori kerja ke direktori home Anda di Google Colab. `!yolo task=detect mode=predict model={HOME}/runs/detect/train/weights/best.pt conf=0.25 source={dataset.location}/test/images save=True`, bagian ini menggunakan program *YOLOv8* untuk melakukan deteksi objek. Deteksi objek, mencari dan mengenali objek tertentu dalam gambar. Mode prediksi, menggunakan model yang sudah dilatih sebelumnya untuk memprediksi objek dalam gambar baru. `model={HOME}/runs/detect/train/weights/best.pt`, menentukan lokasi file bobot model terlatih Anda. Bobot model, parameter yang dipelajari model selama proses pelatihan dan berisi informasi penting untuk melakukan deteksi objek. `conf=0.25`, mengatur ambang kepercayaan minimum. Model hanya akan menampilkan deteksi dengan skor kepercayaan minimal 25% (0.25). Skor kepercayaan, Nilai antara 0 dan 1 yang menunjukkan seberapa yakin model terhadap deteksi objek. `source={dataset.location}/test/images`, Menentukan lokasi gambar yang ingin Anda deteksi objeknya. Ganti `{dataset.location}` dengan nama folder aktual dataset Anda. `save=True`, menyimpan hasil deteksi ke dalam gambar baru. Gambar ini akan memiliki kotak pembatas yang digambar di sekitar objek yang terdeteksi beserta labelnya.



```
!yolo task=detect mode=predict model={HOME}/runs/detect/train/weights/best.pt conf=0.25 source={dataset.location}/test/images save=True
```

Gambar 9. Tampilan Source Code Untuk Menjalankan Deteksi Objek

### 3.2. Performance Metric

Evaluasi pelatihan dengan epoch atau iterasi 50 dapat dilihat pada Gambar 10, dimana pada 13 kelas jenis tanaman yaitu Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat, dengan nilai image 360 merujuk pada gambar panorama 360 derajat. Gambar panorama 360 derajat adalah gambar yang mencakup seluruh lingkaran visual sehingga dapat dilihat ke segala arah secara vertikal

maupun horizontal. Untuk jumlah instances pada setiap kelas berbeda-beda karena model deteksi objek dilatih untuk mengenali dan menempatkan objek-objek yang muncul dalam gambar panorama 360 derajat sehingga dari masing-masing kelas secara berturut-turut model mampu mengenali sebanyak 36, 22, 15, 13, 25, 42, 20, 19, 8, 12, 15, 29, 113, dimana hasil ini bergantung pada karakteristik objek, kerumitan dataset, ukuran dataset pelatihan, kualitas anotasi, parameter pelatihan seperti jumlah Epoch dapat mempengaruhi jumlah instances. Untuk mAP50 (mean average precision) atau metrik yang mencakup precision dan recall pada bagian ambang batas, mAP dihitung dengan cara menghitung nilai rata-rata dari precision sehingga mAP pada rentang nilai ambang deteksi 0,5 hingga 0,95 (mAP50-95) dapat memberikan gambaran lebih tentang evaluasi kinerja model secara menyeluruh di berbagai tingkat ambang deteksi, sehingga nilai mAP yang didapat pada masing-masing kelas di atas 0.27%.

```

/content
Ultralytics YOLOv8.0.196 Python-3.10.12 torch-2.2.1+cu121 CUDA:0 (Tesla T4, 15102MiB)
Model summary (fused): 168 layers, 11118615 parameters, 0 gradients, 28.5 GFLOPs
val: Scanning /content/datasets/live-Pharmacy-2/valid/labels.cache... 242 images, 0 backgrounds, 0 corrupt: 100% 242/242 [00:00<], 71
/user/lib/python3.10/multiprocessing/popen_fork.py:66: RuntimeWarning: os.fork() was called. os.fork() is incompatible with multithread
self.pid = os.fork()
Class      Images  Instances  Box(P  R      mAP50  mAP50-95): 100% 16/16 [00:00<0:00, 1.96it/s]
all        242     369       0.596  0.585  0.688  0.273
Bawang Merah  242     36       0.54  0.361  0.399  0.198
Bawang Putih  242     22       0.811  0.682  0.7   0.392
Daun Salam    242     15       0.623  0.8   0.784  0.375
Daun Sirih    242     13       0.776  0.801  0.896  0.467
Jahe          242     25       0.545  0.64  0.604  0.266
Jambu Biji    242     42       0.706  0.686  0.73  0.367
Jeruk Nipis   242     20       0.786  0.737  0.805  0.385
Jintan        242     19       0.61  0.684  0.695  0.266
Kunyit        242     8        0.122  0.375  0.177  0.0414
Lengkuas      242     12       0.545  0.583  0.476  0.215
Lidah Buaya   242     15       0.422  0.487  0.547  0.23
Temulawak    242     29       0.67  0.276  0.555  0.267
Tomat         242     113      0.596  0.496  0.535  0.227

Speed: 2.8ms preprocess, 14.2ms inference, 0.8ms loss, 4.8ms postprocess per image
Results saved to runs/detect/val
Learn more at https://docs.ultralytics.com/index/val
    
```

**Gambar 10.** Tampilan Parameter Hasil Training Dengan Google Collab

Ringkasan dari hasil ini dapat dilihat pada Tabel 1, pada kolom precision terdapat nilai masing-masing kelas 0.54, 0.811, 0.623, 0.776, 0.545, 0.706, 0.786, 0.61, 0.122, 0.545, 0.422, 0.67, 0.595. Nilai ini merupakan hasil pengukuran deteksi positif yang dilakukan oleh model adalah benar - benar relevan atau akurat. Ini memberikan informasi tentang seberapa baik model membatasi jumlah false positive, yaitu objek yang sebenarnya tidak ada tetapi model mendeteksinya.

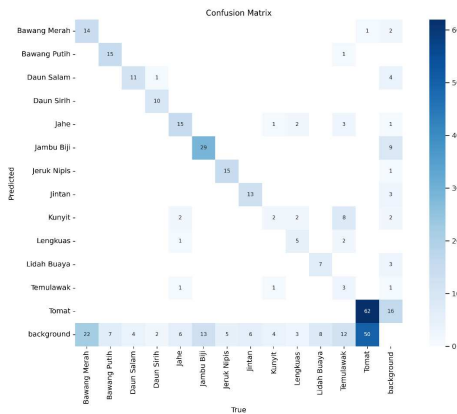
**Tabel 1.** Rincian Hasil Training Model Epoch 50

| Kelas        | Precision | recall | mAP50 |
|--------------|-----------|--------|-------|
| Bawang merah | 0.54      | 0.361  | 0.399 |
| Bawang putih | 0.811     | 0.682  | 0.7   |
| Daun salam   | 0.623     | 0.8    | 0.784 |
| Daun sirih   | 0.776     | 0.801  | 0.896 |
| Jahe         | 0.545     | 0.64   | 0.604 |
| Jambu biji   | 0.706     | 0.686  | 0.73  |
| Jeruk nipis  | 0.786     | 0.737  | 0.805 |
| Jintan       | 0.61      | 0.684  | 0.695 |
| Kunyit       | 0.122     | 0.375  | 0.177 |
| Lengkuas     | 0.545     | 0.583  | 0.476 |
| Lidah buaya  | 0.422     | 0.487  | 0.547 |



| Kelas     | Precision | recall | mAP50 |
|-----------|-----------|--------|-------|
| Temulawak | 0.67      | 0.276  | 0.555 |
| Tomat     | 0.595     | 0.496  | 0.535 |

Hasil evaluasi confusion matrix yang digunakan untuk membandingkan hasil prediksi model dengan ground truth (kebenaran sebenarnya) yang didapat dari perlakuan train data train data, terdapat 13 kelas jenis tanaman yaitu Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat. Untuk yang bernilai true positive (TP), model membuat prediksi benar sesuai dengan kenyataan ditandai dengan Predicted dan True sama, masing-masing kelas bernilai 14, 15, 11, 10, 15, 29, 15, 13, 2, 5, 7, 3, 62. Hal ini dapat dilihat pada Gambar 11 tampilan confusion matrix.



Gambar 11. Funfusion Matrix

#### 4. Kesimpulan

Hasil penelitian ini menegaskan bahwa algoritma *YOLOv8* dalam mengidentifikasi dan memodelkan 13 objek tumbuhan dengan total 1204 gambar diperoleh berupa nilai persentase precision sebesar 59%, recall sebesar 58%, dan nilai MAP sebesar 60%, sehingga nilai rata-rata akurasi di atas 59%. Selain itu, memberikan kontribusi penting dalam bidang pengenalan citra dan pengolahan data visual yang diperoleh melalui penerapan algoritma *YOLOv8* pada masalah identifikasi objek dalam citra. Agar model dapat lebih General dalam pengenalan objek, disarankan untuk memperluas dataset baik menambah objek Lidah Buaya, Daun Sirih, Lengkuas, Temulawak, Jintan, Jahe, Bawang Merah, Bawang Putih, Daun Salam, Jambu Biji, Kunyit, Jeruk Nipis, Tomat, ataupun menambah objek lain. Hal ini dapat meningkatkan kemampuan model dalam mengenali dan mengklasifikasikan objek yang lebih beragam.

#### Daftar Pustaka

[1] A. Muliantara and N. A. S. ER, "Pengembangan Pengklasifikasi jenis Tanaman Menggunakan pendekatan backpropagation Dan Nguyen-Widrow," Jurnal Ilmu Komputer,

- <https://ojs.unud.ac.id/index.php/jik/article/view/18779> (accessed May 6, 2024).
- [2] Bagus Janapriya, A. A. G. (2023). Pengenalan Jenis Rambu Lalu Lintas menggunakan Metode YOLO V5. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 11(4). <https://doi.org/10.24843/jik.2023.v11.i04.p32>.
  - [3] Anggotra, P., & Muliantara, A. (2023). Deteksi Relief Candi Borobudur Menggunakan Metode Template Matching. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 12(1). <https://doi.org/10.24843/jik.2023.v12.i01.p03>.
  - [4] Keraf, M. P. S., Sinlae, A. A. J., & Batarius, P. (2024). Pemodelan Warna Pada Dataset Baru Citra Bunga Lantana Camara Menggunakan Algoritma YOLOv5. *Jurnal Mnemonic*, 7(1), 108-117.
  - [5] Saputra, D. H., & Imran, B. (2023). Object Detection Untuk Mendeteksi Citra Buah-Buahan Menggunakan Metode Yolo. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 2(2), 70-80.
  - [6] Lusiana, L., Wibowo, A., & Dewi, T. K. (2023). Implementasi Algoritma Deep Learning You Only Look Once (YOLOv5) Untuk Deteksi Buah Segar Dan Busuk. *Paspalum: Jurnal Ilmiah Pertanian*, 11(1), 123-130.
  - [7] Prayogo, D. M., Gunadi, K., & Setyati, E. (2020). Pengenalan Jenis Bunga Anggrek Menggunakan Metode Color Local Binary Pattern dan Support Vector Machine. *Jurnal Infra*, 8(1), 242-248.
  - [8] HWB, N., Mailoa, E., & Purnomo, H. D. (2020). Deteksi Buah untuk Klasifikasi Berdasarkan Jenis dengan Algoritma CNN Berbasis YOLOv3. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(3), 476-481.

Halaman ini sengaja dibiarkan kosong

## Pemanfaatan Asap Detektor untuk Pembuatan Smart Garasi Berbasis IoT

Jamrud Ivan Hartono<sup>1</sup>, I Gede Surya Rahayuda<sup>2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>jamrud.hartono@email.com  
<sup>2</sup>igedesuryarahayuda@unud.ac.id

### Abstract

*As a response to increasing safety and efficiency problems, utilizing the Packet Tracer simulation, this research aims to create a smart garage that provides a new solution to the public to provide warnings to avoid fires, vehicles are not left in closed garages for too long so that things don't happen. dangerous things and to increase garage efficiency. This research uses experimental and quantitative methods, which are the methods used to design a system. With this system, garages will avoid theft and fires to quickly apply sensor advances in various situations, with a focus on vehicles that stay too long in the garage and smartphone-enabled garages, which is an important problem in many homes. Utilizes Packet Tracer simulation, emphasizing factors such as placement, sensitivity settings, and effective use. This article provides valuable insight into potential opportunities for future research and innovation in the field of smoke detection. This research highlights the importance of proper use and maintenance in maximizing the capabilities of smoke detectors.*

**Keywords:** Smoke Detector, Garage, Packet Tracer, Experimental Evaluation.

### 1. Pendahuluan

Garasi adalah ruangan yang dirancang khusus untuk menyimpan dan melindungi kendaraan bermotor seperti mobil, sepeda motor, atau sepeda. Garasi biasanya terletak di dekat rumah atau bangunan lainnya dan berfungsi sebagai tempat parkir yang aman. Garasi adalah ruangan yang penting dan berfungsi sebagai tempat parkir, penyimpanan, dan bahkan sebagai ruang tambahan. Bagi pemilik properti, Garasi memberikan rasa aman, perlindungan, dan kesejahteraan. Keamanan berbasis IoT adalah sistem yang memungkinkan pengguna memantau dan mengelola keamanan mereka sendiri melalui perangkat elektronik yang terhubung ke internet, seperti komputer, tablet, dan ponsel pintar. Sistem ini menggabungkan sensor detektor asap dengan perangkat elektronik berupa sirene yang ditempatkan di garasi untuk memberikan akses yang lebih praktis, aman, dan efisien. Dalam penelitian ini, implementasi asap detektor akan disimulasikan menggunakan Packet Tracer, sebuah platform simulasi yang umumnya digunakan untuk merancang dan menguji jaringan komputer. Asap detektor merupakan perangkat elektronik yang dirancang untuk mendeteksi asap dalam suatu area tertentu. Ketika asap terdeteksi, perangkat tersebut dapat memberikan sinyal ke sirene dan membuka garasi secara otomatis. Apabila asap tersebut itu berkurang di titik tertentu, sirene dan pintu garasi otomatis tertutup kembali. Jika ingin membuka garasi atau menyalakan sirene tanpa adanya asap kita bisa juga menyalakannya melalui smartphone. Dengan memanfaatkan teknologi asap detektor, diharapkan dapat meningkatkan tingkat keamanan secara signifikan dan membuat garasi menjadi lebih praktis. Melalui pemanfaatan garasi berbasis IoT, pengguna dapat membuat dan mengelola garasi sendiri dari jarak jauh menggunakan aplikasi yang terhubung ke sistem. Hal ini cukup membantu ketika pengguna ingin memberikan akses kepada orang lain tanpa harus memberikan bukti fisik dan menjaga keamanan agar tidak ada terjadinya kebakaran atau kendaraan yang hidup terlalu lama. Selain itu, sistem ini juga dapat memberikan notifikasi berupa sirene kepada pengguna ketika pintu garasi terbuka atau tertutup, sehingga pengguna dapat memastikan keamanan garasi mereka. Dengan demikian, diharapkan penelitian ini dapat

memberikan kontribusi yang berarti dalam pemahaman dan penerapan teknologi asap detektor sebagai alat yang bisa membangun smart garasi, serta mendorong masyarakat menggunakan garasi dengan lebih aman dan nyaman.

## 2. Metode Penelitian

Penelitian ini menggunakan metode penelitian eksperimental untuk mengevaluasi efektivitas system smart garasi menggunakan Paket Tracer dengan melibatkan DLC 100 Home Gateway, Asap Detector, Smartphone PT, Siren, Garage Door dan Old Car. Untuk pengujian seperti yang ditunjukkan pada Gambar 1 di bawah Konfigurasinya adalah sebagai berikut:



Gambar 1. Desain Sistem

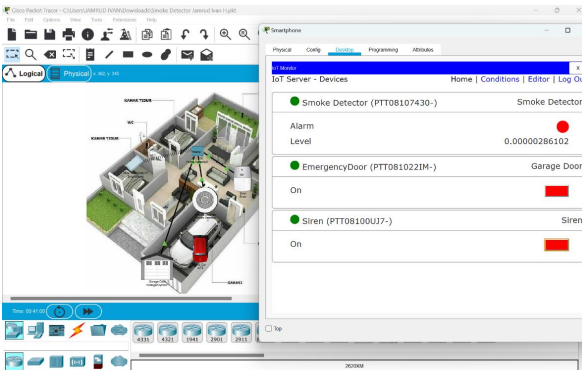
- a. DLC Home Gateway
  - Membuat halaman web dengan nama pengguna dan kata sandi untuk menghubungkan dan mendapatkan kendali sistem.
  - Registrasi dapat dilakukan pada router ini.
  - Alamat IP ditetapkan sebagai 1.1.1.1 secara dinamis.
- b. Smartphone
  - Hubungkan ke sistem dengan masuk ke web browser dan memasukkan IP server registrasi dan login menggunakan ID dan Password.
  - Alamat Ip ditetapkan sebagai 1.0.0.10 secara dinamis.
- c. Detektor Asap
  - Detektor asap dapat menemukan jenis asap apa pun, misalnya: Ketika menghidupkan kendaraan, detektor asap akan mendeteksinya. Dan dalam proyek ini, ketika tingkat asap melampaui 0,09, kondisi tertentu akan terpicu seperti siren dinyalakan dan pintu garasi terbuka. Jika tingkat asap di antara 0.03 – 0.04 siren berhenti dan pintu garasi akan kembali tertutup.
  - Melalui penggunaan pengaturan konfigurasi I/O lanjutan, yaitu pengaturan adaptor jaringan (PT-IOT-NM-1CFE).
  - Alamat IP dinamis diberikan menggunakan DHCP.
- d. Siren
  - Sirene adalah alat yang digunakan untuk peringatan akan bahaya.

- Terhubung ke Home Gateway menggunakan pengaturan lanjutan dalam konfigurasi I/O yaitu (PT-IOT-NM-1CFE) pengaturan adaptor jaringan.
  - Alamat IP dinamis diberikan menggunakan DHCP.
- e. Old Car
- Old car disini berguna untuk menimbulkan asap atau sebagai simulai kendaraan yang ada dalam garasi.
- f. Pintu garasi
- Pintu masuk atau keluar kendaraan disebut pintu garasi. Dalam kasus kami, hal ini sangat penting karena pintu garasi berukuran besar dan dapat membantu udara keluar ketika terjadi kebakaran, melepaskan karbon dioksida dan gas Deteksi Asap dan Pencegahan Kebakaran lainnya ke udara dan membantu setiap orang mengambil udara bersih jika ada yang terjebak di dalam rumah.
  - Melalui penggunaan pengaturan konfigurasi I/O lanjutan, yaitu pengaturan adaptor jaringan (PT-IOT-NM-1CFE).
  - Alamat IP dinamis diberikan menggunakan DHCP.

### 3. Hasil dan Pembahasan

Berikut adalah hasil untuk mengetahui apakah smart garasi dengan asap detektor sudah sesuai dengan yang dikehendaki serta sudah bekerja dengan baik dan juga untuk mengetahui apakah komponen smart garasi dengan asap detektor bekerja dengan baik dan tanpa kendala.

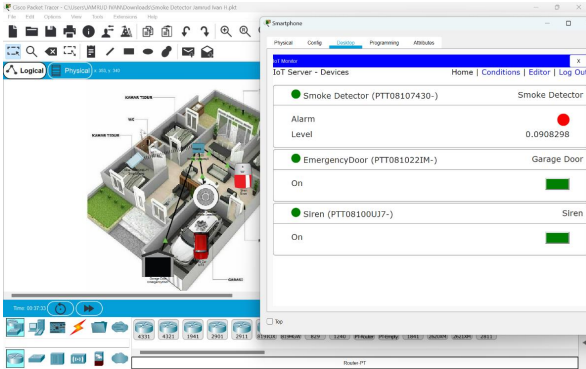
#### 3.1 Saat Level Asap Kurang dari 0.09



Gambar 2. Sistem Ketika Tidak Mendeteksi Asap

Secara teknis, ketika detektor asap mencatat tingkat asap di bawah 0,9, ini menandakan bahwa smart garasi belum mengidentifikasi asap mobil atau keberadaan asap yang menimbulkan kebakaran, dan akibatnya, tidak ada respons atau tindakan otomatis yang akan dimulai oleh sistem.

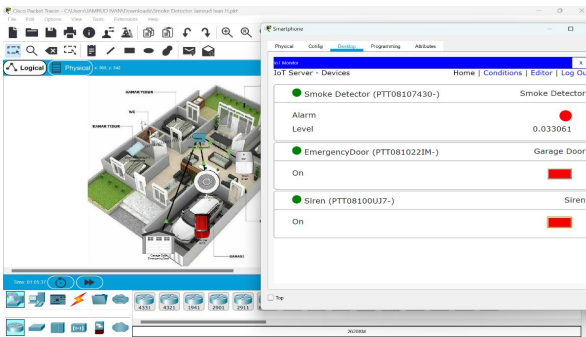
### 3.2 Saat Level Asap Lebih dari 0.09



Gambar 3. Sistem Ketika Mendeteksi Asap

Dalam Smart Garasi, ketika detektor asap mendeteksi kadar asap melebihi 0,09, hal ini menunjukkan bahwa sistem deteksi asap telah mengidentifikasi adanya asap. Selanjutnya, sistem akan menjalankan instruksi yang telah diprogram sebelumnya, seperti mengaktifkan mekanisme pembukaan pintu garasi dan membunyikan siren. Dalam hal ini siren bisa dimatikan melalui smartphone, tapi hanya pintu garasi saja yang tidak bisa ditutup jika level asap tidak turun sampai 0.03 – 0.04.

### 3.3 Saat Level Asap di Antara 0.03 – 0.04

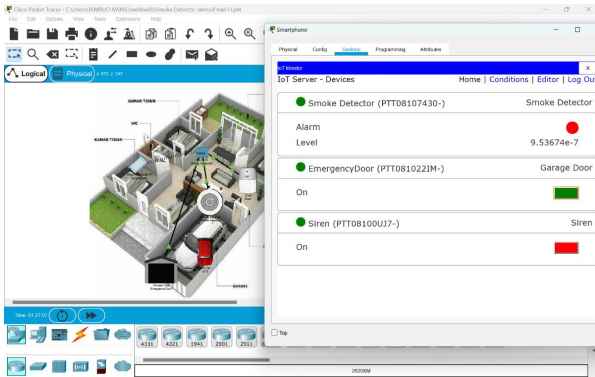


Gambar 4. Sistem Ketika Menutup Pintu Garasi dan Mematikan Siren di Level Tertentu

Ketika detektor asap mendeteksi kadar asap sudah menurun di antara 0.03 – 0.04 maka pintu

garasi akan menutup dan mematikan siren secara otomatis. Terlihat cara ini lebih efisien jika harus menutup pintu garasi dan mematikan siren secara manual.

### 3.4 Saat Membuka Pintu Garasi Tanpa Asap



Gambar 5. Membuka Pintu Garasi Lewat Smartphone

Saat kita ingin membukanya tanpa adanya asap atau kita dari luar ingin masuk ke garasi, kemungkinan tidak ada asap dari dalam, jadi Smart Garasi ini juga bisa memakai smartphone untuk membuka pintu garasi tanpa adanya asap.

## 4. Kesimpulan

Proyek "Pemanfaatan Asap Detektor Untuk Pembuatan Smart Garasi berbasis IoT" secara efektif memamerkan sistem deteksi asap untuk pintu garasi dan alat untuk peringatan akan bahaya. Sistem ini mengintegrasikan detektor asap, mekanisme peringatan, dan mendeteksi keberadaan asap dengan cepat, memberi tahu pemilik kepentingan terkait. Cisco Packet Tracer berperan penting dalam memfasilitasi desain dan simulasi jaringan untuk proyek ini. Selain itu, proyek ini menunjukkan skalabilitas, kemampuan beradaptasi terhadap beragam tata letak bangunan, dan modularitas untuk integrasi komponen yang efisien. Prinsip ini dirancang untuk dapat membuat kemajuan teknologi di kehidupan Masyarakat. Dengan adanya sistem Smart Garasi ini, diharapkan dapat meningkatkan keamanan dan kenyamanan penghuni bangunan. Meskipun demikian, perawatan dan pemeliharaan sistem secara rutin dan berkala tetap diperlukan agar sistem dapat berfungsi dengan baik dan dapat dipercaya dalam memberikan peringatan dan kenyamanan pengguna.

## Daftar Pustaka

- [1] Anusha, V., Madhavi, S., & Gowda, N. C. (2023). Smoke Detection and Fire Prevention Using Cisco Packet Tracer. *International Journal of Computational Learning & Intelligence*, 2(4), 136-142.
- [2] Haeruddin, H., Candra, B., Lee, D., Adiyasa, F., Hadi, H., Sepbianto, S., & Richtan, J. (2022). Rancangan Sistem Keamanan Rumah Berbasis IoT dengan Cisco Packet Tracer. *Telcomatics*, 7(1), 30-41.
- [3] K. Istana, "Home Décor," Pinterest, 2024. <https://id.pinterest.com/pin/17029304830106789> (accessed May 24, 2024).



- [4] Pakarti, Y., & Rahmadewi, R. (2023). Rancang Bangun Deteksi Kebakaran dengan Smoke Detector Addressable Berbasis Outseal PLC Mega V. 3 Standart. *Jurnal Ilmiah Wahana Pendidikan*, 9(11), 259-268.
- [5] Subhana, R. S., & Rahayudaa, I. G. S. Memanfaatkan Motion Sensor untuk Pencegahan Pencurian Barang.

## Penerapan Steganografi untuk Pengamanan Konten Gambar dalam Media Sosial

Gede Eka Putra Wijaya<sup>1</sup>, I Gusti Agung Gede Arya Kadnyanan<sup>2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>deeka2684@gmail.com  
<sup>2</sup>gungde@unud.ac.id

### Abstract

*The rapid development of technology has significantly impacted various aspects of life, especially social media. Social media platforms are not only used for interaction but also for building branding through content uploads. However, this raises concerns about unauthorized use of content that violates Intellectual Property Rights (IPR). Digital watermarking using steganography, specifically the Least Significant Bit (LSB) technique, can protect content by hiding messages without degrading visual quality. This study tests the insertion of text into digital images using two sample logos of different sizes. The test results show Mean Square Error values ranging from 1.8 to 2 and Peak Signal-to-Noise Ratio values between 94.9 and 95.3, indicating excellent visual quality and almost imperceptible differences. This technique effectively protects content from unauthorized use.*

**Keywords:** *Steganography, Least Significant Bit (LSB), Social Media content, Digital Watermarking, Intellectual Property Rights (IPR)*

### 1. Pendahuluan

Perkembangan era yang cepat menyebabkan dampak yang begitu signifikan, terutama dalam domain teknologi. Media sosial merupakan bagian penting dari kehidupan sehari-hari masyarakat di era digital yang semakin maju ini. Mereka tidak hanya menjadi *platform* untuk berinteraksi dan berbagi informasi, tetapi juga sering digunakan untuk menyimpan dan membagikan konten atau karya mereka untuk membangun *branding* dengan mengunggah karya mereka. Namun, seiring dengan kemajuan teknologi, ada kekhawatiran tentang penggunaan konten tanpa seizin pemilik konten yang berkaitan dengan Hak Kekayaan Intelektual (HKI). Isu hak kekayaan intelektual telah menjadi sangat penting dalam era digital yang berkembang pesat. Hak kekayaan intelektual adalah hak yang diberikan kepada pencipta untuk melindungi karyanya yang bersifat intelektual. Perlindungan hak kekayaan intelektual menjadi sangat penting dalam kehidupan karena tantangan yang dihadapi meningkat seiring dengan penggunaan teknologi digital. Perlindungan kekayaan intelektual (hak cipta, paten, merek dagang, desain industri, dan bentuk kekayaan intelektual lainnya) juga sangat penting untuk mengembangkan inovasi dan kreativitas serta mendorong pembangunan ekonomi [1]. *Digital watermarking* salah satu cara untuk mengamankan konten di sosial media, steganografi merupakan suatu metode yang biasa digunakan untuk penyisipan pesan kedalam suatu media sebagai *covernya*. Steganografi merupakan metode menyembunyikan pesan dalam media digital sedemikian rupa sehingga orang lain tidak menyadari ada sesuatu pesan di dalam media tersebut [2]. Dalam hal media sosial, steganografi menjadi semakin relevan karena memungkinkan pengguna melindungi konten pribadi mereka dengan menyisipkan identitas pada konten tanpa merusak visualisasi dari konten tersebut. Salah satu teknik steganografi yang dapat digunakan adalah *Least Significant Bit (LSB)*. Steganografi dengan Teknik LSB memiliki beberapa keunggulan dibandingkan dengan metode steganografi lainnya. Satu keuntungannya adalah sangat mudah untuk diterapkan dan tidak membutuhkan algoritma yang kompleks. Keuntungan lainnya dari steganografi LSB adalah kapasitas penyematan yang tinggi, yang berarti bahwa sejumlah besar data dapat disembunyikan dalam satu gambar. Selain itu, steganografi LSB tidak mempengaruhi kualitas visual gambar

secara signifikan, sehingga pengamat sulit menemukan pesan tersembunyi [3].

## 2. Metode Penelitian

### 2.1. Pengumpulan Data

Pada penelitian ini data yang digunakan oleh sistem yang dikembangkan adalah data citra atau gambar yang akan pengguna unggah ke media sosial. Data uji yang digunakan merupakan 2 sampel karya digital berupa logo.

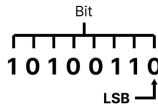
Tabel 1. Data gambar yang diuji

| Karya 1   | Karya 2   |
|---|---|
|  |  |
| 1080 x 1080 piksel  | 1080 x 1350 piksel  |

Pada tabel 1 terdapat dua gambar yang digunakan untuk penelitian ini, karya 1 merupakan gambar dengan ukuran 1080 x 1080 piksel, dan karya 2 merupakan gambar dengan ukuran 1080 x 1350 piksel.

### 2.2. Least Significant Bit (LSB)

Metode ini memanipulasi bit terakhir yang tidak memiliki nilai berarti atau terkecil pada citra gambar.



Gambar 1. Least Significant Bit (LSB)

Metode steganografi LSB memungkinkan penyembunyian informasi rahasia dalam gambar digital tanpa mengubah penampilannya secara signifikan. Misalkan kita menggunakan gambar berukuran 1080 x 1080 piksel sebagai contoh. Dalam gambar ini, setiap piksel memiliki tiga komponen warna RGB (*red, green, blue*), direpresentasikan masing-masing oleh 8 bit, sehingga total 24 bits per piksel. Dengan demikian, kita dapat menyimpan 3-bit pesan dalam setiap piksel dengan mengubah bit dari setiap komponen warna. Dengan total 1.166.400 piksel dalam gambar, kita dapat menyembunyikan total 3.499.200-bit atau sekitar 437.400-byte data yang disisipkan. Pesan rahasia, dalam bentuk urutan bit, akan disisipkan ke dalam bit terkecil (LSB) dari setiap

nilai piksel dalam gambar. Pesan ini akan tersembunyi di dalam gambar tanpa mengganggu penampilan visualnya secara signifikan. Proses ekstraksi pesan melibatkan pengestrakan LSB dari setiap nilai piksel dalam gambar, dan urutan bit yang ditemukan akan dikonversi kembali menjadi pesan yang dapat dibaca. Dengan demikian, metode ini memungkinkan penyembunyian pesan yang cukup besar dalam gambar berukuran 1080x1080 piksel tanpa mengurangi kualitas visualnya secara nyata [4].

### 2.3. Analisis Kebutuhan

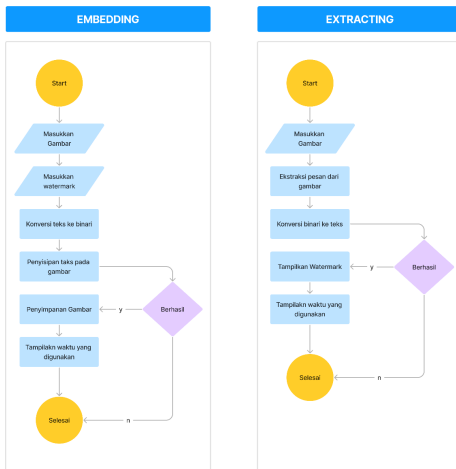
#### 2.3.1. Kebutuhan *Input*

- a. Media yang digunakan untuk *cover image* adalah gambar yang dengan format \*.png.
- b. Pesan yang dimasukkan adalah teks.

#### 2.3.2. Kebutuhan *Output*

- a. File gambar yang telah disisipkan teks dengan format \*.png.
- b. Pesan yang sebelumnya telah di sisipkan atau plain teks yang telah di ekstrak setelah proses dekripsi gambar stegano.

### 2.4. Perancangan Sistem



Gambar 2. Flowchart sistem.

Pada gambar 2 dijelaskan mengenai alur atau *flowchart* dari sistem yaitu proses *embedding* dan proses pengestrakan.

## 2.5. Pengujian Sistem

Pengujian sistem dilakukan dengan mengukur kemiripan dua buah citra, citra pertama adalah citra sebelum dilakukan penyisipan dan citra kedua adalah citra setelah dilakukan penyisipan. Parameter yang digunakan adalah *Mean Square Error* (MSE) dan *Peak Signal-to-Noise Ratio* (PSNR). Parameter ini merupakan parameter yang sering digunakan sebagai indikator untuk mengukur kemiripan dua buah gambar. Semakin mirip kedua citra maka nilai MSE nya semakin mendekati nilai nol, untuk PSNR gambar dikatakan memiliki tingkat kemiripan yang rendah jika nilai PSNR di bawah 30 dB. Berikut merupakan persamaan yang digunakan untuk menghitung kedua parameter tersebut [5].

### 2.5.1. Mean Square Error (MSE)

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [f(i, j) - g(i, j)]^2 \quad (1)$$

### 2.5.2. Peak Signal-to-Noise Ratio (PSNR)

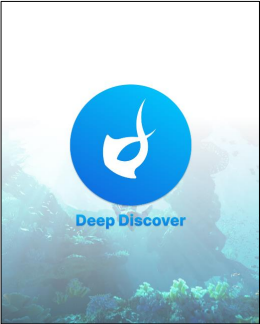
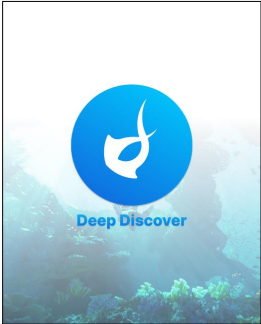
$$10 \log_{10} \frac{255^2}{MSE} \quad (2)$$

## 3. Hasil dan Diskusi

### 3.1. Pengujian Citra

Setelah dilakukan proses *embedding* atau penyisipan teks kedalam gambar, akan dilakukan perbandingan antara gambar sebelum penyisipan dan setelah penyisipan.

Tabel 2. Perbandingan gambar sebelum dan setelah penyisipan

|   | No Sebelum   | Sesudah  |
|---|--|--|
| 1 |  |  |



Terlihat dari tabel 2, tidak terdapat perbedaan yang signifikan antara kedua gambar jika dilihat secara langsung antara gambar yang belum disisipkan pesan dan gambar yang sudah disisipkan pesan.

### 3.2. Pengujian Akurasi

Berikut ini merupakan hasil pengukuran gambar sebelum dan setelah penyisipan menggunakan parameter MSE dan PSNR.

**Tabel 3.** Hasil pengukuran MSE dan PSNR

| No | Gambar  | MSE | PSNR |
|----|---------|-----|------|
| 1  | Karya 1 | 1,8 | 95,3 |
| 2  | Karya 2 | 2,0 | 94,9 |

Dari tabel diatas dapat dilihat hasil dari pengukuran menggunakan parameter MSE dan PSNR, pada gambar karya 1 mendapatkan nilai MSE sebesar 1,8 dan PSNR sebesar 95,3 kemudian untuk gambar karya 2 mendapatkan nilai MSE sebesar 2,0 dan PSNR sebesar 94,9.

## 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan bahwa penyisipan watermarking berupa teks pada gambar konten media sosial berhasil dilakukan menggunakan metode *Least Significant Bit (LSB)* dan mampu memberi perlindungan yang cukup efektif dengan menyembunyikan informasi tanpa merusak estetika dari gambar asli. Dengan adanya informasi yang tersembunyi ini memastikan konten gambar dapat terlindungi dari penggunaan konten yang tidak tepat, karena dapat diidentifikasi secara akurat dan dibedakan berdasarkan informasi watermark yang telah disisipkan. Dalam penelitian ini citra yang digunakan adalah gambar yang akan di posting kemedias sosial, dalam pengujian menggunakan dua sampel gambar. Pengujian yang digunakan untuk mengukur akurasi dari hasil penyisipan adalah menggunakan parameter *Mean Square Error (MSE)* dan *Peak Signal-to-Noise Ratio (PSNR)*. Berdasarkan hasil pengujian dari kedua sample gambar, gambar karya 1 mendapatkan nilai MSE sebesar 1,8 dan PSNR sebesar 95,3. Kemudian untuk gambar karya 2 mendapatkan nilai MSE 2,0 dan PSNR sebesar 94,9. Dari hasil tersebut menunjukkan rata-rata perbedaan kuadrat per piksel antara gambar asli dan gambar yang dimodifikasi adalah 1,8 hingga 2 dan PSNR sebesar 94,9 hingga 95,3 menunjukkan kualitas yang sangat baik, juga menunjukkan perbedaan antara gambar asli dan gambar yang dimodifikasi hampir tidak terlihat walaupun setelah dilakukan penyisipan.

**Daftar Pustaka**

- [1] M. E. Darnia, C. D. Monica, M. Munawardi, and R. Aprillia, "Perlindungan Hak Kekayaan Intelektual di Era Digital," *JERUMI J. Educ. Relig. Humanit. Multidisciplinary*, vol. 1, no. 2, pp. 411–419, 2023, doi: 10.57235/jerumi.v1i2.1378.
- [2] M. Syahril and H. Jaya, "Aplikasi Steganografi Pengamanan Data Nasabah di Standard Chartered Bank Menggunakan Metode Least Significant Bit dan RC4," *Sensasi*, pp. 505–509, 2019, [Online].  
Available: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C505>
- [3] M. Miftahul Amri, M. Waeno, and M. Zain Musa, "LSB Steganography to Embed Creator's Watermark in Batik Digital Arts," *Eng. Sci. Lett.*, vol. 2, no. 01, pp. 27–32, 2023, doi: 10.56741/esl.v2i01.301.
- [4] A. Khuzaifi, F. Fauziah, and I. Fitri, "Teknik Steganography untuk Menyisipkan Pesan pada Sebuah Citra Menggunakan Metode Least Significant Bit (LSB)," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 3, pp. 417–423, 2022, doi: 10.35870/jtik.v6i3.461.
- [5] A. Pamungkas, "Cara Menghitung Nilai MSE, RMSE, dan PSNR pada Citra Digital," *pemrogramanmatlab.com*. [Online].  
Available: <https://pemrogramanmatlab.com/2017/06/04/cara-menghitung-nilai-mse-rmse-dan-psnr-pada-citra-digital/>

# Analisis Keamanan Seed Phrase pada Wallet Cryptocurrency Menggunakan Brute Force

I Kadek Adi Sentana<sup>a1</sup>, I Gusti Ngurah Anom Cahyadi Putra<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>kadekadi1398@gmail.com  
<sup>2</sup>anom.cp@unud.ac.id

## Abstract

*Cryptocurrency wallets serve as digital repositories for securing various digital assets. The security of these wallets is often bolstered by the utilization of seed phrases, which act as a primary means of access and recovery. This research delves into the examination of wallet security, specifically focusing on assessing the resilience against brute force attacks. Brute force methods entail systematically attempting numerous combinations to crack the seed phrase and gain unauthorized access to the wallet contents. Through meticulous experimentation and analysis, this study aims to provide insights into the effectiveness of seed phrase-based security mechanisms in thwarting brute force intrusion attempts. By evaluating different wallet implementations and their susceptibility to brute force attacks, this research contributes to enhancing the robustness of cryptocurrency wallet security measures.*

**Keywords:** wallet, security, brute force, seed phrase, cryptocurrency

## 1. Pendahuluan

Cryptocurrency merupakan bentuk mata uang digital yang sedang mengalami peningkatan signifikan belakangan ini. Berbeda dengan uang kertas fiat yang umumnya digunakan, cryptocurrency tidak mencatat jejak kepemilikan atau transaksi pengguna. Cryptocurrency ini didasarkan pada dua hal, yaitu blockchain yang berperan sebagai buku besar untuk aset dan transaksi, kemudian kriptografi yang berfungsi sebagai sarana untuk melindungi aset. [3] Pengguna cryptocurrency membutuhkan sebuah wallet sebagai tempat untuk menerima dan mengirim koin dari orang lain sebagai pembayaran atas barang atau jasa yang mereka beli atau jual menggunakan aset cryptocurrency. Wallet ini berfungsi sebagai alat penyimpanan dan manajemen yang memungkinkan pengguna untuk mengelola transaksi mereka dalam ekosistem cryptocurrency. Pada dasarnya, ada tiga jenis wallet cryptocurrency yang bisa digunakan oleh pengguna cryptocurrency, cold wallet yang merupakan wallet perangkat keras, hot wallet merupakan wallet yang dihosting di internet, dan warm wallet yang dapat diunduh di komputer pribadi, laptop, dan perangkat seluler pengguna. [4] Meskipun cryptocurrency memiliki keunggulan seperti biaya transaksi yang rendah, eliminasi biaya bank, dan anonimitas dalam berbelanja, namun juga memiliki kelemahan seperti fluktuasi harga yang tinggi, transaksi yang tidak dapat dikembalikan, serta rentan terhadap serangan seperti time jacking, pembajakan ganda, dan penambangan egois. Kekhawatiran serius lainnya adalah resiko kehilangan aset jika seed phrase wallet yang mengelola aset tersebut terlupakan atau hilang. Pada saat ini, terdapat berbagai macam mata uang kripto yang diperdagangkan di pasar dan nilainya terus berubah sesuai dengan permintaan pasar. Terdapat banyak pilihan dompet mata uang kripto yang dapat diakses secara daring, yang memungkinkan pengguna untuk mengelola aset kripto mereka dan melakukan transaksi. Proses pemasangan dan pengaturan dompet pada perangkat komputer atau seluler menghasilkan sebuah seed phrase yang di dalamnya berisi serangkaian 12-24 kata bahasa Inggris yang mewakili entropi dari seed phrase. Ini adalah terjemahan entropi ke dalam format yang mudah dibaca, tetapi mempertahankan informasi yang sama yang terdiri dari 12 hingga 24 kata, yang disarankan untuk disimpan dengan aman sebagai frasa cadangan atau pemulihan. [2] Penelitian ini akan menguji keamanan sebuah seed phrase dalam sebuah dompet



kripto dengan menggunakan serangan brute force. Frase benih tersebut akan diuji dengan mencoba berbagai kombinasi kata-kata yang sesuai dengan standar BIP 39, yang merupakan daftar kata-kata yang diusulkan untuk menciptakan seed phrase. Tujuan pengujian adalah untuk mengevaluasi tingkat keamanan seed phrase tersebut dan melihat seberapa sulitnya bagi penyerang untuk menebak atau merekonstruksi seed phrase yang tepat.

## 2. Metode Penelitian

### 2.1 Gambaran Umum

Pengujian ini akan menggunakan metode serangan brute force. Serangan ini melibatkan pencocokan suatu pola dengan semua kemungkinan kata antara 0 dan nm untuk menemukan keberadaan suatu pola dalam teks. [1] Dalam konteks ini, pola yang dicocokkan adalah seed phrase yang digunakan dalam sebuah wallet cryptocurrency. Seed phrase yang akan digunakan berjumlah 12 kata yang terdiri dari 2048 kata yang telah ditetapkan dalam standar BIP 39. Proses pencocokkan pola ini akan dilakukan dengan menguji setiap kemungkinan kata dari daftar 2048 kata tersebut untuk melihat apakah pola yang sesuai dengan seed phrase dari dompet kripto dapat ditemukan.

- Pertama, pilih seed phrase yang akan diuji. Dalam kasus ini, seed phrase terdiri dari 12 kata yang harus diputuskan untuk wallet cryptocurrency tertentu, seed phrase yang digunakan adalah 'squirrel adjust thunder journey cheese universe pumpkin fever unique ostrich wine taxi'
- Siapkan daftar kata yang akan digunakan dalam serangan brute force. Dalam konteks ini, daftar kata terdiri dari 2048 kata yang telah ditetapkan dalam standar BIP 39 untuk wallet cryptocurrency.
- Lakukan perulangan kombinasi kata dari daftar kata yang telah disiapkan. Proses ini dilakukan untuk menguji setiap kemungkinan kata dari daftar 2048 kata tersebut.
- Cocokkan kata yang diuji dengan seed phrase dari wallet cryptocurrency. Proses ini bertujuan untuk memastikan apakah kata yang diuji sesuai dengan seed phrase yang digunakan dalam wallet.
- Kata yang cocok dengan seed phrase, maka pola tersebut dianggap ditemukan.

### 2.2 Serangan Yang digunakan

#### a. Brute Force

Brute force adalah pendekatan langsung untuk menyelesaikan masalah, biasanya didasarkan pada pernyataan masalah dan definisi konsep yang terlibat. Algoritma brute force bekerja dengan cara yang sangat sederhana, langsung, dan jelas. Dalam pencocokan string, terdapat dua istilah utama: teks dan pola. Teks adalah kata yang dicari dan akan dicocokkan dengan pola. Algoritma Brute Force digunakan untuk memeriksa setiap kemungkinan posisi string dalam teks mulai dari karakter pertama hingga karakter terakhir. Setelah memeriksa karakter pertama, string akan bergeser satu posisi ke kanan, atau karakter akan bergeser ke karakter kedua, ketiga, dan seterusnya. Perbandingan karakter pada teks dapat diselesaikan pada posisi manapun selama proses pencarian, sehingga tidak memerlukan tahap proses tambahan. Sebagai contoh, jika kita memiliki teks "backbone" dan pola yang akan dicocokkan adalah "bone", maka pada karakter pertama "b" akan terjadi kecocokan karena huruf pertama pada pola sama dengan huruf pertama pada teks. Namun, pada karakter kedua "o", terjadi ketidakcocokan karena huruf kedua pada pola tidak sama dengan huruf kedua pada teks. Maka proses pergeseran dilakukan dengan bergeser satu posisi ke kanan. Proses ini akan terus berlanjut sampai seluruh karakter pada pola cocok dengan karakter pada teks. [5]

### 2.3 Perancangan System

Perancangan sistem dijelaskan dengan flowchart. Secara garis besar system berjalan hanya dengan memanggil kata yang tersedia pada BIP 39 dan kemudian mencocokkan pola kata

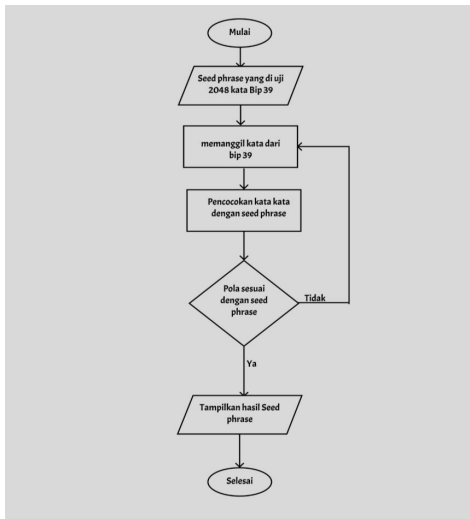
dengan pola kata dari seed phrase yang digunakan yang dijelaskan pada gambar 1.

## 2.4 Pengujian System

Pengujian system bertujuan untuk mengetahui seberapa jauh tingkat keamanan seed phrase dalam menahan serangan brute force, dan menguji apakah system yang dibuat telah berjalan dengan baik tanpa mengalami error.

## 3. Hasil dan Pembahasan

### 3.1 Perancangan Sistem



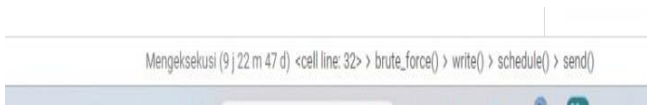
Gambar 1. Flowchart System

Pada gambar 1 ditunjukkan flowchart proses brute force dimulai dengan memanggil pola kata yang ada pada BIP 39, kemudian akan dilakukan proses pencocokan dengan pola kata yang ada pada seed phrase, nantinya jika kata yang dipanggil tidak sesuai dengan seed phrase maka akan diulangi kembali untuk memanggil kata yang ada pada BIP 39, proses ini akan terus berulang hingga pola kata yang dicocokkan sesuai dengan seed phrase, ketika pola berhasil dicocokkan maka system akan berhenti.

### 3.2 Pengujian Sistem

```
*** Streaming output truncated to the last 5000 lines.
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid into
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid invest
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid invite
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid involve
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid iron
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid island
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid isolate
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid issue
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid item
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid ivory
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid jacket
Trying password: abandon ability able about above absent absorb abstract absurd abuse afraid jaguar
```

Gambar 2. Proses pencocokan



Gambar 3. Waktu Proses pencocokan

Pada gambar 2, kita melihat proses pencocokan kata dari daftar kata BIP 39 dengan pola kata seed phrase yang digunakan. Namun, pada gambar 3, terlihat bahwa meskipun proses tersebut telah berlangsung selama 9 jam, belum ada kata yang cocok ditemukan.



Gambar 4. Error akses memory yang berlebihan

Pada gambar 4 ditunjukkan bahwa browser yang digunakan mengalami eror yang dikarenakan jumlah ram memori yang digunakan begitu besar saat menjalankan proses brute force pada seed phrase dengan kombinasi 12 kata

```
***
Trying password: abandon ability able about above absent absorb abstract absurd access source physical
Trying password: abandon ability able about above absent absorb abstract absurd access source piano
Trying password: abandon ability able about above absent absorb abstract absurd access source picnic
Trying password: abandon ability able about above absent absorb abstract absurd access source picture
Trying password: abandon ability able about above absent absorb abstract absurd access source place
Trying password: abandon ability able about above absent absorb abstract absurd access source place
Trying password: abandon ability able about above absent absorb abstract absurd access source pigeon
Trying password: abandon ability able about above absent absorb abstract absurd access source pill
Trying password: abandon ability able about above absent absorb abstract absurd access source pilot
Trying password: abandon ability able about above absent absorb abstract absurd access source plink
Trying password: abandon ability able about above absent absorb abstract absurd access source plowman
Trying password: abandon ability able about above absent absorb abstract absurd access source pipe
Trying password: abandon ability able about above absent absorb abstract absurd access source pistol
Trying password: abandon ability able about above absent absorb abstract absurd access source pitch
Trying password: abandon ability able about above absent absorb abstract absurd access source pizza
Trying password: abandon ability able about above absent absorb abstract absurd access source place
Trying password: abandon ability able about above absent absorb abstract absurd access source planet
Trying password: abandon ability able about above absent absorb abstract absurd access source planet
Buffered data was truncated after reaching the output size limit.
```

Gambar 5. Error dikarenakan ukuran data

Pada gambar 5 diperlihatkan program mengalami eror yang dikarenakan data yang mencapai ukuran batas keluaran.

```
# Kata-kata yang akan digunakan
wordlist = ['play', 'game', 'friend', 'adjust', 'thunder', 'journey', 'cheese', 'universe', 'pumpkin', 'fever', 'taxi']

# seedphrase target yang akan dicoba
target_password = 'taxi adjust thunder game play'

# Panjang kata sandi yang akan dicoba
password_length = 5
```

Gambar 6. Pengurangan jumlah kata dari BIP 39

Di gambar 6, dilakukan pengecilan jumlah kata dari daftar BIP yang sebelumnya mencakup 2048 kata menjadi hanya 11 kata yang dipilih. Selain itu, dilakukan juga pemangkasan panjang pola kata yang sedang diuji yang semulanya 12 menjadi 5 kata.

```

Trying password: taxi adjust friend fever thunder
Trying password: taxi adjust friend fever journey
Trying password: taxi adjust friend fever cheese
Trying password: taxi adjust friend fever universe
Trying password: taxi adjust friend fever pumpkin
Trying password: taxi adjust thunder play game
Trying password: taxi adjust thunder play friend
Trying password: taxi adjust thunder play journey
Trying password: taxi adjust thunder play cheese
Trying password: taxi adjust thunder play universe
Trying password: taxi adjust thunder play pumpkin
Trying password: taxi adjust thunder play fever
Trying password: taxi adjust thunder game play
Password found: taxi adjust thunder game play
'taxi adjust thunder game play'

8s completed at 3:33 PM
```

Gambar 7. Output dari password seed phrase dan waktu yang dihabiskan

Gambar 7 menunjukkan hasil akhir dari pencarian seed phrase yang berhasil ditemukan setelah mengurangi jumlah kata yang diambil dari daftar BIP 39 dan memperpendek panjang seed phrase yang diuji. Selain itu, di bagian bawah gambar, disajikan waktu yang diperlukan untuk menemukan pola seed phrase tersebut, yang mencapai 9 detik.

#### 4. Kesimpulan

Dari penelitian yang telah dilakukan, ditemukan bahwa upaya melakukan serangan brute force terhadap seed phrase yang terdiri dari 12 kata, dengan menggunakan 2048 kata yang tersedia dari BIP 39 sangat sulit untuk berhasil dengan kata lain wallet cryptocurrency masih cukup aman. Kompleksitas masalah ini terjadi karena jumlah kombinasi kata yang luar biasa besar, yang dihasilkan dari 2048 pangkat 12 kata yang digunakan. Proses pengujian pada sistem menunjukkan bahwa meskipun telah berjalan selama 9 jam, belum ada hasil yang memuaskan dalam menemukan seed phrase yang sesuai. Kemudian terjadi error ketika memori dan data yang diperlukan untuk proses ini telah mencapai batasnya. Namun, ketika jumlah kata yang digunakan dari BIP 39 dikurangi, proses brute force tersebut dapat diselesaikan dalam waktu yang sangat singkat, kurang dari 10 detik. Penelitian ini menggambarkan bahwa serangan brute force pada wallet cryptocurrency memungkinkan, namun, hal tersebut hanya dapat dicapai dengan menggunakan perangkat yang canggih dan memerlukan waktu yang cukup lama.

#### Daftar Pustaka

- [1] Barutu, C., & Abdi, N. (n.d.). Brute Force Algorithm Implementation of Dictionary Search 1. <http://ejournal.seaninstitute.or.id/index.php/InfoSains>
- [2] Ledger Academy. (n.d.). Understanding BIP-39: The Origin of Your Seed Phrase. Retrieved May 10, 2024, from <https://www.ledger.com/academy/bip-39-the-low-key-guardian-of-your-crypto-freedom>
- [3] Shaik, C. (2020). Securing Cryptocurrency Wallet Seed Phrase Digitally with Blind Key Encryption. *International Journal on Cryptography and Information Security*, 10(4), 1–10. <https://doi.org/10.5121/ijcis.2020.10401>
- [4] Shaik, C. (2020). Unforgettable User Defined Seed Phrase for Cryptocurrency Wallets. *International Journal on Cryptography and Information Security*, 10(4), 11–20. <https://doi.org/10.5121/ijcis.2020.10402>
- [5] Sinaga, A. (2021). Aditya Sinaga 1, Nuraisana nuraisana 2 [Sistem Pendukung Keputusan Pemilihan Karyawan Tetap pada Trinity Teknologi Nusantara Dengan Metode Moora. 4(1), 6–15

# Perancangan UI/UX Aplikasi Kesehatan Mental “MindEase” Berbasis Mobile dengan Metode Design Thinking

Pande Komang Bhargo Anantha Yogiswara<sup>a1</sup>, I Gusti Ngurah Anom Cahyadi Putra<sup>a2</sup>,

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>yogiswara.2208561067@student.unud.ac.id  
<sup>2</sup>anom.cp@unud.ac.id

## Abstract

The rising global concern over mental health disorders necessitates accessible and quality mental health services. In Indonesia, where 6.1% of the population aged 15 and above experience such disorders, addressing the increasing prevalence becomes imperative. However, barriers like resource scarcity, limited geographic access, and social stigma hinder individuals from accessing professional help. More than 80% of affected individuals lack access to quality, affordable mental health services. In response, the mobile application "MindEase" emerges as an innovative solution, integrating consultation, assessment, education, and training in one platform. Designed to alleviate stress and anxiety, MindEase aims to provide users with a comfortable environment for managing their mental health. Aligned with the Sustainable Development Goals, particularly SDG #3 and #10, MindEase contributes to promoting better mental health and reducing healthcare disparities. Using the design thinking method, the UI/UX of MindEase was developed, resulting in satisfactory usability, as indicated by a System Usability Scale score of 82.8125. However, future research should involve mental health professionals for more comprehensive evaluations.

**Keywords:** User Interface, User Experience, Mobile Application, Mental Health, System Usability Scale (SUS)

## 1. Pendahuluan

Gangguan kesehatan mental seperti depresi, kecemasan, dan kondisi terkait lainnya telah menjadi masalah yang semakin mengkhawatirkan secara global. Di Indonesia pada tahun 2023, statistik menunjukkan bahwa sekitar 6,1% dari populasi yang berusia 15 tahun ke atas mengalami gangguan kesehatan mental [1]. Angka prevalensi yang terus meningkat menyoroti pentingnya akses yang lebih besar terhadap layanan kesehatan mental yang terjangkau dan berkualitas. Namun, berbagai hambatan seperti kurangnya sumber daya, keterbatasan akses geografis, dan stigma sosial-budaya menghalangi individu untuk mendapatkan bantuan profesional yang mereka butuhkan. Lebih dari 80% dari orang yang terpengaruh tidak memiliki kemampuan untuk mencapai perawatan kesehatan mental yang baik dan terjangkau [2]. Dalam konteks ini, aplikasi MindEase muncul sebagai solusi inovatif. MindEase adalah aplikasi kesehatan mental terpadu yang menyediakan konsultasi, asesmen, edukasi, dan pelatihan dalam satu platform. Nama "MindEase" sendiri mencerminkan tujuan aplikasi tersebut. Gabungan kata "mind" yang berarti pikiran atau mental, dan "ease" yang berarti kemudahan atau kenyamanan, memberikan gambaran bahwa MindEase bertujuan untuk memberikan kemudahan atau kenyamanan bagi pengguna dalam mengelola kesehatan mental mereka. Aplikasi ini dirancang untuk membantu meredakan stres dan kecemasan, serta menciptakan kondisi yang lebih nyaman bagi pengguna dalam memantau dan meningkatkan kesehatan mental mereka. MindEase juga sejalan dengan upaya global untuk mencapai Tujuan Pembangunan Berkelanjutan (SDGs), terutama SDG #3 tentang "Kehidupan Sehat dan Sejahtera" dan SDG #10 tentang "Mengurangi Ketimpangan" [3]. Dengan menyediakan akses yang lebih luas terhadap layanan kesehatan mental, MindEase

berkontribusi dalam mempromosikan kesehatan mental yang lebih baik dan mengurangi disparitas dalam perawatan kesehatan.

### 1.1. User Interface

UI (User Interface) merujuk pada semua elemen visual yang terpapar kepada pengguna ketika menggunakan suatu aplikasi atau situs web. Ini meliputi segala komponen seperti tombol, ikon, teks, serta susunan umum dari antarmuka yang digunakan pengguna untuk berinteraksi dengan sistem. Antarmuka pengguna memberikan fasilitas untuk input, yang memungkinkan pengguna untuk mengatur sistem, dan output, yang memungkinkan sistem memberikan informasi kepada pengguna sebagai umpan balik [4]. UI (User Interface) adalah istilah yang digunakan untuk menggambarkan semua komponen yang dirancang khusus untuk berfungsi sebagai alat interaksi dalam suatu perangkat informasi. Ini termasuk semua elemen visual yang terlihat di layar perangkat, yang memfasilitasi interaksi antara pengguna dan sistem tersebut.

### 1.2. User Experience

UX (User Experience) merujuk pada cara pengguna merasakan dan menanggapi penggunaan suatu produk, sistem, atau layanan [5]. Ini melibatkan pemahaman mendalam tentang bagaimana pengguna berinteraksi dengan produk dari awal hingga akhir, termasuk proses pencarian informasi, penggunaan, dan pengalaman setelah penggunaan selesai. Tujuan dari desain UX adalah untuk menciptakan pengalaman yang menyenangkan dan memuaskan bagi pengguna, yang kemudian dapat memengaruhi cara pengguna memahami, mempertahankan, dan merasakan kepuasan terhadap produk, sistem, atau layanan tersebut.

### 1.3. Aplikasi

Aplikasi adalah program komputer yang dibuat untuk menjalankan instruksi yang diberikan oleh pengguna, dengan tujuan memecahkan masalah tertentu menggunakan metode komputasi yang sesuai dengan kebutuhan yang ada [6].

## 2. Metode Penelitian

### 2.1. Design Thinking

Dalam contoh penelitian ini, pendekatan design thinking diterapkan untuk menghasilkan solusi atas permasalahan yang telah dijelaskan sebelumnya dalam bentuk aplikasi mobile. Design thinking adalah metode inovasi yang berorientasi pada manusia, di mana desainer berfokus pada integrasi kebutuhan individu, teknologi, dan kebutuhan bisnis dalam proses perancangan [2]. Metode ini terdiri dari lima tahapan, dengan penjelasan setiap tahapannya sebagai berikut.

#### 2.1.1. Empathize

Empathize adalah tahapan pertama dalam proses design thinking yang bertujuan untuk mendapatkan pemahaman yang mendalam tentang kebutuhan pengguna, termasuk motivasi, perilaku, dan faktor psikologis lainnya [7]. Pada tahap ini, data dikumpulkan melalui observasi, wawancara, dan pengamatan langsung untuk mendapatkan pemahaman yang komprehensif tentang pengguna yang menjadi target, yang dalam konteks ini adalah individu yang berusia 15 tahun ke atas.

#### 2.1.2. Define

Setelah memahami pengguna, tahap ini melibatkan pembentukan pemahaman yang jelas tentang masalah yang dihadapi oleh pengguna. Proses ini mencakup penguraian dan penyimpulan informasi yang diperoleh selama tahap Empathize untuk mengidentifikasi aspek yang paling penting. Dalam tahap ini, data dari hasil wawancara akan disusun menjadi pernyataan menggunakan pendekatan How Might We (HMW) [8].

### 2.1.3. Ideate

Pada tahap ketiga dari proses design thinking, para desainer siap untuk memulai generasi ide. Menciptakan ide-ide kreatif diharapkan dapat mengatasi masalah yang telah diidentifikasi sebelumnya. Saat ini merupakan waktu untuk berpikir secara inovatif, tanpa ada batasan atau penilaian, dan merangsang munculnya ide-ide baru.

### 2.1.4. Prototype

Tahap prototype dalam design thinking adalah langkah keempat dalam proses desain, di mana para peneliti menciptakan model atau prototipe dari ide-ide yang telah dipilih selama tahap ideate [9]. Prototype ini adalah representasi awal dari solusi yang memungkinkan untuk dilakukan pengujian dan evaluasi.

### 2.1.5. Test

Tahap akhir melibatkan pengujian prototipe dengan pengguna untuk memperoleh umpan balik langsung. Meskipun tahap ini merupakan langkah terakhir, namun dapat berulang dan kembali ke tahap-tahap perancangan sebelumnya jika ditemukan kesalahan [9]. Dalam penelitian ini, pengujian kegunaan akan dilakukan menggunakan metode System Usability Scale (SUS). System Usability Scale adalah metode yang mudah dan efektif untuk mengevaluasi kegunaan. Metode ini terdiri dari kuesioner yang terdiri dari 10 pertanyaan dengan lima opsi jawaban yang dapat dipilih oleh responden, mulai dari Sangat setuju hingga Sangat tidak setuju [10].

## 3. Hasil dan Diskusi

### 3.1. Empathize

Dalam contoh penelitian ini, data dari tahap Empathize diperoleh melalui wawancara dengan sepuluh narasumber. Hasil dari tahap Empathize tersebut dapat ditemukan dalam Gambar 1.

|  |   |  |  |   |
|--|---|--|--|---|
| Kalau lagi stress biasanya<br>dengerin lagu atau ke<br>coffee shop<br><small>Widya, 2024</small> | Ragu untuk menggunakan<br>aplikasi kesehatan mental<br>karena khawatir tentang<br>keamanan data pribadi<br>mereka<br><small>Widya, 2024</small> | Akses yang terbatas, jadi<br>pengguna ragu sudah<br>memasukkan aplikasinya<br>tapi tidak bisa digunakan<br><small>Widya, 2024</small>                                    | Sulit dalam memantau<br>kemajuan dalam<br>memperbaiki kesehatan<br>mental<br><small>Widya, 2024</small>                      | Bisa konsultasi dengan<br>muhlis, tepat, dan to the<br>point<br><small>Widya, 2024</small>                                |
| Pingin chat ke<br>seseorang untuk<br>mempertanyakan aspek unik<br><small>Widya, 2024</small>     | Bisa memberikan<br>rekomendasi kegiatan bila<br>memiliki waktu luang, agar<br>tidak gabut<br><small>Widya, 2024</small>                         | Sering lupa bila sedang<br>manggakan aplikasi<br>tersebut jadi kalau bisa<br>ada fitur pengingat seperti<br>"bagaimana harimu<br>sekarang"<br><small>Widya, 2024</small> | gatau sedang stress<br>sampai titik dimana,<br>mungkin adanya fitur<br>mengukur tingkat stress<br><small>Widya, 2024</small> | karna lagi stress, semoga<br>aplikasinya mudah<br>digunakan dan membuat<br>stress berkurang<br><small>Widya, 2024</small> |

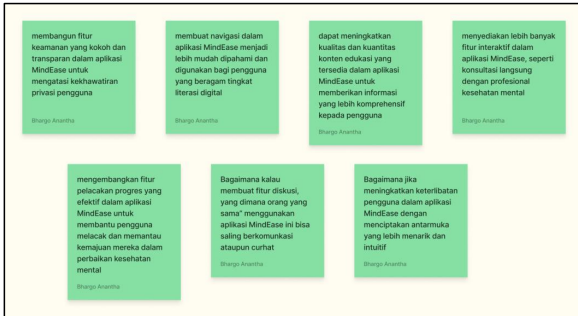
Gambar 1. Hasil Empathize

Dilihat dari hasil wawancara yang tertera pada Gambar 1, terungkap bahwa pengguna memiliki beberapa keinginan fitur pada aplikasi MindEase. Mereka menginginkan fitur-fitur yang memungkinkan mereka untuk melakukan kegiatan yang dapat membantu mereka untuk mengekspresikan perasaan dan emosi mereka, karena dianggap sebagai cara yang efektif dalam mengatasi masalah mental. Selain itu, terdapat harapan untuk adanya fitur yang dapat mengukur tingkat stres, karena beberapa responden mengaku kesulitan dalam mengidentifikasi apakah mereka sedang mengalami stres atau tidak. Aplikasi juga diharapkan dapat memberikan rekomendasi kegiatan bagi pengguna, sehingga mereka dapat melakukan aktivitas yang bermanfaat untuk meningkatkan kesehatan mental mereka.



### 3.2. Define

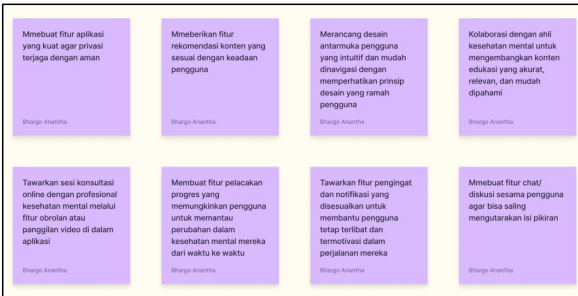
Dalam tahap define, kebutuhan yang telah diidentifikasi dari hasil wawancara akan diartikan menjadi daftar kebutuhan menggunakan pendekatan "how might we". Hasil dari pendekatan ini akan disajikan dalam Gambar 2.



Gambar 2. Hasil Define

### 3.3. Ideate

Dari hasil analisis pertanyaan yang dirumuskan dalam tahap Define, setelah mengidentifikasi akar permasalahan dan pertanyaan yang telah diformulasikan, solusi yang diusulkan oleh peneliti dapat ditemukan dalam Gambar 3.



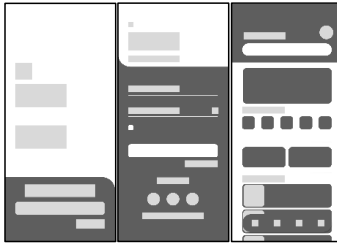
Gambar 3. Hasil Ideate

### 3.4. Prototype

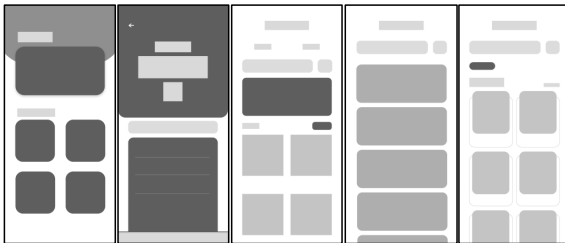
#### 3.4.1. Wireframe

Wireframe dianggap sebagai desain low fidelity karena hanya menggunakan warna hitam dan

putih saja. [11]. Wireframe biasanya digambarkan dalam bentuk sketsa atau diagram sederhana yang menunjukkan posisi dan hubungan antara elemen-elemen seperti tombol, teks, gambar, dan bidang input. Tujuan dari wireframe adalah untuk memberikan gambaran awal tentang interaksi pengguna dengan antarmuka dan tata letak informasi, sehingga desainer dan pengembang dapat mengidentifikasi masalah potensial dan melakukan perbaikan pada struktur sebelum melanjutkan ke tahap desain yang lebih rinci. Di bawah ini adalah beberapa contoh hasil wireframe yang telah disusun:



**Gambar 4.** Wireframe Landing Page menuju Home Page



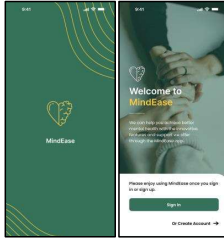
**Gambar 5.** Wireframe Fitur dari MindEase



**Gambar 6.** Wireframe Halaman Profile

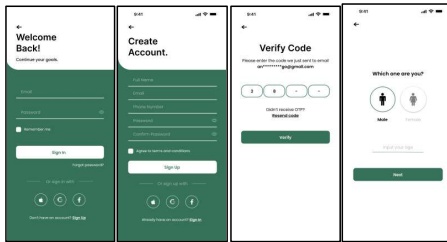
### 3.4.2. Prototype

#### a. Tampilan On Boarding Screen



Gambar 7. Tampilan On Boarding Screen

#### b. Tampilan Sign In & Sign Up



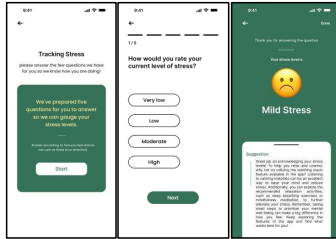
Gambar 8. Tampilan Proses Sign In dan Sign Up

#### c. Tampilan Home Page



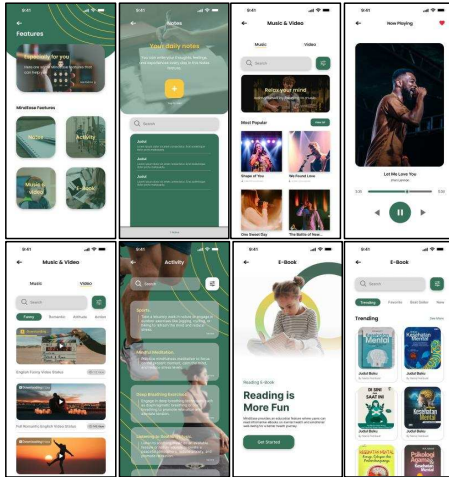
Gambar 9. Tampilan Home Page

d. Tampilan Tracking Stress



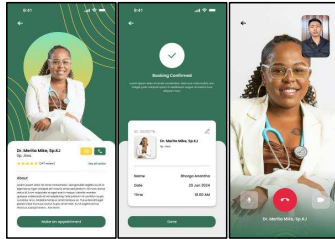
Gambar 10. Tampilan Proses Tracking Stress

e. Tampilan Fitur



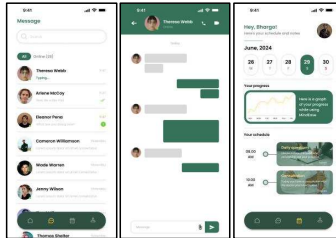
Gambar 11. Fitur-fitur pada MindEase

f. Tampilan Konsultasi Dokter



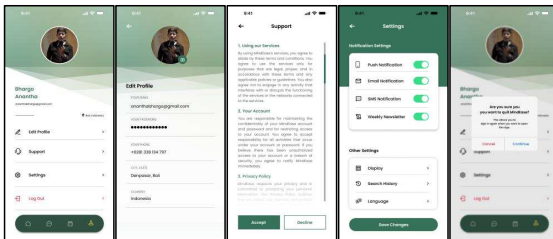
Gambar 12. Booking & Konsultasi Dokter

g. Tampilan Chat, Jadwal dan Progress Kemajuan



Gambar 13. Tampilan Chat, Jadwal dan Progress Kemajuan

h. Tampilan Profile



Gambar 15. Tampilan Profile

### 3.5. Test

Pengujian atau tahap Test dilakukan melalui uji kegunaan dengan menerapkan metode System Usability Scale (SUS). Metode ini melibatkan penggunaan kuesioner yang terdiri dari 10 pertanyaan dengan lima pilihan jawaban untuk responden, mulai dari sangat tidak setuju hingga sangat setuju. Detail kuesioner SUS dapat ditemukan dalam Tabel 1.

Tabel 1. Pertanyaan SUS

| No. | Pertanyaan  |
|-----|---|
| 1   | Saya pikir saya akan sering menggunakan aplikasi ini  |
| 2   | Saya rasa aplikasi ini seharusnya tidak serumit ini   |
| 3   | Saya pikir aplikasi ini mudah untuk digunakan   |
| 4   | Saya pikir saya akan membutuhkan bantuan dari orang teknis untuk dapat menggunakan aplikasi ini |
| 5   | Saya menemukan bahwa berbagai fungsi di aplikasi ini terintegrasi dengan baik                   |
| 6   | Saya pikir terlalu banyak ketidaksesuaian di dalam aplikasi ini                                 |
| 7   | Saya rasa kebanyakan orang akan belajar menggunakan aplikasi ini dengan sangat cepat            |
| 8   | Saya menemukan aplikasi ini sangat susah untuk digunakan  |
| 9   | Saya merasa sangat percaya diri / nyaman dalam menggunakan aplikasi ini                         |
| 10  | Saya perlu mempelajari banyak hal sebelum saya dapat menggunakan aplikasi ini                   |

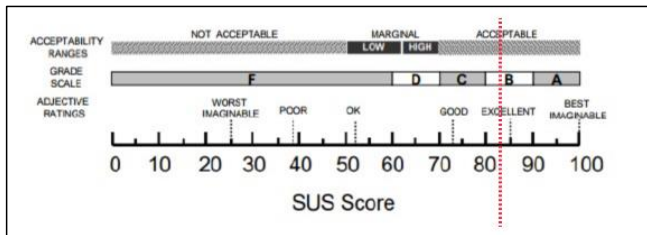
Aturan penilaian untuk System Usability Scale (SUS) terdiri dari beberapa langkah yang harus diikuti. Langkah pertama, untuk pertanyaan dengan nomor ganjil, nilai yang diberikan oleh pengguna dikurangi satu. Kedua, untuk setiap pertanyaan dengan nomor genap, nilai yang diberikan oleh pengguna dikurangi dari lima. Selanjutnya, hasil penjumlahan nilai dari setiap pertanyaan dikalikan dengan 2,5 untuk mendapatkan skor SUS. Proses ini diterapkan untuk setiap responden. Untuk mendapatkan skor rata-rata, skor SUS dari semua responden dijumlahkan dan kemudian dibagi dengan jumlah total responden. [10]. Uji kegunaan melibatkan 40 responden yang berusia 17 tahun ke atas. Respon dari setiap responden dicatat dan terdokumentasi dalam Tabel 2.

Tabel 2. Hasil Responden

| Responden | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | Skor SUS |
|-----------|----|----|----|----|----|----|----|----|----|-----|----------|
| R1        | 5  | 1  | 5  | 1  | 5  | 1  | 4  | 1  | 5  | 1   | 97.5     |
| R2        | 4  | 3  | 4  | 2  | 4  | 2  | 3  | 2  | 4  | 2   | 70       |
| R3        | 4  | 2  | 5  | 1  | 5  | 2  | 5  | 1  | 5  | 1   | 92.5     |
| R4        | 5  | 1  | 5  | 2  | 5  | 3  | 4  | 2  | 4  | 3   | 80       |
| R5        | 5  | 1  | 5  | 1  | 5  | 1  | 5  | 1  | 5  | 1   | 100      |
| R6        | 4  | 3  | 3  | 2  | 4  | 3  | 4  | 3  | 3  | 1   | 65       |
| R7        | 5  | 2  | 5  | 1  | 5  | 1  | 5  | 1  | 5  | 2   | 95       |
| R8        | 5  | 3  | 5  | 1  | 5  | 3  | 4  | 3  | 4  | 3   | 75       |

| Responden             | P1  | P2  | P3  | P4  | P5  | P6  | P7  | P8  | P9  | P10 | Skor SUS       |
|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------------|
| R9                    | 5   | 2   | 4   | 2   | 5   | 2   | 5   | 2   | 4   | 2   | 82.5           |
| R10                   | 3   | 2   | 5   | 1   | 3   | 2   | 4   | 1   | 4   | 2   | 77.5           |
| R11                   | 5   | 2   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 2   | 95             |
| R12                   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 1   | 100            |
| R13                   | 4   | 2   | 4   | 2   | 4   | 2   | 5   | 1   | 4   | 1   | 82.5           |
| R14                   | 4   | 3   | 3   | 1   | 3   | 3   | 5   | 2   | 4   | 1   | 72.5           |
| R15                   | 5   | 2   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 1   | 97.5           |
| R16                   | 4   | 2   | 5   | 1   | 5   | 2   | 5   | 1   | 5   | 2   | 90             |
| R17                   | 5   | 1   | 4   | 1   | 5   | 1   | 4   | 1   | 5   | 1   | 95             |
| R18                   | 4   | 1   | 5   | 2   | 4   | 1   | 5   | 1   | 4   | 2   | 87.5           |
| R19                   | 5   | 2   | 5   | 1   | 5   | 1   | 5   | 2   | 5   | 1   | 95             |
| R20                   | 3   | 1   | 4   | 1   | 5   | 2   | 3   | 3   | 4   | 2   | 75             |
| R...                  | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ...            |
| R40                   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 1   | 5   | 5   | 90             |
| <b>Skor Akhir SUS</b> |     |     |     |     |     |     |     |     |     |     | <b>82.8125</b> |

Berdasarkan data dalam Tabel 2, total skor yang telah dihitung adalah sebesar 82.8125. Setelah menghitung nilai SUS, langkah berikutnya adalah menentukan grade berdasarkan indikator penilaian SUS yang terlampir [12]. Dari gambar indikator penilaian SUS dibawah, skor sebelumnya sebesar 82.8125 diklasifikasikan sebagai grade B atau tingkat kepuasan yang baik dan dapat diterima.



Gambar 16. Hasil Indikator Penilaian SUS

#### 4. Kesimpulan

Penelitian ini menyoroti peran penting aplikasi MindEase dalam upaya pencegahan gangguan kesehatan mental pada remaja Indonesia melalui desain antarmuka yang menarik. Fitur-fitur yang disematkan dalam MindEase dirancang untuk memberikan dukungan yang komprehensif dalam meningkatkan kesehatan mental pengguna. Meskipun hasil pengujian usabilitas menunjukkan tingkat penerimaan yang baik dengan skor SUS sebesar 82.8125, penting untuk dicatat bahwa keterlibatan ahli psikologi atau praktisi kesehatan mental dalam contoh penelitian

ini sangat minim. Dengan demikian, untuk meningkatkan standar evaluasi konten aplikasi di masa mendatang, keterlibatan ahli tersebut menjadi sangat penting. Diharapkan penelitian yang akan datang dapat memberikan pemahaman yang lebih luas tentang dampak dan efektivitas MindEase dalam meningkatkan kesejahteraan mental remaja di Indonesia.

#### Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia, "Menjaga Kesehatan Mental Para Penerus Bangsa," [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20231012/3644025/menjaga-kesehatan-mental-para-penerus-bangsa/>. [Accessed May. 5, 2023].
- [2] World Health Organization, "The WHO special initiative for mental health (2019-2023): universal health coverage for mental health," World Health Organization, No. WHO/MSD/19.1, 2019.
- [3] United Nations. (n.d.). Goal 3: Good health and well-being. [Online]. Available: <https://sdgs.un.org/goals/goal3>
- [4] R. N. Isnaini, I. D. G. Arikesa, R. I. Nasution, and M. F. Hidayat, "Penggunaan User Interface (UI) Aplikasi Google Classroom Pada Siswa Tingkat SMP di Denpasar Selatan," *Seminar Nas. Desain*, vol. 1, pp. 1–7, 2021.
- [5] F. Fariyanto, S. Suaidah, and F. Ulum, "Perancangan aplikasi pemilihan kepala desa dengan metode ux design thinking (Studi kasus: kampung kuripan)," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 2, pp. 52–60, 2021.
- [6] H. Abdurahman, A. R. Riswaya, and A. Id, "Aplikasi pinjaman pembayaran secara kredit pada bank yudha bhakti," *Jurnal Computech & Bisnis*, vol. 8, no. 2, pp. 61–69, 2014.
- [7] I. P. Sari et al., "Implementasi Metode Pendekatan Design Thinking dalam Pembuatan Aplikasi Happy Class Di Kampus UPI Cibiru," *Edsence: Jurnal Pendidikan Multimedia*, vol. 2, no. 1, pp. 45–55, 2020.
- [8] H. Ilham, B. Wijayanto, and S. P. Rahayu, "Analysis and Design of User Interface/User Experience with the Design Thinking Method in the Academic Information System of Jenderal Soedirman University," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 1, pp. 17–26, 2021.
- [9] M. F. Ardiansyah and P. Rosyani, "Perancangan UI/UX Aplikasi Pengolahan Limbah Anorganik Menggunakan Metode Design Thinking," *LOGIC: Jurnal Ilmu Komputer dan Pendidikan*, vol. 1, no. 4, pp. 839–853, 2023.
- [10] W. Buana and B. N. Sari, "Analisis User Interface Meningkatkan Pengalaman Pengguna Menggunakan Usability Testing pada Aplikasi Android Course," *DoubleClick: Journal of Computer and Information Technology*, vol. 5, no. 2, pp. 91–97, 2022.
- [11] A. Hidayat and H. M. Fauziyyah, "Perancangan Desain Antarmuka Aplikasi Pembelajaran Online Berbasis Mobile Menggunakan Metode Design Thinking," *Jurnal Teknik Informatika (Jutekin)*, vol. 10, no. 1.
- [12] D. P. Kesuma, "Penggunaan metode System Usability Scale untuk mengukur aspek Usability pada media pembelajaran daring di Universitas XYZ," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 3, pp. 1615–1626, 2021.



# Pengembangan Interface dan Experience Aplikasi Manajemen Kegiatan Notelife Menggunakan Pendekatan Design Thinking

Ni Luh Gede Cahaya Putri Mahadewi<sup>a1</sup>, I Ketut Gede Suhartana<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana  
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>mahadewi.2208561110@student.unud.ac.id  
<sup>2</sup>ikg.suhartana@unud.ac.id

## Abstract

*The COVID-19 pandemic has presented major challenges to global productivity, especially for office workers who have switched to Work from Home (WFH). This research focuses on developing the "Notelife" application using a design thought approach to overcome these productivity challenges. The design thinking method is applied through five stages: Empathize, Define, Ideate, Prototype, and Test. The Empathize stage involved interviews with five office workers to understand their management activity needs. The Define stage produces user personas and pain points based on interview results. The Ideate stage creates solution ideas, and the Prototype stage produces the "Notelife" application design including task, file, and time management features. Testing was carried out using the System Usability Scale (SUS) on 20 respondents, and the Notelife application achieved an average SUS value of 82.625, exceeding the average value of 68 for system usability. In conclusion, the design thought approach succeeded in producing a Notelife application design that suits user needs and provides a positive user experience in managing daily activities.*

**Keywords:** Productivity Enhancement, Design thinking, System Usability Scale (SUS), User Interface, User Experience

## 1. Pendahuluan

Pandemi COVID-19 telah menghadirkan tantangan yang signifikan terhadap produktivitas global. Resesi global yang terjadi sebagai akibat langsung dari pandemi ini merupakan yang terdalam sejak Perang Dunia Kedua dan tren penurunan yang sudah ada sebelumnya semakin diperparah. Penurunan produktivitas tidak hanya menghasilkan penurunan pendapatan per kapita, namun juga memperlambat pertumbuhan ekonomi secara keseluruhan. Dampaknya tidak berhenti disitu saja, pandemi ini berpotensi meninggalkan bekas yang lama terhadap perekonomian, dengan investasi yang menurun, perdagangan yang mundur, serta rantai pasok global yang terputus akibat pembatasan mobilitas. Tidak hanya itu, penurunan kualitas sumber daya manusia juga menjadi perhatian, karena hilangnya jam belajar dan meningkatnya jumlah pengangguran[1]. Dalam konteks ini, penelitian oleh Suyono, Bambang, dan Hermawan (2013) menyoroti dampak usia terhadap produktivitas tenaga kerja di sektor jasa. Mereka menemukan bahwa semakin tua usia tenaga kerja, semakin menurun produktivitasnya. Tenaga kerja yang lebih tua cenderung memiliki produktivitas yang lebih rendah. Melihat kondisi ini, manajemen kegiatan setiap individu menjadi hal yang perlu disorot untuk meningkatkan produktivitas tersebut[4]. Pekerja kantoran menjadi salah satu segmen yang terkena dampak besar dari penurunan produktivitas selama pandemi COVID-19. Peralihan secara besar-besaran dari yang awalnya bekerja ke kantor menjadi bekerja di rumah atau biasa disebut Work from Home (WFH) telah menimbulkan tantangan baru bagi produktivitas mereka. Meskipun WFH memberikan fleksibilitas yang lebih besar, namun banyak pekerja yang mengalami kesulitan dalam memisahkan waktu kerja dan waktu pribadi, serta menghadapi gangguan dari lingkungan rumah. Selain itu, kurangnya interaksi langsung dengan rekan kerja dapat menghambat kolaborasi dan komunikasi, yang merupakan

elemen penting dalam lingkungan kantor yang produktif. Keterbatasan infrastruktur dan teknologi juga bisa menjadi penghalang, terutama bagi pekerja yang tidak memiliki akses yang memadai atau keterampilan teknologi yang cukup. Untuk mengatasi tantangan produktivitas yang dihadapi akibat pandemi dan mengoptimalkan pengelolaan waktu serta sumber daya, dibutuhkanlah sebuah aplikasi yang dapat menjadi solusi praktis. "Notelife" adalah salah satu aplikasi yang dirancang khusus untuk melakukan manajemen kegiatan menggunakan pendekatan design thinking. Fokus utama pengembangan adalah pada desain antarmuka pengguna (UI) dan pengalaman pengguna (UX) dari aplikasi ini, dengan penekanan khusus pada kemudahan penggunaan dan keterampilan navigasi yang intuitif. Dengan basis mobile, "Notelife" dirancang untuk memberikan aksesibilitas yang maksimal, sehingga pengguna dapat dengan mudah mengelola jadwal, membuat catatan, dan mengatur tugas-tugas mereka dimanapun dan kapan pun. Ini menjadi langkah strategis dalam membantu individu dan organisasi meningkatkan produktivitas dan efisiensi mereka dalam mengelola kegiatan sehari-hari. Penelitian ini menggunakan pendekatan design thinking karena mengakui keunggulan-keunggulan yang ditawarkan oleh metode ini. Pertama, pendekatan ini fokus pada pengguna dan kebutuhan mereka. Dengan memprioritaskan pemahaman mendalam terhadap pengguna, solusi yang dikembangkan menjadi lebih sesuai dengan kebutuhan mereka dan lebih efektif dalam mencapai tujuan yang diinginkan. Selain itu, pendekatan design thinking juga mempromosikan kerjasama lintas disiplin, melibatkan tim yang terdiri dari individu dengan latar belakang yang beragam, seperti desainer, ilmuwan data, dan ahli bisnis. Kolaborasi semacam ini memungkinkan integrasi ide dan pendekatan yang lebih luas, menghasilkan solusi yang lebih holistik dan efektif. Terakhir, metode ini dikenal karena sifatnya yang iteratif. Artinya, solusi dikembangkan melalui serangkaian tahap yang berulang, dengan setiap tahap yang memperbaiki dan meningkatkan solusi sebelumnya. Pendekatan ini memastikan bahwa solusi yang dihasilkan menjadi lebih baik seiring waktu dan lebih efektif dalam mencapai tujuan penelitian[2][5]. Dengan demikian, penerapan design thinking dalam penelitian ini menjadi kunci untuk mencapai solusi yang inovatif dan efektif.

## 2. Metode Penelitian

### 2.1 Design thinking

Design thinking merupakan metode kolaborasi yang mengumpulkan banyak ide dari disiplin ilmu untuk memperoleh sebuah solusi. Design thinking tidak hanya berfokus pada apa yang dilihat dan dirasakan, namun juga berfokus pada pengalaman pengguna (user)[3]. Dimulai dengan fokus empati kepada riset calon pengguna untuk benar-benar memahami manusia (people centered), melihat permasalahan dari point of view yang berbeda dan penentuan solusi yang bervariasi (highly creative), melakukan pengujian oleh tim serta mendapati feedback dari sebuah produk sampel yang diberikan pada calon pengguna (hands on), hingga menghasilkan produk yang sesuai dengan kebutuhan pengguna (iterative). Terdapat 5 tahapan dalam metode design thinking, yaitu:



Gambar 1. Lima Tahapan Design thinking

a. Empathize

Empathize merupakan tahapan yang dilakukan dengan observasi atau wawancara kepada pengguna dengan melihat dan memahami berbagai hal dari sudut pandang pengguna. Tujuan dari tahap empathize adalah untuk mengetahui pandangan dan kebutuhan dari target pengguna. Pada tahapan ini peneliti melakukan wawancara terstruktur kepada pengguna secara langsung kepada lima pekerja kantor. Data atau informasi yang dikumpulkan berupa informasi kebutuhan dan keinginan pengguna terhadap manajemen kegiatan. Pertanyaan wawancara dapat dilihat pada Tabel 1.

Tabel 1. Daftar Pertanyaan Wawancara

| No | Pertanyaan   |
|----|--|
| 1. | Bagaimana anda biasanya mengatur kegiatan anda?  |
| 2. | Menurut anda, apakah cara anda tersebut merupakan cara yang efektif untuk meningkatkan produktivitas anda? |
| 3. | Apa saja hambatan yang anda temui dalam mengatur kegiatan anda?  |
| 4. | Menurut anda, apakah hal yang paling penting dalam manajemen kegiatan? Berikan alasannya!                  |
| 5. | Apakah ada hal yang membantu anda dalam manajemen kegiatan?  |

b. Define

Define merupakan mendefinisikan dan menganalisis permasalahan yang didapatkan dari proses empathize untuk menciptakan solusi yang tepat kedalam bentuk pain points, user persona, dan problem statement. User persona didapatkan dari pengolahan data dari responden yang mengikuti sesi pengujian. User persona berfokus pada data personal seperti karakter, demografis, tingkah laku, motivasi, tujuan, dan kesulitan pengguna yang dapat membantu mengenali pengguna saat brainstorming [7].

c. Ideate

Ideate merupakan tahapan untuk melakukan brainstorming ide. Pada tahapan ini akan menghasilkan ide dari permasalahan yang dihadapi sehingga output dari tahapan ini adalah berupa rancangan solusi atau produk atau wireframe. Wireframe adalah kerangka kerja untuk mengatur elemen pada halaman website atau halaman aplikasi [6].

d. Prototype

Prototype merupakan tahapan mengimplementasikan segala ide dan solusi design yang didapatkan pada proses Ideate kedalam sebuah prototyping sehingga dapat menjadi sebuah produk design yang siap diuji kepada responden yang mengikuti sesi testing sebelum dilanjutkan proses pengembangan.

e. Test

Test merupakan tahapan untuk memastikan apakah ide solusi design sudah sesuai dengan kebutuhan pengguna. Dalam tahapan ini, akan mendapatkan banyak insight dan feedback dari pengguna sehingga perlunya iterasi terhadap ide solusi design untuk dapat menyempurnakan sesuai dengan kebutuhan pengguna. Pada tahapan ini peneliti menggunakan salah satu alat pengujian usability yaitu System Usability Scale (SUS) ketika sudah mendapatkan insight dan feedback dari pengguna.

System Usability Scale (SUS) adalah alat pengujian usability yang dikembangkan oleh John Brooke pada 1986. Terdiri dari sepuluh pertanyaan, SUS dirancang untuk cepat dan mudah dijawab oleh responden dengan skor berkisar antara 0 hingga 100. Nilai rata-rata SUS adalah 68, jika skor sistem berada di bawah angka tersebut maka menunjukkan adanya masalah yang memengaruhi tingkat kegunaan sistem. Berikut daftar pertanyaan SUS pada tahap test Aplikasi Notelife dapat dilihat pada Tabel 2.

**Tabel 2.** Daftar Pertanyaan SUS

| No | Pertanyaan   |
|----|--|
| 1  | Saya berpikir akan menggunakan sistem ini lagi                                     |
| 2  | Saya merasa sistem ini rumit untuk digunakan                                       |
| 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 serasi 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         |

### 3. Hasil dan Pembahasan

Proses implementasi metode Design thinking dalam antarmuka pengguna (user interface) serta pengalaman pengguna (user experience) aplikasi manajemen kegiatan Notelife telah berhasil dilaksanakan dengan mengikuti kelima tahapan design thinking. Berikut hasil dan pembahasan dari setiap tahapan:

#### 3.1 Empathize

Hasil yang didapat dari tahapan empathize adalah mengetahui pandangan pengguna yang akan menjadi dasar pengembangan aplikasi manajemen kegiatan Notelife. Tabel 3 berisi tentang kesimpulan hasil wawancara dari lima pekerja kantor yang telah diwawancarai oleh peneliti.

**Tabel 3.** Kesimpulan Hasil Wawancara

| No | Kesimpulan Hasil Wawancara  |
|----|---|
| 1. | Dari hasil wawancara, orang memiliki kecenderungan sering lupa akan pekerjaan yang diberikan dikarenakan terdapat pekerjaan lain yang cukup padat sehingga seringkali pekerjaan tersebut dibuat mepet dengan deadline   |
| 2. | Dari hasil wawancara, orang cenderung kesulitan untuk mencari berkas atau materi yang diperlukan dikarenakan terkadang materi yang diberikan sudah kadaluarsa, jauh tenggelam oleh berkas lainnya, atau tempat menyimpan yang berantakan  |
| 3. | Dari hasil wawancara, orang cenderung mengalami kesulitan dalam mengatur waktu kapan harus istirahat, fokus bekerja, dan selesai bekerja dikarenakan jika mereka sedang bekerja terkadang mereka akan tetap berusaha bekerja hingga mereka benar benar lelah sehingga mempengaruhi kesehatan mereka |

#### 3.2 Define

Hasil yang didapat dari tahapan Define adalah user persona yang meliputi karakteristik, demografis, perilaku, motivasi, tujuan, dan kesulitan pengguna yang diidentifikasi melalui pengolahan data dari hasil wawancara yang diperoleh pada tahap empathize. User persona dapat dilihat seperti pada Tabel 4.

Tabel 4. User persona

|   |  |
|---|--|
| <b>Nama</b>                                       | : Tegar Maheswara  |
| <b>Usia</b>                                       | : 25 Tahun   |
| <b>Profesi</b>                                    | : Marketing Executive  |
| <b>Latar Belakang</b>                             | : Tegar adalah seorang profesional muda yang bekerja sebagai eksekutif pemasaran di sebuah perusahaan besar. Dia memiliki jadwal yang padat dengan berbagai proyek dan tugas yang harus diselesaikan setiap hari. Tegar sering merasa terbebani dengan deadline yang ketat dan tugas-tugas yang beragam. Dia juga sering kali lupa akan tugas-tugas yang diberikan karena kesibukannya yang menyita perhatian.   |
| <b>Tantangan Utama</b>                            | : <ol style="list-style-type: none"> <li>1. <b>Lupa akan tugas:</b> Karena jadwal yang padat, Tegar sering lupa akan tugas-tugas yang harus diselesaikan, terutama jika tugas tersebut tidak urgent atau memiliki deadline yang masih cukup lama.</li> <li>2. <b>Kesulitan dalam mencari berkas:</b> Tegar sering kesulitan dalam mencari berkas atau materi yang diperlukan untuk tugasnya. Kadang-kadang materi tersebut sudah kadaluarsa atau tenggelam di antara berkas-berkas lainnya.</li> <li>3. <b>Kesulitan dalam mengatur waktu:</b> Tegar sulit untuk mengatur waktu istirahat dan fokus bekerja. Dia cenderung terus bekerja bahkan saat sudah lelah, yang berdampak negatif pada kesehatannya.</li> </ol> |
| <b>Keinginan dari Aplikasi Manajemen Kegiatan</b> | : <ol style="list-style-type: none"> <li>1. <b>Pengingat tugas:</b> Tegar menginginkan fitur yang dapat mengingatkannya tentang tugas-tugas yang harus diselesaikan, terutama yang memiliki deadline mendekati.</li> <li>2. <b>Manajemen berkas yang efisien:</b> Tegar membutuhkan fitur untuk mengatur dan menyimpan berkas-berkasnya secara terstruktur agar mudah ditemukan kembali saat diperlukan.</li> <li>3. <b>Manajemen waktu yang efektif:</b> Tegar ingin memiliki alat yang membantunya mengatur waktu dengan lebih efektif, termasuk mengingatkan untuk istirahat dan mengakhiri jam kerja secara tepat waktu.</li> </ol>  |

Setelah user persona didapatkan dari hasil wawancara, ditemukan pain points yang menjadi masalah utama bagi pengguna. Pain points dapat dilihat seperti pada Tabel 5.

Tabel 5. Pain points

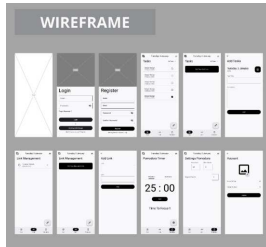
| No Pain points   | Solution   |
|--|--|
| 1. <b>Lupa akan tugas karena beban kerja yang padat:</b> <ul style="list-style-type: none"> <li>• Orang sering lupa akan tugas yang diberikan karena beban kerja yang padat, menyebabkan tugas sering kali dikerjakan dengan deadline yang mepet.</li> </ul> | <b>Manajemen tasks yang efektif:</b> <ul style="list-style-type: none"> <li>• Buat fitur manajemen tugas yang memungkinkan pengguna untuk menambahkan, mengedit, dan menghapus tugas dengan mudah.</li> <li>• Tampilkan tugas berdasarkan status untuk membantu pengguna mengatur waktu dan fokus pada tugas yang belum diselesaikan.</li> <li>• Berikan pengingat atau notifikasi untuk tugas yang mendekati</li> </ul> |

| No | Pain points  | Solution   |
|----|--|--|
|    |  | deadline agar pengguna tidak lupa.   |
| 2. | <b>Kesulitan dalam menemukan materi yang dibutuhkan:</b> <ul style="list-style-type: none"><li>Orang mengalami kesulitan mencari berkas atau materi yang diperlukan karena materi tersebut seringkali sudah kadaluarsa, atau tersembunyi di antara berkas yang berantakan.</li></ul> | <b>Manajemen link yang terorganisir:</b> <ul style="list-style-type: none"><li>Buat fitur manajemen link yang memungkinkan pengguna untuk menyimpan dan mengelola tautan ke materi atau sumber daya penting.</li><li>Fitur copy link yang dapat digunakan untuk menyalin link ke clipboard untuk manajemen hal yang lebih dinamis.</li></ul> |
| 3. | <b>Kesulitan dalam manajemen waktu dan istirahat:</b> <ul style="list-style-type: none"><li>Orang kesulitan mengatur waktu untuk istirahat, fokus bekerja, dan mengakhiri sesi kerja.</li></ul>  | <b>Pomodoro timer untuk manajemen waktu:</b> <ul style="list-style-type: none"><li>Integrasikan Pomodoro Timer untuk membantu pengguna mengatur waktu kerja dan istirahat dengan lebih efisien.</li><li>Berikan opsi untuk menetapkan durasi kerja dan istirahat sesuai preferensi pengguna.</li></ul>                                       |

Berdasarkan user persona dan pain points yang telah didapatkan maka dapat ditentukan problem statementnya adalah banyak pekerja sering mengalami kesulitan dalam mengelola tugas, berkas, dan waktu secara efisien dalam kehidupan sehari-hari. Tantangan utamanya adalah pertama, lupa akan tugas. Banyak orang cenderung lupa akan tugas-tugas yang diberikan karena kesibukan yang tinggi dan banyaknya tugas yang harus diselesaikan dalam waktu yang terbatas. Kedua, kesulitan dalam mencari berkas. Orang sering mengalami kesulitan dalam mencari berkas atau materi yang diperlukan untuk tugas mereka karena berbagai alasan, seperti berkas yang sudah kadaluarsa atau terkubur di antara berkas-berkas lainnya. Ketiga, kesulitan dalam mengatur waktu. Banyak orang kesulitan dalam mengatur waktu istirahat, fokus bekerja, dan waktu selesai bekerja. Kebiasaan untuk terus bekerja bahkan saat sudah lelah dapat berdampak negatif pada kesehatan dan produktivitas kerja mereka. Dengan demikian, diperlukan solusi yang dapat membantu individu dalam berbagai profesi dan kehidupan sehari-hari untuk mengatasi tantangan ini melalui pendekatan yang terintegrasi dan efektif dalam manajemen tugas, berkas, dan waktu yaitu dengan melakukan pengembangan aplikasi manajemen kegiatan.

### 3.3 Ideate

Dari tahap sebelumnya yaitu tahap Define, didapatkan ide dan solusi untuk mengatasi tantangan produktivitas dalam penelitian, yaitu melalui pengembangan aplikasi manajemen kegiatan. Peneliti mengembangkan aplikasi manajemen kegiatan "Notelife" yang dirancang khusus untuk membantu dalam manajemen kegiatan guna mengatasi permasalahan produktivitas seperti pada penelitian yang telah dilakukan. Untuk mewujudkan ide dan solusi tersebut maka langkah pertama yang dilakukan oleh peneliti adalah membuat rancangan awal berupa sketsa tampilan antarmuka aplikasi dalam bentuk wireframe. Wireframe aplikasi Notelife dapat dilihat pada Gambar 1.



Gambar 1. Wireframe Aplikasi Notelife

### 3.4 Prototype

Setelah mewujudkan ide dan solusi ke dalam bentuk wireframe, maka langkah berikutnya adalah tahapan pembuatan desain prototype. Prototype ini terdiri dari kumpulan desain high-fidelity yang nantinya akan dihubungkan antara satu dengan yang lain sehingga menjadi prototype utuh yang dapat dijalankan. Berikut merupakan penjelasan prototype aplikasi manajemen kegiatan Notelife.

a. Desain tampilan splash screen dan login register

Tampilan splash screen berisikan logo dari Notelife yang berbentuk daun melambangkan kehidupan serta didalamnya berisikan baris yang berisikan tulisan yang disimbolkan sebagai note. Kemudian terdapat tampilan login yang digunakan seperti pada umumnya serta tampilan register jika belum memiliki akun. Desain tampilan splash screen dan login register aplikasi Notelife dapat dilihat pada Gambar 2.



Gambar 2. Desain tampilan splash screen dan login register

b. Desain fitur tasks list

Fitur tasks list merupakan fitur penyimpanan list tugas ataupun pekerjaan beserta deadline dari tugas atau pekerjaan tersebut. Fitur ini berguna untuk membantu pengguna mengatur waktu dan fokus pada tugas yang belum diselesaikan. Desain fitur tasks list aplikasi Notelife dapat dilihat pada Gambar 3.



Gambar 3. Desain fitur tasks list

c. Desain fitur link management

Fitur link management merupakan fitur penyimpanan link berkas ataupun link materi. Fitur

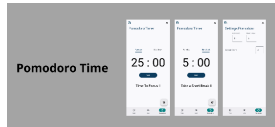
ini berguna untuk menyimpan hal-hal tersebut lebih terstruktur dan mudah dicari. Desain fitur link management aplikasi Notelife dapat dilihat pada Gambar 4.



Gambar 4. Desain fitur link management

d. Desain fitur pomodoro time

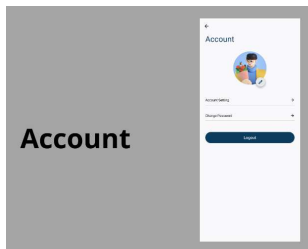
Fitur pomodoro time merupakan fitur teknik pomodoro untuk pengaturan waktu yang memunculkan rasa urgensi. Teknik ini memiliki sistem belajar ataupun mengerjakan suatu kerjaan yang berdurasi singkat namun memiliki intensitas tinggi, umumnya 25 menit belajar dan lima menit istirahat. Namun, setelah mencapai empat interval belajar, waktu istirahat menjadi lebih panjang sekitar 15 hingga 30 menit. Desain fitur pomodoro time dapat dilihat pada Gambar 5.



Gambar 5. Desain fitur pomodoro time

e. Desain tampilan account

Tampilan account aplikasi Notelife terdapat 3 fitur yaitu account setting, change password, dan logout. Pada fitur account setting memungkinkan pengguna untuk mengatur dan mengelola informasi akun mereka, kemudian pada fitur change password memungkinkan pengguna untuk mengganti kata sandi akun mereka, serta pada fitur logout memungkinkan pengguna untuk keluar dari akun mereka. Desain tampilan account aplikasi Notelife dapat dilihat pada Gambar 6.



Gambar 6. Desain tampilan account



### 3.5 Test

Setelah menyelesaikan tahapan prototype, maka dilanjutkan pengujian usability terhadap 20 orang untuk aplikasi Notelife ini. Berikut hasil pengujian yang dapat dilihat pada Tabel 6.

**Tabel 6.** Hasil Survei Sebelum Diolah

| SKOR SUS |    |    |    |    |    |    |    |    |    |     |
|----------|----|----|----|----|----|----|----|----|----|-----|
| Resp.    | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
| R1       | 4  | 2  | 4  | 2  | 5  | 2  | 5  | 1  | 4  | 2   |
| R2       | 4  | 2  | 3  | 1  | 5  | 1  | 4  | 2  | 4  | 1   |
| R3       | 4  | 1  | 4  | 1  | 5  | 1  | 4  | 2  | 4  | 1   |
| R4       | 3  | 1  | 5  | 2  | 5  | 1  | 4  | 1  | 5  | 1   |
| R5       | 3  | 1  | 4  | 1  | 4  | 1  | 5  | 2  | 4  | 1   |
| R6       | 4  | 1  | 4  | 2  | 5  | 2  | 4  | 1  | 5  | 1   |
| R7       | 4  | 1  | 5  | 2  | 5  | 1  | 5  | 1  | 4  | 1   |
| R8       | 4  | 2  | 3  | 2  | 5  | 1  | 4  | 2  | 4  | 3   |
| R9       | 4  | 2  | 5  | 2  | 5  | 2  | 4  | 2  | 5  | 3   |
| R10      | 5  | 3  | 4  | 1  | 4  | 2  | 4  | 2  | 4  | 2   |
| R11      | 4  | 3  | 3  | 1  | 4  | 1  | 5  | 1  | 4  | 2   |
| R12      | 3  | 3  | 4  | 2  | 4  | 1  | 4  | 2  | 4  | 2   |
| R13      | 4  | 3  | 3  | 1  | 4  | 2  | 4  | 2  | 4  | 1   |
| R14      | 5  | 3  | 4  | 2  | 4  | 1  | 4  | 2  | 5  | 1   |
| R15      | 4  | 1  | 5  | 2  | 3  | 2  | 5  | 1  | 5  | 1   |
| R16      | 5  | 2  | 4  | 2  | 5  | 2  | 4  | 2  | 4  | 1   |
| R17      | 3  | 2  | 4  | 2  | 4  | 1  | 5  | 1  | 4  | 2   |
| R18      | 4  | 1  | 3  | 1  | 3  | 1  | 4  | 1  | 4  | 2   |
| R19      | 3  | 1  | 4  | 2  | 4  | 2  | 5  | 2  | 5  | 1   |
| R20      | 5  | 1  | 4  | 2  | 4  | 1  | 4  | 1  | 5  | 1   |

Setelah hasil survei dikumpulkan, maka skor diproses dengan aturan SUS sehingga mendapatkan hasil seperti pada Tabel 7.

**Tabel 7.** Hasil Survei Setelah Diolah

| SKOR SUS |    |    |    |    |    |    |    |    |    |     |     |      |
|----------|----|----|----|----|----|----|----|----|----|-----|-----|------|
| Resp.    | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Sum | x2.5 |
| R1       | 3  | 3  | 3  | 3  | 4  | 3  | 4  | 4  | 3  | 3   | 33  | 82.5 |
| R2       | 3  | 3  | 2  | 4  | 4  | 4  | 3  | 3  | 3  | 4   | 33  | 82.5 |
| R3       | 3  | 4  | 3  | 4  | 4  | 4  | 3  | 3  | 3  | 4   | 35  | 87.5 |
| R4       | 2  | 4  | 4  | 3  | 4  | 4  | 3  | 4  | 4  | 4   | 36  | 90   |
| R5       | 2  | 4  | 3  | 4  | 3  | 4  | 4  | 3  | 3  | 4   | 34  | 85   |

| SKOR SUS                 |    |    |    |    |    |    |    |    |    |     |               |      |
|--------------------------|----|----|----|----|----|----|----|----|----|-----|---------------|------|
| Resp.                    | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Sum           | x2.5 |
| R6                       | 3  | 4  | 3  | 3  | 4  | 3  | 3  | 4  | 4  | 4   | 35            | 87.5 |
| R7                       | 3  | 4  | 4  | 3  | 4  | 4  | 4  | 4  | 3  | 4   | 37            | 92.5 |
| R8                       | 3  | 3  | 2  | 3  | 4  | 4  | 3  | 3  | 3  | 2   | 30            | 75   |
| R9                       | 3  | 3  | 4  | 3  | 4  | 3  | 3  | 3  | 4  | 2   | 32            | 80   |
| R10                      | 4  | 2  | 3  | 4  | 3  | 3  | 3  | 3  | 3  | 3   | 31            | 77.5 |
| R11                      | 3  | 2  | 2  | 4  | 3  | 4  | 4  | 4  | 3  | 3   | 32            | 80   |
| R12                      | 2  | 2  | 3  | 3  | 3  | 4  | 3  | 3  | 3  | 3   | 29            | 72.5 |
| R13                      | 3  | 2  | 2  | 4  | 3  | 3  | 3  | 3  | 3  | 4   | 30            | 75   |
| R14                      | 4  | 2  | 3  | 3  | 3  | 4  | 3  | 3  | 4  | 4   | 33            | 82.5 |
| R15                      | 3  | 4  | 4  | 3  | 2  | 3  | 4  | 4  | 4  | 4   | 35            | 87.5 |
| R16                      | 4  | 3  | 3  | 3  | 4  | 3  | 3  | 3  | 3  | 4   | 33            | 82.5 |
| R17                      | 2  | 3  | 3  | 3  | 3  | 4  | 4  | 4  | 3  | 3   | 32            | 80   |
| R18                      | 3  | 4  | 2  | 4  | 2  | 4  | 3  | 4  | 3  | 3   | 32            | 80   |
| R19                      | 2  | 4  | 3  | 3  | 3  | 3  | 4  | 3  | 4  | 4   | 33            | 82.5 |
| R20                      | 4  | 4  | 3  | 3  | 3  | 4  | 3  | 4  | 4  | 4   | 36            | 90   |
| <b>Total Keseluruhan</b> |    |    |    |    |    |    |    |    |    |     | <b>1652.5</b> |      |
| <b>Rata-rata</b>         |    |    |    |    |    |    |    |    |    |     | <b>82.625</b> |      |

Setelah hasil survei diolah menggunakan metode SUS, diperoleh hasil 82.625 yang berada diatas 68 yang merupakan nilai rata-rata dalam pengujian metode SUS.

#### 4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa pendekatan Design thinking berhasil diterapkan dalam proses perancangan UI/UX aplikasi Notelife. Dengan menggunakan metode ini, produk yang dihasilkan adalah desain prototype yang sangat memperhatikan kebutuhan pengguna. Melalui tahapan-tahapan Design thinking, peneliti berhasil mengidentifikasi masalah yang dihadapi pengguna dan memahami kebutuhan mereka. Selanjutnya, ide-ide dan gagasan dikembangkan sebagai solusi untuk masalah tersebut. Pengujian dilakukan menggunakan metode kuesioner System Usability Scale (SUS). Hasil pengujian menunjukkan bahwa aplikasi Notelife mendapat nilai rata-rata sebesar 82.625, melebihi nilai rata-rata 68 yang merupakan standar kegunaan sistem menurut metode SUS. Hal ini menunjukkan bahwa tidak ada masalah yang signifikan yang memengaruhi kegunaan sistem. Dengan demikian, dapat disimpulkan bahwa pendekatan Design thinking dalam perancangan aplikasi Notelife telah berhasil menghasilkan desain yang sesuai dengan kebutuhan pengguna dan memberikan pengalaman pengguna yang positif.

#### Daftar Pustaka

- [1] C. Abdini, "Bagaimana Indonesia bisa mengatasi penurunan produktivitas perekonomian akibat COVID-19," *The Conversation*, Sep. 20, 2021. <https://theconversation.com/bagaimana-indonesia-bisa-mengatasi-penurunan-produktivitas-perekonomian-akibat-covid-19-168098> (accessed May 09, 2024).
- [2] C. Nieberding, "The Data Product Design Thinking Process," *Towards Data Science*, May 18, 2019. Accessed: May 09, 2024. [Online]. Available: <https://towardsdatascience.com/the-data-product-design-thinking-process-6b3eba561b2b>

- [3] I. P. Sari, A. H. Kartina, A. M. Pratiwi, F. Oktariana, M. F. Nasrulloh, and S. A. Zain, "Implementasi Metode Pendekatan Design Thinking dalam Pembuatan Aplikasi Happy Class Di Kampus UPI Cibiru," *Edsence: Jurnal Pendidikan Multimedia*, vol. 2, no. 1, pp. 45–55, Jun. 2020, doi: 10.17509/edsence.v2i1.25131.
- [4] I. Ukkas, "Faktor-Faktor yang mempengaruhi produktivitas tenaga kerja industri kecil kota Palopo," *Kelola: Journal of Islamic Education Management*, vol. 2, no. 2, Oct. 2017, doi: 10.24256/kelola.v2i2.440.
- [5] Melanie, "Design Thinking: How to link methodology and Data Science?" *Data Science Courses | DataScientest*, Apr. 15, 2024. <https://datascientest.com/en/design-thinking-how-to-link-methodology-and-data-science> (accessed May 09, 2024).
- [6] Ningrum, "Exabytes Indonesia," Tutorial Digital Marketing, Website, &#38; Bisnis Online - Exabytes, Nov. 08, 2022. <https://exabytes.co.id/blog/penjelasan-wireframe/> (accessed May 09, 2024).
- [7] U. Ambarwati, "Mengenal calon pengguna lebih dekat #1 — User Persona," #BelajarDesain, Mar. 09, 2020. Accessed: May 09, 2024. [Online]. Available: <https://medium.com/belajar-desain/mengenal-calon-pengguna-lebih-dekat-1-user-persona-699414e20270>

Halaman ini sengaja dibiarkan kosong

# Perancangan User Interface pada Aplikasi 'LiteraLink' dengan Menggunakan Metode Design Thinking

Lusia Elvira Sue Sare<sup>1</sup>, I Ketut Gede Suhartana<sup>2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>virasare@gmail.com  
<sup>2</sup>kg.suhartana@unud.ac.id

## Abstract

*The level of reading interest among Indonesian people will be 66.7 points in 2023. Increasingly advanced technological developments have led to an increase in the use of digital media, including reading practices which now tend to be digital. Previous research has involved the use of design thinking in creating solutions from a process of empathy for a particular human-centered need. So the design thinking method was used in designing the "LiteraLink" application. Before starting a series of steps in design thinking, it is important to first determine the target user in accordance with the objectives of this application. The results of the interface design test using SUS are shown in Table 2 and produced a score of 80.16. Based on the SUS assessment indicators, the score is categorized as grade B or excellent so that the results of this interface design are acceptable. Based on the results of the research that has been carried out, it can be concluded that the application of the design thinking method in designing the interface of the digital reading application in the 'LiteraLink' application is successful implemented so that this application is considered good and has the potential to increase interest in reading and support literacy among the community, especially children and teenagers.*

**Keywords :** User Interface, Design Thinking, Membaca, User Experience, System Usability Scale

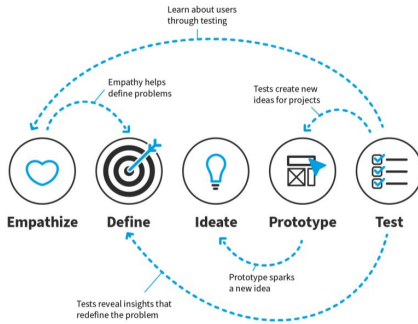
## 1. Pendahuluan

Minat membaca menjadi salah satu dari fondasi utama dalam meningkatkan kualitas pendidikan suatu negara. Berdasarkan data yang diperoleh Perpustakaan Nasional (Perpusnas), Tingkat Kegemaran Membaca (TGM) masyarakat Indonesia memiliki nilai 63,9 poin pada 2022 menjadi 66,7 poin pada tahun 2023, meskipun peningkatannya tidak signifikan. Rendahnya minat baca akan mengakibatkan kurangnya kemampuan berpikir kritis. Padahal, kemampuan ini sangat penting untuk mengembangkan minat, bakat, dan potensi yang ada di dalam diri seseorang [1]. Membaca merupakan salah satu tuntutan dalam kehidupan masyarakat modern. Membaca membuka jendela pengetahuan dengan menggali informasi dari teks yang dibaca, menyediakan akses ke pengetahuan yang relevan, serta memperkenalkan konsep dan gagasan yang sebelumnya belum terpikirkan. Kegiatan membaca tidak hanya dilakukan dengan buku fisik yang biasa dipegang dengan tangan, tidak harus dilakukan di perpustakaan dan bisa membaca di mana saja [2]. Perkembangan teknologi yang semakin maju telah menyebabkan peningkatan penggunaan media digital, termasuk dalam praktik membaca yang kini cenderung bersifat digital. Penelitian terdahulu telah melibatkan penggunaan design thinking dalam menciptakan solusi dari proses empati terhadap suatu kebutuhan tertentu yang berpusat pada manusia sehingga dapat menuju suatu inovasi berdasarkan kebutuhan penggunaannya[3]. Oleh karena itu, dalam rangka untuk meningkatkan minat membaca dan menjadi pendukung literasi bagi pengguna, diperlukan antarmuka pengguna (User Interface) yang optimal dalam aplikasi 'LiteraLink'. Hal ini bertujuan untuk menciptakan pengalaman pengguna (User Experience) yang memuaskan, dengan menggunakan pendekatan design thinking yang berpusat pada kebutuhan pengguna. Pendekatan ini melibatkan lima tahapan utama, yakni Empathize, Define, Ideate, Prototype, dan Test.

## 2. Metode Penelitian

### 2.1. Design Thinking

Penelitian ini akan menggunakan metode design thinking sebagai pendekatan utama dalam merancang aplikasi "LiteralLink". Metode ini memungkinkan pendekatan yang berorientasi pada pengguna dalam mengidentifikasi masalah, menghasilkan ide-ide kreatif, dan menguji prototype aplikasi. Sebelum memulai serangkaian langkah dalam design thinking, penting untuk terlebih dahulu menentukan target pengguna yang sesuai dengan tujuan dari aplikasi ini. Dengan cara ini, proses selanjutnya dalam lima tahapan design thinking dapat dijalankan dengan lebih lancar dan efektif.



**Gambar 1.** Design Thinking Process (Interaction Design Foundation)

Berdasarkan Gambar 1 diketahui bahwa design thinking mempunyai lima tahapan, penjelasan untuk setiap tahapan adalah sebagai berikut :

- a. Empathize  
Empathize menjadi langkah krusial dalam mendapatkan pemahaman yang dalam terhadap pengguna, memungkinkan kita untuk merasakan, memahami, dan menemukan solusi yang tepat. Pada tahap ini, pengumpulan data dapat dilakukan melalui wawancara, observasi maupun penggabungan data dari kedua metode tersebut[4].
- b. Define  
Tahap Define melibatkan analisis dan pemahaman terhadap hasil yang diperoleh dari tahap empathize sebelumnya. Fokus utama dari tahapan ini adalah menentukan pertanyaan masalah atau point of view yang akan menjadi titik pusat dalam penelitian[4].
- c. Ideate  
Ideate merupakan tahapan dimana ide-ide dan gagasan-gagasan baru dicari sebagai dasar dalam pembuatan prototype rancangan. Ide dan gagasan tersebut diharapkan dapat dijadikan solusi untuk menghadapi permasalahan pengguna [4].
- d. Prototype  
Prototype adalah tahap dimana rancangan awal produk dibuat untuk mendapatkan pemahaman yang lebih baik. Tujuan utamanya adalah untuk mendeteksi kesalahan sejak dini dan membuka peluang untuk inovasi. Rancangan awal ini kemudian akan diuji coba kepada pengguna untuk

memperoleh masukan yang relevan guna menyempurnakan desain aplikasi [4].

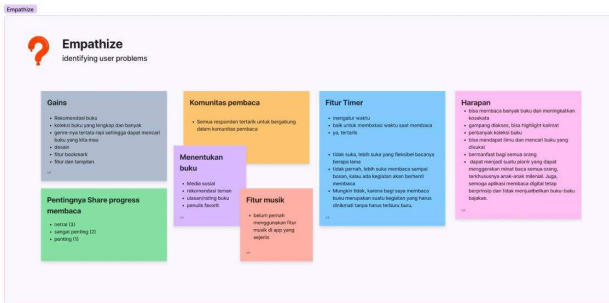
e. Test

Tahap terakhir, Test, merupakan tahapan dimana pengujian dilakukan untuk mengumpulkan masukan dari pengguna terhadap berbagai rancangan prototype. Masukan dari pengguna akan menjadi dasar untuk menyempurnakan desain atau prototype yang telah dibuat sebelumnya. Sehingga tahap ini memungkinkan adanya iterasi dan kembali ke tahapan sebelumnya jika diperlukan [4].

### 3. Hasil dan Pembahasan

#### 3.1. Empathize

Tahapan ini melibatkan observasi dan penyebaran kuesioner menggunakan Google Form untuk memahami dan menemukan solusi serta fitur yang sesuai dengan kebutuhan pengguna. Responden, yang berusia antara 15 - 50 tahun, merupakan pengguna yang pernah menggunakan aplikasi membaca digital sejenis. Responden diminta untuk menjawab beberapa pertanyaan terkait dengan pengembangan aplikasi 'Literalink' ini. Adapun hasil empathize yang berhasil dibuat berdasarkan hasil kuesioner disajikan pada Gambar 2 berikut.



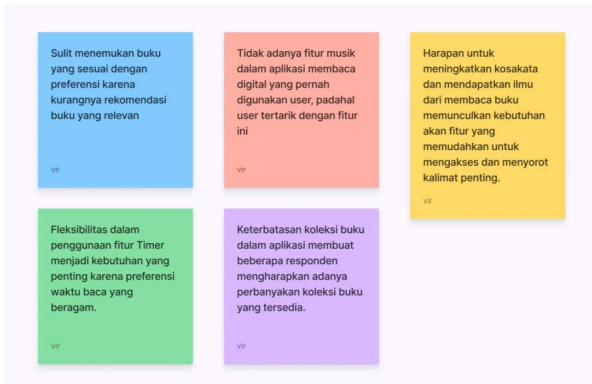
Gambar 2. Hasil Proses Empathize

Berdasarkan gambar tersebut, hasil analisis responden menunjukkan bahwa pengguna menginginkan akses mudah ke rekomendasi buku yang beragam, fitur bookmark, dan kemampuan untuk berbagi progres membaca. Pengguna juga tertarik pada keanggotaan dalam komunitas pembaca. Namun, fitur tambahan seperti musik dan timer belum menjadi prioritas bagi sebagian besar responden. Hal ini menunjukkan bahwa fokus utama pengguna adalah pada fitur yang mendukung pengalaman membaca yang nyaman dan bermanfaat.

#### 3.2. Define

Berdasarkan hasil analisis responden, dapat didefinisikan bahwa pengguna aplikasi literasi mencari platform yang menyediakan rekomendasi buku yang komprehensif, fitur bookmark yang memudahkan pengguna untuk melacak progres membaca, dan kemampuan untuk berbagi pengalaman membaca dengan komunitas. Pengguna juga menginginkan akses mudah ke ulasan buku dan rekomendasi dari teman serta penulis favorit pengguna. Selain itu, pengguna mengharapkan aplikasi literasi dapat meningkatkan minat membaca dan kosakata pengguna melalui beragam fitur yang mempermudah akses dan pemahaman terhadap materi bacaan. Pada tahap ini untuk menentukan masalah dan titik penelitian dibuatlah pain points yang merupakan

kumpulan beberapa poin-poin penting yang menjadi fokus utama perancangan selanjutnya, berikut disajikan 5 pain points yang didapat berdasarkan hasil empathize sebelumnya.



Gambar 3. Pain Points

Oleh karena itu, dalam perancangan aplikasi literasi, fokus utama haruslah pada penyediaan konten berkualitas, fitur interaktif yang mendukung pengalaman membaca, dan kemudahan berbagi informasi dengan komunitas pembaca.

### 3.3. Ideate

Tahap ideate melibatkan proses brainstorming, dimana berbagai ide dan gagasan dibuat untuk menanggapi masalah yang telah diidentifikasi. Tujuannya adalah untuk menghasilkan solusi yang kreatif dan beragam untuk mengatasi pain points yang ada. Beberapa ide tersebut antara lain implementasi fitur rekomendasi buku yang disesuaikan dengan preferensi pengguna, integrasi fitur musik sesuai dengan genre buku yang sedang dibaca, serta penambahan fitur komunitas pembaca untuk meningkatkan interaksi antar pengguna dan mendukung budaya literasi. Ide-ide ini akan dievaluasi lebih lanjut untuk menentukan yang paling efektif dan layak untuk diimplementasikan dalam aplikasi.

### 3.4. Prototype

#### a. Wireframe dan Wireflow

Wireframe/wireflow dalam aplikasi 'LiteraLink' berfungsi sebagai representasi visual yang menyajikan tata letak, alur, dan interaksi antarmuka pengguna. Dengan menggunakan wireframe/wireflow memahami cara navigasi dalam aplikasi dan bagaimana setiap fitur saling berinteraksi akan lebih mudah. Dengan menggunakan wireframe/wireflow pengembangan aplikasi akan lebih efisien dan optimal sesuai dengan kebutuhan pengguna. Berikut disajikan beberapa alur proses dalam aplikasi ini:

- Alur dari tampilan splash screen ke halaman beranda

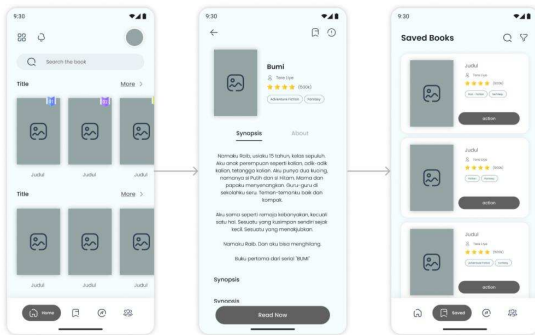




Gambar 4. Wireflow tampilan splash screen ke halaman beranda

Setelah pengguna membuka aplikasi, pengguna akan disambut dengan tampilan awal logo dan merek aplikasi. Setelah itu, pengguna akan dialihkan pada halaman onboarding yang memberikan informasi singkat tentang fitur dan manfaat aplikasi. Setelah menyelesaikan onboarding, pengguna akan secara otomatis diarahkan ke halaman beranda utama, dimana pengguna dapat mulai menelusuri koleksi buku, melihat rekomendasi, dan mengakses fitur-fitur utama aplikasi.

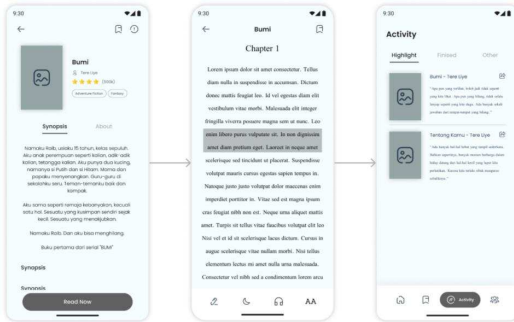
- Alur menyimpan buku yang ingin dibaca



Gambar 5. Wireflow menyimpan buku yang ingin dibaca

Setelah pengguna menemukan buku yang ingin dibaca, pengguna dapat menyimpannya dengan menekan tombol "Simpan" yang terletak di dekat judul buku. Buku yang disimpan akan otomatis ditambahkan ke daftar buku yang ingin dibaca, yang dapat diakses melalui fitur "Buku Disimpan" di menu utama aplikasi.

- Alur memberi highlight untuk kalimat tertentu

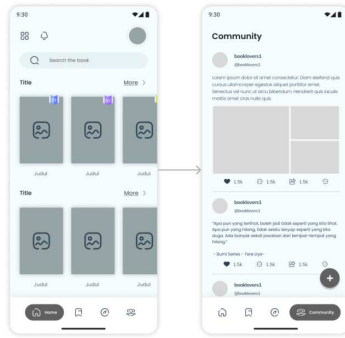


Gambar 6. Wireflow memberi highlight untuk kalimat tertentu

Untuk memberi highlight pada kalimat tertentu, pengguna dapat terlebih dahulu masuk dan memilih buku mana yang hendak dibaca lalu pengguna dapat menyorot teks yang ingin disorot dengan menekan dan menahan jari di atas kalimat yang diinginkan. Setelah menyorot, opsi untuk menambahkan highlight akan muncul, dan pengguna dapat memilih warna highlight yang diinginkan. Setelah selesai, kalimat tersebut akan ditandai dengan warna yang dipilih, memungkinkan pengguna untuk dengan mudah mengidentifikasi dan kembali ke bagian yang menarik bagi pengguna. Hasil highlight yang dibuat akan disimpan pada bagian aktivitas, sehingga pengguna dapat melihat highlight penting yang ditandainya.

- Alur bergabung di komunitas

Untuk bergabung di komunitas, pengguna dapat memilih opsi "Komunitas" dari menu utama aplikasi. pengguna akan dibawa ke halaman komunitas di mana pengguna dapat melihat daftar berbagai grup dan forum pembaca. Pengguna dapat dapat berpartisipasi dalam diskusi, membagikan ulasan buku, dan terlibat dalam kegiatan lainnya yang diselenggarakan oleh komunitas. Berikut disajikan alur bagi pengguna yang ingin menggunakan komunitas.

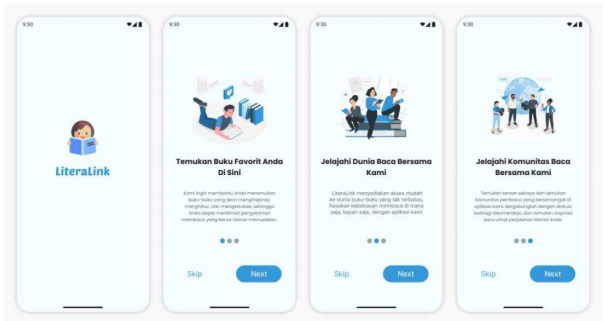


Gambar 7. Wireflow bergabung di komunitas

**b. Prototype**

Prototype aplikasi 'LiteraLink' akan dibuat berdasarkan konsep wireframe/wireflow yang telah dirancang sebelumnya. Prototype ini akan menjadi representasi visual interaktif dari antarmuka pengguna yang telah dirancang, memungkinkan pengguna untuk merasakan pengalaman nyata dalam menggunakan aplikasi. Melalui prototype ini dapat diuji fungsionalitas, navigasi, fitur-fitur utama dalam konteks yang lebih nyata sebelum aplikasi sebenarnya dikembangkan. Berikut beberapa tampilan antarmuka dari beberapa proses yang ada:

- Tampilan Splash Screen dan Onboarding Page

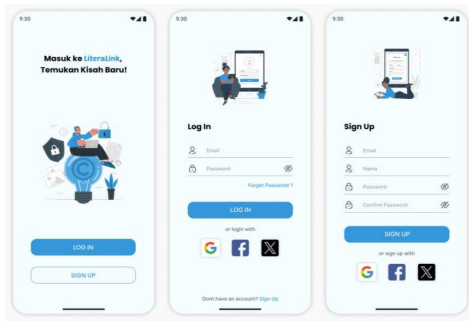


Gambar 8. Landing Page LiteraLink App

Splash Screen menjadi halaman pertama yang muncul ketika pengguna membuka aplikasi, memberikan kesan awal yang menarik dan mencerminkan identitas visual

aplikasi. Sedangkan onboarding page pada aplikasi ini memberikan pengguna informasi singkat tentang fitur-fitur utama aplikasi dan cara penggunaannya seperti yang terlihat pada Gambar 8, sehingga pengguna lebih memahami cara menggunakan aplikasi sebelum mulai menjelajahi lebih jauh.

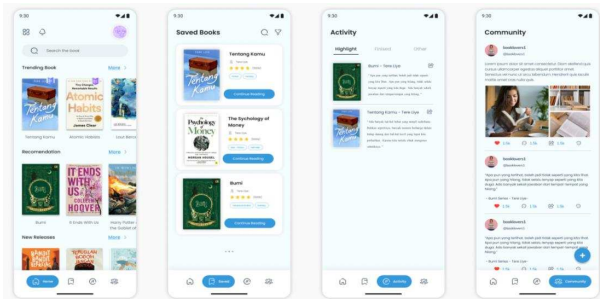
- Tampilan Log In dan Sign Up.



Gambar 9. Tampilan Log In dan Sign Up

Tampilan log in akan memungkinkan pengguna yang telah memiliki akun masuk ke dalam aplikasi dengan memasukkan email dan kata sandi, dan juga terdapat pilihan untuk melakukan log in melalui aplikasi pihak ketiga. Lalu, untuk sign up akan memungkinkan pengguna yang belum mempunyai akun dapat mengisi formulir pendaftaran, seperti nama lengkap, email, dan kata sandi yang diinginkan untuk membuat akun baru.

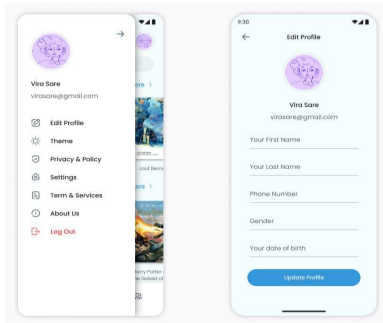
- Tampilan Fitur Utama



Gambar 10. Tampilan Halaman Navigasi

Berdasarkan tampilan yang disajikan pada Gambar 10, aplikasi 'LiteraLink' memiliki empat halaman navigasi, yaitu halaman beranda, halaman untuk menyimpan buku yang diinginkan, halaman aktivitas pengguna, dan juga halaman komunitas. Berikut lebih jelasnya akan disajikan tampilan dari masing-masing halaman navigasi.

- **Tampilan Beranda**  
Tampilan beranda adalah halaman utama aplikasi 'LiteraLink' yang menampilkan rekomendasi buku, kategori buku, dan pilihan buku populer atau yang sedang trending. Pengguna dapat menelusuri berbagai buku dan memulai membaca langsung dari halaman beranda ini.
- **Tampilan Buku yang Disimpan**  
Tampilan buku yang disimpan merupakan halaman yang berisi daftar buku yang disimpan oleh pengguna. Pengguna dapat menandai buku yang menarik dan menyimpannya untuk dibaca nanti. Halaman ini memudahkan pengguna untuk mengakses buku-buku yang telah pengguna pilih sebelumnya.
- **Tampilan Aktivitas**  
Pada halaman ini ditampilkan aktivitas terkini pengguna, seperti highlight yang ditambahkan, buku yang selesai dibaca, atau ulasan yang dibagikan.
- **Tampilan Komunitas**  
Komunitas adalah halaman yang memungkinkan pengguna untuk terhubung dengan komunitas pembaca lainnya. Di sini, pengguna dapat berpartisipasi dalam diskusi, berbagi rekomendasi buku, atau pengalaman membaca. Halaman ini menciptakan ruang bagi pengguna untuk berinteraksi dan berbagi minat membaca pengguna.
- **Tampilan Sidebar dan Edit Profile.**



**Gambar 11.** Tampilan Sidebar dan Edit Profile

Sidebar pada aplikasi 'LiteraLink' merupakan bagian menu yang terletak di sisi kiri layar. Sidebar ini menyediakan akses cepat ke berbagai fitur dan menu penting dalam aplikasi. Pengguna dapat mengakses pengaturan, tema, bantuan, dan juga informasi terkait aplikasi melalui sidebar ini.

### 3.5 Test

Penelitian ini menggunakan metode System Usability Scale (SUS) untuk mengukur kegunaan pada prototype aplikasi secara sistematis. Skala ini berkisar dari 1 (sangat tidak setuju) hingga 5 (sangat setuju) dan diterapkan melalui kuesioner yang disebarluaskan melalui platform Google Form kepada para responden yang berpartisipasi. Setiap tanggapan dievaluasi menggunakan usability scale. Setelah semua tanggapan dinilai, rata-rata dari skor usability dihitung. Jika rata-rata skor usability lebih besar atau sama dengan 85, maka pengujian dianggap selesai dan aplikasi siap untuk dipublikasikan dan digunakan oleh pengguna [5].

**Tabel 1.** Pertanyaan System Usability Scale

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

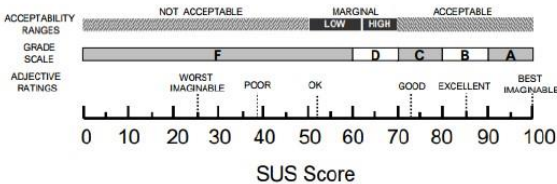
Setelah didapatnya data responden dengan skala 1 sampai 5 poin berdasarkan pertanyaan tersebut, selanjutnya dilakukan perhitungan skor SUS. Untuk pertanyaan bernomor ganjil (1,3,5,7, dan 9), kontribusi skor dihitung dengan mengurangi 1 dari skor yang diberikan. Sedangkan untuk pertanyaan bernomor genap (2,4,6,8, dan 10), kontribusinya dihitung dengan mengurangi skor dari 5. Setelah semua kontribusi skor dari setiap pertanyaan dijumlahkan, totalnya dikalikan dengan 2,5 untuk mendapatkan skor akhir SUS. Responden yang berpartisipasi dalam pengujian usability ini terdiri dari 15 orang, dengan rentang usia 15 tahun hingga 50 tahun, berasal dari berbagai kalangan, mulai dari pelajar, mahasiswa/mahasiswi maupun para pekerja. Sehingga hasil perhitungan jawaban dari responden disajikan pada tabel berikut.

**Tabel 2.** Hasil Perhitungan SUS

| Responden | Kuesioner SUS |    |    |    |    |    |    |    |    |     | Nilai Skor |
|-----------|---------------|----|----|----|----|----|----|----|----|-----|------------|
|           | Q1            | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |            |
| R1        | 3             | 4  | 4  | 4  | 3  | 4  | 3  | 4  | 3  | 0   | 80         |
| R2        | 2             | 3  | 1  | 3  | 3  | 3  | 3  | 3  | 3  | 0   | 60         |

| Responden         | Kuesioner SUS |    |    |    |    |    |    |    |    |     | Nilai Skor   |
|-------------------|---------------|----|----|----|----|----|----|----|----|-----|--------------|
|                   | Q1            | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |              |
| R3                | 2             | 2  | 3  | 3  | 3  | 3  | 2  | 2  | 2  | 2   | 60           |
| R4                | 4             | 4  | 4  | 1  | 4  | 4  | 4  | 4  | 4  | 0   | 82,5         |
| R5                | 4             | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 0   | 90           |
| R6                | 4             | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 3   | 97,5         |
| R7                | 3             | 4  | 4  | 4  | 3  | 3  | 4  | 4  | 3  | 3   | 87,5         |
| R8                | 3             | 3  | 3  | 3  | 3  | 3  | 3  | 3  | 3  | 3   | 75           |
| R9                | 3             | 4  | 4  | 3  | 2  | 2  | 4  | 4  | 3  | 4   | 82,5         |
| R10               | 3             | 3  | 3  | 4  | 2  | 3  | 3  | 4  | 3  | 3   | 77,5         |
| R11               | 4             | 3  | 3  | 4  | 3  | 4  | 4  | 3  | 3  | 2   | 82,5         |
| R12               | 4             | 0  | 4  | 4  | 4  | 4  | 4  | 2  | 4  | 3   | 82,5         |
| R13               | 4             | 3  | 2  | 3  | 3  | 4  | 4  | 3  | 3  | 3   | 80           |
| R14               | 4             | 3  | 3  | 3  | 2  | 4  | 4  | 2  | 4  | 4   | 82,5         |
| R15               | 4             | 3  | 3  | 4  | 4  | 2  | 3  | 2  | 4  | 4   | 82,5         |
| <b>Total Skor</b> |               |    |    |    |    |    |    |    |    |     | <b>80,16</b> |

Seperti yang terlihat pada Tabel 2, total skor akhir setelah dikalkulasi adalah 1202,5 yang kemudian dibagi dengan jumlah responden sehingga menghasilkan skor 80,16. Berdasarkan skor SUS yang diperoleh, langkah selanjutnya adalah menentukan grade pada indikator penilaian SUS. Indikator penilaian SUS dapat dilihat pada Gambar 12 di bawah ini, yang akan membantu dalam menentukan kategori yang sesuai.



**Gambar 12.** Indikator Penilaian SUS

Berdasarkan gambar indikator penilaian SUS tersebut, dengan hasil akhir sebesar 80,16 aplikasi ini masuk dalam kategori excellent. Skor ini menunjukkan bahwa antarmuka pengguna yang dirancang berhasil memenuhi harapan dan kebutuhan pengguna dengan baik.

#### 4. Kesimpulan

Hasil penelitian menunjukkan bahwa penerapan metode design thinking dalam merancang antarmuka dari aplikasi membaca digital pada aplikasi 'LiteraLink' berhasil diimplementasikan. Melalui lima tahapan metode design thinking yang digunakan menciptakan antarmuka pengguna yang optimal. Dengan nilai 80,16 pada System Usability Scale (SUS), aplikasi ini dinilai baik dan berpotensi meningkatkan minat membaca serta mendukung literasi di kalangan masyarakat, khususnya anak-anak dan remaja.

#### Daftar Pustaka

- [1] Prayoga, R.A., Budiarto, H. A, Afif, M. F., Pradipta, A. S., and Lestari, A. S., "Peningkatan Minat Baca Siswa Melalui Pekan Literasi dan Pembiasaan 15 Menit Membaca: Studi Kasus MI Mulyadarama Girimukti", *Jurnal Warta LPM*, Vol. 26, No.4, pp.1410-9344, 2023
- [2] Imanda, E.T. and Mukhayaroh, A., "Metode Design Thinking Perancangan User Interface dan User Experience Aplikasi "KuyBaca"", *Journal of Students Research*, Vol. 4, No. 1, pp.23-36, 2023
- [3] Razi, A.A, Mutiaz, I. R., and Setiawan, P., "Penerapan Metode Design Thinking Pada Model Perancangan UI/UX Aplikasi Penanganan Laporan Kehilangan dan Temuan Barang Tercecer", *Jurnal Desain Komunikasi Visual, Manajemen Desain dan Periklanani*, Vo.03, No.02, pp. 2477-6106, 2018
- [4] Widhiana, I. B. A, and Astuti, L. G, "Perancangan User Interface dan User Experience Website KosiN dengan Pendekatan Design Thinking", *Jurnal Nasional Teknologi Informasi dan Aplikasinya*. Vol.2, No.2, pp.2986-3929, 2024
- [5] Charisa, Fini, "Perancangan User Interface dan Experience Aplikasi Skripsiku Dengan Fitur Auto Layout Menggunakan Metode Usability Testing", *Teknik Informatika*. 2022



# Pengamanan File Video dengan Enkripsi dan Deskripsi Menggunakan Algoritma Salsa20 dan RSA

I Putu Herdy Juniawan<sup>a1</sup>, I Komang Ari Mogi<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana

Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia

<sup>1</sup>juniawan.2208561033@student.unud.ac.id

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

## Abstract

*In the face of escalating security challenges in the advanced digital era, the exchange and storage of video files have become integral to daily life, impacting everything from business communication to personal entertainment. While the internet facilitates easy access to multimedia content, it also brings heightened risks of data theft and privacy breaches. Cryptography offers a solution by encrypting sensitive data. This research employs two encryption algorithms, Salsa20 and RSA, to protect video files. Salsa20 ensures fast and secure encryption, maintaining data integrity and confidentiality. RSA facilitates secure key exchange, limiting access to authorized recipients. The combined use of these algorithms effectively secures video files, guarding against data theft and privacy breaches. This offline encryption system offers flexible accessibility without compromising security. Additionally, it preserves the original file extension, adding an extra layer of protection, while decryption enables seamless access and playback of video content.*

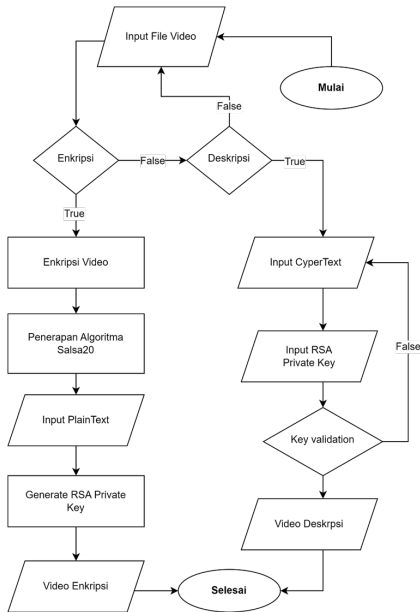
**Keywords:** Video Encryption, Salsa20, RSA, Data Security, Cryptography.

## 1. Pendahuluan

Dalam menghadapi tantangan keamanan dalam era digital yang semakin maju, Penting untuk menyadari bahwa berbagi dan menyimpan file video kini menjadi bagian tak terpisahkan dari kehidupan sehari-hari, mempengaruhi berbagai aspek dari komunikasi bisnis hingga hiburan pribadi. Seiring dengan kemajuan teknologi, internet telah memfasilitasi akses yang lebih cepat dan mudah terhadap konten multimedia, namun, di sisi lain, meningkatnya konektivitas ini juga menimbulkan risiko yang tak terhindarkan terkait dengan pencurian data dan pelanggaran privasi[1]. Kriptografi telah muncul sebagai solusi yang ampuh untuk menjaga data sensitif agar tidak diakses oleh pihak yang tidak berwenang. Dengan menggunakan teknik-teknik kriptografi yang canggih, informasi yang disimpan dalam format video dapat dienkripsi, membuatnya sulit untuk diakses oleh pihak yang tidak berwenang[2]. Enkripsi mengacu pada proses mengubah data ke dalam format yang tidak dapat dibaca tanpa memiliki kunci enkripsi yang cocok., sedangkan deskripsi adalah proses mengembalikan data ke keadaan aslinya menggunakan kunci dekripsi yang tepat. Selain itu, dengan meningkatnya kesadaran akan keamanan informasi, implementasi kriptografi dalam pertukaran dan penyimpanan file video telah menjadi semakin penting. Dalam jurnal ini peneliti menerapkan dua algoritma enkripsi yaitu Salsa20 dan RSA. Algoritma Salsa20 adalah algoritma enkripsi arus yang terkenal karena kecepatan dan keamanannya. Dengan menggunakan Salsa20, peneliti dapat mengenkripsi file video secara efisien dan efektif, menjaga integritas dan kerahasiaan data. Namun, untuk mengamankan proses pertukaran kunci enkripsi Salsa20, peneliti juga menerapkan algoritma kunci publik RSA sebagai enkripsi tambahan. RSA menyediakan mekanisme yang aman untuk pertukaran kunci, memastikan bahwa hanya penerima yang dimaksud yang dapat mendapatkan akses ke kunci enkripsi yang diperlukan[3].

## 2. Metode Penelitian

### 2.1. Desain Sistem

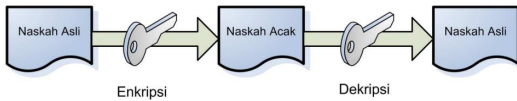


Gambar 1. Flowchart Sistem

Untuk alur cara kerja *Flowchart* Sistem tahapan dimulai dengan memasukkan file video yang akan dienkripsi atau didekripsi. Pengguna kemudian memilih apakah ingin melakukan enkripsi atau dekripsi. Jika opsi enkripsi dipilih, proses dilanjutkan dengan penggunaan algoritma *Salsa20* untuk mengenkripsi video tersebut. Selain itu, pengguna diminta untuk memasukkan *plaintext* yang nantinya akan dijadikan sebagai kunci dekripsi juga, untuk memastikan keamanan tambahan. Setelah itu, kunci privat RSA dihasilkan untuk keamanan tambahan. Hasil akhir dari proses ini adalah video yang terenkripsi dengan aman, sehingga terhindar dari akses tanpa izin. Di sisi lain, jika opsi dekripsi dipilih, pengguna diminta memasukkan *Ciphertext* (teks terenkripsi) dan kunci privat RSA. Validasi kunci dilakukan, dan jika benar, proses lanjut ke tahap dekripsi. Video kemudian didekripsi, dan hasilnya adalah video yang terdekripsi sepenuhnya, siap untuk ditonton kembali. Dalam kesimpulan, *flowchart* ini memberikan gambaran yang jelas mengenai langkah-langkah yang harus diikuti untuk mengamankan sebuah video melalui enkripsi serta mengembalikan video tersebut ke bentuk semula melalui dekripsi.

## 2.2. Kriptografi

Kriptografi adalah teknik-teknik pengamanan data, Informasi diubah menggunakan suatu kunci enkripsi sehingga sulit untuk dibaca oleh pihak yang tidak memiliki kunci dekripsi yang sesuai. Proses dekripsi kemudian menggunakan kunci tersebut untuk mengembalikan data ke bentuk aslinya[4]. Sementara itu, asal-usul kata "kriptografi" berasal dari bahasa Yunani, menggabungkan kata "cryptos" yang berarti "secret" (tersembunyi) dan "graphein" yang berarti "writing" (penelitian). Oleh karena itu, kriptografi dapat diartikan sebagai "secret writing" (penelitian rahasia) [4].



Gambar 2. Kriptografi

Gambar diatas menggambarkan hasil dari proses enkripsi dan dekripsi. Secara umum, proses enkripsi melibatkan pengacakan teks asli (plaintext) menjadi teks teracak (ciphertext) yang sulit dibaca bagi siapa pun yang tidak memiliki kunci dekripsi [5].

## 2.3. Algoritma Salsa20

Algoritma Salsa20 adalah sebuah algoritma kriptografi simetris yang dikembangkan oleh Daniel J. Bernstein pada tahun 2005. Algoritma ini digunakan untuk melakukan enkripsi dan dekripsi data dengan kecepatan tinggi serta tingkat keamanan yang tinggi. Langkah-langkah utama dalam algoritma Salsa20 melibatkan inialisasi, pembangkitan kunci, XOR, iterasi, dan output.

Tabel 1. Initial State Salsa20

|     |      |       |       |
|-----|------|-------|-------|
| Key | Cons | Nonce | Nonce |
| Pos | Pos  | Cons  | Key   |
| Key | Key  | Key   | Cons  |

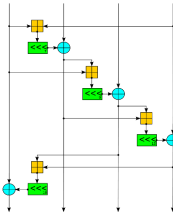
$$b \oplus = (a + d) \lll 7; \quad (1)$$

$$c \oplus = (b + a) \lll 9; \quad (2)$$

$$d \oplus = (c + b) \lll 13; \quad (3)$$

$$a \oplus = (d + c) \lll 18; \quad (4)$$

Algoritma Salsa20 merupakan salah satu kandidat dalam proyek eSTREAM. Algoritma ini menerima empat masukan, yaitu 8-byte block counter, 8-byte nonce (nomor pesan unik) yang biasanya berupa angka acak atau angka pseudorandom, dan 32 byte kunci. Keempat masukan untuk Algoritma Salsa20 diatur seperti yang ditunjukkan pada Tabel 1. Setiap kotak matriks dalam Gambar 1 memiliki ukuran 4 byte. Struktur ini dikenal sebagai initial state. Setiap nilai indeks dari matriks dalam initial state diubah menjadi bentuk little endian[6].

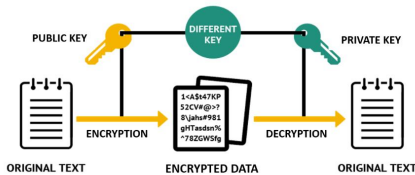


Gambar 3. Fungsi quarter rounds

Langkah selanjutnya setelah proses konversi ke bentuk little endian adalah pembentukan keystream. Dalam proses ini, algoritma Salsa20 melakukan total 20 putaran, di mana setiap putaran melibatkan fungsi quarter round. Persamaan 1-4 diterapkan pada setiap putaran dengan perbedaan: pada putaran ganjil, fungsi quarter round hanya memproses bagian kolom dari struktur, sementara pada putaran genap, fungsi quarter round memproses bagian baris. Hasil dari putaran terakhir ditambahkan dengan keadaan awal dalam bentuk little endian. Kemudian, hasil operasi sebelumnya disimpan dalam bentuk little endian sebagai keystream [6]. Keystream adalah komponen kunci krusial dalam proses enkripsi dan dekripsi algoritma Salsa20, diperlukan untuk operasi XOR dengan teks plaintext atau ciphertext. Setelah inisialisasi dengan kunci dan vektor inisialisasi, algoritma menghasilkan aliran kunci yang sesuai dengan panjang data yang akan dienkripsi atau didekripsi. Data dienkripsi dengan cara di-XOR-kan dengan aliran kunci yang dihasilkan [6]. Algoritma Salsa20 melakukan serangkaian iterasi menggunakan fungsi operasi dan permutasi untuk menghasilkan keystream. Setelah iterasi selesai, aliran kunci yang dihasilkan digunakan untuk melakukan enkripsi atau dekripsi data. Keamanan algoritma Salsa20 bergantung pada kekuatan kunci yang digunakan dan ketangguhan fungsi operasi dan permutasi di setiap putaran.

#### 2.4. Algoritma RSA (Rivest Shamir Adleman)

Algoritma RSA adalah sebuah algoritma kriptografi kunci publik yang dikembangkan oleh Ron Rivest, Adi Shamir, dan Leonard Adleman pada tahun 1978. Dalam algoritma ini, pesan dienkripsi menggunakan kunci publik, dan hanya dapat didekripsi dengan kunci privat yang sesuai. RSA didasarkan pada kesulitan dalam memecahkan faktorisasi dari bilangan-bilangan besar. Algoritma ini memiliki berbagai aplikasi, termasuk dalam pengamanan komunikasi digital dan penandatanganan digital[7].



Gambar 4. Algoritma RSA

Proses pembuatan sepasang kunci RSA (Rivest Shamir Adleman) melibatkan langkah-langkah sebagai berikut:

**a. Inisialisasi:**

- Pilih dua bilangan prima acak, sebagai contoh  $P = 11$  dan  $q = 19$ .
- Hitung nilai  $n$ :  $n = p \times q = 11 \times 19 = 209$
- Hitung nilai  $\phi(n)$ :  $\phi(n) = (p-1) \times (q-1) = 10 \times 18 = 180$

**b. Kunci Publik:**

Pilih bilangan acak  $e$  yang relatif prima dengan  $\phi(n)$ , misalnya  $e = 7$ .

**c. Kunci Pribadi:**

Temukan bilangan  $d$  yang memenuhi syarat  $d \times e \equiv 1 \pmod{\phi(n)}$ .  
Gunakan algoritma extended Euclidean untuk mencari  $d$ :

1. Inisialisasi:  $\phi(n) = 180, e = 7$ .
2. Langkah 1:  $180 = 7 \times 25 + 5$ .
3. Langkah 2:  $7 = 5 \times 1 + 2$ .
4. Langkah 3:  $5 = 2 \times 2 + 1$ .
5. Langkah 4: Terapkan substitusi balik:  $1 = 5 - 2 \times 2 = (180 - 7 \times 25) \times 2 = 175 \times 2 - 180 \times 2$
6. Langkah 5: Hasilkan  $d$  dengan memodulo  $\phi(n)$ :  $d = 175 \pmod{180} = 175$

Hasil dari proses ini adalah kunci publik  $(n, e) = (209, 7)$  dan kunci pribadi  $(n, d) = (209, 175)$ . Ini menjelaskan langkah-langkah awal dalam pembuatan sepasang kunci RSA, termasuk inisialisasi, pemilihan kunci publik, dan perhitungan kunci pribadi[2].

**3. Hasil dan Diskusi**

**3.1. Uji Coba Sistem**





**a. Proses Enkripsi Video**

Proses enkripsi video ini bertujuan untuk mengevaluasi kemampuan sistem dalam mengenkripsi file video menggunakan algoritma Salsa20, dengan tambahan enkripsi RSA. Tahapan-tahapan yang dilakukan adalah sebagai berikut:

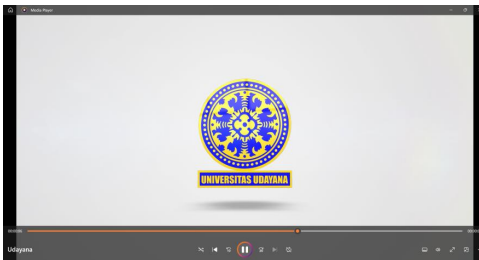
```
D:\Herdy\SNATIA>python salsa20.py
=====
Masukkan path file video: D:\Herdy\SNATIA\Udayana.mp4
=====
Ukuran file: 22367946 bytes
=====
Pilih mode:
(e) Enkripsi
(d) Deskripsi
e
=====
Masukkan plaintext yang akan dienkripsi: 12345
=====
File berhasil dienkripsi.
Ukuran file asli : 22367946 bytes
File terenkripsi tersimpan sebagai: D:\Herdy\SNATIA\Udayana_encrypted.mp4
Kunci terenkripsi : HEr2ciQ1h8PCdRv1vZKAB87970wtz7RIEgn5QaasBDGwHk5
U10GxY18cg9Ww/zjWjJqOfs3v/wwv/ae8cbadepZwML5MAym35Y9TEGfG150atjJd8z0uAYW07
a4BEYSxa6J7w18MGlcclycfxNHu/K4j@pyusVITrsi6LL7GIJpngNjRkGS1+S8UtR8R4E1D/nD
HNnPSFdaQunVjEf6g+jn5RWzQ3F7IHGRzG==
Waktu yang diperlukan : 0.71 detik
=====
```

Gambar 5. Proses Enkripsi Video

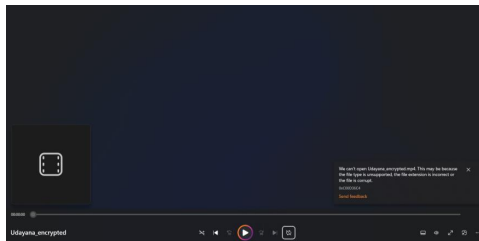
- **Persiapan File Video:** File video dengan format .mp4 dipersiapkan untuk proses enkripsi.
- **Input Path Video:** Pengguna diminta untuk memasukkan lokasi file video yang akan dienkripsi ke dalam program.
- **Enkripsi dengan Salsa20:** Program akan menginisiasi proses enkripsi video menggunakan algoritma Salsa20.
- **Pemasukan Plaintext:** Pengguna diminta untuk memasukkan plaintext yang akan digunakan untuk mendekripsi file video nanti.
- **Generasi RSA Key:** Program akan menghasilkan kunci RSA dalam bentuk `public_key.pem` dan `private_key.pem`, yang dapat digunakan untuk proses dekripsi video.
- **Penyelesaian Proses Enkripsi:** Setelah proses enkripsi selesai, program akan secara otomatis membuat file video dengan tambahan awalan 'encrypted\_' pada nama file aslinya.

|   |                  |               |           |
|---|------------------|---------------|-----------|
|  private_key.pem       | 10/05/2024 19:18 | PEM File      | 2 KB      |
|  public_key.pem        | 10/05/2024 19:18 | PEM File      | 1 KB      |
|  Udayana_encrypted.mp4 | 10/05/2024 19:18 | MP4 File      | 21.844 KB |
|  encrypted_key.txt     | 10/05/2024 19:18 | Text Document | 1 KB      |

Gambar 6. Hasil Enkripsi Video



Gambar 7. Video Sebelum Enkripsi

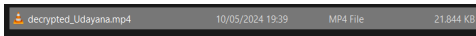


Gambar 7. Video Sesudah Enkripsi

b. Proses Enkripsi Video

Berikut adalah deskripsi untuk setiap tahapan dalam proses deskripsi file video:

- **Persiapan File Video:** Tahap ini melibatkan pemilihan file video yang akan dideskripsi. File video dengan format .mp4 dipersiapkan untuk proses deskripsi.
- **Input Path Video:** Pengguna diminta untuk menyediakan path atau lokasi dari file video yang akan dideskripsi. Path ini akan digunakan oleh program untuk mengakses file video tersebut.
- **Input CyperText:** Pengguna akan diminta untuk menginputkan ciphertext yang diperoleh dari proses enkripsi sebelumnya. Ciphertext ini adalah hasil dari proses enkripsi menggunakan algoritma Salsa20 dan RSA.
- **Input Path RSA Private Key:** Pengguna juga diminta untuk memasukkan path atau lokasi dari file Private Key yang digunakan dalam proses enkripsi sebelumnya. Private Key ini diperlukan untuk mendeskripsi ciphertext menjadi plaintext.
- **Validasi Kunci:** Program akan melakukan validasi terhadap kedua kunci yang dimasukkan oleh pengguna, yaitu ciphertext dan RSA Private Key. Jika keduanya sesuai, program akan melanjutkan ke tahap selanjutnya.
- **Penyelesaian Proses Deskripsi:** Setelah proses validasi kunci selesai, program akan melakukan proses deskripsi terhadap file video menggunakan kunci yang telah divalidasi. Hasil deskripsi akan disimpan dengan nama file baru yang memiliki awalan "decrypted\_".



Gambar 8. Hasil Deskripsi Video



Gambar 9. Video Sesudah Deskripsi

3.2. Hasil Uji Coba Sistem

Tabel 1. Hasil Pengujian Deskripsi Video

| No | Bulir Pengujian         | Output yang diinginkan  | Output yang keluar   | Keterangan |
|----|-------------------------|---|--|------------|
| 1  | Membaca data video      | Sistem dapat membaca data video dari file explorer.                     | Sistem berhasil membaca data video dari file explorer.                     | Sesuai     |
| 2  | Plaintext ke Ciphertext | Sistem dapat membaca plaintext dan mengubahnya dalam bentuk ciphertext. | Sistem berhasil membaca plaintext dan mengubahnya dalam bentuk ciphertext. | Sesuai     |

| No | Bulir Pengujian          | Output yang diinginkan  | Output yang keluar   | Keterangan |
|----|--------------------------|---|--|------------|
| 3  | Generate key             | Sistem mampu menghasilkan sepasang kunci RSA yaitu public key dan <i>private key</i> .  | Sistem berhasil mendapatkan sepasang kunci RSA yaitu public key dan <i>private key</i> .   | Sesuai     |
| 4  | Enkripsi data video      | Sistem mampu mengenkripsi data video sesuai algoritma RSA(Rivest Shamir Adleman) dan Salsa20  | Sistem berhasil mengenkripsi data video sesuai algoritma RSA(Rivest Shamir Adleman) dan Salsa20.   | Sesuai     |
| 5  | Dekripsi data video      | Sistem dapat mendeteksi data video yang telah terenkripsi untuk mendekripsikan kembali menggunakan <i>private key</i> dan <i>ciphertext</i> | Sistem berhasil mendeteksi data video yang telah terenkripsi untuk mendekripsikan kembali menggunakan <i>private key</i> dan <i>ciphertext</i> | Sesuai     |
| 6  | Deteksi kesalahan sistem | Sistem dapat mendeteksi kesalahan pengguna seperti pengguna tidak memasukkan file video dan kesalahan pada kunci deskripsi                  | Sistem berhasil mendeteksi kesalahan pengguna seperti pengguna tidak memasukkan file video dan kesalahan pada kunci deskripsi                  | Sesuai     |

#### 4. Kesimpulan

Berdasarkan pada penelitian dan pembahasan diatas, peneliti menyimpulkan bahwa penerapan Enkripsi dan Deskripsi pada File Video menggunakan Algoritma Salsa20 dan RSA adalah solusi yang efektif dalam menjaga kerahasiaan dan integritas data video, serta melindunginya dari ancaman pencurian data dan pelanggaran privasi. Dalam program ini, kombinasi metode enkripsi Salsa20 dan RSA telah terbukti efisien, menghasilkan lapisan keamanan yang kuat bagi data video. Metode RSA memberikan tambahan keamanan tingkat tinggi sementara Salsa20 mempercepat proses enkripsi secara signifikan. Sistem Enkripsi dan Dekripsi ini mampu beroperasi secara offline, memastikan aksesibilitas yang fleksibel tanpa mengorbankan keamanan. Proses enkripsi dalam sistem ini mampu mengamankan konten data video tanpa mengubah ekstensi aslinya, sehingga menghasilkan lapisan perlindungan yang lebih dalam. Sedangkan, proses dekripsi mempertahankan ekstensi asli dan mengembalikan isi data video ke bentuk semula, memungkinkan data video tersebut dapat diakses dan diputar kembali dengan lancar.

#### Daftar Pustaka

- [1] F. Zuli and A. Irawan, "Implementasi Kriptografi Dengan Algoritma Blowfish Dan Riverst Shamir Adleman (RSA) Untuk Proteksi File," vol. 9, no. 1, pp. 5–13, 2017, doi: 10.22441/fifo.v9i1.2568.
- [2] Yusmaifany, Tommy, and R. Siregar, "Aplikasi Enkripsi Data Video Menggunakan Metode Rsa Dan Blowfish Berbasis Web," vol. 2, no. 3, pp. 2024–535, 2024.
- [3] A. Ramadhan, A. Kusyanti, and P. H. Trisnawan, "Implementasi Algoritme Enkripsi Salsa20 untuk Pengamanan Data Video Surveillans secara Real-Time," vol. 5, no. 2, pp. 477–484, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] M. Qamal, "Kriptografi File Citra Menggunakan Algoritma Tea (Tiny Encryption Algorithm)," vol. 5, no. 2, pp. 11–43, 2014.
- [5] S. Kromodimoeljo, *Teori dan Aplikasi Kriptografi*. 2009.



- [6] M. Thareq, P. Beyri, A. Kusyanti, and F. A. Bakhtiar, "Implementasi Algoritme Salsa20 untuk Pengamanan Search Keyword Dokumen Terenkripsi," vol. 4, no. 10, pp. 3531–3541, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] U. Indriani, O. Alfina, and N. Syahputri, "Journal of Machine Learning and Data Analytics (MALDA) Penerapan Algoritma RSA Dalam Keamanan File Ms Word," vol. 1, no. 2, pp. 95–100, 2022.

Halaman ini sengaja dibiarkan kosong

# Implementasi Algoritma Modified K-Nearest Neighbor untuk Klasifikasi Indeks Kualitas Udara Perkotaan di Berbagai Negara

Ni Made Ayu Pranasanthi Dewi<sup>a1</sup>, I Made Widiartha<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>gekana744@gmail.com  
<sup>2</sup>madewidiartha@unud.ac.id

## Abstract

*One important component for the sustainability of living organisms is the availability of clean air. Air pollution has detrimental repercussions such as deteriorating air quality, which can have harmful impacts on the environment. As a result, modeling and monitoring air quality are essential steps in reducing air pollution, and the Air Quality Index allows for the review of this monitoring. This study's objective is to apply the Modified K-Nearest Neighbor method in classifying urban Air Quality Index data from various countries. The "World Air Quality Index by City and Coordinates" Kaggle website provided the data that were used. The data were then processed and tested using the Modified K-Nearest Neighbor model, this builds on the K-Nearest Neighbor technique. 16,695 data in total were divided into 80% training data (13,356 data) and 20% testing data (3,339 data) in order to conduct testing. Evaluation was performed by comparing the performance of the Modified K-Nearest Neighbor method with the traditional K-Nearest Neighbor method. The Modified K-Nearest Neighbor approach with  $K=1$  produced an accuracy rate of 99.73% during testing, whereas the K-Nearest Neighbor method produced an accuracy rate of 99.64%. Using the K-Fold Cross Validation, K-Nearest Neighbor method perform highest mean score of 99,04% and Modified K-Nearest Neighbor perform highest mean score of 99,46%.*

**Keywords:** Air Pollution, Air Quality Index, Classification, K-Nearest Neighbor, Modified K-Nearest Neighbor.

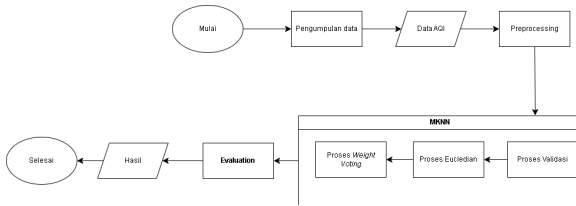
## 1. Pendahuluan

Salah satu komponen yang memiliki peran penting dalam kelangsungan makhluk hidup adalah ketersediaan udara. Adanya kontaminan yang terdapat dalam udara memberikan efek negatif berupa penurunan kualitas udara. Kualitas udara atau *air quality* merupakan kadar kandungan udara berdasarkan konsentrasi polutan di lokasi tertentu [1], semakin tinggi kontaminan yang terdapat di suatu daerah maka berbanding lurus dengan tingkat polusi udara di daerah tersebut. Polusi udara dapat berasal dari berbagai aktivitas seperti bencana alam, emisi industri, pembakaran, serta aktivitas manusia lainnya. Komponen utama polutan udara meliputi partikel halus ( $PM_{2.5}$ ), Nitrogen Dioksida ( $NO_2$ ), Sulfur Dioksida ( $SO_2$ ), Karbon Monoksida (CO), dan Ozon ( $O_3$ ) [2]. Menurut WHO, dampak berbahaya dari polusi udara menyebabkan sekitar 7 juta kematian setiap tahunnya dan berkontribusi pada berbagai masalah kesehatan seperti penyakit pernapasan, penyakit kardiovaskular, dan kanker [3]. Kabut asap yang disebabkan oleh kebakaran hutan dan lahan atau karhutla mengakibatkan setidaknya 6.025 warga menderita infeksi saluran pernapasan akut (ISPA) [4]. Oleh karena itu pemantauan dan pemodelan kualitas udara menjadi sangat penting dalam upaya untuk menanggulangi polusi udara, pemantauan ini dapat ditinjau dari Indeks Kualitas Udara. *Air Quality Index* (AQI) atau Indeks Kualitas Udara adalah suatu sistem yang digunakan untuk mengukur, menilai, dan melaporkan kualitas udara pada suatu wilayah. Meninjau uraian tersebut maka sangat penting dilakukan deteksi indeks kualitas udara secara berkala sehingga dapat dilakukan pengontrolan yang sesuai untuk menghindari risiko kerusakan lingkungan lebih lanjut yang dapat memengaruhi kesehatan

masyarakat. Dalam menyelesaikan permasalahan ini, salah satu pendekatan yang dapat digunakan adalah dengan mengaplikasikan metode klasifikasi menggunakan model *Machine Learning*. *Machine Learning* (ML) atau Pembelajaran Mesin adalah cabang dari *Artificial Intelligence* (AI) yang fokus belajar dari data (*learn from data*), yaitu fokus pada pengembangan sistem yang mampu belajar secara "mandiri" tanpa harus berulang kali diprogram manusia [5]. Metode klasifikasi menggunakan ML dapat dilakukan dengan beberapa algoritma seperti *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan *Support Vector Machine* (SVM). Berdasarkan penelitian sebelumnya yang berjudul "*Comparative Analysis of Restock Needs Bottled Water using K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and the Naive Bayes Algorithm*" mengenai analisis perbandingan kebutuhan pengisian ulang air minum botol dengan dengan algoritma *K-Nearest Neighbor*, *Support Vector Machine*, dan *Naive Bayes* menunjukkan akurasi bahwa metode klasifikasi menggunakan algoritma *K-Nearest Neighbor* mampu memberikan hasil terbaik dengan nilai akurasi 88,20% dibandingkan dengan algoritma *Naive Bayes* (66,20%) dan *Support Vector Machine* (84,51%) [6]. Algoritma *K-Nearest Neighbor* merupakan metode yang menerapkan prinsip "*nearest neighbor*" dimana prediksi untuk sebuah instansi baru didasarkan pada label atau nilai-nilai yang dimiliki oleh tetangga terdekatnya dalam ruang fitur yang ditentukan dari sejumlah K tetangga terdekat yang akan dipertimbangkan untuk melakukan prediksi [3]. Algoritma *K-Nearest Neighbor* memiliki kelemahan, salah satunya yaitu kelas objek baru ditentukan berdasarkan *voting* mayoritas kelas pada K jarak terdekat [7]. Maka dari itu dilakukan beberapa modifikasi untuk mengembangkan performa algoritma ini, yaitu dengan penambahan proses klasifikasi baru yang mencakup perhitungan nilai validitas untuk menilai validitas data latih dan perhitungan suara berat untuk menentukan bobot masing-masing terdekat. Pengembangan algoritma *K-Nearest Neighbor* ini selanjutnya disebut dengan algoritma *Modified K-Nearest Neighbor*. Hasil penelitian terdahulu terkait dengan proses klasifikasi tingkat kualitas udara di Provinsi DKI Jakarta menggunakan algoritma *Naive Bayes* mendapatkan hasil rata rata *accuracy* sebesar 88%, *recall* 96%, *precision* 85%, dan *f1-score* 90% [8]. Penelitian topik serupa dengan menerapkan algoritma *Random Forest Classifier* memperoleh hasil terbaik pada akurasi 90% [9]. Penelitian serupa berjudul "*Klasifikasi Data Mining untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-NN)*" memperoleh hasil akurasi tertinggi MKNN sebesar 75,35% [3]. Pada penelitian berjudul "*Feature Selection using Information Gain on the K-Nearest Neighbor (KNN) and Modified K-Nearest Neighbor (MKNN) Methods for Chronic Kidney Disease Classification*" mengenai klasifikasi data penyakit ginjal kronis yang membandingkan algoritma *Modified K-Nearest Neighbor* (MKNN) dan *K-Nearest Neighbor* (KNN) dengan 400 data yang terbagi menjadi 2 kelas diperoleh hasil akurasi tertinggi MKNN sebesar 98%, lebih tinggi 1,39% dibanding KNN dengan akurasi tertinggi 96,61% [10]. Kemudian penelitian selanjutnya dengan judul "*Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Kanker Payudara*" memperoleh akurasi dengan K=1 sebesar 97,61% [11]. Berdasarkan uraian diatas, penelitian ini mencoba untuk memberikan penyelesaian klasifikasi indeks kualitas udara dengan penerapan algoritma *Modified K-Nearest Neighbor* (MKNN) sebagai pengoptimasian dari penelitian sebelumnya. Penelitian ini menerapkan beberapa tahapan baru dalam algoritma *K-Nearest Neighbor* (KNN) sebagai modifikasi untuk dapat memberikan hasil akurasi yang lebih tinggi dan menggunakan pengujian variasi nilai K serta *K-fold Cross Validation*. Tujuan penelitian ini yaitu untuk mengimplementasikan dan menganalisis algoritma MKNN dalam mengklasifikasikan indeks kualitas udara pada sejumlah 174 negara di seluruh dunia.

## 2. Metode Penelitian

Penelitian ini mengimplementasikan algoritma *Modified K-Nearest Neighbor* untuk proses klasifikasi data indeks kualitas udara menggunakan *input* berupa data *Air Quality Index* (AQI) dari 174 negara. *Output* yang akan dihasilkan adalah hasil evaluasi dari algoritma *Modified K-Nearest Neighbor* dan *K-Nearest Neighbor* tradisional. Alur penelitian yang digunakan terdiri dari 5 tahapan dapat dilihat pada gambar 1, dimulai dari proses pengumpulan data, melakukan *preprocessing*, membagi data, melakukan klasifikasi dengan algoritma *Modified K-Nearest Neighbor* untuk menghasilkan hasil evaluasi berupa nilai akurasi.



Gambar 1. Alur penelitian

2.1. Pengumpulan Data

Data yang diaplikasikan dalam penelitian ini berupa data sekunder yaitu dataset *AQI and Lat Long of Countries* yang bersumber dari situs *Kaggle* yang berjumlah 16695 data. Dataset ini terakhir diperbarui pada tahun 2023. Dalam dataset ini terdapat informasi mengenai tingkat polusi udara di berbagai negara di dunia yang ditinjau dari *Air Quality Index* masing-masing negara. Karakteristik data ini terdiri dari 14 kolom dengan atribut berupa *country, city, AQI category, AQI value, AQI CO category, AQI CO value, AQI NO2 category, AQI NO2 value, AQI Ozone value, AQI Ozone category, AQI PM2.5 value, AQI PM2.5 category, latitude, serta longitude* dengan format file berupa *csv*. Data ini menampilkan indeks kualitas udara dari sejumlah 174 negara dengan tampilan 5 data teratas pada gambar 2. Fitur yang digunakan dalam penelitian ini berupa *AQI value, AQI CO category, AQI CO value, AQI NO2 category, AQI NO2 value, AQI Ozone value, AQI Ozone category, AQI PM2.5 value, AQI PM2.5 category* dan terdapat 5 kelas *output* dalam atribut *AQI category* yang digunakan sebagai label yaitu *good, moderate, unhealthy for sensitive group, unhealthy, very unhealthy dan hazardous*. Pada tabel 1 merupakan tabel atribut dan domain yang digunakan dalam penelitian ini.

Gambar 2. Atribut dataset *Air Quality Index*

| Country            | City             | AQI Value | AQI Category | CO AQI Value | CO AQI Category | Ozone AQI Value | Ozone AQI Category | NO2 AQI Value | NO2 AQI Category | PM2.5 AQI Value | PM2.5 AQI Category | lat      | lng      |
|--------------------|------------------|-----------|--------------|--------------|-----------------|-----------------|--------------------|---------------|------------------|-----------------|--------------------|----------|----------|
| Russian Federation | Praskoweya       | 51        | Moderate     | 1            | Good            | 36              | Good               | 0             | Good             | 51              | Moderate           | 44.7444  | 44.2051  |
| Brazil             | Presidente Dutra | 41        | Good         | 1            | Good            | 5               | Good               | 1             | Good             | 41              | Good               | -5.2900  | -44.4900 |
| Brazil             | Presidente Dutra | 41        | Good         | 1            | Good            | 5               | Good               | 1             | Good             | 41              | Good               | -11.2958 | -41.9869 |
| Italy              | Priolo Gargallo  | 66        | Moderate     | 1            | Good            | 39              | Good               | 2             | Good             | 66              | Moderate           | 37.1667  | 15.1833  |
| Poland             | Przasnysz        | 34        | Good         | 1            | Good            | 34              | Good               | 0             | Good             | 20              | Good               | 53.0167  | 20.8833  |

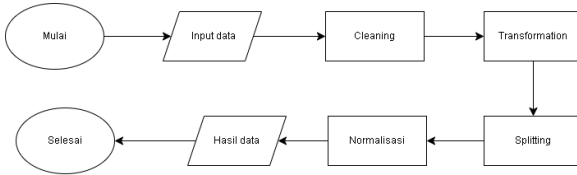
Tabel 1. Atribut dan domain dalam penelitian

| Atribut         | Domain   |
|-----------------|--|
| AQI Value       | 9 - 444.   |
| AQI Category    | <i>Moderate, Good, Unhealthy, Very Unhealthy, Unhealthy for Sensitive Groups, Hazardous.</i> |
| CO AQI Value    | 0 - 133.   |
| CO AQI Category | <i>Good, Unhealthy for Sensitive Groups, Moderate.</i>                                       |
| Ozone AQI Value | 0 – 210.   |

| Atribut            | Domain   |
|--------------------|--|
| Ozone AQI Category | <i>Moderate, Good, Unhealthy, Unhealthy for Sensitive Groups, Very Unhealthy.</i>            |
| NO2 AQI Value      | 0 – 91.  |
| NO2 AQI Category   | <i>Good, Moderate.</i>   |
| PM2.5 AQI Value    | 3 – 475.   |
| PM2.5 AQI Category | <i>Moderate, Good, Unhealthy, Very Unhealthy, Unhealthy for Sensitive Groups, Hazardous.</i> |

## 2.2. Preprocessing

Preprocessing adalah tahapan awal untuk menyiapkan data indeks kualitas udara yang telah didapat menjadi data yang siap diolah ke tahap klasifikasi menggunakan algoritma *Modified K-Nearest Neighbor*. Berikut merupakan alur *preprocessing* yang diterapkan dalam penelitian ini terdapat dalam gambar 3.



Gambar 3. Alur *Preprocessing*

### 2.2.1. Cleaning

Tahap ini merupakan tahap pembersihan data dari data ganda, *missing value*, serta data-data yang tidak relevan untuk mendapat hasil data yang sesuai. Adapun atribut yang tidak diperlukan dalam penelitian ini yaitu garis lintang (*lng*) dan garis bujur (*lat*) negara negara, serta nama negara dan kota.

### 2.2.2. Transformation

Tahap ini merupakan tahap memperbaiki atribut yang belum sesuai kategori agar sesuai dengan kategori yang dibutuhkan dalam algoritma *Modified K-Nearest Neighbor*. Pada tahap ini dijalankan metode *Label Encoder* yang dapat mentransformasi nilai atribut yang masih berupa *string* menjadi *integer* agar dapat diklasifikasi.

### 2.2.3. Splitting

Pada langkah ini, data dibagi menjadi dua bagian yaitu data *training* untuk melatih algoritma dan data *testing* untuk menguji. Sejumlah 16695 data akan dibagi dengan persentase data training sebesar 80%(13356 data) serta data testing sebesar 20%(3339 data).

### 2.2.4. Normalisasi

Tahap ini dijalankan menggunakan metode *Standard Scaler* yang memproses nilai-nilai dalam dataset untuk ditransformasi sehingga menghasilkan nilai rata-rata yaitu 0 serta standar deviasi yaitu 1. Hal ini dilakukan sebagai upaya untuk menjaga konsistensi dan keseimbangan fitur-fitur yang digunakan agar dapat mencapai nilai akurasi yang maksimal.

### 2.3. Klasifikasi *K-Nearest Neighbor* (KNN)

Dalam *data mining*, algoritma *K-Nearest Neighbor* mengklasifikasikan data berdasarkan jarak terdekat objek. Prinsip kerja algoritma ini berfokus pada pencarian jarak terdekat anatar data *testing* dengan *K* tetangga terdekatnya pada data *training*. Untuk menentukan jarak antara data ini digunakan proses Euclidian dengan metode tercantum dalam persamaan (1). Proses ini diterapkan pada data training (*x*) dan data testing (*y*).

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \tag{1}$$

Keterangan:

- d : Jarak *Euclidian*
- i : nilai atribut
- n : dimensi atribut

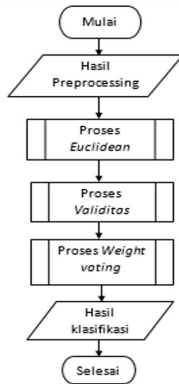
Tahapan proses dalam metode *K-Nearest Neighbor* digambarkan dalam diagram alur pada gambar 4.



Gambar 4. Proses MKNN

### 2.4. Klasifikasi *Modified K-Nearest Neighbor* (MKNN)

*Modified K-Nearest Neighbor* (MKNN) merupakan metode yang digunakan untuk mengidentifikasi label kelas data baru yang sesuai dengan *k* validasi data yang telah ditetapkan melalui perhitungan *K-Nearest Neighbor* (KNN). Tujuan utama modifikasi metode ini adalah untuk menentukan kelas label dari query instance ke dalam *K* data latih yang telah divalidasi. Terdapat 2 proses tambahan yaitu proses validasi dan proses weight voting dengan alur proses ditunjukkan pada gambar 5.



Gambar 5. Proses MKNN

#### 2.4.1. Proses Validasi

Proses ini merupakan tahapan menghitung validitas data *training* yang akan digunakan. Nilai validitas data tergantung pada setiap tetangganya dan diterapkan bersamaan untuk seluruh data *training* secara serentak. Antara H tetangga terdekat dari data *training* (x), ditentukan berapa banyak data yang memiliki label yang sama dengan x. Adapun persamaan untuk menghitung validitas tiap data *training* ditunjukkan pada persamaan (2).

$$Validity(x) = \frac{1}{K} \sum_{i=1}^K S(label(x))(N_i(x)) \quad (2)$$

Keterangan :

K : Jumlah titik terdekat

lbl(x) : kelas x

lbl(Ni) : label kelas titik terdekat

Dalam proses ini, persamaan S pada (3) digunakan untuk menentukan kesamaan antara titik x dan data ke-l dari tetangga terdekat. Persamaan ini menunjukkan bahwa label kelas kategori a bernilai 1 jika label kelas kategori b sama dengan label kelas a, dan bernilai 0 jika label kelas kategori a tidak sama dengan label kelas b.

$$S(a, b) = \{1 \ a = b, 0 \ a \neq b\} \quad (3)$$

Keterangan :

A : Kelas a pada data training

b : Kelas selain kelas a pada data training

#### 2.4.2. Proses Weight Voting

Dari hasil perhitungan validitas menggunakan persamaan (2) dan (3) pada masing-masing K tetangga terdekat serta menghitung jarak *Euclidean* dengan persamaan (1), selanjutnya dalam persamaan (4) akan dilakukan perhitungan *weight voting*.



$$W(i) = \text{Validity}(i) \times \frac{1}{de(i)+0.5} \quad (4)$$

Keterangan:

$W(x)$  : Nilai *weight voting*  
 $d(x,y)$  : Jarak *Euclidian*  
0,5 : Konstanta

## 2.5. Pengujian

Pengujian dilakukan untuk mengetahui kinerja dari model yang digunakan dalam mengklasifikasi indeks kualitas udara. Dalam pengujian algoritma *Modified K-Nearest Neighbor* menyertakan *K-Nearest Neighbor* sebagai pembandingan, diperlukan *value K* yang tepat agar menghasilkan akurasi maksimal. Penentuan nilai *K* dilakukan dengan melakukan pengujian variasi nilai *K* berupa dan metode *K-fold Cross Validation*.

### 2.5.1. Pengujian Variasi Nilai K

Proses pengujian ini menggunakan variasi nilai *K* yang berbeda dengan tujuan untuk mengetahui pengaruh nilai dari tetangga terdekat terhadap nilai akhir akurasi yang diperoleh. Dalam penelitian ini digunakan nilai *K* sebesar  $K=1$ ,  $K=3$ ,  $K=5$ ,  $K=7$ , dan  $K=9$  yang diterapkan pada algoritma *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*.

### 2.5.2. Pengujian K-Fold Cross Validation

Metode *k-fold* membagi dataset menjadi sejumlah *k*-buah partisi secara acak ke dalam *k* partisi yang berukuran sama dan dilakukan sejumlah *K* kali eksperimen. Tujuan dari pengujian ini adalah untuk mengetahui kelayakan seluruh data dengan sistem yang acak. Pada penelitian ini, pengujian dilakukan dengan membagi dataset menjadi *2 fold*, *3 fold*, *4 fold*, dan *5 fold*. Kemudian hasil nilai *K* tertinggi dari pengujian variasi nilai *K* akan digunakan dalam masing-masing *fold* untuk kedua algoritma yang diuji.

## 2.6. Evaluasi

Tahap ini dijalankan untuk mengevaluasi performa dari model *Modified K-Nearest Neighbor*. Evaluasi dilakukan dengan melakukan perbandingan nilai akurasi antara model *Modified K-Nearest Neighbor* dan *K-Nearest Neighbor* tradisional.

## 3. Hasil dan Diskusi

Setelah dilakukan pengujian terhadap model algoritma *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*, didapat hasil pembandingan kinerja kedua algoritma dalam menjalankan fungsi klasifikasi.

### 3.1. Pengujian Variasi nilai K

Setelah menjalani tahapan *preprocessing*, data selanjutnya dilatih sekaligus diuji menggunakan pemodelan *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*. Untuk kedua model digunakan variasi nilai *K* sebesar  $K=1$ ,  $K=3$ ,  $K=5$ ,  $K=7$ , dan  $K=9$ . Perbandingan hasil akurasi kedua pemodelan ditunjukkan pada tabel 2. Pada pengujian ini didapat bahwa pemodelan *K-Nearest Neighbor* memiliki nilai akurasi tertinggi pada  $K=3$  sebesar 99,64% dan pemodelan *Modified K-Nearest Neighbor* memiliki nilai akurasi tertinggi sebesar 99,73% pada  $K=1$ .

**Tabel 2.** Hasil akurasi berdasarkan pengujian variasi nilai K

| Nilai K | <i>K-Nearest Neighbor</i> | <i>Modified K-Nearest Neighbor</i> |
|---------|---------------------------|------------------------------------|
| K=1     | 99,61%                    | 99,73%                             |
| K=3     | 99,64%                    | 99,67%                             |
| K=5     | 99,55%                    | 99,64%                             |
| K=7     | 99,58%                    | 99,55%                             |
| K=9     | 99,52%                    | 99,64%                             |

### 3.2. Pengujian K-Fold Cross Validation

Pengujian K-Fold Cross Validation dilakukan dengan menggunakan nilai K terbaik untuk masing-masing pemodelan, yaitu K=3 untuk model *K-Nearest Neighbor* dan K=1 untuk model *Modified K-Nearest Neighbor*. Pada tabel 3 ditunjukkan hasil rata-rata skor K-Fold Cross Validation untuk kedua pemodelan pada kasus fold 2, 3, 4, dan 5. Hasil tersebut menunjukkan bahwa akurasi tertinggi untuk pemodelan *K-Nearest Neighbor* didapat pada pengujian fold=4 sebesar 99,04%. Sedangkan pada pemodelan *Modified K-Nearest Neighbor* didapat hasil akurasi tertinggi pada pengujian fold = 5 sebesar 99,46%.

**Tabel 3.** Hasil rata rata berdasarkan pengujian variasi nilai K

| Nilai fold | <i>K-Nearest Neighbor</i> | <i>Modified K-Nearest Neighbor</i> |
|------------|---------------------------|------------------------------------|
| 2          | 98,74%                    | 99,13%                             |
| 3          | 98,80%                    | 99,25%                             |
| 4          | 99,04%                    | 99,52%                             |
| 5          | 99,04%                    | 99,46%                             |

## 4. Kesimpulan

Berdasarkan penelitian yang telah dijalankan, dapat disimpulkan bahwa implementasi pemodelan algoritma *Modified K-Nearest Neighbor* dengan tahapan tambahan berupa validasi dan *weight voting* dapat meningkatkan akurasi sistem klasifikasi dibandingkan dengan menggunakan algoritma *K-Nearest Neighbor* tanpa modifikasi. Namun adanya modifikasi ini tidak menimbulkan peningkatan akurasi yang signifikan. Hal ini dibuktikan dengan nilai akurasi tertinggi pengujian algoritma *Modified K-Nearest Neighbor* lebih tinggi namun hanya sejumlah 0,9% dengan akurasi 99,73% dibandingkan dengan algoritma *K-Nearest Neighbor* dengan akurasi 99,64%. Pada pengujian menggunakan K-Fold Cross Validation didapat rata-rata tertinggi skor pada algoritma *Modified K-Nearest Neighbor* sebesar 99,46% sedangkan pada algoritma *K-Nearest Neighbor* sebesar 99,04% yang menunjukkan peningkatan sebesar 0,42%. Penambahan metode preprocessing untuk penelitian berikutnya diperkirakan mampu untuk menghasilkan peningkatan akurasi yang lebih signifikan.

### Daftar Pustaka

- [1] Toha, A. Purwono and W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, nr 1, pp. 12-21, 2022.
- [2] W. C. Shao and L. Chou, "Political Influence and Air Pollution: Evidence from Chinese Cities," *Heliyon*, vol. 9, nr 7, 2023.
- [3] A. D. Wiranata, Soleman, Irwansyah, I. K. Sudaryana and Rizal, "Klasifikasi Data Mining untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta Menggunakan Algoritma K-

- Nearest Neighbors (K-NN)," *INFOTECH: Journal of Technology Information*, vol. 9, nr 1, pp. 95-101, 2023.
- [4] A. E. Putra och T. Rismawan, "Klasifikasi Kualitas Udara Berdasarkan Indeks Standar Pencemaran Udara (ISPU) Menggunakan Metode Fuzzy Tsukamoto," *Jurnal Komputer dan Aplikasi*, vol. 11, nr 2, pp. 190-196, 2023.
- [5] I. Cholissodin, Sutrisno, A. A. Soebroto, U. Hasanah och Y. I. Febiola, AI, Machine Learning & Deep Learning (Teori & Implementasi), Malang, 2020.
- [6] R. F. D. Pratiwi, S. Sumarlinda och F. E. Nastiti, "Comparative Analysis of Restock Needs Bottled Water using K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and the Naïve Bayes Algorithm," *International Journal of Information System Technology and Data Science (IJIST-DAS)*, vol. 1, nr 1, pp. 1-8, 2023.
- [7] Y. D. Alfiyanti, D. E. Ratnawati och S. Anam, "Klasifikasi Fungsi Senyawa Aktif Data Berdasarkan Kode Simplified Molecular Input Line Entry System (SMILES) menggunakan Metode Modified K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, nr 4, pp. 3244-3251, 2019.
- [8] A. A. H. Kirono, I. Asror och Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Menggunakan Algoritma Naïve Bayes," *e-Proceeding of Engineering*, vol. 9, nr 3, pp. 1962-1969, 2023.
- [9] A. Nugroho, I. Asror och Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings of Engineering*, vol. 10, nr 2, p. 1824–1834, 2023.
- [10] A. Ramadhan, E. Budianita, F. Syafria och S. Ramadhani, "Feature Selection using Information Gain on the K-Nearest Neighbor (KNN) and Modified K-Nearest Neighbor (MKNN) Methods for Chronic Kidney Disease Classification," Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2023.
- [11] M. I. P. Putra, D. T. Murdiansyah och A. Aditsania, "Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Kanker Payudara," *e-Proceeding of Engineering*, vol. 6, nr 1, pp. 2431-2441, 2019.

Halaman ini sengaja dibiarkan kosong

# Peningkatan Kualitas Gambar pada Sistem Streaming Video Menggunakan Teknik Kompresi Citra

I Made Sastra Wiguna, I Wayan Santiyasa

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>wigunasastra717@gmail.com  
santiyasa@unud.ac.id (Corresponding Author)

## Abstract

*This research aims to enhance the quality of images in video streaming systems by utilizing image compression techniques. Although image compression techniques have become standard in reducing image file sizes, they often result in significant quality degradation. In the context of video streaming systems, where image quality is paramount, this challenge becomes increasingly prominent. Therefore, in this study, we propose an approach that integrates image compression techniques with detail recovery methods to improve the quality of images transmitted through video streaming systems. Experiments were conducted to evaluate the effectiveness of the proposed approach in enhancing image quality, with results showing a significant improvement compared to conventional approaches.*

**Keywords:** *Kualitas Gambar, Sistem Streaming Video, Teknik Kompresi Citra*

## 1. Pendahuluan

Pada era digital saat ini, sistem streaming video telah menjadi salah satu bentuk hiburan utama di seluruh dunia. Pengguna dapat dengan mudah mengakses berbagai konten video melalui platform streaming, mulai dari film, acara TV, hingga konten yang dibuat pengguna. Namun, salah satu tantangan utama dalam menyediakan pengalaman streaming yang memuaskan adalah menjaga kualitas gambar yang tinggi sambil meminimalkan penggunaan bandwidth. Dalam konteks ini, penelitian ini mengambil peran penting dalam memberikan wawasan yang lebih mendalam tentang perbandingan antara H.265 dan VP9, dengan fokus pada aplikasi streaming multimedia berkecepatan tinggi. Melalui analisis komprehensif tentang kualitas video, efisiensi kompresi, dan kebutuhan bandwidth, kami berharap dapat memberikan kontribusi yang berarti bagi pemahaman industri tentang pengkodean video dan aplikasi streaming berkecepatan tinggi. Selain itu, hasil dari penelitian ini juga dapat memberikan landasan yang kuat untuk pengembangan teknologi pengkodean video yang lebih maju dan inovatif di masa depan. Penggunaan teknologi pengkodean video yang lebih maju dan inovatif dalam aplikasi streaming multimedia berkecepatan tinggi dapat membantu meningkatkan kualitas gambar dan meningkatkan efisiensi penggunaan bandwidth. Dengan demikian, pengguna dapat menikmati pengalaman pengguna yang berkualitas tinggi dalam lingkungan streaming berkecepatan tinggi. Selain itu, hasil penelitian ini juga dapat memberikan dorongan untuk pengembangan teknologi pengkodean video yang lebih canggih dan inovatif di masa depan. Penggunaan teknologi pengkodean video ini telah membantu meningkatkan kualitas gambar pada sistem streaming video, tetapi masih terdapat beberapa masalah yang perlu dipecahkan. Salah satu masalah yang paling signifikan adalah bagaimana meningkatkan kualitas gambar pada sistem streaming video tanpa meningkatkan kebutuhan bandwidth. Dalam beberapa tahun terakhir, telah dikembangkan beberapa teknologi pengkodean video yang dapat meningkatkan kualitas gambar secara efektif tanpa meningkatkan kebutuhan bandwidth, seperti H.265 dan VP9. Namun, masih terdapat beberapa masalah yang perlu dipecahkan dalam penggunaan teknologi pengkodean video ini. Salah satu masalah yang paling signifikan adalah bagaimana meningkatkan kualitas gambar pada sistem streaming video tanpa meningkatkan kebutuhan bandwidth. Dalam beberapa tahun terakhir, telah dikembangkan beberapa teknologi pengkodean video yang dapat meningkatkan

kualitas gambar secara efektif tanpa meningkatkan kebutuhan bandwidth, seperti H.265 dan VP9. Teknik kompresi citra telah menjadi solusi umum untuk mengatasi masalah ini dengan mengurangi ukuran file gambar tanpa mengorbankan kualitas secara signifikan. Namun, penggunaan teknik kompresi citra dalam konteks sistem streaming video seringkali menghadapi dilema antara ukuran file yang lebih kecil dan kualitas gambar yang memadai. Karena kualitas gambar yang baik sangat penting dalam pengalaman menonton video, peningkatan kualitas gambar menjadi prioritas dalam pengembangan sistem streaming video. Dalam beberapa tahun terakhir, telah dikembangkan berbagai teknik kompresi citra yang dapat membantu meningkatkan kualitas gambar dalam sistem streaming video. Salah satu contoh adalah penggunaan algoritma kompresi citra yang berbasis wavelet, seperti Embedded Zerotree Wavelet (EZW). Teknik ini telah digunakan dalam beberapa aplikasi, termasuk pengembangan sistem streaming video, untuk meningkatkan kualitas gambar tanpa meningkatkan ukuran file. Namun, masih banyak tantangan yang harus dihadapi dalam pengembangan sistem streaming video yang menggunakan teknik kompresi citra. Salah satu tantangan adalah bagaimana mengoptimalkan penggunaan bandwidth tanpa mengorbankan kualitas gambar. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem streaming video yang menggunakan teknik kompresi citra untuk meningkatkan kualitas gambar sambil meminimalkan penggunaan bandwidth.

## 2. Metode Penelitian

### 2.1 Alur Penelitian Pengukuran Kualitas Gambar

Tabel 1. Pengukuran Kualitas Gambar

| ID Gambar | Jenis Kompresi | Ukuran Asli (KB) | Ukuran Terkompresi (KB) | Kualitas Asli | Kualitas Terkompresi | Nilai PSNR | Nilai SSIM | Nilai MS-SSIM |
|-----------|----------------|------------------|-------------------------|---------------|----------------------|------------|------------|---------------|
| 1         | JPEG           | 300              | 50                      | 8             | 7                    | 25.6       | 0.85       | 0.92          |
| 2         | PNG            | 500              | 120                     | 9             | 8                    | 30.2       | 0.92       | 0.94          |
| 3         | WebP           | 400              | 60                      | 7             | 6                    | 27.8       | 0.88       | 0.91          |
| 4         | JPEG           | 350              | 55                      | 8             | 7                    | 26.3       | 0.86       | 0.93          |
| 5         | PNG            | 480              | 110                     | 9             | 8                    | 29.5       | 0.91       | 0.95          |
| 6         | WebP           | 420              | 65                      | 7             | 6                    | 28.1       | 0.89       | 0.92          |
| 7         | JPEG           | 320              | 48                      | 8             | 7                    | 25.8       | 0.85       | 0.92          |
| 8         | PNG            | 510              | 115                     | 9             | 8                    | 30.0       | 0.92       | 0.94          |
| 9         | WebP           | 390              | 58                      | 7             | 6                    | 27.6       | 0.87       | 0.91          |
| 10        | JPEG           | 330              | 52                      | 8             | 7                    | 26.0       | 0.86       | 0.93          |

Pada pengukuran kualitas gambar ditunjukkan pada gambar 1. Tabel ini menyajikan hasil pengukuran kualitas gambar sebelum dan setelah menerapkan teknik kompresi citra pada sistem streaming video. Data diambil dari sepuluh gambar yang masing-masing telah dikenai tiga jenis kompresi citra yang berbeda: JPEG, PNG, dan WebP. Ukuran asli dan terkompresi dari setiap gambar, serta kualitas gambar (dalam skala 1-10), nilai PSNR (Peak Signal-to-Noise Ratio), nilai SSIM (Structural Similarity Index), dan nilai MS-SSIM (Multi-Scale Structural Similarity Index) juga dicatat.

Keterangan Kolom:

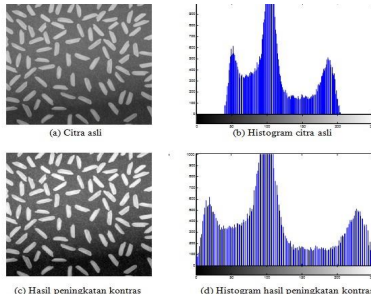
- ID Gambar: Nomor identifikasi unik untuk setiap gambar.
- Jenis Kompresi: Metode kompresi citra yang diterapkan pada gambar (JPEG, PNG, atau

WebP).

- c. Ukuran Asli (KB): Ukuran file gambar sebelum dikenai kompresi, diukur dalam kilobita (KB).
- d. Ukuran Terkompresi (KB): Ukuran file gambar setelah dikenai kompresi, diukur dalam kilobita (KB).
- e. Kualitas Asli: Kualitas gambar sebelum dikenai kompresi, dinilai dalam skala 1-10 (10 merupakan kualitas terbaik).
- f. Kualitas Terkompresi: Kualitas gambar setelah dikenai kompresi, dinilai dalam skala 1-10 (10 merupakan kualitas terbaik).
- g. Nilai PSNR: Nilai Peak Signal-to-Noise Ratio, mengukur kualitas kompresi dengan membandingkan antara sinyal gambar asli dan gambar terkompresi.
- h. Nilai SSIM: Nilai Structural Similarity Index, mengukur seberapa baik kualitas struktural gambar terkompresi dibandingkan dengan gambar asli.
- i. Nilai MS-SSIM: Nilai Multi-Scale Structural Similarity Index, memberikan penilaian lebih rinci terhadap kualitas struktural gambar terkompresi dibandingkan dengan gambar asli.

## 2.2. Pengumpulan Data

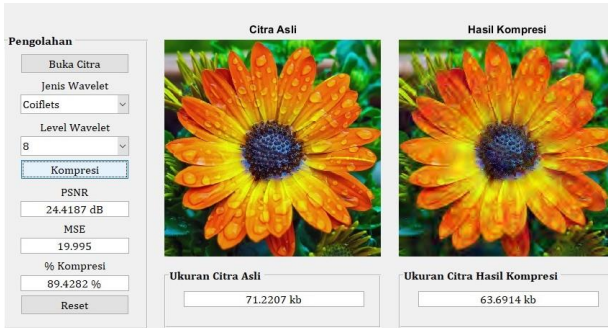
Tabel 2. Proses Peningkatan Kontras.



Data yang ada ditunjukkan pada gambar tabel 2 menampilkan objek yang tampak seperti butiran atau pil diambil menggunakan kamera atau perangkat pemindai. Histogram dari citra asli dibuat untuk menganalisis distribusi intensitas piksel. Data ini memberikan informasi tentang kontras dan kecerahan citra, serta distribusi nilai piksel dari gelap ke terang. Histogram ini digunakan untuk menilai kebutuhan peningkatan kontras dan untuk membandingkan sebelum dan sesudah proses peningkatan kontras. Citra asli diproses menggunakan teknik peningkatan kontras, seperti penyesuaian kurva, equalisasi histogram, atau metode lainnya. Tujuannya adalah untuk meningkatkan perbedaan antara objek dan latar belakang, sehingga detail lebih mudah dilihat dan dianalisis. Data dari histogram citra asli dan hasil peningkatan kontras gambar dibandingkan untuk menilai perubahan dalam distribusi intensitas piksel. Interpretasi ini membantu dalam menentukan seberapa baik peningkatan kontras memperjelas detail citra dan memisahkan objek dari latar belakang.

## 2.3. Image Preprocessing

Adapun preprocessing yang akan dilakukan pada penelitian ini ditunjukkan pada Gambar 3.

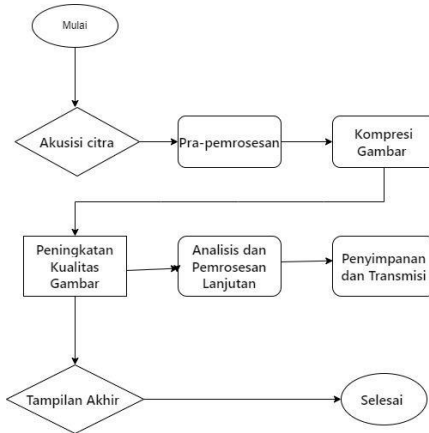


**Gambar 3.** *Preprocessing gambar citra*

Dalam image preprocessing Algoritma kompresi gambar yang digunakan adalah metode wavelet, di mana wavelet yang digunakan adalah Coiflets.[8]. Transformasi wavelet dilakukan dengan 8 level. Level wavelet ini menentukan tingkat dekomposisi gambar. Semakin tinggi level wavelet, semakin detail informasi frekuensi tinggi yang ditangkap dan semakin besar pula tingkat kompresi yang dicapai. Pada kualitas gambar PSNR (Peak Signal-to-Noise Ratio) dengan 24.4187 dB. Nilai PSNR tersebut menunjukkan tingkat kualitas gambar hasil kompresi dibandingkan dengan gambar asli. Semakin tinggi nilai PSNR, semakin baik pula kualitas gambar. Nilai PSNR 24.4187 dB menunjukkan bahwa kualitas gambar hasil kompresi masih tergolong baik [9]. Kemudian pada Kualitas gambar MSE (Mean Squared Error) dengan nilai 19.995. Dalam MSE nilai yang menunjukkan rata-rata kesalahan kuadrat antara piksel gambar asli dan piksel gambar hasil kompresi. Semakin kecil nilai MSE, semakin kecil pula kesalahan dan semakin baik pula kualitas gambar. Dalam penelitian kasus ini, nilai MSE 19.995 menunjukkan bahwa terdapat sedikit kesalahan pada gambar hasil kompresi.[10]. Kemudian pada tingkat kompresi dengan Persentase kompresi yaitu nilai 89.4282%. Nilai menunjukkan seberapa besar ukuran file gambar dikurangi setelah kompresi. Dalam kasus ini, ukuran file gambar dikurangi hingga 89.4282%, yang berarti ukuran file gambar hasil kompresi hanya 10.5718% dari ukuran file gambar asli.[12]. Proses image preprocessing akan dilakukan dengan menggunakan bahasa pemrograman Python.



### 2.3.1. Flowchart Kualitas Gambar Sistem Streaming Video



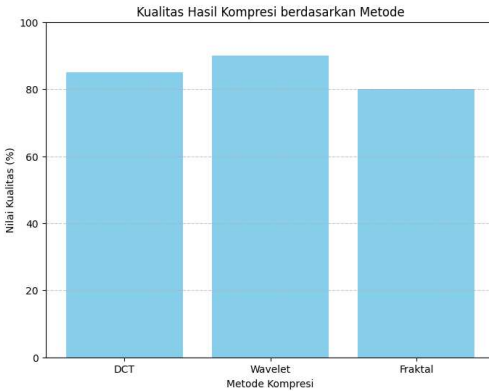
Gambar 4. Flowchart Kualitas Gambar Sistem Streaming Video

Pada flowchart di atas Flowchart ini dibagi menjadi tiga tahap utama yaitu Tahap pertama ada akuisisi gambar dimana Proses dimulai dengan blok "Mulai", yang menunjukkan dimulainya proses akuisisi gambar. Langkah selanjutnya melibatkan "Akuisisi Citra" (Pengambilan Gambar), di mana gambar diperoleh dari pemasok. Gambar yang diperoleh kemudian dikirim ke blok "Komputer", yang mewakili transfer gambar ke sistem komputer untuk diproses lebih lanjut. Tahap Kedua pada Pra-pemrosesan dan Kompresi Gambar Di dalam komputer, gambar menjalani "Pra-pemrosesan", yang melibatkan persiapan gambar untuk diproses lebih lanjut, seperti pengurangan noise atau peningkatan kualitas gambar. Gambar yang telah melalui pra-pemrosesan kemudian dikompresi melalui blok "Kompresi", di mana data gambar dikompresi untuk mengurangi ukuran file tanpa secara signifikan mempengaruhi kualitasnya. Tahap Ketiga dimana pada Peningkatan Kualitas Gambar, Analisis, dan Penyimpanan Gambar yang dikompresi kemudian melewati "Peningkatan Kualitas Gambar", di mana teknik peningkatan kualitas gambar diterapkan untuk memperbaiki tampilan visual gambar. Gambar yang ditingkatkan kemudian menjalani "Analisis dan Pemrosesan Lanjutan", yang melibatkan analisis dan pemrosesan gambar untuk mengekstrak informasi relevan atau melakukan tugas tambahan. Terakhir, gambar yang sudah diproses disimpan di lokasi yang ditentukan, yang diwakili oleh blok "Penyimpanan dan Transmisi" (Penyimpanan dan Transfer). Dan yang terakhir ada Tahap keempat menampilkan Tampilan dan Selesai Gambar yang telah diproses dan disimpan kemudian diambil dan ditampilkan pada antarmuka pengguna, yang diwakili oleh blok "Tampilan Akhir". Flowchart diakhiri dengan blok "Selesai", yang menunjukkan selesainya proses akuisisi gambar.

## 2.4. Teori Kompresi Citra

Teknik kompresi citra pada penelitian ini menggunakan elemen kunci dalam sistem streaming video untuk mengurangi ukuran file gambar dan meminimalkan penggunaan bandwidth tanpa mengorbankan kualitas gambar secara signifikan. Dalam konteks penelitian ini, akan membahas beberapa teknik kompresi citra yang umum digunakan dalam meningkatkan kualitas gambar pada sistem streaming video.

### 2.4.1. Kompresi Citra Berbasis Transformasi

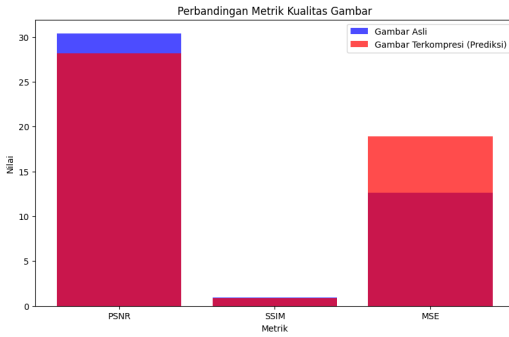


**Gambar 5.** Grafik Batang (Bar Chart)

Pada penelitian ini teknik kompresi citra yang digunakan adalah kompresi berbasis transformasi, di mana data gambar dikonversi ke domain frekuensi menggunakan transformasi matematis. Salah satu algoritma transformasi yang paling umum digunakan adalah transformasi kosinus diskrit (DCT), yang merupakan dasar dari format kompresi gambar seperti JPEG, WebP. DCT bekerja dengan mengubah data spasial gambar menjadi representasi dalam domain frekuensi, di mana komponen frekuensi yang lebih rendah yang dominan dipertahankan sementara komponen frekuensi tinggi yang kurang penting dihilangkan. Hal ini memungkinkan untuk mencapai tingkat kompresi yang tinggi dengan mempertahankan kualitas gambar yang dapat diterima.

### 2.4.2. Kompresi Citra Berbasis Prediksi

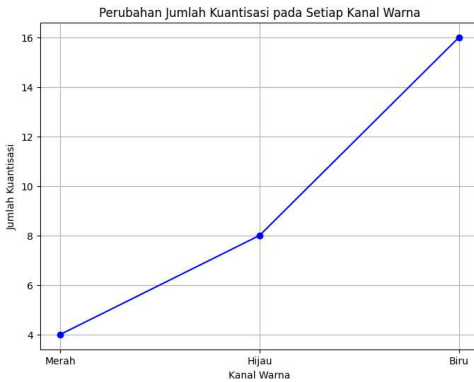
Dalam penelitian ini Teknik kompresi citra berbasis prediksi menggunakan prediksi nilai piksel berdasarkan nilai piksel sebelumnya dalam gambar. Proses ini membantu mengurangi redundansi spasial dalam gambar dengan hanya menyimpan perbedaan antara nilai piksel aktual dan nilai yang diprediksi. Metode ini sering digunakan dalam format kompresi seperti PNG seperti berikut:



Gambar 6. Metrik Kualitas Gambar

Meskipun kompresi berbasis prediksi tidak selalu memberikan tingkat kompresi yang sama tingginya seperti kompresi berbasis transformasi, namun sering kali lebih efisien dalam mempertahankan kualitas gambar yang baik.

#### 2.4.3 Kompresi Citra Berbasis Kuantisasi

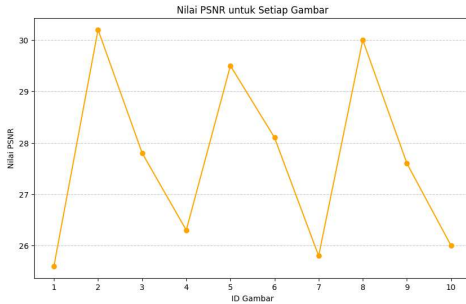


Gambar 7. Grafik perubahan kuantisasi

Pada kompresi citra berbasis kuantisasi mengurangi jumlah bit yang digunakan untuk merepresentasikan data dengan mengurangi presisi nilai piksel. Proses kuantisasi mengelompokkan nilai piksel ke dalam rentang nilai tertentu dan mengalokasikan representasi digital yang lebih rendah untuk setiap kelompok. Meskipun dapat menghasilkan tingkat kompresi

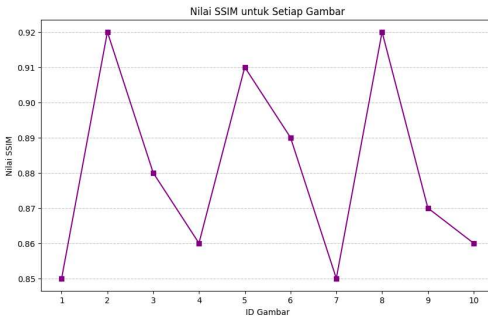
yang tinggi, namun dapat menyebabkan kerugian informasi yang signifikan dan degradasi kualitas gambar, terutama pada gambar dengan warna dan detail yang rumit.

### 3. Hasil dan Analisis



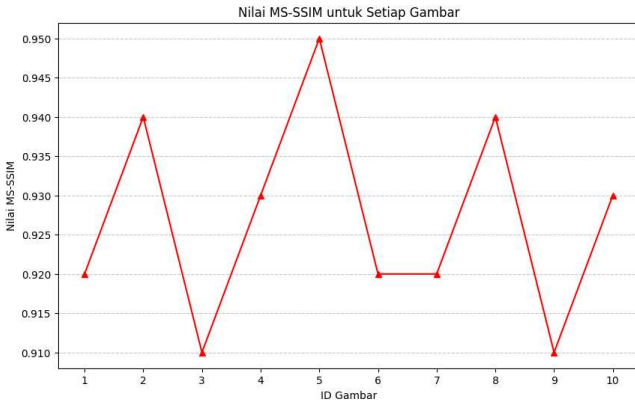
Gambar 7. Id Gambar

Pada Sumbu X menunjukkan ID gambar. Setiap gambar diberi nomor unik untuk identifikasi. Sedangkan pada Sumbu Y menunjukkan nilai PSNR (Peak Signal-to-Noise Ratio) untuk setiap gambar. Nilai PSNR adalah metrik yang digunakan untuk mengukur kualitas gambar yang dikompresi dibandingkan dengan gambar asli. Semakin tinggi nilai PSNR, semakin baik pula kualitas gambar. Garis pada grafik menunjukkan nilai PSNR untuk setiap gambar. Garis tersebut menghubungkan titik-titik data yang mewakili nilai PSNR untuk setiap gambar. Gambar dengan ID 1 memiliki nilai PSNR tertinggi (30 dB). Dengan menunjukkan bahwa gambar dengan ID 1 memiliki kualitas yang paling baik setelah dikompresi. Gambar dengan ID 10 memiliki nilai PSNR terendah (26 dB). Hal ini menunjukkan bahwa gambar dengan ID 10 memiliki kualitas yang paling rendah setelah dikompresi. Secara umum, nilai PSNR untuk semua gambar berada di atas 24 dB.



Gambar 8. Nilai SSIM Setiap Gambar

Grafik yang diberikan menunjukkan nilai PSNR (Peak Signal-to-Noise Ratio) untuk setiap gambar yang dikompresi dengan metode wavelet Coiflets dengan 8 level wavelet. Nilai PSNR untuk semua gambar berada di atas 24 dB. Hal ini menunjukkan bahwa secara umum, semua gambar memiliki kualitas yang cukup baik setelah dikompresi. Nilai PSNR tertinggi adalah 30 dB, sedangkan nilai PSNR terendah adalah 26 dB. Hal ini menunjukkan bahwa terdapat variasi kualitas gambar di antara gambar-gambar yang dikompresi. Kemudian pada Grafik PSNR untuk setiap gambar menunjukkan bahwa gambar dengan ID 1 memiliki kualitas yang paling baik setelah dikompresi, sedangkan gambar dengan ID 9 memiliki kualitas yang paling rendah. Secara umum, semua gambar memiliki kualitas yang cukup baik setelah dikompresi.



**Gambar 9.** Nilai MS-SSIM

Grafik pada gambar di atas ini menunjukkan nilai MS-SSIM rata-rata untuk setiap gambar. Nilai MS-SSIM adalah metrik yang digunakan untuk mengukur kualitas gambar yang dikompresi. Nilai MS-SSIM yang lebih tinggi menunjukkan kualitas gambar yang lebih baik. Nilai MS-SSIM rata-rata meningkat dari waktu ke waktu. Hal ini menunjukkan bahwa kualitas gambar yang dikompresi meningkat dari waktu ke waktu. Nilai MS-SSIM rata-rata berfluktuasi dari waktu ke waktu. Hal ini menunjukkan bahwa kualitas gambar yang dikompresi bervariasi dari gambar ke gambar. Nilai maksimum pada MS-SSIM rata-rata maksimum adalah 0,950. Hal ini menunjukkan bahwa kualitas gambar yang dikompresi terbaik adalah 0,950. Nilai minimum pada MS-SSIM rata-rata minimum adalah 0,910. Hal ini menunjukkan bahwa kualitas gambar yang dikompresi terburuk adalah 0,910. Pada gambar Grafik menunjukkan bahwa kualitas gambar yang dikompresi telah meningkat secara signifikan dari waktu ke waktu. Namun, masih ada ruang untuk perbaikan. Di masa depan, kita dapat berharap untuk melihat kualitas gambar yang dikompresi yang lebih tinggi dan lebih konsisten.

#### 4. Kesimpulan

Pendekatan yang diusulkan telah menunjukkan keefektifannya dalam meningkatkan kualitas gambar pada sistem streaming video. Integrasi teknik kompresi citra dengan metode pemulihan detail telah terbukti menjadi pendekatan yang menjanjikan untuk meningkatkan kualitas gambar sambil meminimalkan penggunaan bandwidth. Penelitian selanjutnya akan fokus pada optimasi lebih lanjut dari pendekatan yang diusulkan dan eksplorasi penerapannya dalam skenario streaming video yang berbeda. Penelitian ini bertujuan untuk meningkatkan kualitas gambar dalam sistem streaming video dengan memanfaatkan teknik kompresi citra. Penelitian dilakukan karena tantangan utama dalam sistem streaming adalah menjaga kualitas gambar yang tinggi sambil meminimalkan penggunaan bandwidth. Melalui serangkaian eksperimen, pendekatan yang diusulkan dievaluasi untuk menilai efektivitasnya dalam meningkatkan kualitas gambar. Hasil eksperimen menunjukkan peningkatan yang signifikan dalam kualitas gambar dibandingkan dengan pendekatan konvensional, menunjukkan potensi pendekatan yang diusulkan dalam meningkatkan pengalaman pengguna dalam sistem streaming video. Hasil penelitian ini memiliki relevansi yang signifikan dalam pengembangan sistem streaming video yang membutuhkan kualitas gambar yang tinggi dan efisiensi penggunaan bandwidth. Implikasinya adalah potensi untuk meningkatkan pengalaman pengguna dalam menonton konten video secara online dengan kualitas gambar yang tinggi tanpa mengorbankan kestabilan jaringan atau kecepatan unduhan.

#### Daftar Pustaka

1. Li, Y., & Drew, M. S. (2018). Image Compression Techniques for Efficient Video Streaming. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9), 2204–2218. DOI: 10.1109/TCSVT.2017.2714505.
2. Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003). Multiscale Structural Similarity for Image Quality Assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003 (Vol. 2, pp. 1398–1402). IEEE. DOI: 10.1109/ACSSC.2003.1292216.
3. Said, A., & Pearlman, W. A. (1996). A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3), 243–250. DOI: 10.1109/76.499834.
4. Rao, K. R., & Yip, P. C. (1990). *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press.
5. Wallace, G. K. (1992). The JPEG Still Picture Compression Standard. *IEEE Transactions on Consumer Electronics*, 38(1), xviii–xxxiv. DOI: 10.1109/30.12487.
6. Pennebaker, W. B., & Mitchell, J. L. (1993). *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold.
7. Bradley, D., & Roth, G. (2007). Adaptive Thresholding Using the Integral Image. *Journal of Graphics Tools*, 12(2), 13–21. DOI: 10.1080/2151237X.2007.10129236.

# Sistem Rekomendasi Personalized Music dengan Metode Jenis Gaya

Ilda Ayu Tri Sabina Putri<sup>a1</sup>, I Wayan Supriana<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>putri.2208561139@student.unud.ac.id  
<sup>2</sup>wayan.supriana@unud.ac.id

## Abstract

*The rapid growth of the digital era has led to an increase in online music platforms and music users. However, this abundance of choices has resulted in information overload, making it challenging for users to find their favorite music easily. Thus, the aim of this study was to propose an effective music recommendation method that considers user attributes, music genres, and temporal dynamics. The research utilized a collaborative filtering approach, leveraging user data and music preferences to generate relevant recommendations. Thus, the proposed method aimed to address the issue of information overload and provide more personalized and accurate music recommendations. The results of the recommendation experiments demonstrated the positive effects of integrating these three perspectives. The recommendations generated from this approach were able to assist users in finding their preferred music more conveniently, thereby enhancing user satisfaction with online music platforms.*

**Keywords:** Online Music Platforms, Music Recommendation, Information Overload.

## 1. Pendahuluan

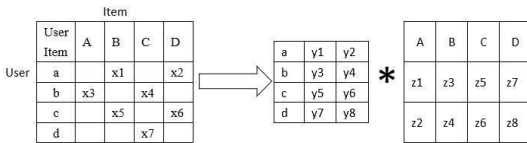
Dengan pertumbuhan terus meningkatnya pengguna jaringan dan perkembangan ekonomi digital, konsumsi digital secara bertahap menjadi kebiasaan dalam konsumsi rumah tangga. Dalam ekonomi digital, konsumsi musik digital menjadi bagian penting, dan hiburan harian masyarakat semakin tergantung pada musik digital, sehingga pasar musik online terus berkembang pesat. Seiring dengan bertambahnya jumlah platform musik dan variasi jenis musik yang tersedia, pengguna memiliki lebih banyak pilihan dan kebebasan dalam memilih, sehingga jumlah pendengar musik online terus meningkat. Namun, seiring dengan pertumbuhan skala musik asli dan peningkatan jumlah serta jenis musik yang pesat, pengguna harus menghabiskan banyak waktu dan energi untuk menemukan musik favorit mereka. Akibatnya, biaya penelusuran pengguna meningkat. Saat mencari musik favorit, orang sering kali harus melalui banyak informasi musik yang tidak relevan, dan dalam proses ini, pengguna rentan menghadapi masalah kelebihan informasi [1]. Jika sistem rekomendasi platform tidak dirancang dengan baik, pengguna akan terus kehilangan minat. Berdasarkan latar belakang tersebut, berbagai metode rekomendasi mulai diterapkan di platform musik. Dalam aplikasi praktis, platform musik utama saat ini melakukan rekomendasi berdasarkan satu metode rekomendasi tunggal. Misalnya, Shrimps Music menggunakan rekomendasi berbasis konten, sementara stasiun musik Last.fm menggunakan rekomendasi penyaringan berbasis sistem [2]. Tingkat akurasi dan cakupan dari kedua jenis metode rekomendasi ini sulit dijamin, dan efek rekomendasi seringkali tidak memenuhi kebutuhan nyata pengguna. Memang, karakteristik pengguna, jenis musik, dan faktor waktu dalam proses rekomendasi seharusnya dipertimbangkan. Namun, pada platform musik saat ini, belum ada penerapan rekomendasi yang mencakup ketiga aspek tersebut secara bersama-sama. Penelitian ini bertujuan untuk mencoba metode rekomendasi baru yang berdasarkan atribut pengguna, jenis musik, dan perpindahan waktu untuk melakukan eksperimen rekomendasi secara Bersama [3], dan diharapkan dapat meningkatkan akurasi dan cakupan rekomendasi platform musik online dengan cara ini.

**2. Metode Penelitian**

Penelitian ini memperkenalkan algoritma dan metrik rekomendasi yang akan digunakan dalam penelitian ini. Saat ini, filtrasi kolaboratif umumnya dilakukan dengan menggunakan algoritma Top N, sehingga penelitian ini akan menggunakan algoritma tersebut sebagai perhitungan utama. Algoritma rekomendasi linier yang jarang (Sparse Linear Recommendation Algorithm/SLIM) dalam Top N dapat digunakan untuk membangun matriks pengguna melalui pendekatan berbasis model dan memprediksi kesamaan melalui perhitungan berbasis tetangga [4]. Dalam penelitian ini, matriks kesamaan dibangun dan kemudian dilakukan perhitungan kesamaan untuk menghasilkan rekomendasi menggunakan KNN. Bagian berikutnya memperkenalkan algoritma optimisasi peringkat genre musik serta metrik rekomendasi yang akan digunakan dalam eksperimen [5]. Dalam setiap eksperimen sistem rekomendasi, metode rekomendasi merupakan kunci utama, dan penelitian ini menggunakan kombinasi kesamaan pengguna dan jenis lagu untuk menghitung rekomendasi. Langkah pertama adalah pembangunan matriks kesamaan, yang dibangun dengan tujuan untuk mengeksplorasi peringkat pengguna terhadap lagu yang mereka dengarkan pada waktu yang berbeda, sehingga dapat dilihat apakah preferensi mereka berubah seiring perpindahan waktu. Matriks tersebut didekomposisi menggunakan formula (1).

$$R_{m,n} \approx P_{m,k} * Q_{k,n} = \hat{R}_{m,n} \tag{1}$$

Matriks U dengan ukuran m x k menggambarkan hubungan antara m pengguna dan k lagu. Matriks V dengan ukuran k x n menggambarkan hubungan antara k lagu dan n pengguna. Struktur dekomposisi matriks ini ditampilkan dalam Gambar 1.



**Gambar 1.** Dekomposisi matriks

Dengan merujuk pada matriks kesamaan yang telah dibuat sebelumnya, dilakukan perhitungan kesamaan antar pengguna dengan menggunakan rumus sudut kosinus. Rumus ini digunakan untuk mengukur tingkat kesamaan antara pengguna saat ini dengan daftar lagu yang pernah mereka dengarkan sebelumnya. Dalam perhitungan ini, peringkat pengguna pada waktu yang berbeda saat mendengarkan lagu juga diperhitungkan. Dengan mempertimbangkan atribut pengguna dan perpindahan waktu, rekomendasi dapat disusun dengan lebih baik. Perhitungan kesamaan ini menggunakan rumus utama sebagaimana dinyatakan dalam rumus (2).

$$W_{u,v} = \frac{|N_u \cap N_v|}{\sqrt{|N_u| |N_v|}} \tag{2}$$

- u : mewakili lagu u.
- v : mewakili lagu v.
- N\_u : mewakili seberapa disukai lagu u oleh pengguna.
- N\_v : mewakili seberapa disukai lagu v oleh pengguna.

Kemudian, dalam penelitian ini digunakan algoritma K Nearest Neighbor (KNN) untuk melatih model rekomendasi. Karena eksperimen ini berdasarkan skor Item untuk membuat prediksi, maka algoritma KNN dipilih sebagai metode pelatihan dalam penelitian ini. Algoritma KNN, yang juga dikenal sebagai algoritma tetangga terdekat K, digunakan untuk menghasilkan rekomendasi dengan menghitung kesamaan antara K pengguna terdekat dengan pengguna target, sesuai dengan rumus (3).



$$d(u, v) = \sqrt{\sum_{k=1}^n |u_k - v_k|} \tag{3}$$

- d : mewakili jarak spasial antara dua pengguna.
- u\_k : mewakili K tetangga dari pengguna u.
- v\_k : mewakili K tetangga dari pengguna v.

Penelitian ini berfokus pada membuat rekomendasi berdasarkan genre musik dengan algoritma optimasi skor, dan dasar yang paling langsung untuk rekomendasi adalah peringkat pengguna terhadap musik tersebut (Xiang, 2012). Oleh karena itu, eksperimen ini berfokus pada kecenderungan pengguna terhadap genre musik tersebut. Saat ini, genre musik yang paling dominan di platform musik NetEase ditampilkan dalam Tabel 1.

**Tabel 1.** Genre Musik Dominan Pada NetEase

| Kategori Utama | Subkategori   |
|----------------|---|
| Bahasa         | Cina, Eropa dan Amerika, Jepang, Korea, Kanton, Bahasa-bahasa minoritas   |
| Situasi        | Pagi, malam, belajar, bekerja, istirahat makan siang, teh sore, naik kereta bawah tanah, mengemudi, olahraga, perjalanan, berjalan, bar   |
| Tema           | ACG (Anime, Komik, Game), Children (Anak-anak), Campus (Kampus), Game, Post 70s (Setelah tahun 70-an), Post 80s (Setelah tahun 80-an), Post 90s (Setelah tahun 90-an), Internet Song (Lagu Internet), KTV, Classical (Klasik) |
| Gaya           | Pop, Rock, Folk, Elektronik, Dance, Rap, Musik Ringan, Jazz, Country, R&B/Soul, Klasik, Etnik, Inggris, Metal, Punk, Blues, Reggae, Musik Dunia, Latin, Alternatif/Indie, New Age, Kuno, Post-rock, Bossa Nova                |
| Emosi          | Nostalgia, segar, romantis, seksi, sentimental, penyembuhan, relaksasi, kesepian, emosional, bersemangat, bahagia, tenang, nostalgia  |

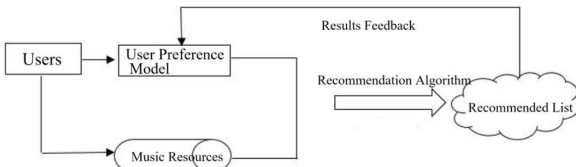
Data dalam penelitian ini utamanya berasal dari platform musik NetEase cloud, yang memiliki jumlah pendengar yang lebih besar dan informasi lagu yang lebih lengkap. Program crawler yang ditulis dalam bahasa Python digunakan untuk mengumpulkan 66.291 informasi data dari platform musik NetEase cloud. Bidang utama dari data yang dikumpulkan dalam penelitian ini adalah id daftar lagu, genre musik dari daftar lagu, jumlah kali lagu diputar, artis, waktu rilis lagu, durasi lagu, dan bidang-bidang terkait lainnya.

Data dalam penelitian ini masih perlu dalam format sistem rekomendasi, sehingga untuk memudahkan partisipasi dalam perhitungan dan menyesuaikan dengan kebutuhan library Python surprise saat ini dan kerangka kerja sistem rekomendasi yang paling umum, format data diproses menjadi format dasar dari dataset movielens. Data yang telah diproses ditampilkan dalam Gambar 4.

| Website_id           | 01_kind  | 02_kind            | create_time | favorite_count | share_count | comment_count | play_count | playlist_count |
|----------------------|----------|--------------------|-------------|----------------|-------------|---------------|------------|----------------|
| http://www.375661395 | Subject  | After 90           | 2016/5/5    | 378664         | 8295        | 5243          | 2223996    | 178            |
| http://www.373224005 | Language | Chinese            | 2016/5/9    | 439451         | 8161        | 1441          | 2170396    | 210            |
| http://www.554473977 | Subject  | List               | 2017/1/1    | 283964         | 2592        | 2955          | 29803614   | 100            |
| http://www.498780823 | Language | Chinese            | 2016/10/30  | 464891         | 6006        | 2561          | 29403924   | 200            |
| http://www.440561923 | Language | Europe and America | 2016/8/10   | 546976         | 4139        | 2650          | 20364706   | 100            |
| http://www.758104743 | Emotion  | Nostalgia          | 2017/8/10   | 209546         | 1171        | 959           | 1968308    | 300            |
| http://www.497639728 | Language | Video Soundtrack   | 2016/10/29  | 262802         | 1485        | 718           | 19142126   | 50             |
| http://www.375779419 | Subject  | Excitement         | 2016/5/5    | 500907         | 18925       | 3743          | 18364244   | 59             |
| http://www.331298286 | Emotion  | Nostalgia          | 2016/4/6    | 509601         | 5323        | 1521          | 17436442   | 92             |
| http://www.417794815 | Emotion  | Chinese            | 2016/7/8    | 239929         | 2131        | 995           | 17951472   | 101            |
| http://www.330634870 | Language | Europe and America | 2016/4/5    | 475196         | 3379        | 2642          | 17258234   | 181            |
| http://www.504347934 | Language | Chinese            | 2016/11/7   | 462240         | 3291        | 1241          | 17223906   | 83             |
| http://www.374012053 | Language | Europe and America | 2016/5/2    | 287285         | 1901        | 1060          | 16491452   | 104            |
| http://www.427850291 | Style    | Rock and Roll      | 2016/7/22   | 478410         | 3533        | 2861          | 16243550   | 55             |
| http://www.818563810 | Scene    | Evening            | 2017/7/21   | 337444         | 11287       | 22277         | 15164669   | 41             |
| http://www.111664488 | Style    | Electronic         | 2016/2/18   | 683050         | 13507       | 3216          | 15085941   | 76             |
| http://www.153864251 | Style    | Electronic         | 2016/1/14   | 313006         | 2957        | 1646          | 14168671   | 344            |
| http://www.308909592 | Subject  | Video Soundtrack   | 2016/2/28   | 394877         | 4910        | 2209          | 14157364   | 100            |
| http://www.399722525 | Language | Europe and America | 2016/6/12   | 433802         | 6576        | 2366          | 14107973   | 67             |
| http://www.458329079 | Language | Europe and America | 2016/9/5    | 393350         | 2902        | 1276          | 14069202   | 15             |
| http://www.517438933 | Emotion  | Sentimental        | 2016/11/23  | 348668         | 3017        | 829           | 13973394   | 78             |
| http://www.698720887 | Language | Europe and America | 2017/4/16   | 178391         | 866         | 517           | 13914202   | 101            |
| http://www.367649303 | Emotion  | Alone              | 2016/4/22   | 341332         | 2477        | 1888          | 13870766   | 117            |
| http://www.307592523 | Language | Europe and America | 2016/2/25   | 191953         | 1886        | 2091          | 13322389   | 44             |
| http://www.449559997 | Style    | World Music        | 2016/8/23   | 354823         | 3601        | 2423          | 13312112   | 76             |

Gambar 3. Contoh potret data sebagian

Id menggambarkan nilai id dari daftar lagu; kind menggambarkan jenis gaya dari daftar lagu, yang digunakan untuk perhitungan penilaian gaya kemudian; create\_time, yang mewakili waktu pembuatan, digunakan untuk perhitungan migrasi waktu kemudian; dan data pada kolom terakhir adalah jumlah koleksi, jumlah berbagi, jumlah komentar, jumlah pemutaran, dan jumlah lagu, yang akan digunakan dalam penelitian lebih lanjut di masa mendatang.



Gambar 4. Pemodelan

Sistem rekomendasi musik online mirip dengan model sistem rekomendasi tradisional yang terdiri dari tiga komponen utama: modul informasi pengguna, model sumber daya musik (pemodelan objek rekomendasi), dan algoritma rekomendasi musik personal. Hal ini dapat dilihat pada Gambar 4. Fokus dari eksperimen penelitian ini adalah pada modul sumber daya musik, yang terutama berisi beberapa informasi tentang daftar lagu seperti ID daftar lagu, jumlah lagu, penyanyi, dan gaya musik. Rekomendasi utamanya didasarkan pada informasi gaya musik di dalamnya. Pada saat yang sama, dari tiga metode sebelumnya, dapat dilihat bahwa rekomendasi berbasis konten memiliki data yang terbatas di domain musik dan sulit untuk menemukan preferensi baru pengguna. Rekomendasi berbasis aturan asosiasi mempertimbangkan objek yang lebih sedikit dan cakupannya lebih rendah. Rekomendasi musik berbasis jaringan sosial membutuhkan komputasi yang besar dan struktur yang kompleks serta seringkali memerlukan integrasi data lintas platform. Oleh karena itu, masing-masing dari ketiga metode tersebut memiliki kelebihan dan kekurangannya sendiri. Dalam penelitian ini, algoritma rekomendasi berbasis filtrasi kolaboratif digunakan untuk menghindari kekurangan ketiga metode tersebut dengan mempertimbangkan atribut pengguna, jenis gaya, dan migrasi waktu. Selain itu, algoritma ini memiliki volume komputasi yang relatif kecil dan digunakan di berbagai platform musik utama saat ini.

### 3. Hasil dan Diskusi

Ada tiga metode eksperimen umum yang sering digunakan dalam sistem rekomendasi, yaitu Studi Pengguna, Eksperimen Offline, dan Eksperimen Online. Dalam penelitian ini, metode eksperimen yang digunakan adalah eksperimen offline. Dataset yang ada dibagi menjadi dua bagian berdasarkan perbandingan 8:2, yaitu Training Set dan Test Set. Algoritma rekomendasi berbasis filtrasi kolaboratif terdiri dari dua jenis, yaitu algoritma berbasis pengguna (UserCF) dan algoritma berbasis produk (ItemCF). Pada awal eksperimen, sebagian kecil data digunakan untuk membandingkan kedua algoritma tersebut. Penelitian ini menggunakan nilai  $K = 10$ , dan hasilnya ditampilkan dalam Tabel 2.

**Tabel 2.** UserCF and ItemCF

|            | UserCF   | ItemCF  |
|------------|----------|---------|
| Recall     | 0.431348 | 0.43487 |
| Precision  | 0.37293  | 0.40143 |
| Popularity | 3.45781  | 3.27159 |
| Coverage   | 0.51246  | 0.53841 |

Berdasarkan tabel di atas, terlihat bahwa indikator-alat dari algoritma filtrasi kolaboratif berbasis pengguna (UserCF), seperti Recall, Precision, Popularity, dan Coverage, memiliki nilai yang lebih rendah dibandingkan dengan algoritma filtrasi kolaboratif berbasis produk (ItemCF). Oleh karena itu, dalam penelitian ini, algoritma filtrasi kolaboratif berbasis produk dipilih. Dalam penelitian ini, dilakukan percobaan dengan variasi nilai  $K$  pada algoritma ItemCF, yaitu  $K=5, 10, 20$ , dan  $30$ . Hasil percobaan tersebut kemudian ditampilkan dalam Tabel 3.

**Tabel 3.** ItemCF dengan nilai  $K$  yang berbeda

| K          | 5       | 10      | 20      | 30      |
|------------|---------|---------|---------|---------|
| Recall     | 0.27364 | 0.30142 | 0.31921 | 0.39281 |
| Precision  | 3.10323 | 0.32157 | 0.38805 | 3.11451 |
| Popularity | 0.51348 | 0.45841 | 0.35421 | 0.51967 |
| Coverage   | 0.30142 | 0.31921 | 0.39281 | 0.45841 |

Berdasarkan tabel di atas, terlihat bahwa terdapat pola peningkatan dan penurunan tingkat recall dan akurasi seiring dengan peningkatan nilai  $K$ . Titik tertinggi untuk tingkat recall dan akurasi terjadi saat  $K$  bernilai  $20$ , dengan nilai masing-masing sebesar  $0.45841$  dan  $0.38805$ . Oleh karena itu, dalam penelitian ini dapat disimpulkan bahwa algoritma lebih akurat saat  $K$  bernilai  $20$ . Penelitian ini melakukan klasifikasi terlebih dahulu terhadap jenis lagu, kemudian melakukan perhitungan peringkat pengguna untuk jenis lagu yang berbeda guna memberikan rekomendasi. Namun, untuk mengkaji pengaruh nilai  $K$  yang berbeda, dilakukan perhitungan terpisah menggunakan algoritma UserCF dengan nilai  $K 5, 10, 20$ , dan  $30$ . Hasil percobaan ini ditampilkan dalam Tabel 4.

**Tabel 4.** UserCF dengan nilai  $K$  yang berbeda

| K          | 5       | 10      | 20      | 30      |
|------------|---------|---------|---------|---------|
| Recall     | 0.31121 | 0.33429 | 0.34957 | 0.34281 |
| Precision  | 0.15481 | 0.16249 | 0.27543 | 0.16329 |
| Popularity | 3.09541 | 3.11428 | 3.13027 | 3.14438 |

| K        | 5        | 10      | 20      | 30      |
|----------|----------|---------|---------|---------|
| Coverage | 0.509128 | 0.51459 | 0.50129 | 0.49499 |

Berdasarkan perbandingan eksperimental, dapat dilihat bahwa ketika nilai K adalah 5, 10, 20, dan 30, kedua tabel menunjukkan pola yang sama. Hal ini menunjukkan bahwa ItemCF dan UserCF memiliki sifat yang serupa ketika nilai K mengambil keempat nilai tersebut. Selain itu, pada saat K=20, Recall dan Precision sistem lebih tinggi, menandakan bahwa nilai K tersebut adalah yang paling sesuai. Dengan demikian, eksperimen menunjukkan bahwa ketika nilai K diambil dengan tepat, akurasi dan cakupan rekomendasi dari algoritma rekomendasi berbasis filtrasi kolaboratif dengan klasifikasi gaya musik lebih tinggi. Dengan kata lain, ketika kita menggunakan nilai K yang tepat, tingkat ketepatan dan ketersediaan rekomendasi sistem meningkat.

#### 4. Kesimpulan

Dapat disimpulkan bahwa penelitian ini menemukan bahwa efek rekomendasi dapat dioptimalkan dengan menyesuaikan nilai K menggunakan algoritma KNN. Selain itu, penelitian ini menemukan bahwa pengaturan algoritma rekomendasi gabungan dapat mengoptimalkan sistem rekomendasi hingga batas tertentu. Rekomendasi gabungan berdasarkan perspektif-perspektif seperti atribut pengguna, jenis gaya, dan migrasi waktu dapat mengkompensasi kekurangan metode rekomendasi lainnya dan meningkatkan akurasi. Dalam sistem rekomendasi musik online, algoritma penilaian jenis gaya digunakan untuk menghitung skor tertinggi pada lagu-lagu yang direkomendasikan, dan tiga perspektif yaitu atribut pengguna, gaya musik, dan migrasi waktu dianalisis, dan sistem rekomendasi dibangun dengan menggabungkan ketiga perspektif. Meskipun demikian, penelitian ini juga memiliki beberapa kelemahan. Misalnya, kecerdasan buatan seperti deep learning telah mulai diterapkan dalam bidang sistem rekomendasi, namun penelitian ini tidak mempertimbangkan dan menguji teknologi tersebut. Hal ini juga menjadi arah pengembangan masa depan dari sistem rekomendasi musik online.

#### Daftar Pustaka

- [1] L. Chen, Y. Yang, N. Wang, K. Yang, and Q. Yuan, "How Serendipity Improves User Satisfaction with Recommendations? A Large-Scale User Evaluation," in *The World Wide Web Conference*, in WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 240–250. doi: 10.1145/3308558.3313469.
- [2] Y. Cheng and X. Bu, "Research on Key Technologies of Personalized Education Resource Recommendation System Based on Big Data Environment," *J. Phys. Conf. Ser.*, vol. 1437, no. 1, 2020, doi: 10.1088/1742-6596/1437/1/012024.
- [3] S. Gopal Krishna Patro et al., "A Conscious Cross-Breed Recommendation Approach Confining Cold-Start in Electronic Commerce Systems," *IEEE Access*, vol. 11, no. August, pp. 82857–82870, 2023, doi: 10.1109/ACCESS.2023.3274844.
- [4] F. Abbas, "One Size Does Not Fit All: Modeling Users' Personal Curiosity in Recommender Systems." [Online]. Available: <https://arxiv.org/abs/1907.00119>
- [5] E. J. Graham et al., "Optimization of a combined power plant CO 2 capture and direct air capture concept for flexible power plant operation," 2024, doi: 10.1039/d4ee00309h.

## Perancangan Desain UI/UX Aplikasi Somnia untuk Manajemen Pola Tidur Berbasis Mobile

Ni Putu Meita Kartika Dewi<sup>a1</sup>, Ida Ayu Gde Suwiprabayanti Putra<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>kartikadewimeita@gmail.com  
<sup>2</sup>agsuwiprabayantiputra@unud.ac.id

### Abstract

*Sleep is an essential basic need for human health and well-being. However, in the midst of a busy modern lifestyle, achieving ideal sleep becomes a challenge due to several factors. Consequently, lack of sleep can cause a negative impact on both physical and mental health. Based on these problems, this research aims to design the sleep management application "Somnia" using the prototype method. The application design is then evaluated using the System Usability Scale method. The evaluation results show that the average value of user satisfaction reaches 85.5, indicating that the design results is considered acceptable, with the adjective category being excellent, and received grade A. It was concluded, the design of the Somnia application is believed to have met user needs.*

**Keywords:** Sleep Management Application, User Interface, User Experience, Prototype, System Usability Scale.

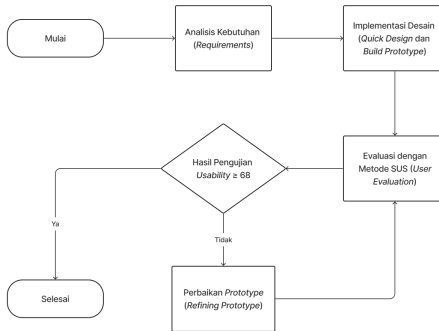
### 1. Pendahuluan

Tidur merupakan salah satu kebutuhan dasar manusia yang esensial bagi kesehatan dan kesejahteraan secara keseluruhan. Namun, di tengah era modern dengan segala kesibukan dan gaya hidup yang serba cepat, mencapai tidur yang ideal seringkali menjadi tantangan yang besar bagi banyak individu. Faktor-faktor seperti tekanan kerja yang tinggi, tingkat stress yang kronis, atau kebiasaan tidur yang kurang baik dapat mengganggu kualitas tidur seseorang. Akibatnya, kurang tidur dapat membawa dampak negatif bagi kesehatan fisik dan kesejahteraan mental, seperti penurunan konsentrasi, peningkatan risiko terhadap berbagai penyakit, dan gangguan suasana hati yang berkelanjutan [1]. Memperbaiki kualitas tidur dapat dicapai dengan membiasakan diri untuk menjalani pola tidur yang sehat. Langkah-langkah yang dapat dilakukan seperti menetapkan jadwal tidur yang konsisten, menciptakan lingkungan tidur yang nyaman, dan relaksasi sebelum tidur [1]. Kemajuan teknologi, telah memberikan solusi alternatif melalui aplikasi manajemen tidur yang dapat membantu individu dalam mengoptimalkan pola tidur mereka. Berdasarkan penelitian sebelumnya [2], Badriyyah dan Putra telah melakukan perancangan desain pada sebuah aplikasi pengelola pola tidur dengan menggunakan metode *design thinking*. Meskipun penelitian tersebut berhasil mencapai tingkat *usability* sebesar 70, masih terdapat ruang untuk perbaikan yang signifikan. Hal ini dapat dicapai melalui penerapan desain yang lebih baik, pemahaman yang lebih mendalam terhadap kebutuhan pengguna, serta pertimbangan terhadap metode lain yang dapat digunakan. Dengan adanya permasalahan tersebut, terbentuklah latar belakang yang menjadikan penelitian ini berfokus pada perancangan *User Interface (UI)* dan *User Experience (UX)* pada sebuah aplikasi mobile yang diberi nama "Somnia" dengan menggunakan metode *prototype*. Aplikasi ini bertujuan untuk memberikan dukungan dalam perencanaan dan manajemen waktu tidur secara efektif. Tidak hanya itu, aplikasi ini juga menyediakan fitur-fitur yang memungkinkan pengguna untuk melacak dan menganalisis pola tidur, menikmati berbagai suara menenangkan sebelum tidur, membantu pengguna menjaga konsistensi tidur mereka, serta menawarkan gamifikasi. Diharapkan bahwa hasil dari penelitian ini akan menghasilkan rancangan desain yang tidak hanya memenuhi kebutuhan pengguna, tetapi juga dapat meningkatkan kepuasan dan memberikan pengalaman tidur yang lebih

menyenangkan bagi pengguna secara keseluruhan.

## 2. Metode Penelitian

Penelitian ini menggunakan metode *prototype* untuk merancang desain aplikasi yang akan diuji oleh calon pengguna. *Prototype*, yang dalam bahasa Indonesia diartikan sebagai purwarupa atau rupa awal, adalah model awal dari suatu sistem yang dapat memberikan gambaran tentang bagaimana sistem tersebut akan terlihat secara keseluruhan pada tahap akhir. Metode *prototype* merupakan sebuah teknik dalam pengembangan sistem yang menggunakan *prototype* untuk memberikan gambaran kepada calon klien atau pengguna mengenai sistem yang akan dibangun oleh tim pengembang [3].



Gambar 1. Alur Penelitian

Adapun alur penelitian ini dimulai dengan tahap analisis kebutuhan (*requirements*), kemudian dilanjutkan dengan implementasi desain (*quick design* dan *build prototype*). Selanjutnya, dilakukan evaluasi dengan metode *System Usability Scale* (SUS). Apabila hasil pengujian mencapai nilai lebih dari atau sama dengan 68, maka penelitian dinyatakan selesai. Namun jika belum mencapai nilai tersebut, maka dilakukan perbaikan *prototype* hingga mencapai nilai lebih dari atau sama dengan 68.

### 2.1. Analisis Kebutuhan (*Requirements*)

Tahapan pertama dari metode *prototype* dimulai dengan pengumpulan kebutuhan dan analisis sistem. Pada tahapan ini dilakukan identifikasi kebutuhan sistem dan gambaran umum dari sistem yang akan dikembangkan [4]. Informasi ini dikumpulkan berdasarkan kebutuhan dan harapan pengguna terhadap sistem.

### 2.2. Implementasi Desain (*Quick Design dan Build Prototype*)

Tahapan kedua dalam metode *prototype* melibatkan perancangan desain secara cepat dan sederhana. Proses perancangan desain ini didasarkan pada pengumpulan kebutuhan dan analisis sistem pada tahap pertama. Tahap ini berfokus pada implementasi fitur-fitur yang menggambarkan konsep dan fungsi utama dari sistem tersebut. Tujuan dari tahap ini adalah untuk menciptakan representasi visual awal dari konsep sistem yang akan dibangun [4].

### 2.3. Evaluasi dengan Metode SUS (*User Evaluation*)

Pada tahap ini, sistem yang telah dibuat dalam bentuk *prototype* disajikan kepada pengguna untuk dinilai. Evaluasi oleh pengguna pada penelitian ini menggunakan metode *System Usability Scale* (SUS). SUS pertama kali diciptakan oleh John Brooke pada tahun 1986 [5]. Hingga saat ini, SUS telah umum digunakan untuk menguji kegunaan berbagai teknologi, mulai dari *software* hingga *hardware*. Perlu diingat, nilai SUS bukanlah nilai mutlak dari kegunaan suatu sistem, akan tetapi sebuah nilai yang mengukur kegunaan yang dirasakan oleh pengguna [5]. Kuesioner SUS terdiri dari sepuluh pertanyaan yang dijawab dengan lima pilihan tanggapan, mulai dari 1 (sangat tidak setuju) hingga 5 (sangat setuju). Setelah mengumpulkan respon pengguna, dilakukan perhitungan untuk mendapatkan nilai akhir SUS. Nilai yang berupa angka dalam skala 0-100 ini akan merepresentasikan tingkat kegunaan dari sistem yang sedang dievaluasi. Berikut adalah daftar pertanyaan yang digunakan dalam metode SUS.

**Tabel 1.** Daftar Pertanyaan Kuesioner SUS

| No | Pertanyaan   |
|----|--|
| 1  | Saya berpikir akan menggunakan sistem ini lagi                                     |
| 2  | Saya merasa sistem ini rumit untuk digunakan                                       |
| 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 serasi 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         |

Untuk menghitung nilai SUS, terdapat beberapa aturan sebagai berikut.

- Untuk pertanyaan bernomor ganjil, kurangi jawaban pengguna dengan 1 ( $x-1$ ).
- Untuk pertanyaan bernomor genap, kurangi 5 dengan jawaban pengguna ( $5-x$ ).
- Semua nilai hasil dari aturan 1 dan 2 akan berada dalam rentang 0-4, dimana angka 4 menunjukkan respon yang paling positif.
- Jumlahkan semua nilai tersebut dan kalikan dengan 2.5. Hal ini akan mengubah skala awal yang berkisar dari 0-40 menjadi 0-100.

Aturan perhitungan nilai tersebut berlaku pada satu responden. Untuk mencari nilai rata-rata dari masing-masing responden adalah dengan menggunakan rumus sebagai berikut [5].

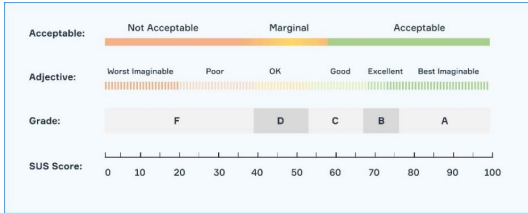
$$\bar{x} = \frac{\sum x}{n} \quad (1)$$

Keterangan:

- $\bar{x}$  = nilai rata-rata  
 $\sum x$  = jumlah nilai *System Usability Scale*  
 $n$  = jumlah responden

Hasil akhir kemudian disesuaikan dengan interpretasi nilai SUS. Hasil pengujian dikategorikan berdasarkan tiga aspek penilaian yang meliputi *acceptability*, *adjective rating*, dan *grade scale*. *Acceptability* adalah indikator untuk mengukur tingkat penerimaan pengguna terhadap *prototype*,

*grade scale* digunakan sebagai standar untuk menilai kualitas sebuah *prototype*, dan *adjective rating* digunakan sebagai pengukuran untuk mengevaluasi penilaian terhadap *prototype*.



Gambar 2. Interpretasi Nilai SUS

#### 2.4. Perbaikan *Prototype* (*Refining Prototype*)

Tahapan ini dilakukan untuk memperbaiki *prototype* berdasarkan hasil umpan balik pengguna yang diperoleh dari tahap sebelumnya. Proses ini akan terus berulang secara iteratif hingga kebutuhan pengguna terpenuhi. Nilai rata-rata dari banyak penelitian (standar) dengan metode SUS adalah 68, maka jika nilai di atas 68 akan dianggap di atas rata-rata dan nilai di bawah 68 dianggap di bawah rata-rata [5]. Nilai akhir di bawah 68 menandakan adanya masalah pada kegunaan sistem dan diperlukan perbaikan.

### 3. Hasil dan Diskusi

#### 3.1. Analisis Kebutuhan (*Requirements*)

Aplikasi ini dikembangkan dengan tujuan membantu pengguna meningkatkan kualitas tidur mereka. Pada aplikasi ini, memuat fitur-fitur yang dirancang untuk mendorong pengguna mencapai tujuan mereka.

##### a. Kebutuhan Fungsional

- Aplikasi dapat menyediakan fitur pengingat dan alarm untuk menetapkan dan mencapai target tidur.
- Aplikasi dapat menyediakan analisis data tidur yang terperinci untuk memantau kualitas tidur.
- Aplikasi dapat menampilkan statistik yang mencatat rata-rata tidur pengguna per minggu, bulan, dan tahun.
- Aplikasi dapat menyediakan musik atau suara menenangkan untuk membantu pengguna merasa lebih rileks sebelum tidur.
- Aplikasi dapat digunakan untuk mengatur preferensi tidur pengguna.
- Aplikasi dapat menyediakan *achievements* untuk memotivasi pengguna mencapai tujuan tidur mereka.

##### b. Kebutuhan Non Fungsional

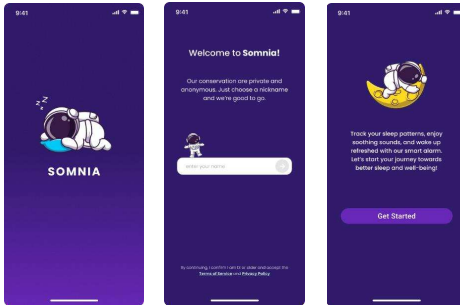
- Aplikasi ini dapat diakses melalui berbagai platform perangkat mobile.
- Aplikasi harus mematuhi regulasi serta standar kebijakan dan persyaratan terakrit.

#### 3.2. Implementasi Desain (*Quick Design dan Build Prototype*)

##### a. Tampilan *Splash Screen* dan *Onboarding Screen*



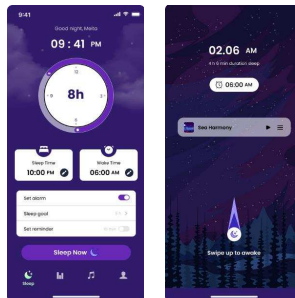
Dalam tampilan *splash screen*, pengguna akan melihat logo dan nama aplikasi. Setelah itu, pengguna akan diarahkan ke *onboarding screen*, dimana pengguna diminta untuk memasukkan nama mereka. Setelah nama dimasukkan, layar akan menampilkan informasi singkat tentang aplikasi sehingga membantu pengguna memahami tujuan dan manfaat dari menggunakan aplikasi Somnia.



Gambar 3. *Splash Screen dan Onboarding Screen*

#### b. Tampilan Pengingat Tidur dan Alarm

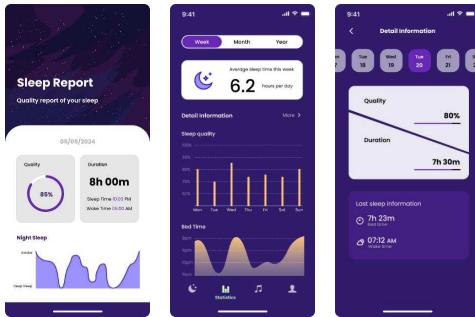
Dalam tampilan pengingat tidur dan alarm, pengguna akan menemukan informasi tentang target tidur mereka, waktu tidur yang direncanakan, waktu bangun, dan opsi untuk mengatur alarm pengingat tidur. Setelah pengguna menekan tombol *Sleep Now*, pola tidur mereka akan dianalisis. Ketika sudah waktunya bangun tidur, pengguna dapat men-*swipe up* tombol *wake up* untuk menghentikan alarm.



Gambar 4. *Pengingat Tidur dan Alarm*

**c. Tampilan Analisis dan Statistik Data Tidur**

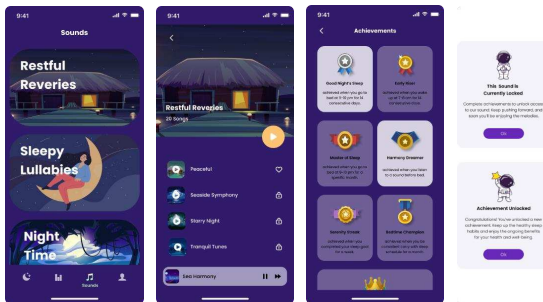
Dalam tampilan analisis, pengguna akan melihat informasi mengenai pola tidur mereka setelah mereka bangun, termasuk kualitas dan durasi tidur. Sedangkan pada tampilan statistik, tersedia informasi rata-rata waktu tidur serta informasi kualitas tidur, yang dapat dilihat per minggu, bulan, atau tahun. Pengguna dapat menjelajahi informasi lebih detail untuk mendapatkan pemahaman yang lebih baik tentang pola tidur mereka.



Gambar 5. Analisis dan Statistik Tidur

**d. Tampilan Daftar Musik Tidur dan Achievements**

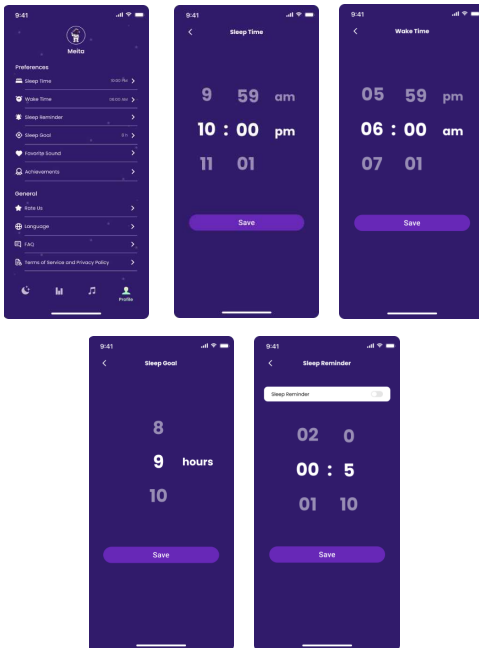
Dalam tampilan daftar musik tidur, pengguna dapat melihat musik-musik yang dapat mereka dengarkan sebelum tidur. Beberapa musik akan terkunci, dengan tujuan mendorong pengguna untuk mencapai suatu pencapaian (*achievements*) tertentu sebelum musik tersebut dapat diakses. Kemudian, pada tampilan *achievements*, pengguna dapat melihat daftar pencapaian yang telah mereka raih serta pencapaian yang masih belum tercapai.



Gambar 6. Daftar Musik Tidur dan Achievements

**e. Tampilan Menu Profil dan Edit Preferensi**

Dalam tampilan menu profil, pengguna dapat melihat nama yang telah mereka inputkan sebelumnya. Pengguna juga dapat mengakses menu-menu preferensi dan menu general. Kemudian, dalam tampilan edit preferensi, pengguna dapat mengedit preferensi sesuai kebutuhan, seperti menyesuaikan waktu tidur dan waktu bangun, mengedit pengingat tidur, menentukan *sleep goal*, serta mengatur preferensi lainnya.



**Gambar 7.** Menu Profil dan Edit Preferensi Pengguna

**3.3. Evaluasi dengan Metode SUS (User Evaluation)**

Pengujian desain aplikasi Somnia dilakukan dengan menyebarkan kuesioner yang sesuai dengan metode *System Usability Scale* (SUS) kepada 10 responden. Setiap responden diminta untuk menjawab serangkaian pertanyaan yang dirancang untuk mengukur aspek kegunaan dan kepuasan pengguna terhadap desain aplikasi Somnia. Nilai dari responden kemudian dihitung dan dianalisis berdasarkan aturan perhitungan SUS. Rincian hasil penilaian dari masing-masing responden menggunakan aturan perhitungan SUS dapat dilihat sebagai berikut.

Tabel 2. Hasil Evaluasi *System Usability Scale*

| Responden                       | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | Skor SUS    |
|---------------------------------|----|----|----|----|----|----|----|----|----|-----|-------------|
| R1                              | 4  | 4  | 4  | 3  | 4  | 4  | 3  | 4  | 3  | 2   | 87.5        |
| R2                              | 4  | 4  | 4  | 4  | 3  | 3  | 3  | 4  | 4  | 3   | 90          |
| R3                              | 4  | 3  | 4  | 4  | 3  | 3  | 4  | 3  | 4  | 3   | 87.5        |
| R4                              | 3  | 4  | 4  | 4  | 3  | 3  | 4  | 4  | 4  | 4   | 92.5        |
| R5                              | 3  | 3  | 3  | 3  | 3  | 4  | 3  | 4  | 3  | 1   | 75          |
| R6                              | 4  | 3  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 2   | 92.5        |
| R7                              | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 4  | 1   | 92.5        |
| R8                              | 3  | 4  | 4  | 4  | 3  | 3  | 4  | 3  | 3  | 4   | 87.5        |
| R9                              | 3  | 3  | 3  | 4  | 2  | 2  | 4  | 3  | 4  | 2   | 75          |
| R10                             | 3  | 3  | 3  | 3  | 4  | 3  | 3  | 3  | 3  | 2   | 75          |
| <b>Nilai Rata-Rata Skor SUS</b> |    |    |    |    |    |    |    |    |    |     | <b>85.5</b> |

Tabel 2 menunjukkan hasil evaluasi menggunakan metode *System Usability Scale* (SUS), dimana dalam perhitungan tersebut didapatkan nilai rata-rata sebesar 85,5. Nilai ini telah melampaui nilai minimum kelayakan (standar) yang sering dijumpai dalam penelitian dengan metode SUS yakni 68, serta telah melampaui nilai dari penelitian sebelumnya [2] yang memiliki nilai rata-rata sebesar 70. Berdasarkan interpretasi nilai SUS, hasil penelitian ini diklasifikasikan sebagai *acceptable*, dengan kategori *adjective* yaitu *excellent*, dan memiliki *grade A*.

### 3.4. Perbaikan *Prototype* (*Refining Prototype*)

Hasil evaluasi menunjukkan nilai SUS mencapai 85,5 dan telah melampaui nilai minimum kelayakan yaitu 68. Oleh karena itu, dalam penelitian ini tidak diperlukan perbaikan *prototype* dan hasil desain aplikasi dapat dilanjutkan ke tahap pengembangan berikutnya oleh tim pengembang.

## 4. Kesimpulan

Aplikasi *Somnia* merupakan aplikasi mobile yang dirancang menggunakan metode *prototype*, dengan tujuan untuk membantu pengguna dalam manajemen pola tidur dan meningkatkan kualitas tidur mereka. Dari hasil evaluasi desain aplikasi *Somnia*, didapatkan nilai rata-rata kepuasan pengguna mencapai 85,5. Hal ini menunjukkan bahwa desain aplikasi *Somnia* dinilai *acceptable*, dengan kategori *adjective* yaitu *excellent*, dan mendapat *grade A*. Dengan demikian, desain aplikasi *Somnia* diyakini telah memenuhi kebutuhan pengguna dan layak untuk diimplementasikan dalam tahapan pengembangan berikutnya.

## Daftar Pustaka

- [1] E. Savitrie, "Kurang Tidur dapat Mempengaruhi Kesehatan Mental", Kemenkes, 2022. [Online]. Available: Direktorat Jenderal Pelayanan Kesehatan (kemkes.go.id). [Accessed 8 Juni 2024]
- [2] S. Badriyyah and A. B. Putra, "Perancangan Aplikasi Pengelola Pola Tidur Dengan Metode *UX Design Thinking*," *Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC)*, vol. 16, no. 2, pp. 111-117, 2023.
- [3] B. Kurniawan and M. Romzi, "Perancangan UI/UX Aplikasi Manajemen Penelitian dan Pengabdian Kepada Masyarakat Menggunakan Aplikasi Figma," *JSIM: Jurnal Sistem Informasi Mahakarya*, vol. 5, no. 1, pp. 1-7, 2022
- [4] N. Renaningtias and D. Apriliani, "Penerapan Metode *Prototype* Pada Pengembangan Sistem Informasi Tugas Akhir Mahasiswa," *Jurnal Rekursif*, vol. 9, no. 1, pp. 92-98, 2021
- [5] E. Susilo, "Cara Menggunakan *System Usability Scale* (SUS) Pada Evaluasi *Usability*", *Edi*

Susilo, 2019. [Online]. Available: Cara Menggunakan System Usability Scale (SUS) Pada Evaluasi Usability ([edisusilo.com](http://edisusilo.com)). [Accessed 9 Juni 2024]

Halaman ini sengaja dibiarkan kosong

# Analisis Sentimen Ulasan Aplikasi M-Paspor Menggunakan TF-IDF dan Support Vector Machine

Ni Luh Putu Happy Nirmala<sup>a1</sup>, Ida Bagus Gede Dwidasmara<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>nirmala.2208561015@student.unud.ac.id  
<sup>2</sup>dwidasmara@unud.ac.id

## Abstract

*In the era of globalization, getting a passport has become an essential requirement as an official document for international travel. The Directorate General of Immigration introduced M-Paspor, a new application with more than 1 million downloads and 29 thousand reviews on the Google Play Store. This research aims to analyze the sentiments of Indonesian people using the M-Paspor application using the Support Vector Machine and TF-IDF methods for weighting, as well as evaluating the model with K-fold Cross Validation in Google Colab with the Python programming language. The SVM method was chosen because of its ability to achieve high classification accuracy, while feature extraction was carried out using the TF-IDF method to determine the weight of the words in the review. The dataset consists of 3,000 review data, with 1,500 negative sentiment review data and 1,500 positive sentiment review data, which underwent a series of preprocessing stages, namely noise removal, case folding, tokenization, normalization, stopwords removal, and stemming. The SVM model used to analyze and get the best combination of parameters C:1, gamma:scale, with the kernel:rbf. Evaluation of the model with K-Fold Cross Validation shows an average accuracy of 83.62%, precision of 84.7%, recall of 83.65%, and F1 score of 83.51%.*

**Keywords:** Sentiment Analysis, M-Paspor, TF-IDF, Support Vector Machine, K-Fold Cross Validation

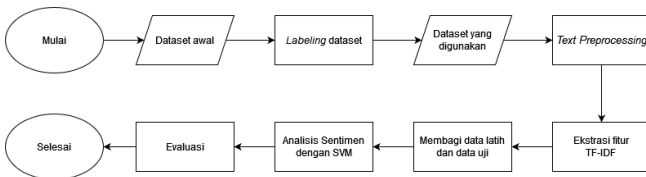
## 1. Pendahuluan

Pada era globalisasi ini, mobilitas manusia menjadi semakin penting. Warga Indonesia tidak terkecuali dari tren ini, semakin banyak yang tertarik untuk berkunjung ke luar negeri. Kebutuhan akan paspor sebagai dokumen resmi untuk melakukan perjalanan internasional menjadi suatu keharusan. Paspor dalam undang-undang diatur pada Undang-Undang Nomor 6 Tahun 2011 tentang Keimigrasian, hingga Peraturan Menteri Hukum dan HAM Nomor 8 Tahun 2014 [1]. Menurut rencana Kementerian Hukum dan HAM pada tahun 2020-2024, salah satu fokus utama adalah meningkatkan layanan keimigrasian dengan menggunakan teknologi [2]. Oleh karena itu, Direktorat Jenderal Imigrasi memperkenalkan M-Paspor, sebuah aplikasi baru yang memudahkan proses pembuatan paspor karena dapat digunakan untuk mendaftar permohonan paspor Indonesia baru atau penggantian paspor. Aplikasi ini sudah diunduh oleh 1 juta lebih masyarakat Indonesia dengan total 29 ribu ulasan di Google Play Store. Apabila proses evaluasi dari umpan balik yang diberikan oleh pelanggan dilakukan secara manual maka tentu saja akan menjadi tidak efektif karena ulasan yang diberikan oleh pelanggan terhitung sangat banyak dengan variasi komentar yang berbeda-beda. Maka dari itu diperlukan suatu metode untuk menganalisis hal tersebut dengan lebih efektif dan akurat, salah satunya adalah dengan analisis sentimen. Analisis sentimen adalah sebuah penelitian komputasional yang berkaitan dengan emosi yang direpresentasikan ke dalam sebuah teks. [3] Sebelumnya, sebuah penelitian terkait analisis sentimen pernah dilakukan yaitu dengan menganalisis sebuah ulasan hotel berbahasa Indonesia, yang juga menggunakan metode support vector machine (SVM) serta TF-IDF [4]. Pada penelitian tersebut dilakukan tahap normalisasi dengan cara memperbanyak kosakata yang ada pada kamus slang Bahasa Indonesia lalu memperoleh akurasi metode SVM dari eksperimen yang dilakukan yaitu sebesar 85%. Tujuan diadakannya penelitian ini tidak lain

adalah untuk menganalisis sentimen masyarakat Indonesia pengguna aplikasi M-Paspor yang memberikan ulasan di Google Play Store menggunakan metode (SVM) serta dengan metode TF-IDF untuk pembobotannya. Dengan adanya analisis sentiment ini, ulasan akan dibagi menjadi ulasan dengan sentimen positif dan sentimen negatif. Beberapa penelitian sebelumnya yang menggunakan dan membandingkan metode SVM dengan metode lainnya menyatakan bahwasanya Support Vector Machine (SVM) seringkali dipilih karena kemampuannya dalam mencapai akurasi klasifikasi yang tinggi sehingga mengungguli metode klasifikasi lainnya. [5] Namun, sebelum itu harus dilakukan tahap ekstraksi fitur yang merupakan salah satu tahapan penting. Penelitian ini melakukan ekstraksi fitur dengan metode TF-IDF untuk mencari bobot dari setiap kata-kata yang ada. Lalu bobot yang ada akan diletakkan di dalam suatu ruang vektor. Setelah seluruh proses tersebut dilakukan, di akhir akan dilakukan proses evaluasi model dengan K-fold Cross Validation. Penelitian ini dilakukan di Google Colab menggunakan Python serta bantuan beberapa library yang digunakan.

## 2. Metode Penelitian

### 2.1 Alur Penelitian



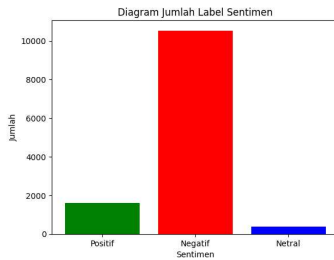
Gambar 1. Alur Penelitian

Alur pada penelitian ini tercakup pada Gambar 1. Langkah awal yang dilakukan adalah mencari dataset ulasan aplikasi M-Paspor yang ada di Google Play Store melalui beberapa situs yang ada. Setelah itu dilakukan labeling dataset untuk mengetahui label dataset yang akan digunakan untuk penelitian. Kemudian proses pengambilan sejumlah dataset yang akan digunakan agar mendapat dataset sebenarnya. Setelah proses tersebut dilakukan tahapan text preprocessing sebelum melalui ekstraksi fitur. Lalu akan dilakukan ekstraksi fitur dengan metode TF-IDF agar data tekstual yang ada selanjutnya dapat dipahami oleh metode analisis sentimen yang akan digunakan. Kemudian dilakukan proses pada dataset tersebut melalui model pembelajaran mesin dengan Support Vector Machine (SVM) untuk mengklasifikasi apakah data tekstual tersebut merupakan analisis sentiment positif atau negatif. Kemudian proses selanjutnya yang akan dilakukan ialah proses evaluasi dengan K-Fold Cross-Validation terhadap analisis sentimen yang telah dilakukan.

### 2.2 Pengumpulan Data

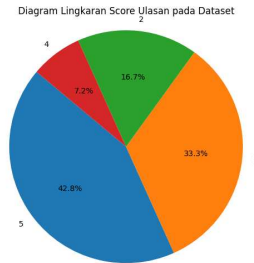
Proses pengumpulan dataset review aplikasi M-Paspor di Google Play Store ini menggunakan data primer yang diambil dari situs Kaggle tepatnya pada link berikut [ini](#). Dataset ulasan yang didapat menggunakan Bahasa Indonesia merupakan data ulasan dari tanggal 30 Desember 2021 hingga tanggal 05 Januari 2024. Data ini terdiri dari 12.499 ulasan yang terdiri dari nama pengguna yang memberikan ulasan, rating bintang (score), tanggal memberikan ulasan dan tentu saja isi ulasan yang diberikan. Kemudian dari dataset awal tersebut dilakukan proses labeling data, untuk ulasan dengan bintang 1 dan 2 diberi label negatif, ulasan dengan bintang 3 diberi label netral, serta ulasan dengan bintang 4 dan 5 diberi label positif.





**Gambar 2.** Diagram Batang Jumlah Label Sentimen pada Dataset Awal

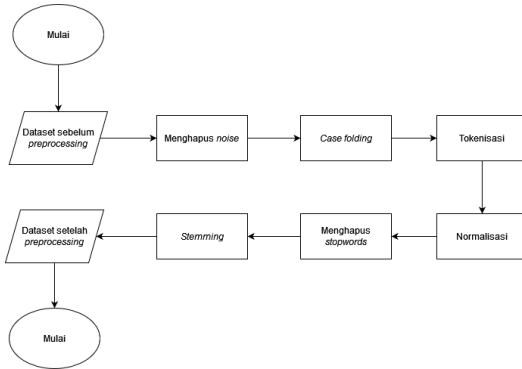
Namun karena setelah dianalisis bahwa ulasan dengan bintang 3 datanya cenderung rancu dengan adanya data yang terkadang netral, positif maupun negatif maka peneliti tidak mengikutsertakan ulasan dengan bintang 3 pada penelitian kali ini dan tidak menggunakan seluruh dataset yang ada, melainkan hanya memilih sejumlah data terbaru saja. Data yang akan dipakai yaitu sejumlah 1000 data ulasan dengan bintang 1, 500 data ulasan dengan bintang 2, 215 data dengan bintang 4 dan 1285 data dengan bintang 5. Persentase score ulasan yang digunakan pada dataset dapat dilihat di diagram lingkaran Gambar 3.



**Gambar 3.** Diagram Lingkaran Score Ulasan pada Dataset yang Digunakan

Data tersebut berasal dari tanggal 30 Januari 2022 hingga tanggal 05 Januari 2024. Sehingga total data yang dipakai pada penelitian ini yaitu 3.000 data ulasan dengan jumlah ulasan sentimen yang negatif sejumlah 1.500 data dan jumlah ulasan sentimen yang positif sejumlah 1.500 data.

### 2.3 Text Preprocessing



Gambar 4. Alur Text Preprocessing

Adapun tahap-tahap dari text preprocessing penelitian kali ini dapat direpresentasikan pada Gambar 4. Proses awal yang dilakukan pada dataset adalah menghapus noise misalnya seperti menghilangkan tanda baca, karakter yang bukan teks, tautan, angka dan yang lainnya. Selanjutnya tahap case folding yaitu proses untuk menyamakan teks yang ada pada dataset agar tidak terpengaruh oleh adanya huruf besar maupun huruf kecil, dataset ini akan dibuat lower text seluruhnya. [6] Kemudian proses tokenisasi ialah proses pengolahan teks yang dapat membagi kalimat menjadi sebuah token yang pada kali ini dibagi menjadi kata perkata.[7] Setelah itu dataset melalui proses normalisasi yaitu perubahan kata-kata yang menggunakan bahasa sehari-hari atau slang yang belum baku agar menjadi kata-kata yang baku dan sesuai aturan penulisannya menggunakan bantuan kamus. [8] Lalu akan dilakukan proses penghapusan stopwords pada dataset yang digunakan seperti kata-kata yang sebenarnya tidak penting atau kurang relevan di dalam dataset dan hanya akan menjadikan proses semakin lama sehingga harus dihapus. Setelah tahap-tahapan tersebut barulah dataset dilakukan proses stemming yaitu tahapan agar kata yang mengandung imbuhan awalan maupun akhiran dapat diubah menjadi kata dasar saja. Setelah itu akan dihasilkan sebuah dataset yang telah melalui tahap-tahap preprocessing menggunakan bahasa pemrograman Python.

### 2.4 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan singkatan dari Term Frequency-Inverse Document Frequency adalah salah satu metode untuk melakukan ekstraksi fitur dan pembobotan dalam sebuah teks berdasarkan frekuensi munculnya sebuah kata untuk mendeteksi pentingnya kata tersebut di sebuah dataset. Pembobotan kata dengan TF-IDF didapat dari hasil kali Term Frequency dengan hasil Inverse Document Frequency berdasarkan query.[9] Nilai TF sendiri berarti frekuensi dari kemunculan suatu kata di dalam dokumen sehingga jika nilai TFnya tinggi berarti kata tersebut sangat penting di dalam dokumen. Sedangkan nilai IDF adalah invers dari frekuensi dari kemunculan suatu kata dari seluruh dokumen yang ada sehingga jika nilai IDFnya tinggi maka berarti jumlah kemunculan katanya sedikit dan berarti sangat penting. [10] Setelah dataset melalui proses preprocessing selanjutnya akan dilakukan ekstraksi fitur dan pembobotan kata dengan metode TF-IDF. Untuk sebuah kata "t" dalam dokumen "d", maka TF-IDFnya didapat menggunakan rumus seperti di bawah ini:

$$TF(t, d) = \frac{\text{Frekuensi munculnya kata } t \text{ di dokumen } d}{\text{Frekuensi kata seluruhnya di dokumen } d} \quad (1)$$

$$IDF(t) = \log \frac{\text{Frekuensi dokumen}}{\text{Frekuensi dokumen yang terdapat kata } t} \quad (2)$$

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

## 2.5 Support Vector Machine (SVM)

SVM adalah singkatan dari Support Vector Machine adalah salah satu algoritma machine learning yang dipakai untuk analisis sentimen. Konsep utama dari SVM adalah mencari pemisah atau hyperplane terbaik yang dapat sebagai pemisah antara dua jenis sentimen data dengan margin maksimum, margin adalah jarak terdekat antara hyperplane dan titik data terdekat dari kedua kelas. SVM berusaha untuk menemukan hyperplane yang paling baik membedakan antara kelas dengan memaksimalkan margin ini, sehingga membuatnya sangat efektif dalam mengklasifikasikan data yang kompleks. Selain itu, SVM juga dapat memproyeksikan data dengan memanfaatkan fungsi kernel sehingga dapat diproyeksikan ke dimensi yang lebih tinggi, yang memungkinkannya untuk menangani kasus-kasus untuk kelas tidak dapat dipisahkan secara linear dalam dimensi yang rendah. Support vector merupakan penyebutan untuk pola yang posisinya terdekat dengan hyperplane. [11]

## 2.6 Evaluasi Model Dengan K-Fold Cross-Validation

Evaluasi model dengan k-fold cross-validation adalah sebuah metode yang dipakai untuk mengukur seberapa baik model machine learning dalam melakukan prediksi. Dalam k-fold cross-validation, data dipisahkan menjadi k subset yang sama besar. Model dilatih menggunakan k-1 subset dan diuji pada subset yang tersisa. Proses ini diulangi k kali, di mana setiap subset digunakan sebagai data uji secara bergantian. Hasil akhir dari evaluasi ini adalah rata-rata dari nilai-nilai metrik evaluasi (seperti akurasi) dari setiap iterasi k-fold cross-validation. [12]

## 3. Hasil dan Pembahasan

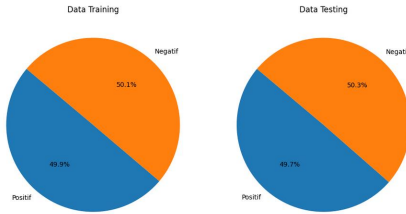
### 3.1 Text Preprocessing

Langkah awal yang dilakukan terhadap dataset yaitu tahap preprocessing kepada 3.000 data yang ada di dataset. Berikut adalah tahapan preprocessing yang dilakukan terhadap dataset:

| Tahap Preprocessing        | Data  |
|----------------------------|---|
| Data sebelum preprocessing | Hari libur, server libur? aplikasi tidak dpat diakses.. Kedepannya harap diperbaiki                       |
| Menghapus noise            | Hari libur server libur aplikasi tidak dpat diakses Kedepannya harap diperbaiki                           |
| Case folding               | hari libur server libur aplikasi tidak dpat diakses kedepannya harap diperbaiki                           |
| Tokenisasi                 | ['hari', 'libur', 'server', 'aplikasi', 'tidak', 'dpat', 'diakses', 'kedepannya', 'harap', 'diperbaiki']  |
| Normalisasi                | ['hari', 'libur', 'server', 'aplikasi', 'tidak', 'dapat', 'diakses', 'kedepannya', 'harap', 'diperbaiki'] |
| Menghapus stopwords        | ['hari', 'libur', 'server', 'aplikasi', 'diakses', 'kedepannya', 'diperbaiki']                            |
| Stemming                   | ['hari', 'libur', 'server', 'aplikasi', 'akses', 'depan', 'perbaik']                                      |

Gambar 5. Hasil Tahap Preprocessing

Tahapan text preprocessing di atas dilakukan di Google Colab menggunakan bahasa pemrograman python. Dalam prosesnya penelitian ini dibantu oleh library pandas, nltk dan juga sastrawi. Serta untuk proses normalisasi menggunakan bantuan kamus kamus.txt. Kemudian dataset yang telah melalui tahap preprocessing dibagi menjadi data training serta data testing. Adapun perbandingan data training dan data testing-nya yaitu 8:2 sehingga jumlah data latihnya menjadi 2.400 dan data ujinya menjadi 600. Untuk persentase jumlah label dataset pada data uji dan data latih, terlihat pada Gambar 6.



Gambar 6. Diagram Lingkaran Persentase Sentimen pada Data Training dan Data Testing

Berikut merupakan code pada tahap text processing dari awal hingga akhir:

```
def casefolding(teks):  
    #menghapus noise  
    dataset = re.sub(r'[A-Za-a0-9]+', '', teks)  
    dataset = re.sub(r'[A-Za-z0-9]+', '', teks)  
    dataset = re.sub(r'http\S+', '', teks)  
    dataset = re.sub(r'[0-9]+', '', teks)  
    dataset = re.sub(r'[-()\"#%/&;<>{}+~-.!?,_]', "", teks)  
  
    #case folding  
    teks = teks.lower()  
  
    return teks
```

Gambar 7. Code Tahap Preprocessing Bagian Menghapus Noise dan Case Folding

```
def tokenisasi(teks):  
    teks = word_tokenize(teks)  
  
    return teks
```

Gambar 8. Code Tahap Preprocessing Bagian Tokenisasi

```
def normalisasi(teks):  
    kamusSlang = eval(open("kamus.txt").read())  
    pattern = re.compile(r'\b( ' + '|'.join(kamusSlang.keys())+r')\b')  
    content = []  
    for kata in teks:  
        filterSlang = pattern.sub(lambda x: kamusSlang[x.group()],kata)  
        content.append(filterSlang.lower())  
    teks = content  
    return teks
```

Gambar 9. Code Tahap Preprocessing Bagian Normalisasi

```
stopwords_indonesia = set(stopwords.words('indonesian'))

def removedStopwords(teks):
    tokens_with_no_stopwords = list()

    for word in teks:
        if word not in stopwords_indonesia:
            tokens_with_no_stopwords.append(word)

    return tokens_with_no_stopwords
```

**Gambar 10.** Code Tahap Preprocessing Bagian Menghapus Stopwords

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemming(teks):
    return [stemmer.stem(word) for word in teks]
```

**Gambar 11.** Code Tahap Preprocessing Bagian Stemming

### 3.2 Ekstraksi Fitur

Tahapan ekstraksi fitur kepada dataset yang sudah menjalani tahap preprocessing adalah dengan TF-IDF. Hasil ekstraksi fitur ini yang akan menjadi input dalam model analisis sentimen dengan metode SVM. Dalam proses ekstraksi fitur ada dua parameter utama yang digunakan yaitu  $max\_df=0.5$  yang berarti kata yang muncul di lebih dari 50% dari semua dokumen akan diabaikan. Kata-kata ini kemungkinan besar adalah kata umum yang tidak relevan. Selain itu terdapat juga  $min\_df=2$  yang berarti kata yang muncul di kurang dari 2 dokumen akan diabaikan. Ini membantu menghilangkan kata-kata yang terlalu spesifik yang mungkin tidak memberikan banyak informasi. Kemudian dataset diubah menjadi representasi TF-IDF yang mana hasilnya adalah Setiap kolom akan mewakili fitur yang dihasilkan oleh TF-IDF, dan setiap barisnya akan mewakili satu dokumen.

### 3.3 Analisis Sentimen

Setelah dataset melalui beberapa tahap proses preprocessing dan ekstraksi fitur pembobotan dengan TF-IDF, akan dilakukan proses analisis sentimen dengan metode Support Vector Machine. Adapun parameter C yang digunakan yaitu 0.1, 1, 10, 100, parameter kernel yang digunakan yaitu kernel linear, rbf, dan poly, sedangkan parameter gamma menggunakan scale dan auto. Pada implementasinya gridsearch digunakan untuk mencari kombinasi hyperparameter yang terbaik. Ternyata kombinasi parameter terbaik yang ditemukan yaitu C yaitu 1, gamma yaitu scale, dengan kernel yaitu rbf. Untuk data latihnya didapat hasil SVM terbaik yaitu mean accuracy 95,29%, precision 95,35%, recall 95,29% dan F1 score 95,29%. Untuk data uji nya didapat hasil SVM terbaik yaitu mean accuracy 84,5%, precision 85,13%, recall 84,5% dan F1 score 84,37%.

### 3.4 Evaluasi Model dengan K-Fold Cross-Validation

Setelah tahap analisis sentimen berhasil dilakukan pada dataset. Maka evaluasi model dengan K-Fold Cross-Validation telah dapat dilakukan untuk mengevaluasi model dengan lebih baik. Metrik yang dipakai adalah accuracy, precision, recall, serta F1 score. Pada penelitian kali ini didapat hasil mean accuracy 83,62%, precision 84,7%, recall 83,65% dan F1 score 83,51%.

#### 4. Kesimpulan

Penelitian ini memiliki tujuan untuk melakukan analisis sentimen dari ulasan pengguna aplikasi M-Paspor menggunakan metode TF-IDF untuk ekstraksi fitur dan pembobotan, serta Support Vector Machine (SVM) untuk analisis sentimen. Dataset yang terdiri dari 3.000 data ulasan, dengan 1.500 data ulasan sentimen negatif dan 1.500 data ulasan sentimen positif, menjalani serangkaian tahapan preprocessing termasuk penghapusan noise, case folding, tokenisasi, normalisasi, penghapusan stopwords, dan stemming. Fitur-fitur dari ulasan diekstraksi dan diberi bobot dengan metode TF-IDF, kemudian model SVM digunakan untuk menganalisis sentimen dan didapat kombinasi parameter terbaik C yaitu 1, gamma yaitu scale, dengan kernel yaitu rbf. Evaluasi model dilakukan dengan K-Fold Cross Validation, menyatakan akurasi rata-rata sebesar 83,62%, precision sebesar 84,7%, recall sebesar 83,65%, dan F1 score yaitu sebesar 83,51%. Hasil penelitian ini menunjukkan potensi metode TF-IDF dan SVM dalam analisis sentimen ulasan khususnya ulasan aplikasi M-Paspor.

#### Daftar Pustaka

- [1] A. Irawan, M. Z. Jamlean, dan H. Ibrahim, "Manajemen Pelayanan Pembuatan Paspor Pada Kantor Imigrasi Kelas I Tpi Merauke," MADANI Jurnal Politik dan Sosial Kemasyarakatan, vol. 14, hlm. 399–415, Agu 2022.
- [2] A. Prameswari, S. Zakaria, dan S. Centia, "Pelayanan Publik Berbasis Electronic Government Melalui Penerapan Aplikasi Mobile Paspor (M-Paspor) Di Kantor Imigrasi Kelas I Tpi Bandung Tahun 2022," Jurnal Administrasi Pemerintahan (JANITRA), vol. 3, hlm. 256–262, Nov 2023.
- [3] F. Gunawan, M. Ali Fauzi, dan P. Pandu Adikara, "Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes Dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi Bca Mobile)," Malang, Agu 2017.
- [4] V. W. D. Thomas dan F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," Jurnal Media Informatika Budidarma, vol. 6, no. 3, hlm. 1767–1774, Jul 2022, doi: 10.30865/mib. v6i3.4218.
- [5] P. A. Octaviani, Y. Wilandari, dan D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (Sd) Di Kabupaten Magelang," Jurnal Gaussian, vol. 3, no. 4, hlm. 811–820, 2014, [Daring]. Tersedia pada: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [6] D. Alita dan A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," Labuhanratu, 2020.
- [7] E. K. Putri dan T. Setiadi, "Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes," Jurnal Sarjana Teknik Informatika, vol. 2, hlm. 73–83, Okt 2014.
- [8] Y. Arum Sari dan P. Pandu Adikara, "Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata," Malang, Mei 2019. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [9] M. A. Rofiqi, Abd. C. Fauzan, A. P. Agustin, dan A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," ILKOMNIKA: Journal of Computer Science and Applied Informatics, vol. 1, no. 2, hlm. 58–64, Des 2019, doi: 10.28926/ilkomnika. v1i2.18.
- [10] A. Ikegami, I. Dewa, M. Bayu, dan A. Darmawan, "Analisis Sentimen dan Pemodelan Topik Ulasan Aplikasi Noice Menggunakan XGBoost dan LDA," Badung, Nov 2022.
- [11] P. A. N. Aryanti dan I. B. M. Mahendra, "Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine," Jurnal Elektronik Ilmu Komputer Udayana (JNATIA), vol. 12, no. 1, hlm. 2654–5101, Agu 2023.
- [12] R. Tuntun, K. Kusriani, dan K. Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation," Jurnal Media Informatika Budidarma, vol. 6, no. 4, hlm. 2111, Okt 2022, doi: 10.30865/mib. v6i4.4681.

# Analisis Sentimen Ulasan Aplikasi Citilink menggunakan Metode Support Vector Machine dengan TF-IDF

David Brave Moarota Zebua<sup>a1</sup>, Ida Bagus Gede Dwidasmara<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>davidzebua8@gmail.com  
<sup>2</sup>dwidasmara@unud.ac.id

## Abstract

*In line with the advancement of the Industry 4.0 era, Indonesian society has been living side by side and is inseparable from the existing technological advancements. One of the conveniences experienced by today's society is that transactions no longer need to be conducted face-to-face in a particular place but can now be done online. In the context of air transportation, technological advancements have been very helpful to the public. Airline applications are one of the most widely used by passengers. In this study, the researchers focused on analyzing public sentiment towards the Citilink application, one of Indonesia's leading airlines. The researchers used the Support Vector Machine (SVM) method enhanced with TF-IDF (Term Frequency-Inverse Document Frequency) text representation to analyze sentiment from user reviews. The stages of this research began with data collection containing reviews from the Citilink application to analyze its sentiment. Then, it proceeded to the data preprocessing stage, where the collected data was cleaned until it became tokens ready for testing. After that, it moved to the weighting stage using Term Frequency-Inverse Document Frequency (TF-IDF). Then it continued to the stage of applying the Support Vector Machine (SVM) model. The last one is the evaluation to measure the accuracy level of the model used. Based on the results of this study, it can be concluded that the Support Vector Machine model that has been adapted to the dataset of Citilink application reviews from Google Playstore and supported by TF-IDF feature extraction successfully classified the sentiment of reviews with high accuracy, reaching 88%. Further evaluation also showed satisfactory values of precision, recall, and F1-Score, namely 90%, 83%, and 85%, respectively. This study shows that the Support Vector Machine model can be an effective instrument in understanding user responses to the performance of the Citilink application.*

**Keywords:** Sentiment Analysis, Citilink, Support Vector Machine, TF-IDF, Confusion Matrix

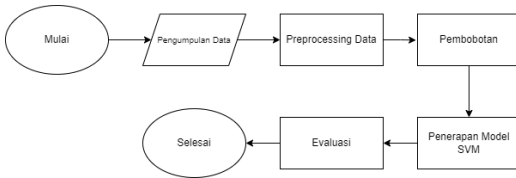
## 1. Pendahuluan

Seiring dengan perkembangan era industri 4.0, masyarakat Indonesia telah hidup berdampingan dan sudah tidak bisa dipisahkan lagi dengan kemajuan teknologi yang ada. Disamping itu, pemanfaatan teknologi yang sangat maju memberikan kemudahan dalam berbagai sektor kehidupan manusia, baik untuk melaksanakan kewajiban, maupun memenuhi kebutuhan. Adapun kemudahan yang dirasakan bagi masyarakat saat ini, dimana sekarang transaksi tidak perlu bertemu secara langsung dalam suatu tempat melainkan sekarang sudah bisa dilakukan secara online [1]. Kemudahan-kemudahan ini mempengaruhi cara masyarakat berperilaku dan memaksa kita untuk terus beradaptasi. Dalam konteks transportasi udara, adanya kemajuan teknologi sangat membantu masyarakat. Semuanya bermula dari munculnya peningkatan yang terjadi pada penerbangan domestik di Indonesia sehingga memunculkan ide untuk mengembangkan sebuah aplikasi online berbasis android yang berfungsi untuk memesan tiket pesawat [2]. Mulai dari pemesanan tiket secara *online* hingga memberikan informasi penerbangan secara *real-time* terus menerus dikembangkan. Aplikasi maskapai penerbangan menjadi salah satu yang paling banyak digunakan oleh penumpang. Di tengah persaingan yang

ketat di industri penerbangan, pendapat masyarakat terhadap aplikasi suatu maskapai dapat menjadi kunci keberhasilan maskapai tersebut. Dalam penelitian ini, peneliti berfokus dalam analisis sentimen pendapat masyarakat terhadap aplikasi Citilink, salah satu maskapai penerbangan terkemuka di Indonesia. Peneliti menggunakan metode Support Vector Machine (SVM) yang diperkuat dengan representasi teks TF-IDF (Term Frequency-Inverse Document Frequency) untuk menganalisis sentimen dari ulasan pengguna. Tujuan dari penelitian ini adalah untuk mengembangkan model yang dapat mengklasifikasikan ulasan pengguna menjadi kategori sentimen positif maupun negatif, sehingga maskapai penerbangan dapat merespons dengan cepat terhadap umpan balik pelanggan dan meningkatkan kualitas layanan mereka.

## 2. Metode Penelitian

### 2.1 Desain Penelitian



Gambar 1. Bagan Alur Penelitian

Tahap pertama dari penelitian ini yaitu pengumpulan data yang berisi ulasan dari aplikasi Citilink untuk menganalisis sentimennya. Kemudian, masuk ke tahap preprocessing data, dimana data-data yang sudah dikumpulkan dibersihkan hingga menjadi token yang siap diuji. Setelah itu, masuk ke tahap pembobotan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Kemudian lanjut ke tahap penerapan model *Support Vector Machine* (SVM). Yang terakhir yaitu evaluasi untuk mengukur tingkat akurasi model yang digunakan.

### 2.2 Pengumpulan Data

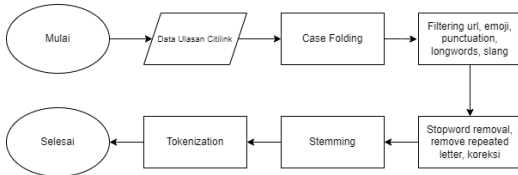
Teknik pengumpulan data yang digunakan pada penelitian ini yaitu menggunakan data sekunder yang diambil melalui *Kaggle*. Dataset yang digunakan merupakan dataset ulasan aplikasi Citilink yang terdapat di Google Playstore, dimana terdapat sebanyak 1471 data. Pada data *Kaggle* terdapat *score* yang dimana dapat diasumsikan *score* 1-3 merupakan ulasan buruk dan *score* 4-5 merupakan ulasan baik, namun tidak terdapat label yang menentukan sentimen.

Tabel 1. Tabel Gambaran Dataset

| Content  | Score |
|--|-------|
| Makin hari pelayanan makin buruk,tiket makin mahal,,m                | 1     |
| Worst airplan ever....   | 1     |
| Tidak bisa mngajukan refund, saat maskapai merubah jdwal penerbangan | 1     |
| Tidak dapat mengubah destinasi Banyak bug                            | 1     |
| Mantap sih aplikasinya   | 5     |



### 2.3 Text Preprocessing



Gambar 2. Bagan Alur Tahap Preprocessing

Pada tahap preprocessing data, tahap pertama yang dilakukan yaitu melakukan *case folding* dimana semua huruf yang ada diubah menjadi huruf kecil. Tahap selanjutnya ialah filtering yang berfungsi untuk menghilangkan URL, simbol, emoji, nomor, kata panjang, dan juga *slang* yang ada pada dataset. Kemudian, masuk ke tahap *stopword removal* untuk menghapus *stopword*, menghapus huruf yang berulang, dan juga koreksi beberapa kesalahan penulisan maupun singkatan. Selanjutnya masuk ke tahap *stemming* terlebih dahulu untuk mengubah tiap kata menjadi kata dasar, dan yang terakhir yaitu tokenisasi untuk mengubah data menjadi token-token terpisah.

### 2.4 Term-Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang umum digunakan untuk merepresentasikan dokumen teks sebagai vektor numerik. Vektor ini kemudian dapat digunakan dalam berbagai tugas pemrosesan bahasa alami (NLP), seperti klasifikasi teks, pengelompokan teks, dan analisis sentimen. *Term Frequency* yaitu frekuensi kemunculan term  $i$  pada dokumen  $j$  dibagi dengan total term pada dokumen  $j$ , dituliskan dalam bentuk [3],

$$tf_{ij} = \frac{f_{d(i)}}{\max_{j \in d} f_{d(j)}} \quad (1)$$

*Inverse Document Frequency* berfungsi mengurangi bobot suatu term jika kemunculannya banyak tersebar diseluruh dokumen, dituliskan dalam bentuk [4],

$$idf(w) = \log \left( \frac{N}{DF(w)} \right) \quad (2)$$

### 2.5 Support Vector Machine

Support Vector Machine (SVM) adalah salah satu algoritma *machine learning* yang memiliki kinerja sangat baik dalam mengklasifikasikan data. Hal yang luar biasa dari SVM ini yaitu dapat digunakan tanpa bergantung pada dimensi dari ruang fitur [5]. SVM mengukur kompleksitas hipotesis berdasarkan margin yang memisahkan bidang dan bukan jumlah fitur. SVM juga dikatakan sebagai pengklasifikasi linear yang didasarkan pada prinsip memaksimalkan margin. SVM menggunakan *hyperplane* secara optimal untuk mengklasifikasikan data ke dalam dua kelompok di ruang dimensi yang lebih tinggi. Margin adalah jarak antara *hyperplane* dan data terdekat dari setiap kelas. Data terdekat ini disebut vektor pendukung. *Hyperplane* adalah pemisah terbaik antara dua kelas yang telah ditentukan sebelumnya [6]. Prinsip dasar dari SVM adalah pengklasifikasi linear, dan kemudian dikembangkan agar dapat bekerja pada masalah non-linear, yaitu dengan menggabungkan konsep trik kernel dalam ruang kerja berdimensi tinggi.

### 2.6 Confusion Matrix

*Confusion matrix* dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan

analisis apakah *classifier* tersebut baik dalam mengenali *tuple* dari kelas yang berbeda. Nilai dari *True-Positive* dan *True-Negative* memberikan informasi kepada *classifier* dalam melakukan klasifikasi data bernilai benar, sedangkan *False-Positive* dan *False-Negative* memberikan informasi ketika *classifier* salah dalam melakukan klasifikasi data [7]. Berikut merupakan rumus untuk melakukan perhitungan *confusion matrix*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{6}$$

### 3. Hasil dan Pembahasan

#### 3.1 Preprocessing Data

Jumlah data yang akan melalui tahap *preprocessing* terdiri dari 974 data positif, dan 497 data negatif. Tahap pertama yaitu *Case Folding*, yang dilakukan untuk mengkonversi semua huruf dalam dataset menjadi huruf kecil. Kemudian terdapat penghapusan URL, simbol, emoji, nomor, kata berukuran panjang, dan bahasa gaul. Selanjutnya dilakukan penghapusan *stopwords* dengan *library* Sastrawi untuk menghapus *stopwords* berbahasa Indonesia, juga koreksi terhadap pengulangan huruf serta kesalahan penulisan. Kemudian yaitu tahap *stemming* yang dilakukan dengan *library* Sastrawi untuk mengubah setiap kata menjadi bentuk dasarnya. Tahap terakhir yaitu tokenisasi dimana hasil *stemming* dibagi menjadi token per kata. Adapun contoh data yang telah melewati tahap *preprocessing* terdapat pada tabel 2.

**Tabel 2.** Contoh Data Preprocessing

| Content  | Case folding   | Cleaned  | Stemmed   | Token   |
|--|--|--|---|---|
| Tidak bisa<br>mngajukan refund,<br>saat maskapai<br>merubah jdwal<br>penerbangan | tidak bisa<br>mngajukan refund,<br>saat maskapai<br>merubah jdwal<br>penerbangan | tidak bisa<br>mengajukan<br>refund maskapai<br>merubah jadwal<br>penerbangan | tidak bisa aju<br>refund<br>maskapai<br>rubah jadwal<br>terbang | ['tidak' 'bisa'<br>'aju' 'refund'<br>'maskapai'<br>'rubah' 'jadwal'<br>'terbang'] |

#### 3.2 Perhitungan Ekstraksi Fitur Term Frequency-Inverse Document Frequency (TF-IDF)

Sesudah tahap *preprocessing* dilakukan, langkah berikutnya yaitu perhitungan ekstraksi fitur TF-IDF yang dimana data akan diubah menjadi vektor numerik. Ekstraksi fitur ini dilakukan menggunakan *library scikit-learn*. Pertama-tama, *TfidfVectorizer* akan mempelajari semua kata unik (*vocabulary*) yang muncul di seluruh data teks. Kemudian *TfidfVectorizer* akan menghitung berapa kali setiap kata (*term*) muncul dalam dokumen tersebut. Setelah itu, *TfidfVectorizer* akan menghitung seberapa umum kata tersebut muncul di seluruh kumpulan dokumen, dimana kata yang jarang muncul di banyak dokumen akan memiliki nilai IDF yang lebih tinggi. Kemudian *TfidfVectorizer* akan menggabungkan nilai TF dan IDF untuk setiap kata dalam setiap dokumen, menghasilkan skor TF-IDF. Terakhir, *TfidfVectorizer* mengubah setiap dokumen menjadi vektor numerik yang berisi skor TF-IDF untuk semua kata dalam *vocabulary*. Dokumen yang serupa akan memiliki vektor yang lebih mirip karena kata-kata penting yang sama akan memiliki skor TF-IDF yang tinggi di kedua dokumen.

```
tfidf_vectorizer = TfidfVectorizer()  
x_train_tfidf = tfidf_vectorizer.fit_transform(x_train)  
x_test_tfidf = tfidf_vectorizer.transform(x_test)
```

Gambar 3. Implementasi Ekstraksi Fitur Term Frequency-Inverse Document Frequency

### 3.3 Pemodelan Support Vector Machine

Setelah ekstraksi fitur TF-IDF, langkah selanjutnya yaitu menerapkan metode klasifikasi SVM. Metode ini merupakan salah satu metode unggulan dalam klasifikasi teks. Sebelum melakukan pemodelan, terlebih dahulu dilakukan *splitting* data menjadi data latih dan juga data uji dengan perbandingan data uji sebesar 30%, dan data latih sebesar 70% yang dapat dilihat pada gambar 4.

```
x_train, x_test, y_train, y_test = train_test_split(df['token'], df['sentiment'], test_size=0.3, random_state=42)  
  
print(f'jumlah data latih: {len(x_train)}')  
print(f'jumlah data uji: {len(x_test)}')
```

jumlah data latih: 1029  
jumlah data uji: 442

Gambar 4. Splitting Data Latih dan Data Uji

Kemudian pemodelan data *Support Vector Machine* dilakukan dengan *library sci-kit learn*. Kali ini, fungsi kernel yang akan digunakan oleh SVM yaitu linear. Dalam hal ini, kernel linear dipilih karena kernel linear berguna untuk data yang dapat dipisahkan secara linier dalam ruang fitur. Kemudian masing-masing nilai *hyperparameter C*, *intercept*, dan juga *support vector* dicetak untuk mengontrol *trade-off* antara kerentanan model terhadap *overfitting* dan margin antara kelas serta membantu dalam pembuatan keputusan klasifikasi.

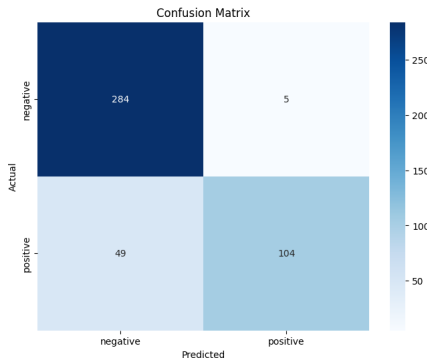
```
svm_model = SVC(kernel='linear', random_state=42)  
  
svm_model.fit(x_train_tfidf, y_train)  
  
print("Parameter model SVM:")  
print(f"Kernel: {svm_model.kernel}")  
print(f"C: {svm_model.C}")  
print(f"Intercept: {svm_model.intercept}")  
print(f"Support Vectors: {svm_model.support_vectors}")
```

Parameter model SVM:  
Kernel: linear  
C: 1.0  
Intercept: [-0.01213951]  
Support Vectors: (0, 147) 0.11840844543151888  
(0, 177) 0.2192686062728066  
(0, 202) 0.16036921698949588  
(0, 215) 0.1413647744818176  
(0, 462) 0.13541252856014313  
(0, 482) 0.2192686062728066

Gambar 5. Pelatihan Model Support Vector Machine dan Sebagian Hasil Cetaknya

Setelah melatih model Support Vector Machine, langkah selanjutnya yaitu menguji model

menggunakan data uji yang telah diproses dengan ekstraksi fitur TF-IDF sebelumnya. Setelah melakukan prediksi terhadap data uji, hasil prediksi akan dibandingkan dengan label aslinya dari data uji yang akan menghasilkan *confusion matrix*. *Confusion matrix* adalah tabel uji yang menyajikan jumlah prediksi benar dan salah yang telah dibuat oleh model klasifikasi pada setiap kelas.



**Gambar 6.** Confusion Matrix Bag-of-Words

Berdasarkan *confusion matrix* pada gambar 6, model SVM berhasil memprediksi sebanyak 284 sentimen negatif, namun terdapat 5 data bersentimen negatif yang diprediksi positif. Model ini juga berhasil memprediksi sebanyak 104 sentimen positif, namun terdapat 49 sentimen positif yang diprediksi negatif. Berdasarkan *confusion matrix* diatas didapatkan nilai *Accuracy*, *Precision*, *F1-Score* pada model Support Vector Machine sebagai berikut:

```

Classification Report:
              precision    recall  f1-score   support

 negative     0.85         0.98         0.91         289
 positive     0.95         0.68         0.79         153

 accuracy                    0.88         442
 macro avg     0.90         0.83         0.85         442
 weighted avg  0.89         0.88         0.87         442
    
```

**Gambar 7.** Hasil Evaluasi

Dari hasil akurasi yang didapatkan model Support Vector Machine yang telah dilatih dengan ekstraksi fitur TF-IDF, didapatkan nilai *accuracy* sebesar 88%, nilai *precision* 90%, nilai *recall* 83%, dan juga nilai *F1-Score* 85%

#### 4. Kesimpulan

Dari hasil penelitian ini, dapat disimpulkan bahwa model Support Vector Machine yang telah disesuaikan dengan dataset ulasan aplikasi Citilink dari Google Playstore dan ditunjang dengan ekstraksi fitur TF-IDF berhasil mengklasifikasikan sentimen ulasan dengan akurasi yang tinggi, mencapai 88%. Evaluasi lanjutan juga menunjukkan nilai *precision*, *recall*, dan *F1-Score* yang memuaskan, yaitu masing-masing sebesar 90%, 83%, dan 85%. Penelitian ini menunjukkan bahwa model Support Vector Machine dapat menjadi instrumen yang efektif dalam memahami tanggapan pengguna terhadap performa aplikasi Citilink.

#### Daftar Pustaka

- [1] D. Septiansari and T. Handayani, "Pengaruh Belanja Online Terhadap Perilaku Konsumtif pada Mahasiswa di Masa Pandemi Covid-19," *Jurnal Ekonomi dan Manajemen Teknologi*, vol. 5, no. 1, pp. 53–65, 2021, doi: 10.35870/emt.v5i1.372
- [2] A. Trisriatri, A. H. Brata, and L. Fanani, "Perbandingan User Interface Aplikasi Mobile Pemesanan Tiket Pesawat Online dengan Design Thinking," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 6, pp. 2113-2120, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1509/550>
- [3] M. Yunus, "TF-IDF (Term Frequency-Inverse Document Frequency) : Representasi Vector Data Text," Medium, 30 April 2020, [Online]. Tersedia: <https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>
- [4] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *ILKOM Jurnal Ilmiah*, vol. 13, no. 2, pp. 101–109, 2021, doi: 10.33096/ilkom.v13i2.721.101-109
- [5] G. Patil, V. Galande, V. Kekan et al., "Sentiment Analysis Using Support Vector Machine," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 1, pp. 2607-2612, 2014, [Online]. Available: <https://www.semanticscholar.org/paper/Sentiment-Analysis-Using-Support-VectorMachine-Patil-Gal/759f15464a7ad372cba9c38a0f8c5caff6d85cf1>
- [6] Styawati, A. R. Isnain, N. Hendrastuty et al., "Comparison of Support Vector Machine and Naive Bayes on Twitter Data Sentiment Analysis," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 1, pp. 56-60, 2021, doi: 10.30591/jpit.v6i1.3245
- [7] A. Mulyani, D. Kurniadi, M. R. Nashrulloh et al., "The Prediction Of Ppa And Kip-Kuliah Scholarship Recipients Using Naive Bayes Algorithm," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 4, pp. 821-827, 2022, doi: 10.20884/1.jutif.2022.3.4.297

Halaman ini sengaja dibiarkan kosong

# Implementasi Ekstraksi Fitur VGG-16 dan Pemodelan LSTM untuk Pembangkitan Caption Gambar Otomatis

Made Pranajaya Dibyacita<sup>a1</sup>, Luh Gede Astuti<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia  
<sup>1</sup>mdpranjaya@gmail.com  
<sup>2</sup>lg.astuti@unud.ac.id

## Abstract

*Image captioning, the task of automatically generating descriptive captions for images, has gained significant attention due to its potential applications in various domains. This paper addresses the challenges associated with integrating computer vision and natural language processing techniques to develop an effective image caption generator. The proposed solution leverages the VGG-16 model for feature extraction from images and an LSTM (Long Short-Term Memory) model for caption generation. The Flickr8k dataset, containing approximately 8000 images with five different captions per image, is utilized for training and evaluation. The methodology encompasses several steps, including data preprocessing, feature extraction, model training, and evaluation. Data preprocessing involves cleaning captions by removing punctuations, single characters, and numerical values, while incorporating start and end sequences. Image features are extracted using the pre-trained VGG-16 model, and similar images are clustered to ensure accurate feature extraction. Subsequently, the captions and corresponding image features are merged and tokenized for model training. The LSTM model is designed with input layers for image features and captions, as well as an output layer for caption generation. Extensive hyperparameter tuning is conducted to optimize the model's performance, involving variations in the number of nodes and layers. The generated captions are evaluated using BLEU scores, where a score closer to 1 indicates higher similarity between predicted and actual captions. The proposed system demonstrates promising results in generating meaningful captions for images, with potential applications in assisting visually impaired individuals, medical image analysis, and advertising industry automation.*

**Keywords:** Image Captioning, Deep Learning, VGG-16, LSTM, NLP, BLEU.

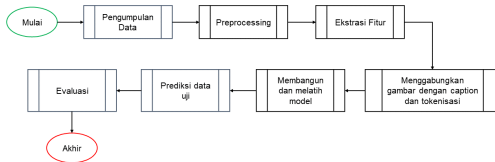
## 1. Pendahuluan

Dalam era digital yang semakin maju, gambar dan foto telah menjadi bagian integral dari kehidupan kita sehari-hari. Tak hanya sebagai media pengalaman visual semata, gambar juga menjadi alat komunikasi yang penting, yang membawa tantangan baru dalam hal pengelolaan dan pemahaman konten visual. Mengatasi tantangan ini memerlukan pendekatan yang inovatif, salah satunya adalah dengan menggabungkan teknologi *computer vision* dan pemrosesan bahasa alami. Dengan demikian, muncul bidang *Image Caption Generation*, yang menggabungkan kedua teknologi tersebut untuk menghasilkan deskripsi otomatis yang relevan dan bermakna untuk gambar. Bidang *Image Caption Generation* merupakan hasil dari perpaduan antara *computer vision* dan pemrosesan bahasa alami. Tujuannya adalah memanfaatkan model *deep learning* untuk secara otomatis menghasilkan deskripsi bahasa yang tepat dan koheren terhadap informasi visual. Keberadaan teknologi ini memiliki makna yang sangat penting, terutama dalam memberikan bantuan bagi orang awam terhadap dunia medis untuk menganalisis gambar hasil rontgen, dan membantu orang dengan gangguan penglihatan, memungkinkan mereka memahami konten gambar melalui deskripsi teks [1]. *Computer vision, machine translation, dan object detection* adalah domain yang dinamis dalam dunia *machine learning*, mengalami pertumbuhan yang signifikan dalam dekade terakhir [2]. Pertumbuhan ini telah menghasilkan berbagai kerangka kerja (framework) yang memfasilitasi implementasi

*caption generation*. Metode tradisional sebagian besar mengandalkan *feature extractor* yang dirancang secara manual dan model bahasa berbasis aturan, tetapi efeknya terbatas [1]. Dalam konteks ini, penggunaan model VGG-16 dan LSTM (*Long Short-Term Memory*) telah menjadi pilihan yang populer dalam mengatasi tantangan *Image Caption Generation*. Model VGG-16 digunakan untuk ekstraksi fitur dari gambar, sementara LSTM digunakan untuk menghasilkan deskripsi bahasa yang relevan. Pendekatan ini, di mana model VGG-16 digunakan untuk ekstraksi fitur dari gambar dan LSTM untuk menghasilkan deskripsi bahasa yang relevan, telah membuka peluang baru dalam menangani kompleksitas visual dan linguistik. Tiga langkah utama dalam menciptakan model *image captioning* melibatkan ekstraksi fitur dari gambar dan caption untuk mendukung model, menggunakan fitur-fitur ini untuk melatih model, dan menggunakan model yang telah dilatih untuk menghasilkan caption berdasarkan atribut gambar input. Hasil dari penggunaan teknologi ini dievaluasi menggunakan skor *Bilingual Evaluation Understudy* (BLEU) sebagai tolak ukur untuk membandingkan dan menilai efektivitas model. Penelitian menunjukkan peningkatan yang signifikan dalam akurasi skor BLEU dibandingkan dengan pendekatan dasar, menegaskan keberhasilan dan relevansi pendekatan yang diusulkan. Selain itu, model yang dikembangkan juga memiliki kemampuan unik untuk menciptakan caption gambar secara otomatis, yang membedakannya dari metode dasar yang hanya meminjam caption yang sudah ada. Dengan demikian, teknologi *Image Caption Generation* terbukti menjadi alat yang berharga dalam memahami dan mengelola konten visual di era digital saat ini.

## 2. Metode Penelitian

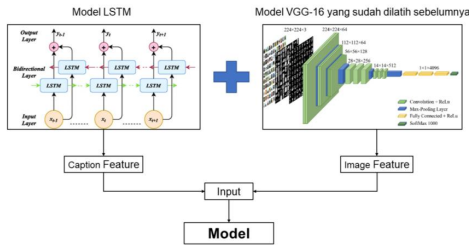
Penelitian ini bertujuan untuk mengeksplorasi integrasi antara ekstraksi fitur menggunakan model VGG-16 dan pemodelan LSTM dalam konteks pembangkitan caption gambar otomatis, dengan tools yang digunakan yaitu *Jupyter Notebook* dengan bahasa python. Langkah-langkah yang diterapkan dalam penelitian ini meliputi pengumpulan data, preprocessing, ekstraksi fitur menggunakan VGG-16, menggabungkan gambar dengan caption dan tokenisasi, membangun dan melatih model, prediksi data uji, dan evaluasi yang akan diuraikan sebagai berikut:



Gambar 1. Diagram Alur

Dalam pendekatan tersebut, penelitian ini juga melibatkan pelatihan model menggunakan dataset gambar Flickr8K [3]. Untuk prediksi, model ini menggunakan arsitektur jaringan saraf *Long Short-Term Memory* (LSTM), bersama dengan kombinasi caption dataset Flickr8K dan atribut gambar yang diekstraksi menggunakan VGG. Aspek inti dari proses prediksi gambar melibatkan ekstraksi fitur VGG dan penggunaan model LSTM yang telah dilatih untuk pembangkitan caption seperti yang dapat dilihat pada Gambar 2.





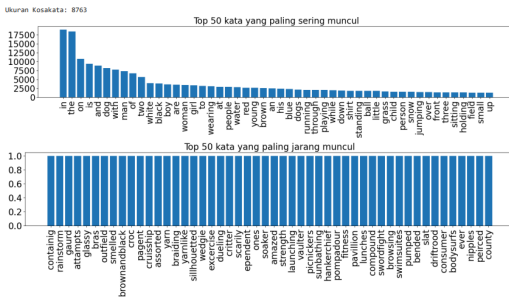
Gambar 2. Kerangka Model

### 2.1 Pengumpulan Data

Dataset yang digunakan adalah Flickr8k, terdiri dari 8.092 gambar dengan hingga lima keterangan deskriptif [4]. Data ini berasal dari layanan berbagi foto Flickr dengan lisensi yang sesuai. Setiap gambar dilengkapi dengan caption dalam bahasa Inggris, mendukung penelitian dalam pemrosesan bahasa alami dan visi komputer, termasuk *image captioning*, pengenalan objek, dan pemahaman konten gambar. Dataset ini digunakan sebagai sumber daya penting bagi pemula dalam machine learning dan AI, memberikan titik acuan dasar untuk proyek deskripsi gambar berbasis kalimat [5]. Selain itu, model juga menggunakan data dari VGG-16 yang telah dilatih sebelumnya untuk ekstraksi fitur gambar yang dapat diakses [disini](#). Untuk pembagian data untuk latihan, uji, dan validasi adalah 60%, 20%, dan 20% dari total dataset secara berturut-turut.

### 2.2 Tahap Preprocessing Data

Langkah preprocessing data melibatkan pembersihan caption untuk menghilangkan ekspresi reguler, angka, dan kata-kata tidak relevan (*stop words*). Ini termasuk penghapusan tanda baca, karakter tunggal, dan nilai numerik. Setelah pembersihan, dilakukan analisis untuk menemukan kata-kata paling umum dan paling jarang dalam dataset. Data caption yang telah dibersihkan diperbarui dalam DataFrame dengan nama file gambar yang terkait. Selanjutnya, peneliti melakukan perhitungan frekuensi kemunculan kata-kata dan memvisualisasi 50 kata teratas dan 50 kata terendah.



Gambar 3. Top 50 kata paling sering dan paling jarang muncul

### 2.3 Tahap Ekstraksi Fitur

Pada penelitian, dilakukan ekstraksi fitur menggunakan VGG-16 yang telah dilatih sebelumnya pada dataset besar yang berisi gambar-gambar bervariasi. Dalam konteks pembangkitan caption gambar, model ini digunakan untuk mengekstraksi fitur-fitur penting dari gambar yang akan digunakan sebagai input untuk model pembangkit caption [6]. Dalam penelitian ini, langkah pertama yang dilakukan adalah memuat model beserta bobot yang telah dilatih sebelumnya. Setelah itu, lapisan terakhir dari model yang digunakan untuk klasifikasi akan dihapus, sehingga model hanya akan menghasilkan vektor fitur dari gambar yang diberikan sebagai input. Dalam proses ini, VGG-16 akan menghasilkan vektor fitur dengan dimensi 4096 untuk setiap gambar yang diproses.

Model: "functional\_1"

| Layer (type)               | Output Shape          | Param #     |
|----------------------------|-----------------------|-------------|
| input_layer (InputLayer)   | (None, 224, 224, 3)   | 0           |
| block1_conv1 (Conv2D)      | (None, 224, 224, 64)  | 1,792       |
| block1_conv2 (Conv2D)      | (None, 224, 224, 64)  | 36,028      |
| block1_pool (MaxPooling2D) | (None, 112, 112, 64)  | 0           |
| block2_conv1 (Conv2D)      | (None, 112, 112, 128) | 73,856      |
| block2_conv2 (Conv2D)      | (None, 112, 112, 128) | 147,664     |
| block2_pool (MaxPooling2D) | (None, 56, 56, 128)   | 0           |
| block3_conv1 (Conv2D)      | (None, 56, 56, 256)   | 285,168     |
| block3_conv2 (Conv2D)      | (None, 56, 56, 256)   | 580,080     |
| block3_conv3 (Conv2D)      | (None, 56, 56, 256)   | 580,080     |
| block3_pool (MaxPooling2D) | (None, 28, 28, 256)   | 0           |
| block4_conv1 (Conv2D)      | (None, 28, 28, 512)   | 1,180,160   |
| block4_conv2 (Conv2D)      | (None, 28, 28, 512)   | 2,359,808   |
| block4_conv3 (Conv2D)      | (None, 28, 28, 512)   | 2,359,808   |
| block4_pool (MaxPooling2D) | (None, 14, 14, 512)   | 0           |
| block5_conv1 (Conv2D)      | (None, 14, 14, 512)   | 2,359,808   |
| block5_conv2 (Conv2D)      | (None, 14, 14, 512)   | 2,359,808   |
| block5_conv3 (Conv2D)      | (None, 14, 14, 512)   | 2,359,808   |
| block5_pool (MaxPooling2D) | (None, 7, 7, 512)     | 0           |
| flatten (Flatten)          | (None, 25088)         | 0           |
| fc1 (Dense)                | (None, 4096)          | 102,764,544 |
| fc2 (Dense)                | (None, 4096)          | 16,781,512  |

Total params: 194,200,544 (512.16 MB)  
 Trainable params: 134,200,544 (512.16 MB)  
 Non-trainable params: 0 (0.00 B)

Gambar 4. Menghapus Lapisan Terakhir Model

Sebelum gambar dapat diproses oleh VGG-16, gambar harus diubah ke dalam format yang sesuai dengan input yang diharapkan oleh model. Dalam kasus ini, gambar harus diubah ukurannya menjadi 224x224 piksel dan dinormalisasi agar nilai pikselnya berada dalam rentang 0 hingga 1. Proses ini dilakukan dengan menggunakan fungsi-fungsi dari library Keras seperti `load_img`, `img_to_array`, dan `preprocess_input`.

```

In [ ]:
from keras.preprocessing.image import load_img, img_to_array
from keras.applications.vgg16 import preprocess_input
from collections import OrderedDict

images = OrderedDict()
npix = 224 #ukuran gambar ditetapkan pada 224 karena model VGG16 telah dilatih sebelumnya dengan ukuran tersebut.
target_size = (npix,npix,3)
data = np.zeros((len(imgs),npix,npix,3))
for i,name in enumerate(imgs):
    # load gambar dari file
    filename = dir_flickr.jpg + '/' + name
    image = load_img(filename, target_size=target_size)
    # ubah piksel gambar menjadi larik numpy
    image = img_to_array(image)
    #image = preprocess_input(image)

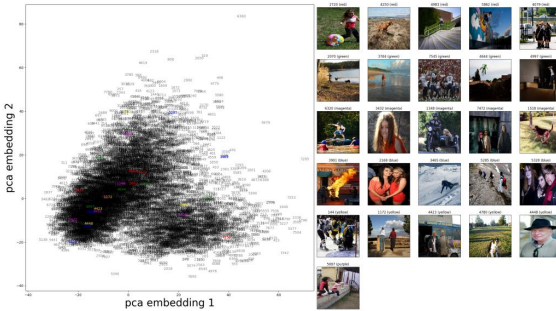
    y_pred = modelvgg.predict(image.reshape((-1,) + image.shape[3]))
    images[name] = y_pred.flatten()

1/1 ----- 1s 911ms/step
1/1 ----- 0s 435ms/step
1/1 ----- 0s 317ms/step
1/1 ----- 0s 323ms/step
    
```

Gambar 5. Ekstraksi Fitur

Setelah fitur diekstraksi menggunakan model VGG-16, dilakukan visualisasi untuk mengelompokkan gambar dengan karakteristik visual serupa menggunakan teknik *Principal Component Analysis* (PCA). PCA digunakan untuk mengurangi dimensi fitur dari 4096 menjadi 2, memungkinkan visualisasi dalam ruang dua dimensi. Cluster gambar dibentuk berdasarkan kesamaan fitur visual, dan beberapa contoh gambar diambil dari setiap kelompok untuk ditampilkan dalam plot. Proses ini penting untuk memverifikasi keberhasilan ekstraksi fitur dan

memahami pola dalam dataset gambar.



**Gambar 6.** Memplot Gambar-gambar yang Mirip dari Dataset

#### 2.4 Tahap Menggabungkan Gambar dengan Caption dan Tokenisasi

Penelitian ini, melakukan penggabungan gambar yang dimana proses ini dimulai dengan pemilihan caption pertama dari setiap gambar dalam dataset. Dalam konteks ini, pemilihan hanya dilakukan terhadap caption pertama untuk menghindari kompleksitas yang mungkin timbul dari penggunaan seluruh caption. Selanjutnya, caption dan gambar yang terkait digabungkan bersama untuk membentuk pasangan data yang akan digunakan dalam pelatihan.

Out[34]:

|    | namafile                  | index | caption  |
|----|---------------------------|-------|--|
| 0  | 1000268201_693b08cb0e.jpg | 0     | startseq child in pink dress is climbing up s... |
| 5  | 1001773457_577c3a7d70.jpg | 0     | startseq black dog and spotted dog are fighti... |
| 10 | 1002674143_1b742ab4b8.jpg | 0     | startseq little girl covered in paint sits in... |
| 15 | 1003163366_44323f5815.jpg | 0     | startseq man lays on bench while his dog sits... |
| 20 | 1007129816_e794419615.jpg | 0     | startseq man in an orange hat starring at som... |

**Gambar 7.** Menggabungkan Gambar dengan Caption

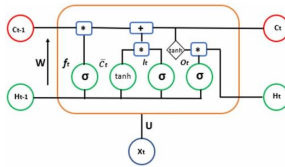
Setelah proses penggabungan, selanjutnya dilakukan tokenisasi caption. Tokenisasi ini diperlukan karena model yang akan digunakan tidak dapat menerima teks mentah sebagai input, melainkan memerlukan representasi vektor kata-kata. Tokenisasi dilakukan menggunakan Tokenizer dari TensorFlow, di mana setiap kata dalam caption dikonversi menjadi urutan angka yang merepresentasikan kata-kata dalam kamus. Pada tahap ini, juga ditetapkan jumlah kata maksimum dalam kamus, yang dalam kasus ini disetel sebesar 6000 kata.

```
ukuran kosakata : 4476
[[1, 38, 3, 66, 144, 7, 124, 52, 486, 9, 367, 3, 24, 2351, 522, 2], [1, 12, 8, 5, 752, 8, 17, 368, 2], [1, 48, 15, 178, 3, 584, 181, 3, 41, 9, 551, 1198, 11, 55, 213, 3, 1876, 2], [1, 18, 821, 6, 158, 27, 23, 8, 181, 48, 112, 2], [1, 18, 3, 24, 82, 96, 1199, 19, 162, 2]]
```

**Gambar 8.** Tokenisasi Caption

## 2.5 Tahap Membangun dan Melatih Model

Penelitian ini, akan membangun model LSTM yang merupakan tahap krusial dalam proses pembangkitan caption untuk gambar. Model LSTM dipilih karena mampu memperhitungkan keadaan keluaran sel sebelumnya dan masukan sel saat ini untuk menghasilkan keluaran saat ini [7]. Hal ini sangat berguna saat menghasilkan caption untuk gambar-gambar.



Gambar 9. LSTM Network

- \* = Elementwise multiplicant (perkalian elemen demi elemen)
- + = Element-wise addition (penambahan elemen demi elemen)

$$f_t = \sigma (X_t * U_f + H_{t-1} * W_f)$$

$$c_t = \tanh (X_t * U_c + H_{t-1} * W_c)$$

$$i_t = \sigma (X_t * U_i + H_{t-1} * W_i)$$

$$o_t = \sigma (X_t * U_o + H_{t-1} * W_o)$$

$$C_t = f_t * C_{t-1} + i_t * c_t$$

$$H_t = o_t * \tanh (C_t)$$

Keterangan:

- $X_t$  = Input Vector
- $H_{t-1}$  = Sel Output Sebelumnya
- $C_{t-1}$  = Memori Output Sebelumnya
- $H_t$  = Sel Output Saat Ini
- $C_t$  = Sel Memori Saat Ini
- $W, U$  = Weight vector untuk gerbang forget (f), candidate (c), gerbang input (i), gerbang output (o)

Pada model LSTM ini, peneliti akan membangun lapisan input dan output untuk menghasilkan caption dengan variasi jumlah simpul dan lapisan, mulai dari 256 hingga 1024, dengan penyetaan *hyperparameter* yang cermat. Proses ini melibatkan penggunaan fitur gambar dan teks yang telah di-tokenisasi sebelumnya, dengan teks melewati lapisan *embedding* dan LSTM untuk mengekstraksi pola urutan. Dua lapisan LSTM digunakan dengan *dropout* untuk mencegah *overfitting*, diikuti dengan penggabungan hasilnya menggunakan operasi penambahan. Output dari model adalah distribusi probabilitas kata-kata dalam kamus, dikompilasi dengan fungsi *loss categorical\_crossentropy* dan *optimizer adam*.

| Model                       | functional_9  |         |                      |  |
|-----------------------------|---------------|---------|----------------------|--|
| input_layer_3 (Concatenate) | None, 256     | 0       | -                    |  |
| embedding_3 (Embedding)     | None, 30, 640 | 300,400 | input_layer_3[0]...  |  |
| max_pooling_3 (MaxPooling)  | None, 300     | 0       | input_layer_3[0]...  |  |
| FiturCaption (LSTM)         | None, 30, 200 | 300,700 | embedding_3[0][0]... |  |
| propool_3 (Propool)         | None, 30, 200 | 0       | FiturCaption[0]...   |  |
| input_layer_7 (Concatenate) | None, 4096    | 0       | -                    |  |
| FiturCaption2 (LSTM)        | None, 200     | 500,300 | propool_3[0][0]...   |  |

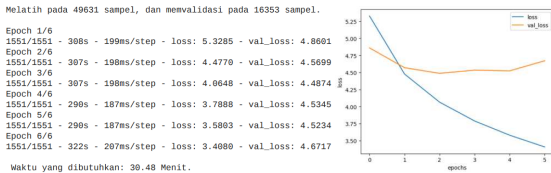
  

|                     |              |           |                     |
|---------------------|--------------|-----------|---------------------|
| FiturGambar (Dense) | (None, 256)  | 1,048,832 | input_layer_7[0]... |
| add_3 (Add)         | (None, 256)  | 0         | FiturCaption2[0]... |
| dense_6 (Dense)     | (None, 256)  | 65,792    | add_3[0][0]         |
| dense_7 (Dense)     | (None, 4476) | 1,150,332 | dense_6[0][0]       |

Total params: 3,405,436 (12.99 MB)  
 Trainable params: 3,405,436 (12.99 MB)  
 Non-trainable params: 0 (0.00 B)

Gambar 10. Membangun Model LSTM

Setelah model dibangun, pelatihan dilakukan menggunakan metode *fit()* dari Keras dengan data latih dan validasi. Selama pelatihan, evaluasi kerugian (*loss*) dipantau untuk mengukur kinerja model. Proses pelatihan dilakukan hingga model mencapai kinerja yang diharapkan atau tidak lagi meningkat secara signifikan. Grafik *loss* dan *val\_loss* digunakan untuk memvisualisasikan kinerja model selama pelatihan, menunjukkan perubahan *loss* pada data latih dan validasi seiring dengan jumlah *epochs*. Perubahan ini membantu memahami apakah model cenderung *underfitting* atau *overfitting*.








Gambar 11. Melatih Model LSTM dan Menampilkan Grafik Loss dan Val\_loss

## 2.6 Tahap Prediksi Data Uji

Penelitian ini, akan menguji pada dataset uji untuk melihat bagaimana kinerjanya dalam menghasilkan caption untuk beberapa gambar. Jika caption yang dihasilkan sudah dapat diterima, langkah selanjutnya adalah menghasilkan caption untuk seluruh dataset uji. Proses ini melibatkan perbandingan antara caption yang dihasilkan oleh model dengan caption sebenarnya dari dataset.

Tabel 1. Hasil Prediksi Data Uji

| Gambar  | Prediksi  |
|---|---|
|  | startseq man in blue shirt is standing on the street endseq |
|  | startseq black and white dog is running in the grass endseq |

| Gambar  | Prediksi  |
|---|---|
|  | startseq black and white dog is running through the snow endseq |
|  | startseq man in blue shirt is riding on the air endseq          |
|  | startseq boy is riding up on the ocean endseq                   |

## 2.7 Tahap Evaluasi

Setelah model dilatih, penting untuk menguji kemampuan prediksi pada dataset uji. Dalam konteks evaluasi teks, metrik tradisional seperti akurasi tidak relevan. Sebagai gantinya, peneliti menggunakan skor BLEU (*Bilingual Evaluation Understudy*), sebuah metrik untuk membandingkan teks kandidat dengan satu atau lebih teks referensi. Sebagai contoh, pada dua hipotesis yang berbeda, peneliti menggunakan skor BLEU untuk mengevaluasi kesamaan antara teks prediksi dan teks referensi. Misalnya, skor BLEU untuk hipotesis pertama adalah 0.603 dan untuk hipotesis kedua adalah 0.544. Berikut adalah rumus yang dapat digunakan untuk menghitung skor BLEU [8]:

$$BLEU = BP \times \exp \left( \sum_{n=1}^N w_n \times \log(P_n) \right)$$

Keterangan:

$BP$  = Faktor penyusutan *Brevity Penalty* yang mengkompensasi kecenderungan sistem untuk menghasilkan teks yang terlalu pendek.

$N$  = Urutan maksimum  $n$ -gram yang dievaluasi (biasanya 4).

$w_n$  = Bobot yang diberikan pada setiap urutan  $n$ -gram.

$P_n$  = Presisi urutan  $n$ -gram, yaitu rasio jumlah urutan  $n$ -gram yang cocok dalam hasil sistem terhadap jumlah urutan  $n$ -gram dalam teks referensi.

**Tabel 2.** Hasil Evaluasi

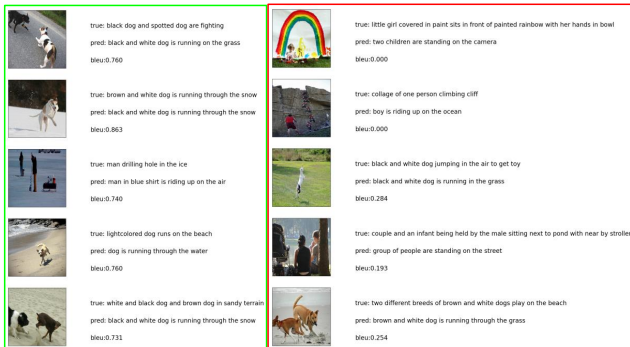
| Hipotesis   | Referensi     | Skor BLEU |
|-------------|---------------|-----------|
| I like dog  | I do like dog | 0.603     |
| I love dog! | I do like dog | 0.544     |

Metrik BLEU digunakan sebagai alat untuk mengevaluasi sejauh mana terjemahan yang dihasilkan oleh model (*hypothesis*) cocok dengan terjemahan referensi (*reference*). Skor BLEU berkisar dari 0 hingga 1, dengan nilai yang lebih tinggi menunjukkan kesamaan yang lebih baik antara terjemahan hypothesis dan referensi. Dalam contoh yang diberikan, hipotesis pertama "I like dog" memiliki skor BLEU sebesar 0.603 ketika dibandingkan dengan referensi "I do like dog". Sementara itu, hipotesis kedua "I love dog!" mendapat skor BLEU sebesar 0.544 terhadap referensi yang sama. Dari skor BLEU ini, dapat disimpulkan bahwa hipotesis pertama lebih mirip

dengan referensi daripada hipotesis kedua.

### 3. Hasil dan Pembahasan

Setelah semua dilakukan semua proses pada tahap sebelumnya, penelitian ini akan menghasilkan caption untuk seluruh data uji dan mengevaluasinya menggunakan skor BLEU. Dalam proses ini, setiap gambar dari data uji diproses secara individual. Untuk setiap gambar, caption yang dihasilkan oleh model dibandingkan dengan caption referensi yang sebenarnya menggunakan metrik BLEU. Skor BLEU yang dihasilkan memberikan gambaran tentang seberapa baik caption yang dihasilkan oleh model cocok dengan caption referensi. Peneliti melakukan iterasi melalui setiap gambar dari data uji. Untuk setiap gambar, caption yang dihasilkan oleh model dibandingkan dengan caption referensi yang sebenarnya. Jika skor BLEU lebih tinggi dari 0.7, caption tersebut dianggap baik dan dimasukkan ke dalam kategori "caption baik". Sebaliknya, jika skor BLEU kurang dari 0.3, caption tersebut dianggap buruk dan dimasukkan ke dalam kategori "caption buruk". Contoh-caption baik dan buruk juga diberikan untuk memberikan pemahaman visual tentang kualitas caption yang dihasilkan oleh model. Dengan melihat contoh-caption baik dan buruk, peneliti dapat memahami secara lebih konkret tentang kekuatan dan kelemahan model dalam menghasilkan caption untuk gambar-gambar tertentu.



Gambar 12. Pembangkitan Caption Baik dan Buruk

Keterangan:

- Hijau = Hasil pembangkitan caption yang baik.
- Merah = Hasil Pembangkitan caption yang buruk.

Peneliti juga menghitung rata-rata skor BLEU dari semua caption setelah semua tahapan selesai dilakukan. Dalam kasus ini, rata-rata skor BLEU adalah 0.415. Ini mengindikasikan bahwa secara keseluruhan, model memiliki tingkat kesesuaian yang cukup baik dengan caption referensi. Untuk menghitung rata-rata BLEU dari sejumlah teks yang dievaluasi, dapat digunakan rumus berikut [8]:

$$\text{Rata - rata BLEU} = \frac{\sum_{i=1}^N \text{BLEU}_i}{N}$$

Keterangan:

$BLEU =$  Skor BLEU untuk teks ke- $i$ .  
 $N =$  Jumlah teks yang dievaluasi.

```
In [71]: print("Rata-rata BLEU {:.4f}".format(np.mean(bleus)))  
  
Rata-rata BLEU 0.415
```

**Gambar 13.** Rata-rata Skor BLEU dari Semua Caption

### 3.1 Dokumentasi

Hasil penelitian tertuang dalam bentuk file PDF yang merupakan konversi dari *Jupyter Notebook* dengan bahasa python yang dapat diakses [disini](#). Dokumen ni mencakup langkah-langkah eksperimen, analisis, dan temuan utama dalam pengembangan model secara lengkap. Penelitian ini diharapkan dapat menjadi sumber wawasan bagi peneliti lain dalam bidang Pembangkitan Caption Gambar Otomatis, memberikan inspirasi untuk eksplorasi lebih lanjut, serta mendorong pengembangan model yang lebih canggih dan akurat.

## 4. Kesimpulan

Berdasarkan paparan penelitian yang telah dilakukan sebelumnya, penelitian ini telah berhasil menerapkan model LSTM pada dataset Flickr8k untuk menghasilkan caption gambar yang memuaskan, meskipun masih terdapat ruang untuk peningkatan. Evaluasi menggunakan skor BLEU menunjukkan bahwa model mencapai rata-rata skor BLEU sebesar 0.415 atau 41.5% untuk seluruh dataset uji yang digunakan untuk mengevaluasi kinerja model. Untuk pengembangan lebih lanjut, disarankan untuk mengeksplorasi arsitektur model yang lebih mutakhir, seperti CNN, RNN, dll. Selain itu, penerapan teknik data augmentasi dapat membantu meningkatkan kinerja model. Integrasi informasi kontekstual tambahan, seperti objek dan aktivitas dalam gambar, juga dapat diperluas untuk meningkatkan keakuratan caption. Dengan penelitian dan pengembangan lebih lanjut, diharapkan dapat tercipta model yang lebih unggul dalam menghasilkan caption gambar yang informatif dan bermanfaat dalam berbagai aplikasi praktis.

### Daftar Pustaka

- [1] H. Wang, Y. Zhang, and X. Yu, "An overview of image caption generation methods," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–13, Jan. 2020. doi:10.1155/2020/3062706.
- [2] A. R. GRIGOREV, *Tensorflow Deep Learning Projects: 10 Real-World Projects on Computer Vision, Machine Translation, Chatbots, and Reinforcement Learning*; 10 Real-World. PACKT Publishing, 2018.
- [3] K. Anitha Kumari, C. Mouneeshwari, R. B. Udhaya, and R. Jasmitha, "Automated image captioning for Flickr8k dataset," *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*, pp. 679–687, 2020. doi:10.1007/978-3-030-24051-6\_62.
- [4] adityajn105, "Flickr 8K dataset," *Kaggle*, 27-Apr-2020. [Online]. Available: <https://www.kaggle.com/datasets/adityajn105/flickr8k>. [Accessed: 01-May-2024].
- [5] B. Jawade, D. D. Mohan, N. M. Ali, S. Setlur, and V. Govindaraju, "NAPReg: Nouns as proxies' regularization for semantically aware cross-modal embeddings," *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023. doi:10.1109/wacv56688.2023.00119.
- [6] T. V. Sneha and Dr. S. J. Rani, "LSTM-VGG-16: A Novel and Modular Model for Image Captioning Using Deep Learning Approaches on Flickr8k dataset," vol. 12, no. 11, pp. 131–141, 2021.
- [7] C. Wang, H. Yang, C. Bartz, and C. Meinel, "Image captioning with deep bidirectional lstms," *Proceedings of the 24th ACM international conference on Multimedia*, Oct. 2016.



- doi:10.1145/2964284.2964299.
- [8] GeeksforGeeks, "NLP - Bleu score for Evaluating Neural Machine Translation - Python," *GeeksforGeeks*, 08-Mar-2024. [Online]. Available: <https://www.geeksforgeeks.org/nlp-bleu-score-for-evaluating-neural-machine-translation-python/>. [Accessed: 01-May-2024].

Halaman ini sengaja dibiarkan kosong

## Implementasi Internet of Things dengan Smart Faucet pada Sistem Irigasi Subak Bali

Ida Bagus Rahadi Putra<sup>a1</sup>, Ngurah Agus Sanjaya ER<sup>a2</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>idabaguserahadi@unud.ac.id  
<sup>2</sup>agus\_sanjaya@unud.ac.id

### Abstract

*This research proposes the implementation of the internet of things with smart faucets in Bali's Subak irrigation system by using algorithms used to implement the internet of things. Sensors including temperature sensors, soil moisture sensors, rainfall sensors, and water level sensors, are installed to monitor real-time environmental and crop conditions. The data collected from these sensors is used in algorithms that are used to optimize water usage according to crop needs and environmental conditions. The results of this study show that there is a reduction in water wastage and an increase in water use efficiency with more accurate control. The optimized irrigation system ensures a water supply that matches the needs of the plants, improves plant growth, and reduces the risk of disease. Users can monitor and control the irrigation system remotely through the website, providing accurate information about the condition of the plants and their water needs. It is an innovative solution to improve crop yields, plant health, and water use efficiency in Bali's Subak irrigation systems.*

**Keywords:** Irrigation, Subak, Bali, Smart Faucet, Internet of Things (IoT)

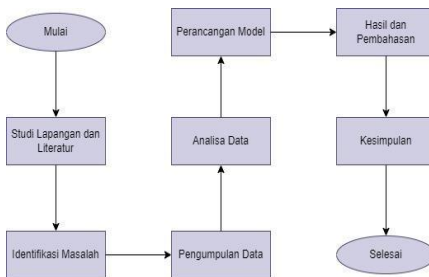
### 1. Pendahuluan

Subak merupakan sistem irigasi tradisional yang unik dan terkenal di Bali, Indonesia. Sistem ini telah menjadi inti dari keberlanjutan pertanian padi di pulau tersebut selama berabad – abad. Sistem Subak ini menerapkan prinsip bahwa sawah – sawah dikelola secara kolektif oleh masyarakat setempat, yang dikenal sebagai petani Subak. Petani Subak ini bekerja sama untuk mengatur aliran air dari mata air gunung melalui jaringan saluran irigasi yang rumit. Fokus utama Subak adalah menjaga keunikan dan khasnya dalam pendistribusian air irigasi kepada anggotanya secara adil, efektif, dan konflik-minimal, dengan menerapkan dua jenis sistem pembagian air, yaitu sistem tektek dan bumbung [1]. Subak tidak hanya berperan sebagai penyedia air untuk tanaman padi, tetapi juga sebagai representasi budaya, kearifan lokal, dan hubungan harmonis antara manusia dan alam. Sistem ini juga memainkan peran penting dalam pelestarian lingkungan dan keberlanjutan ekosistem pertanian di Bali. Berdasarkan hal tersebut, Subak bukan hanya sebuah sistem irigasi, tetapi juga sebuah warisan budaya yang patut dijaga dan dipelihara. Permasalahan kompleks yang dihadapi oleh sistem irigasi Subak saat ini mencakup peningkatan jumlah sawah yang beralih fungsi menjadi wilayah terbangun, serta meningkatnya permintaan akan sumber air bersih. Ironisnya, selama satu dekade terakhir, terjadi penurunan yang cukup berarti dalam luas lahan subak karena banyak lahan sawah yang telah beralih fungsi menjadi penggunaan non-pertanian, dengan tingkat pengurangan mencapai 1000 hektar setiap tahun [2]. Perubahan atau transformasi ini dapat memberikan manfaat ekonomi dengan percepatan pembangunan dan peningkatan kualitas ekonomi, namun juga mengancam keberlangsungan sistem irigasi Subak. Faktor internal seperti keterbatasan kemampuan petani dalam mengatur penggunaan air secara optimal ditambah dengan praktek pengelolaan tradisional Subak, semakin memperumit situasi tersebut. Teknologi informasi dan komunikasi, khususnya Internet of Things (IoT), telah menjadi bagian integral dalam perkembangan pertanian modern. Penerapan teknologi ini memudahkan pengelolaan lahan pertanian dan merupakan aspek penting dalam pengembangan sektor pangan saat ini. Teknologi

IoT memungkinkan penghubungan antara berbagai perangkat melalui koneksi internet guna menjalankan aktivitas pertanian [3]. Fuzzy Logic telah digunakan secara luas di berbagai bidang, termasuk sebagai cara untuk merepresentasikan pemikiran manusia dalam suatu sistem. Hal ini merupakan metode yang efektif untuk memetakan masukan ke dalam keluaran. Fuzzy Logic merupakan ekstensi dari teori himpunan klasik dan menghasilkan output yang dapat bervariasi tergantung pada inputnya. Secara umum, fuzzy logic melibatkan input, proses, dan output, serta dirancang untuk menangani nilai-nilai yang berada di antara kebenaran dan kesalahan. Berbeda dengan himpunan yang bersifat tegas dengan nilai 1 atau 0 (ya atau tidak), Logika Fuzzy adalah jenis logika yang memperhitungkan tingkat kekaburan atau kesamaran antara konsep benar dan salah [4]. Berdasarkan hal tersebut, menggunakan fuzzy logic, keluarannya dapat berupa seluruh rentang antara 0 dan 1, bukan hanya "ya" atau "tidak". Pada penelitian ini akan menggunakan smart faucet dan metode algoritma yang dipergunakan untuk mengimplementasikan internet of things (IoT). Sebelumnya pada tahun 2021 sudah ada penelitian serupa yang diteliti oleh Erlangga Bayu Linggar Sukarta, I Made Sukarsa, dan I Nyoman Piarsa dengan jurnal berjudul "Smart Automatic Water Filler (SAWF) berbasis Internet of Things" dengan memiliki tujuan agar pengguna dapat melakukan kontrol serta memonitor keran air pengisian dan pembuangan melalui bot aplikasi telegram, berbasis Internet of Things (IoT) dengan menggunakan mikrokontroler Arduino NodeMcu [5]. Penelitian dilakukan dengan menggunakan pengujian RMSE yang menyatakan bahwa sensor berfungsi baik dan tidak melenceng jauh dari nilai aslinya [5]. Hasil pengujian prototipe yang dibuat menunjukkan alat berfungsi dengan baik dibuktikan dengan fitur buka tutup keran pengisian air dan keran penguras berfungsi baik dan merespon perintah pengguna [5]. Sehingga pada penelitian ini menggunakan smart faucet dan algoritma fuzzy logic dalam mengimplementasikan internet of things (IoT) pada sistem irigasi Subak Bali.

## 2. Metode Penelitian

Penelitian ini dilakukan beberapa tahapan agar mendapatkan hasil yang baik dan maksimal. Berikut merupakan gambaran dari tahapan yang dilakukan pada penelitian ini:



Gambar 1. Diagram Tahapan Penelitian

### 2.1 Studi Lapangan dan Literatur

Tahap awal melibatkan penelitian tentang teknologi yang telah diterapkan sebelumnya melalui studi lapangan dan literatur untuk mengevaluasi kemajuan teknologi yang digunakan dalam sistem irigasi di berbagai daerah, termasuk di Bali dan pengumpulan referensi data pustaka, dengan membaca dan mencatat, serta juga mengelola bahan penelitian.

### 2.2 Identifikasi Masalah

Tahapan ini merupakan langkah awal dalam penelitian yang akan dilakukan untuk

mendefinisikan masalah yang ada. Langkah ini juga sebagai upaya dalam menentukan kualitas dari sebuah penelitian. Implementasi Internet of Things (IoT) dengan Smart Faucet pada Sistem Irigasi Subak Bali menghadapi beberapa tantangan. Pertama, keterbatasan infrastruktur jaringan dan konektivitas di daerah pertanian dapat mempengaruhi kinerja sistem. Kedua, keandalan sensor – sensor harus diperhatikan agar pengaturan aliran air akurat. Ketiga, perlindungan data dan privasi pengguna menjadi isu penting dalam mengirimkan data melalui jaringan. Terakhir, pelatihan dan edukasi bagi petani dan pengelola irigasi diperlukan untuk memastikan penggunaan sistem IoT yang efektif.

### 2.3 Pengumpulan Data

Dataset dikumpulkan berdasarkan tingkat suhu, kelembaban tanah, curah hujan, dan level air dengan menggunakan sensor yang dipasang pada daerah dekat tanaman. Tahap selanjutnya, dataset yang diperoleh akan dilakukan training menggunakan metode fuzzy logic untuk mendapatkan algoritma dan decision making terhadap kran. Suhu, kelembaban tanah, curah hujan, dan level air merupakan parameter yang diterapkan pada metode ini dengan tingkatan tertentu. Proses pengambilan data menggunakan indeks kelembaban tanah sesuai dengan ketentuan sebagai berikut:

**Tabel 1.** Indeks Kelembaban Tanah

| No | Indeks | Value    |
|----|--------|----------|
| 1  | Kering | 0 – 25   |
| 2  | Sedang | 26 – 50  |
| 3  | Lembab | 51 – 75  |
| 4  | Basah  | 76 – 100 |

Berdasarkan tabel tersebut, diambil indeks yang digunakan sebagai batasan dalam pembuatan aturan fuzzy sebelum proses komputasi dilakukan. Nilai-nilai indeks ini digunakan sebagai parameter untuk membandingkan nilai sensor dan menentukan aturan fuzzy-nya. Selain itu, nilai-nilai keanggotaan suhu juga ditetapkan sebagai parameter pembanding dari sensor suhu yang terhubung ke mikrokontroler, seperti yang tercantum dalam tabel 2 di bawah ini:

**Tabel 2.** Nilai Keanggotaan Suhu

| No | Keanggotaan | Suhu      |
|----|-------------|-----------|
| 1  | Dingin      | $\leq 23$ |
| 2  | Sedang      | 23 – 26   |
| 3  | Panas       | 27 – 40   |

Setelah nilai indeks kelembaban tanah dan nilai keanggotaan suhu ditentukan, langkah berikutnya adalah menetapkan aturan fuzzy-nya. Parameter-parameter ini digunakan

dalam perhitungan di dalam penelitian dan dijadikan sebagai data latihan. Aturan fuzzy dinyatakan dalam bentuk "IF A is a and B is b THEN C is c". Aturan ini digunakan untuk menentukan variabel linguistik dan nilai fuzzy-nya, yang diperlukan untuk melakukan penalaran atau inferensi dan biasanya dibuat oleh ahli atau berdasarkan pengetahuan yang tersedia. Selanjutnya, dilakukan proses Fuzzifikasi, yaitu mengubah nilai tegas (Crisp) menjadi nilai fuzzy dengan menggunakan aturan  $\mu_A(a)$ ,  $\mu_A(b)$ ,  $\mu_B(a)$ , dan  $\mu_B(b)$ .

Data curah hujan dapat diambil dari data sekunder BMKG atau BPS Provinsi Bali. Perhitungan data curah hujan dapat dihitung dengan cara:

$$\text{Rata-rata curah hujan} = \frac{\text{Jumlah total intensitas curah hujan 12 bulan}}{12} \quad (1)$$

Menghitung level air pada suatu lahan melibatkan beberapa langkah dan pengukuran untuk memastikan akurasi dengan mempersiapkan alat Water Level Gauge. Pengukuran level air dapat dilakukan di beberapa titik pengukuran yang representatif di seluruh lahan. Titik-titik ini harus mencakup area yang berbeda untuk mendapatkan gambaran yang akurat tentang variasi level air di seluruh lahan.

- Menentukan titik referensi. Tentukan titik tetap (misalnya, tepi kolam atau sumur) sebagai titik referensi untuk pengukuran.
- Menggunakan water level gauge. Masukkan alat pengukur tinggi air ke dalam air hingga mencapai dasar.
- Membaca kedalaman. Catat kedalaman air yang terbaca pada alat pengukur.
- Mengulang pengukuran. Lakukan pengukuran di beberapa titik yang telah ditentukan.

Langkah selanjutnya adalah menghitung rata – rata kedalaman air, jika menggunakan beberapa titik pengukuran, hitung rata-rata kedalaman air dengan menjumlahkan semua hasil pengukuran dan membaginya dengan jumlah titik pengukuran. Selanjutnya membuat profil kedalaman atau level air dengan data dari beberapa titik pengukuran, buat profil level air untuk lahan tersebut. Ini bisa berupa diagram atau peta yang menunjukkan variasi kedalaman air di seluruh lahan. Identifikasi pola dengan memeriksa apakah ada pola tertentu dalam distribusi level air. Misalnya, area tertentu yang selalu lebih tinggi atau lebih rendah. Bandingkan dengan data historis jika memungkinkan, bandingkan dengan data pengukuran sebelumnya untuk melihat tren atau perubahan.

## 2.4 Analisa Data

### a. Algoritma Fuzzy Logic

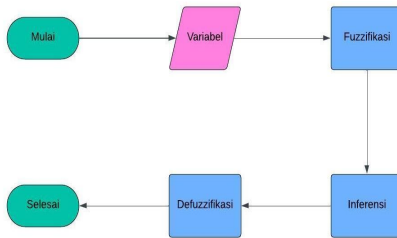
Logika Fuzzy merupakan suatu cara tepat untuk memetakan suatu ruang lingkup input kedalam suatu ruang lingkup output. Teknik ini menggunakan teori yang matematis yaitu himpunan fuzzy. Logika fuzzy sangat berhubungan dengan ketidakpastian yang telah menjadi sifat alamiah pada manusia. Teori Fuzzy pertama kali dibentuk dengan prinsip teori himpunan. Dalam himpunan konvensional (crisp), elemen dari semesta pembicaraan yaitu anggota atau bukan anggota dari sebuah himpunan. Dengan demikian, keanggotaan dari himpunan adalah bernilai tetap. Fuzzy logic sering juga disebut dengan logika samar. Sebelum munculnya teori logika fuzzy, logika ini memiliki nilai benar atau salah yang tegas. Logika Fuzzy yang digunakan pada penelitian ini adalah metode Sugeno dan Tsukamoto. Dalam penelitian ini, digunakan inferensi Sugeno, sehingga langkah agregasi dan defuzzifikasi digabungkan menjadi satu proses dengan menghitung output menggunakan persamaan yang diberikan.

$$\text{Weighted Average} = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i} \quad (2)$$

Berdasarkan persamaan tersebut, nilai  $w_i$  merupakan rata-rata bobot, sedangkan nilai  $z_i$  adalah nilai prediksi. Hasil dari persamaan tersebut akan menentukan apakah solenoid valve akan dibuka atau ditutup, sehingga dapat disesuaikan dengan kondisi tanah. Namun, sebelum diterapkan pada teknologi irigasi yang telah dibangun, model tersebut harus dievaluasi terlebih dahulu untuk memastikan apakah dapat beroperasi dengan baik dan memenuhi kebutuhan lahan atau tidak, dengan menggunakan persamaan berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - F_t)^2} \quad (3)$$

Persamaan tersebut dipergunakan untuk mengevaluasi model yang telah dikembangkan dengan memeriksa tingkat kesalahan dari model tersebut; semakin kecil kesalahan model, semakin baik modelnya, dan dianggap layak untuk diterapkan dan diuji kembali di lingkungan nyata. Di mana  $X_t$  merupakan nilai aktual yang diperoleh dari sensor,  $F_t$  adalah nilai yang diprediksi oleh model, dan  $n$  adalah jumlah data yang diperoleh selama pengujian. Selanjutnya metode Tsukamoto digunakan untuk pengambilan keputusan berdasarkan beberapa kriteria atau variabel masukan. Pada penelitian ini, penulis ingin melakukan decision making terhadap kran berdasarkan tingkat suhu, kelembaban tanah, curah hujan, dan level air. Berikut ini merupakan gambaran cara kerja dari metode Tsukamoto terhadap sistem smart faucet yang dibuat, sebagai berikut:



Gambar 2. Alur Logika Fuzzy

Penjelasan:

- Menentukan variabel input dan output yang terlibat dalam pengambilan keputusan. Variabel input adalah suhu, kelembaban tanah, curah hujan, dan level air, sedangkan variabel output adalah aksi pada solenoid valve yang membuka atau menutup kran dengan mengalirkan arus DC.
- Menentukan Himpunan fuzzy untuk setiap variabel input dan output. Seperti contoh untuk variabel masukan tingkat kelembaban tanah, himpunan fuzzy yang digunakan untuk menggambarkan kondisi adalah kering, sedang, lembab, dan basah. Dimana jika suhu < 26% adalah kering, 26-50% adalah sedang, 51-75% adalah lembab, dan >75% adalah basah. Selanjutnya untuk variabel suhu tanah adalah dingin, sedang, dan panas. Dimana suhu >26°C adalah panas.

- Melakukan inferensi fuzzy dengan menggunakan aturan fuzzy yang telah ditentukan. Sebagai contoh, kelembaban tanah memiliki status kering dan suhu dalam nilai sedang. Maka, solenoid valve akan membuka kran dan air akan disalurkan melalui sistem irigasi subak hingga status kelembaban tanah berubah menjadi lembab.
- Defuzzifikasi output untuk mendapatkan hasil numerik. Sebagai contoh, jika hasil inferensi fuzzy adalah sedang, maka hasil defuzzifikasi dapat ditentukan nilai tengah dari himpunan fuzzy sedang pada variabel keluaran dari aksi solenoid valve yang membuka atau menutup kran.

## 2.5 Perancangan Model

Tahapan ini merupakan tahapan dasar yang dimana pada tahapan tersebut dilakukan desain rancangan dan pengembangan dari solusi yang didapat. Tahapan ini dibedakan menjadi 3 tahapan yaitu tahap komunikasi data, perancangan perangkat keras (hardware), dan perancangan perangkat lunak (software). Berikut merupakan gambaran dan penjelasan dari tahapan perancangan model yang dilakukan pada penelitian ini:

### a. Tahap Komunikasi Data

Tahap pertama pada implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali ini adalah tahap komunikasi data diantara mikrokontroler yang digunakan dengan sistem yang dibuat. Pada implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali ini komunikasi data dengan sistem menggunakan Restful API dan protocol HTTP untuk mengintegrasikan sistem yang diterapkan dan digunakan sebagai output nilai sensor-sensor.



Gambar 3. Tahap Komunikasi Data

Pada penelitian ini komunikasi data antara mikrokontroler dengan sistem ada 3 tahap yaitu input (masukan), process (proses), dan output (keluaran). Berikut merupakan masukan, proses, dan keluaran dari implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali ini:

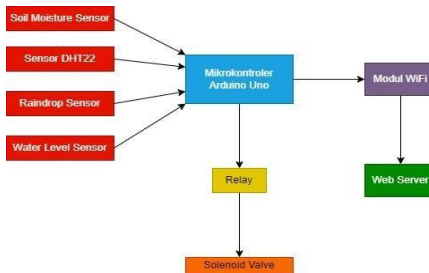
- Masukan nilai dari mikrokontroler Arduino Uno  
Pada tahapan masukan nilai dari mikrokontroler Arduino Uno. Pertama mikrokontroler Arduino Uno akan menerima nilai suhu, kelembaban tanah, curah hujan, dan level air pada sensor-sensor yang sudah terhubung dengan mikrokontroler. Selanjutnya mikrokontroler mengirim nilai data suhu, kelembaban tanah, curah hujan, dan level air ke database server dengan menggunakan protokol HTTP.
- Proses pembentukan Restful API dan Logika Fuzzy  
Pada tahap ini data dari mikrokontroler Arduino Uno akan diterima dan akan disimpan di dalam database server. Kemudian setelah data diterima dan disimpan database server, sistem akan memproses data menjadi format Javascript Object Notation (JSON) dan mengintegrasikan Restful API. Proses logika fuzzy pada sistem ini digunakan untuk menghasilkan aksi pada solenoid valve.



- Keluaran sistem  
Tahap keluaran dalam implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali akan dinyatakan berhasil apabila pada implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali dapat menghasilkan aksi pada solenoid valve yang membuka dan menutup kran secara otomatis, sesuai dengan data-data yang telah diterima.

#### b. Perancangan Perangkat Keras (Hardware)

Dalam mengimplementasikan internet of things (IoT) dengan smart faucet ini, adapun hardware yang digunakan pada penelitian ini, sebagai berikut:



**Gambar 4.** Perancangan Desain Perangkat Keras (Hardware)

Berdasarkan pada perancangan desain perangkat keras (hardware) pada gambar 4 di atas, terdapat penggunaan beberapa alat atau komponen digunakan, yang berfungsi sebagai berikut:

- Penggunaan sensor kelembaban tanah (soil moisture sensor) yang sebagai alat untuk mengukur tingkat kelembaban di dalam tanah.
- Penggunaan sensor suhu DHT22 yang sebagai alat untuk mengukur tingkat suhu dan kelembaban udara di sekitar lingkungan.
- Penggunaan sensor tetesan hujan (raindrop sensor) yang sebagai alat untuk mendeteksi hujan atau curah hujan.
- Penggunaan sensor ketinggian air (water level sensor) yang sebagai alat untuk mengukur atau memantau tingkat air dalam berbagai konteks. Contohnya, pada sistem irigasi Subak Bali.
- Penggunaan mikrokontroler Arduino Uno sebagai alat yang memproses kerja output dari sensor-sensor.
- Penggunaan Relay yang berfungsi sebagai saklar untuk mengendalikan solenoid valve yang dikontrol oleh mikrokontroler Arduino Uno.
- Penggunaan Solenoid Valve yang berfungsi sebagai alat untuk mengatur laju aliran air dengan cara membuka atau menutup jalur aliran.
- Penggunaan Modul WiFi yang berfungsi sebagai sebuah perangkat untuk mengirimkan data ke web server.

- Penggunaan Web Server yang berfungsi sebagai penyedia layanan bagi pengguna internet.

### c. Perancangan Perangkat Lunak (Software)

Pada proses pengimplementasian sistem ini, penulis menggunakan sebuah Website untuk melakukan optimisasi terhadap sistem irigasi Subak Bali berbasis internet of things (IoT) dengan smart faucet berdasarkan sistem yang telah dibuat. Untuk memenuhi kebutuhan akan alat yang diperlukan dalam membangun sistem optimisasi ini agar berjalan secara optimal, maka adapun perangkat lunak yang dibutuhkan dalam penelitian ini yaitu XAMPP, web server apache, web browser, text editor sublime text, database MySQL, dan IDE Arduino Uno. Langkah pertama yang akan dilakukan adalah melakukan penginstalan web server apache dan database MySQL menggunakan aplikasi XAMPP. Selanjutnya masuk kepada pembuatan website, dimana sebelum itu penulis melakukan pembuatan database dan tabel terlebih dahulu agar nantinya dapat menampung output sensor. Adapun dua program yang akan dibuat pada penelitian ini, adalah sebagai berikut:

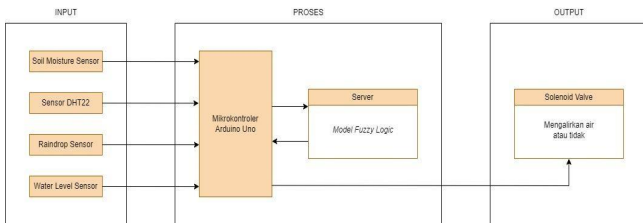
- Program Web  
Program Web dibuat untuk menampilkan nilai dari sensor suhu, kelembaban tanah, curah hujan, dan level air. Selain itu, para petani dapat memantau informasi suhu, kelembaban tanah, curah hujan, dan level air secara rinci per harinya.
- Program Arduino Uno  
Program Arduino Uno dilakukan untuk melakukan pengkodean pada mikrokontroler agar nantinya sensor yang diperoleh dapat ditampilkan pada Website. Selain itu, program Arduino Uno juga dilakukan untuk melakukan optimisasi pada sistem irigasi Subak Bali yang dimana solenoid valve dapat melakukan optimisasi dengan membuka dan menutup kran secara otomatis. Pada proses ini penulis akan melakukan setting jaringan agar mikrokontroler terkoneksi ke server menggunakan jaringan Wi-Fi.

## 3. Hasil dan Pembahasan

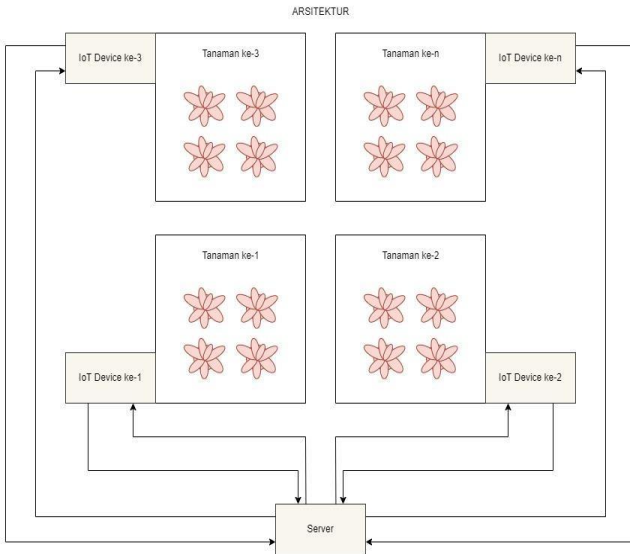
Hasil yang diperoleh dari pengujian terhadap implementasi internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali menggunakan algoritma Fuzzy Logic menunjukkan beberapa aspek yang dapat diperhatikan.

### 3.1 Tahap Penyusunan Desain Teknis

Penyusunan desain teknis akan digunakan sebagai panduan dalam manufaktur sehingga menghasilkan dimensi yang sesuai. Adapun alur sistem berjalan ditunjukkan pada Gambar 5 dengan uraian sebagai berikut:



Gambar 5. Gambaran Umum Sistem Smart Faucet

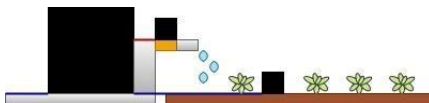


**Gambar 6.** Arsitektur Implementasi Sistem Smart Faucet

Smart Faucet ini memiliki alur kerja sistem yang dimulai dari pengambilan data kelembaban tanah menggunakan soil moisture sensor, suhu udara, curah hujan, dan level air. Soil moisture sensor akan ditanamkan ke tanah yang berdekatan dengan tanaman, sensor suhu dan curah hujan akan dipasang pada kotak control, dan sensor level air akan dicelupkan ke dalam air. Sensor-sensor ini akan mengirimkan data pada mikrokontroler. Setelah data terkirim, akan dilakukan pengiriman data ke server untuk menyimpan data pada database dan akan diolah menggunakan model Tsukamoto untuk menghasilkan pengambilan keputusan. Hasil pengambilan keputusan ini akan dikirimkan ke mikrokontroler untuk menghasilkan aksi pada Solenoid Valve.

- a. Soil Moisture Sensor  
Sensor ini berfungsi untuk mendeteksi tingkat kelembaban tanah di daerah tanaman. Data yang diperoleh akan diolah sesuai dengan kondisi tingkat kondisi tanah.
- b. Sensor Suhu dan Curah Hujan  
Sensor suhu digunakan untuk menangkap data suhu udara dan kelembaban yang berada di lingkungan sekitar. Untuk sensor curah hujan digunakan untuk mendeteksi terjadinya hujan atau tidak pada suatu area.
- c. Sensor Level Air  
Sensor level air digunakan untuk mendeteksi ketinggian air dalam berbagai konteks. Contohnya, pada sistem irigasi Subak Bali.
- d. Solenoid Valve  
Solenoid valve pada prototipe ini berfungsi sebagai perangkat untuk membuka atau menutup kran dengan mengalirkan arus DC.
- e. Kotak Kontrol

Kotak kontrol merupakan komponen yang memuat perangkat kontrol dan pengolahan dari data kelembaban tanah dan level air. Perangkat ini akan mengecek input yang masuk dan memproses sesuai dengan logika fuzzy serta menghasilkan keputusan untuk mengontrol solenoid valve. Sehingga aspek fungsionalitas smart faucet menjadi optimal.



Gambar 7. Desain Skema Prototipe Fungsional Smart Faucet

### 3.2 Algoritma Fuzzy Logic

Penelitian ini mengumpulkan data kelembaban tanah dan suhu di lokasi penelitian dengan rentang kelembaban 0-100 dan suhu 0-40°C, data curah hujan dan level kedalaman air untuk menghitung kebutuhan air irigasi. Data ini digunakan sebagai input untuk aturan Fuzzy dengan output berupa nilai 0 atau 1 untuk mengontrol Solenoid Valve terbuka atau tertutup. Hasil percobaan mengklasifikasikan kondisi wilayah menjadi kering, sedang, lembab, dan basah, kemudian menggabungkan kategori kelembaban dan suhu menggunakan operator AND untuk menghasilkan output berupa Penyiraman atau Berhenti. Langkah selanjutnya adalah menentukan himpunan fuzzy berdasarkan data yang telah dikumpulkan dan mengonversi nilai agar menjadi nilai yang distandarkan menggunakan persamaan kelembaban. Penelitian ini juga melibatkan evaluasi model menggunakan Root Mean Square Error (RMSE) untuk menilai tingkat kesalahan model. Hasil pengujian menunjukkan tingkat error yang optimal atau nol persen, yang normal karena setiap nilai berada dalam rentang yang ditetapkan. Penelitian ini menghasilkan pengembangan aplikasi web sebagai perangkat pemantauan dan pengendalian yang dioperasikan secara otomatis oleh sistem tanpa memerlukan campur tangan petani. Hasil dari algoritma fuzzy logic adalah mencapai akurasi 100% dan 0% kesalahan dalam model yang dikembangkan. Hasil tersebut dapat digunakan untuk mengurangi pemborosan air sehingga dapat diupayakan sebagai langkah efisiensi penggunaan air secara keseluruhan. Output lainnya adalah nutrisi tanaman akan terjaga dan mengurangi resiko akibat kelebihan atau kekurangan air. Pengujian di lapangan nyata diperlukan untuk memastikan keakuratan dan kecocokan dengan kondisi tanah dan tanaman yang sebenarnya. Pengembangan lebih lanjut juga perlu dilakukan, terutama dalam hal pengujian air untuk berbagai jenis tanaman, yang memiliki kebutuhan air yang berbeda.

### 4. Kesimpulan

Penerapan internet of things (IoT) dengan smart faucet pada sistem irigasi Subak Bali menggunakan algoritma Fuzzy Logic memungkinkan petani untuk mengoptimalkan penggunaan air berdasarkan kebutuhan tanaman dan kondisi lingkungan secara real-time. Melalui kontrol yang lebih akurat terhadap pasokan air dan nutrisi tanaman, sistem irigasi yang dioptimalkan dapat mengurangi pemborosan air dan meningkatkan efisiensi penggunaan air secara keseluruhan. Hal ini dapat membantu menjaga kesehatan tanaman dengan memberikan pasokan air yang sesuai dengan kebutuhan mereka, meningkatkan pertumbuhan tanaman, dan mengurangi risiko penyakit akibat kelebihan atau kekurangan air. Hasil panen juga dapat ditingkatkan karena tanaman mendapatkan kondisi lingkungan yang ideal untuk pertumbuhan mereka, menghasilkan bibit yang berkualitas, dan mengurangi risiko kegagalan panen akibat masalah irigasi. Selain itu, pengguna dapat melakukan monitoring dan pengendalian jarak jauh terhadap sistem irigasi melalui website, memberikan kemudahan dalam pengelolaan sistem irigasi dan memberikan informasi yang akurat tentang kondisi tanaman dan kebutuhan air mereka.

**Daftar Pustaka**

- [1] Yusmita, W., Putra, I. G. S. A., & Budiassa, I. W. (2017). Manajemen Irigasi Tradisional pada Sistem Subak Umayu di Desa Talibeng Kecamatan Sidemen Kabupaten Karangasem. *J. Agribisnis Dan Agrowisata (Journal Agribus. Agritourism)*, 6(2), 179–189.
- [2] Norken, I. N. (2015). Manajemen Subak: Permasalahan dan Upaya Pemecahannya. In *Jurusan Teknik Sipil. Universitas Udayana*.
- [3] Sandi, G. H. dan Fatma, Y. (2023). Pemanfaatan Teknologi Internet of Things (IoT) pada Bidang Pertanian. *JATI (Jurnal Mahasiswa Teknik Informatika)*. Universitas Muhammadiyah Riau.
- [4] Setia, B. (2019). Penerapan Logika Fuzzy pada Sistem Cerdas. *Jurnal Sistem Cerdas*, 2(1), 61–66.
- [5] Sukarta, E. B. L., Sukarsa, I. M., & Piarsa, I. N. (2021). "Smart Automatic Water Filler (SAWF) berbasis Internet of Things". *JITTER (Jurnal Ilmiah Teknologi dan Komputer)*. Universitas Udayana.

Halaman ini sengaja dibiarkan kosong

p-ISSN

ISSN 2986-3929



9 772986 392004

e-ISSN

ISSN 3032-1948



9 773032 194009