



Volume 1, Nomor 3, Mei 2023

JNATIA

Jurnal Nasional Teknologi Informasi dan Aplikasinya

Program Studi Informatika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana

Daftar Isi

Rancang Model Ontologi untuk Representasi Pengetahuan Tari Tradisional Indonesia Sang Putu Febri Wira Pratama	781-788
Perancangan User Interface dan User Experience pada Aplikasi Pencari Indeks Gusto Gibeon Ginting, I Gusti Ngurah Anom Cahyadi Putra	789-796
Prediksi Paket Return Menggunakan Metode Decision Tree Menerapkan Algoritma C4.5 Berbasis Website Oskar Maha Kasi, Agus Muliantara	797-802
Isolation Forest dengan Exploratory Data Analysis pada Anomaly Detection untuk Data Transaksi I Made Sudarsana Taksa Wibawa, Anak Agung Istri Ngurah Eka Karyawati	803-810
Perancangan Antar Muka Wearable Sistem Bagi Atlet Ni Wayan Sani Utari Dewi, Cokorda Rai Adi Pramatha	811-818
Implementasi Algoritma KNN untuk Memprediksi Performa Siswa Sekolah I Made Ryan Prana Dhita, Gst. Ayu Vida Mastrika Giri	819-826
Penerapan RMS Contrast sebagai Penentu Citra Terbaik Berdasarkan Tingkat Kontras Shiennyta Florensia Adiriyanto, I Gede Arta Wibawa	827-832
Pengujian Fungsionalitas Sistem Pengamanan Digital Watermarking Kartu Indonesia Sehat Menggunakan Algoritma MSB Ni Putu Intan Cahyani, I Gede Santi Astawa	833-840
Implementasi Sistem Monitoring pada Bunga Anggrek Menggunakan Arduino dengan Aplikasi Berbasis Website Gede Krisnawa Sandhya Wandhana, I Gusti Agung Gede Arya Kadyanan	841-846
Analisis Performa Sistem pada Pemisahan Database Analitik dengan Transaksional I Made Ari Madya Santosa, I Gusti Ngurah Anom Cahyadi Putra	847-856
Analisis Performa Write/Read Kompresi Delta Encoding pada Data Logging Menggunakan Go Benchmark I Putu Gede Mahardika Adi Putra, I Ketut Gede Suhartana	857-864
Purwarupa Perangkat Monitoring Ketinggian Air Tandon Menggunakan TinkerCAD I Kadek Widiarthawan, I Komang Ari Mogi	865-874
Pengenkripsian dan Dekripsi Gambar Menggunakan Algoritma AES dengan MAC untuk Peningkatan Keamanan Ni Wayan Amanda Putri Astawa, I Made Widiartha	875-880
Analisa UI dalam Perancangan Aplikasi Ekspor Hasil Laut "FishGo!" dengan Metode SUS Ketut Agus Cahyadi Nanda, I Putu Gede Hendra Suputra	881-892

Sistem Informasi Prediksi Penjualan E-commerce Menggunakan Analisis Data Historis dan Algoritma MLR I Gusti Putu Wisnu Wardhana, I Wayan Santiyasa	893-900
Sistem Rekomendasi Konten Youtube Berdasarkan Durasi Menonton Menggunakan Algoritma Content-Based Filtering I Gede Ngurah Wahyu Ananta, I Wayan Supriana	901-908
Perhitungan Nilai Besaran Fisis Mammografi Jenis Histopatologi IDC dan ILC Anak Agung Ngurah Frady Cakranegara, Ida Ayu Gde Suwiprabayanti Putra	909-916
Uji Performansi Algoritma LR dan RFR pada Implementasi Sistem Prediksi Harga Rumah I Putu Teddy Dharma Wijaya, Ida Bagus Gede Dwidasmara	917-924
Perancangan Ontologi Semantik: Representasi Digital Oleh - Oleh Khas Bali Ngakan Made Alit Wiradhanta, Ida Bagus Made Mahendra	925-932
Klasifikasi Genre Musik Menggunakan Support Vector Machine Berdasarkan Spectral Features I Gusti Agung Ngurah Diputra Wiraguna, Luh Arida Ayu Rahning Putri	933-940
Analisis Performa Algoritma K-Nearest Neighbor dalam Klasifikasi Tingkat Kerontokan Rambut Gede Dikka Widya Prana, Luh Gede Astuti	941-950
Analisis Sentimen Gambar pada Media Sosial dengan Pendekatan Deep Learning Ronaldito Juan Bantaras T, Made Agung Raharja	951-960
Pengembangan Model Ontologi Cerita Rakyat Bali Ni Made Julia Budiantari, Ngurah Agus Sanjaya ER	961-966
Implementasi Aplikasi "SahabatAnabul" Berbasis UCD Sebagai Website Pangkalan Informasi Anabul Ni Luh Putu Ayu Siwastuti Cayadewi, Made Agung Raharja	967-972
Penyusunan Sistem Rekomendasi Produk Diecast Mobil dengan Metode Content-Based Filtering (CBF) Anak Agung Aditya Nugraha, Ngurah Agus Sanjaya ER	973-976
Sistem Penjualan Merchandise Berbasis Aplikasi Mobile I Dewa Gede Partha Wijaya, I Made Widiartha	977-984
Integrasi Chatbot Berbasis GPT Terhadap Website E-Commerce (Studi Kasus: Toko JSA) Made Putra Teguh Pramana, I Gede Surya Rahayuda	985-994
Implementasi Metode Certainty Faktor untuk Rekomendasi Pembelian Hp Komang Arsa Wiguna, I Gede Arta Wibawa	995-1002
Klasifikasi Musik Daerah Indonesia dengan Metode Machine Learning Gst. Ayu Vida Mastrika Giri, Made Leo Radhitya	1003-1010

Rancang Model Ontologi untuk Representasi Pengetahuan Tari Tradisional Indonesia

Sang Putu Febri Wira Pratama^{a1}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana

Bali, Indonesia

¹febriwiraprtma@email.com

Abstract

Indonesia is the largest archipelagic country in the world with more than 17,000 islands inhabited by approximately 255 million people, a figure that makes Indonesia the fourth most populous country in the world. This implies that there is a lot of cultural diversity in Indonesia, one of which is traditional dance. Traditional dances in various regions in Indonesia are certainly different and have their own characteristics, so they need to be well documented. The ontology knowledge base is an appropriate method used to represent information. In this project, the ontology model was built using the Protégé ontology development tool. We use the method of methontology in the development of the ontology model where this method describes each step-in detail. The ontology model built has 11 classes, 5 object properties, 8 data properties, and 72 individuals. We focus on describing the types of traditional dances from various regions in Indonesia. The testing process in the development of the ontology model by performing SPARQL queries.

Keywords: *Indonesia Traditional Dance, Ontology, Methontology, Protégé, SPARQL query*

1. Pendahuluan

Indonesia sebagai negara kepulauan terbesar di dunia dengan jumlah pulau kurang lebih 17 ribu pulau. Tidak hanya dikenal dengan keindahan alamnya, Indonesia juga dikenal dengan keanekaragaman budaya yang memukau perhatian dunia, salah satunya tari tradisional. Tari tradisional merupakan salah satu jenis kebudayaan di Indonesia yang harus dilestarikan. Keanekaragaman tari tradisional di berbagai daerah harus didokumentasikan dan tergambarkan dengan baik agar masyarakat menjadi lebih mengenal jenis tari tradisional dari daerah lain. Oleh karena itu, maka diperlukan suatu model yang dapat merepresentasikan pengetahuan.

Penggunaan konsep basis pengetahuan Ontologi merupakan solusi yang digunakan oleh penulis. Ontologi merupakan cara merepresentasikan secara eksplisit suatu domain pengetahuan melalui suatu konsep dengan memberikan makna, sifat, dan hubungan pada konsep tersebut sehingga terkumpul dalam domain pengetahuan dan membentuk basis pengetahuan [1]. Adapun alasan mengapa menggunakan ontologi, seperti mampu menjelaskan domain pengetahuan secara eksplisit, menyediakan struktur hierarki konsep untuk menjelaskan domain dan bagaimana mereka terkait [2]. Untuk memodelkan ontologi, penulis dapat menggunakan Ontology Web Language (OWL) yang merupakan turunan dari RDFS [3]. Metode yang digunakan untuk membangun model Ontologi adalah Methontology. Metode Methontology dapat menggunakan kembali ontologi yang dibangun untuk pengembangan sistem lebih lanjut. Metode Methontology memiliki keunggulan dalam penggambaran setiap aktivitasnya dengan jelas.

2. Metode Penelitian

Dalam metode penelitian, metode yang penulis gunakan untuk membangun model ontologi adalah methontology. Methontology adalah metode pengembangan model ontologi berdasarkan hubungan antar objek. Berikut ini adalah langkah-langkah dalam metode metodologi:

2.1. Spesifikasi

Tujuan dari tahap spesifikasi adalah untuk menghasilkan dokumen spesifikasi ontologi informal, semi-formal atau formal yang ditulis dalam bahasa alami, masing-masing menggunakan seperangkat representasi perantara atau menggunakan pertanyaan kompetensi [4].

2.2. Akuisi Pengetahuan

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

2.3. Konseptual

Pada fase ini, penulis akan membangun struktur domain pengetahuan dalam model konseptual yang menggambarkan masalah dan solusinya dalam bentuk kosakata domain yang diidentifikasi dalam aktivitas spesifikasi ontologi. Hal pertama yang penulis lakukan adalah menyusun Glosarium Istilah, memasukkan konsep, contoh, kata kerja, dan property [4].

2.4. Integrasi

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

2.5. Implementasi

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

2.6. Evaluasi

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

2.7. Dokumentasi

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

3. Hasil dan Pembahasan

3.1. Spesifikasi

Pada tahap ini, dihasilkan spesifikasi dokumen ontologi informal, semi-formal, dan formal yang ditulis dalam bahasa alami. Berikut dekripsi dari ontologi Tari Tradisional:

- a. Domain: Tari Tradisional Indonesia
- b. Tanggal: 3 Oktober 2021
- c. Dirancang oleh: Sang Putu Febri Wira Pratama
- d. Diimplementasikan oleh: Sang Putu Febri Wira Pratama
- e. Level Formalitas: Semi-Formal
- f. Ruang Lingkup: Tari Tradisional 34 Provinsi di Indonesia
- g. Sumber Pengetahuan: Internet

3.2. Akusisi Pengetahuan

Tahapan ini bertujuan untuk memperoleh pengetahuan yang dilakukan secara independen dalam proses pengembangan Ontologi. Berikut ini teknik yang digunakan dalam tahapan ini sebagai berikut:

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

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

3.3. Konseptual

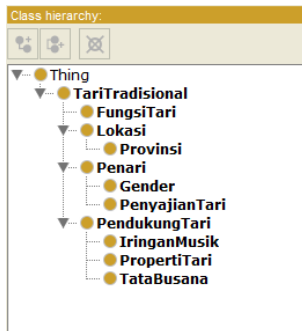
Tahap ini bertujuan untuk merancang konsep dan istilah yang akan digunakan untuk menggambarkan masalah beserta solusinya pada domain ontologi tari tradisional yang nantinya akan dideskripsikan ke dalam bentuk *class*, *object properties*, dan *data properties* dalam ontologi.

3.4. Integrasi

Tahap ini bertujuan untuk mempertimbangkan kembali definisi yang digunakan dalam ontologi yang dibangun, sehingga tidak ada kesalahan dalam pembangunan relasi.

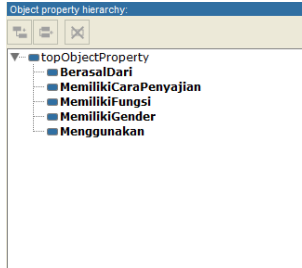
3.5. Implementasi

Pada tahap implementasi model ontologi, peneliti menggunakan aplikasi Protégé 4.3 dalam membangun ontologi. Setiap bagian ontologi didefinisikan sesuai dengan hasil dari setiap tahapan dalam metode METHONTOLOGY. Dimana didapatkan konsep *class* ditunjukkan pada Gambar 1. Relasi atau hubungan antar *class* dalam model ontologi yang didefinisikan sebagai *object properties* ditunjukkan pada Gambar 2. Atribut *class* dan atribut *instance* didefinisikan sebagai *data properties* yang ditunjukkan pada Gambar 3 dan *instance* didefinisikan sebagai individu ditunjukkan pada Gambar 4.

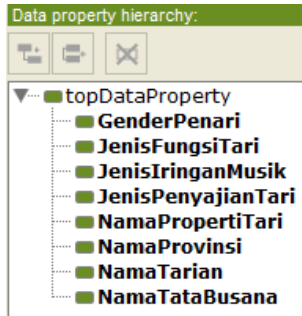


Gambar 1. Hirarki Class dari Model Ontologi Tari Tradisional

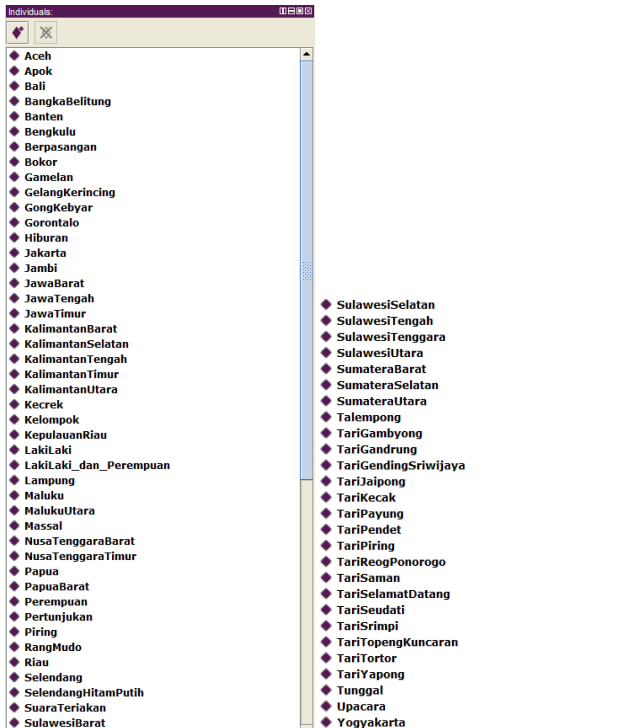
Terdapat 11 *class* yang didefinisikan dalam model ontologi tari tradisional. Kemudian 5 *object properties* dan 8 *data properties* yang telah didefinisikan.



Gambar 2. Hirarki *Object Properties* dari Model Ontologi Tari Tradisional

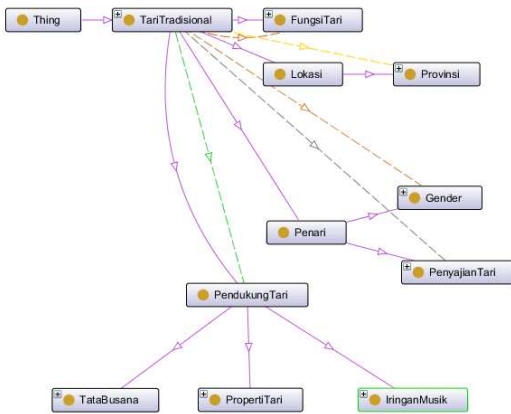


Gambar 3. Hirarki *Data Properties* dari Model Ontologi Tari Tradisional



Gambar 4. Individuals dari Model Ontologi Tari Tradisional

Terdapat 72 individu yang digunakan dalam ontologi tari tradisional. Individu dalam kelas yang diperluas disebut *instance*.



Gambar 5. Ontograf dari Model Ontologi Tari Tradisional

Dari ontograf diatas, dapat diketahui dari model ontologi tari tradisional terdapat 11 class yang terdiri dari 10 sub class yang mana dihubungkan satu sama lain dengan object properties.

3.6. Evaluasi

Pada tahapan evaluasi, penulis melakukan pengujian terhadap model ontologi yang dibangun dengan menggunakan query dalam aplikasi Protégé 4.3, dimana proses query menggunakan SPARQL query. Penulis membuat PREFIX yang baru disebut dengan project dengan menggunakan alamat ontologi IRI. Query dibentuk dari pertanyaan yang diinginkan oleh pengguna. Berikut hasil query yang dilakukan yang ditunjukkan pada Gambar 6.

```

SPARQL query
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX project: <http://www.semanticweb.org/teknoinformatika2021/islurtted-ontology-68/>
SELECT *
WHERE { ?TariTradisional project Menggunakan project Selendang }

TariTradisional
TariJaipong
TariPendet
    
```

Gambar 6. Hasil Pengujian Pertama dengan SPARQL query

Pada Gambar 6, query yang dihasilkan berupa tari tradisional yang menggunakan property "Selendang".

```
SPARQL query
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX project: <http://www.semantic-web.org/2015/03/ontologies/2015/03/Lurified-ontology-04#>

SELECT *
WHERE { ?TariTraditional project:BerasalDari project:046 }
```

TarifKecak:
TariPendet:

Gambar 7. Hasil Pengujian Kedua dengan SPARQL query

Kemudian pengujian juga dilakukan kembali untuk memastikan bahwa ontology telah dirancang dengan baik. Dapat dilihat pada Gambar 7, hasil query tersebut menunjukkan ada 2 *individu* atau *instance* tari tradisional yang berasal dari Bali yakni Tari Kecak dan Tari Pendet.

3.7. Dokumentasi

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

4. Kesimpulan

Berdasarkan pembahasan yang telah dijabarkan, maka dapat disimpulkan sebagai berikut. Model ontologi tari tradisional di Indonesia telah dirancang dengan metode METHONTOLOGY dan menggunakan aplikasi Protégé 4.3 yang menghasilkan 11 *class*, 5 *object properties*, 8 *data properties*, dan 72 *individual*. Pada proses evaluasi, model ontologi ini memberikan hasil yang benar dalam menjawab permintaan dari pengguna yang diubah ke dalam bentuk query. Model ontologi yang dihasilkan, yaitu ontologi tari tradisional di Indonesia dapat digunakan sebagai dasar pengembangan sistem pengetahuan tari tradisional di Indonesia berbasis web semantik.

Daftar Pustaka

- [1] H. R. Badron, Yunizar; Agus, Fahrul; Hatta, "Studi Tentang Pemodelan Ontologi Web Semantik dan Prospek Penerapan pada Bibliografi Artikel Jurnal Ilmiah," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 164–169, 2017, [Online]. Available: <http://journal.student.uny.ac.id/ojs/ojs/index.php/pgsd/article/viewFile/135/130>.
- [2] I. L. Koten and C. R. A. Pramarta, "Semantic Representation of Balinese Traditional Dance," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, p. 411, 2020, doi: 10.24843/jik.2020.v08.i04.p07.
- [3] C. R. A. Pramarta, "Assembly the Semantic Cultural Heritage Knowledge," *J. Ilmu Komput.*, vol. 11, no. 2, p. 83, 2018, doi: 10.24843/jik.2018.v11.i02.p03.
- [4] M. Fernández, A. Gómez-Pérez, and N. Juristo, "METHONTOLOGI: From Ontological Art Towards Ontological Engineering," *Proc. Ontol. Eng. AAAI-97 Spring Symp. Ser.*, pp. 33–40, 1997, [Online]. Available: <http://speech.inesc.pt/~joana/prc/artigos/06c METHONTOLOGY from Ontological Art towards Ontological Engineering -Fernandez, Perez, Juristo -AAAI -1997.pdf>.
- [5] P. R. Ganeswara and C. R. A. Pramarta, "Ontology-based Approach for Klungkung Royal Family," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, p. 497, 2020, doi: 10.24843/jik.2020.v08.i04.p16.

Halaman ini sengaja dibiarkan kosong

Perancangan UI dan UX pada Aplikasi Pencari Indeks di Sekitar Kampus Universitas Udayana

Gusto Gibeon Ginting^{a1}, I Gusti Ngurah Anom Cahyadi Putra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia

¹gustoginting@gmail.com

²anom.cp@unud.ac.id

Abstract

This study aims to design a website-based user interface and user experience. Where it is, proportional to the increase in the number of purchases of existing motorcycles. The website that has been made was tested on the General Public with a total of 20 respondents. To find out the level of satisfaction of respondents in using the motorcycle recommendation website, usability testing was carried out using the system usability scale method. This method measures the usability of a computer system according to the user's subjective point of view by filling out a Likert-scale questionnaire. The website that has been made is tested on the General Public. To find out the level of satisfaction of respondents in using the motorcycle recommendation website, usability testing was carried out using the system usability scale method. This method measures the usability of a computer system according to the user's subjective point of view by filling out a Likert-scale questionnaire.

Keywords: *Indeks, Marketplace, Mobile App, e-marketplace, Applications*

1. Pendahuluan

Bali merupakan salah satu provinsi di Indonesia yang terkenal dengan tempat wisatanya yang beragam dan turisnya yang selalu ramai yang berasal dari berbagai daerah domestik maupun mancanegara. Para pengunjung atau pendatang biasanya memiliki latar belakang tersendiri dalam mengunjungi pulau dengan julukan 1000 pura ini, entah itu berlibur, bekerja, ataupun dengan tujuan Pendidikan. Dengan hal tersebut, tak heran bisnis dengan tipe properti atau sewa penginapan, seperti hotel, villa, ataupun apartemen sangat ramai. Para penyedia layanan ini juga menyediakan beragam pilihan penginapan, mulai dari penginapan harian, mingguan, bulanan atau bahkan hingga jangka panjang seperti tahunan.

Tidak hanya untuk destinasi liburan. Bali juga memiliki kampus unggulan, salah satunya adalah Universitas Udayana. Universitas Udayana memiliki total 6705 mahasiswa baru pada semester ganjil di tahun 2021, Dimana diantaranya merupakan mahasiswa yang berasal dari daerah luar Bali, hal tersebut menjadi kesulitan tersendiri bagi mahasiswa pendatang, mulai dari kurangnya pengetahuan mengenai lokasi atau daerah di Bali dan kurangnya relasi. Untuk menghadapi hal tersebut dibutuhkan suatu aplikasi untuk membantu para pendatang untuk menemukan tempat penginapan. Aplikasi ini memiliki fitur seperti, mencari rumah kost terdekat, filter rating, harga, serta, fasilitas yang diinginkan.

1.1 E-Marketplace

E-marketplace yang berarti adalah ruang informasi bisnis umum, yang harus memenuhi empat sifat *e-marketplace* yaitu distribusi, otonomi, ketergantungan, dan kemunculan [1]. *E-marketplace* adalah sebuah sistem informasi antar organisasi dimana pembeli dan penjual di pasar mengkomunikasikan informasi tentang harga, produk dan mampu menyelesaikan transaksi melalui saluran komunikasi elektronik. Dengan melalui internet para penjual dimudahkan dalam melakukan promosi dan memasarkan produknya dalam jangkauan yang lebih luas [2]. Pengguna

internet yang semakin meningkat berpengaruh pada perkembangan e-commerce di Indonesia, khususnya forum jual-beli online. *E-Commerce* merupakan proses pembelian dan penjualan jasa dan barang-barang secara elektronik dengan transaksi bisnis terkomputerisasi menggunakan internet, jaringan, dan teknologi digital lain [3].

1.2 Indekos

Pengertian indekost merupakan sejenis kamar sewa yang disewa (*booking*) selama kurun waktu tertentu sesuai dengan perjanjian pemilik kamar dan harga yang disepakati. Umumnya pemesanan kamar dilakukan selama kurun waktu satu tahun. Namun demikian ada pula yang hanya menyewakan selama satu bulan, tiga bulan, dan enam bulan, sehingga sebutannya menjadi sewa tahunan, bulanan, tri bulanan, dan tengah tahunan. Perjanjian sewa dan harga biasanya dilakukan melalui kesepakatan secara lisan sesuai dengan perjanjian sewa dan harga yang disepakati [4].

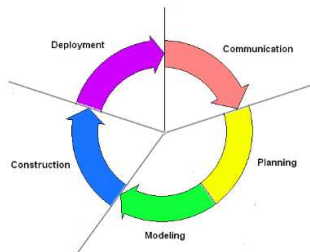
1.3 Android

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

2. Metode Penelitian

Dalam penelitian ini metode yang digunakan adalah metode Prototyping. proses pengembangan sistem sering menggunakan pendekatan prototipe (*prototyping*) dan *Modeling*. Metode ini paling baik digunakan untuk menyelesaikan masalah yang terjadi antara pengguna serta mampu menganalisis masalah yang muncul dari pengguna yang kurang sesuai dengan kebutuhan mereka.

2.1 Prototyping



Gambar 1. Prototyping

Tahapan dalam Model *Prototyping* dirangkum sebagai berikut:

a. Langkah 1: Pengumpulan dan analisis persyaratan

Model *prototyping* dimulai dengan analisis kebutuhan. Pada tahap ini, syarat - syarat dari sistem didefinisikan secara rinci. Pada proses ini, pengguna akan diwawancarai untuk mengetahui tujuan dan keinginan mereka dari sistem yang akan dibuat.

b. Langkah 2: Desain Awal

Tahap kedua adalah desain pendahuluan atau *quick design*. Pada tahap ini dilakukan pembuatan desain sederhana dari sistem yang ingin dibuat. Namun desain yang dibuat hanya gambaran saja mengenai cara kerja dari sistem tersebut kepada pengguna. Desain awal sangat membantu dalam mengembangkan pada metode prototipe.

c. Langkah 3: Membangun Desain Prototipe

Pada tahap ini, sistem akan dirancang berdasarkan informasi yang dikumpulkan dari desain awal. Tahap ini merupakan lanjutan dari gambaran dari sistem yang diperlukan.

d. Langkah 4: Evaluasi

Pada tahap ini, sistem yang dibangun akan diberikan kepada pengguna untuk dilakukan evaluasi awal. Pengguna akan membantu untuk mengetahui kekuatan dan kelemahan dari cara kerja sistem. Komentar dan saran dari pengguna akan dikumpulkan dan diberikan kepada pengembang dari sistem.

e. Langkah 5: Perbaikan Prototipe

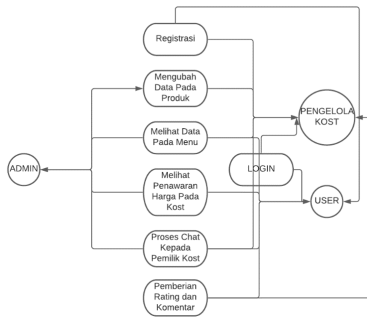
Jika pengguna tidak senang dengan prototipe saat ini, maka diperlukan perbaikan ulang pada prototipe sesuai dengan komentar dan saran pengguna. Tahap ini tidak akan berakhir sampai semua persyaratan yang ditentukan oleh pengguna terpenuhi. Jika pengguna puas dengan prototipe yang dikembangkan, sistem akhir dikembangkan berdasarkan prototipe akhir yang disetujui.

f. Langkah 6: Penerapan Sistem dan Pemeliharaan

Setelah sistem sudah mencapai tahap final dalam pengembangan dengan berdasarkan prototipe akhir, sistem tersebut akan diuji secara menyeluruh dan dikerahkan ke produksi. Sistem menjalani perawatan rutin untuk meminimalkan *downtime* dan mencegah kegagalan skala besar.

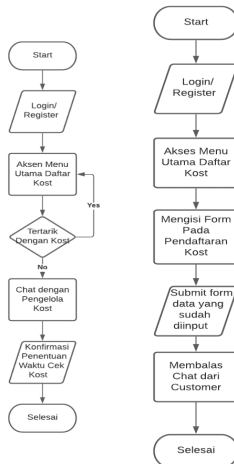
2.2 Modeling

Unified Modeling Language (UML) merupakan salah satu metode yang baik pada sistem berorientasi objek dalam pengembangan sistem. Hal tersebut karena UML Menyediakan sebuah bahasa pemodelan visual yang memungkinkan pengembang dari sistem untuk membuat rancangan yang sesuai dengan standar dari pengembang ataupun pengguna, metode ini mudah dipahami dan dilengkapi dengan mekanisme yang efektif untuk berbagi dan mengkomunikasikan desain mereka dengan orang lain. Alur perancangan sistem yang digunakan adalah *use case diagram* dan diagram aktivitas.



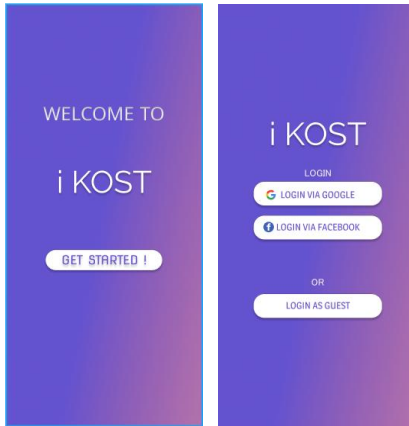
Gambar 2. Perancangan Sistem

Pada use case diagram diatas dapat dijelaskan bahwa, pengelola login terlebih dahulu setelah itu pengelola dapat mengakses data menu, data dalam daftar kost yang ditawarkan kemudian melakukan konfirmasi untuk pengecekan dan *booking* kost. Kemudian pengguna dapat mengakses semua menu yang ada di aplikasi setelah melakukan login terlebih dahulu.



Gambar 3. Use Case User dan Pengelola

3. Hasil dan Pembahasan



Gambar 4. Start Page dan Login Page

3.1 Start Page

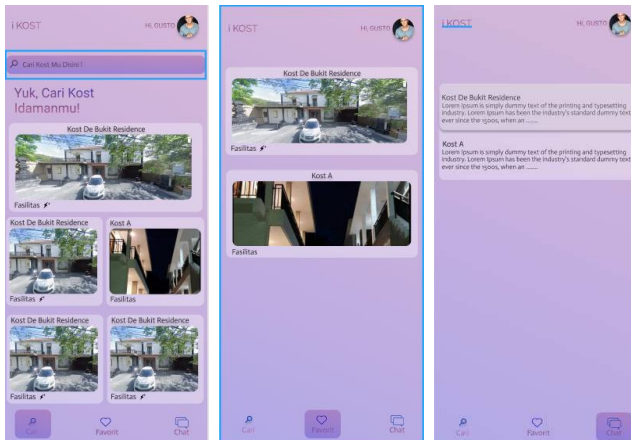
Menu ini merupakan menu awal saat user telah melakukan instalasi aplikasi. Pada menu ini user harus menekan tombol "Get Started" untuk melanjutkan.

3.2 Login Page

Pada menu ini, user akan diberikan beberapa pilihan untuk melakukan login di aplikasi ini. User akan diberikan 3 pilihan, Login melalui Google, Facebook, dan apabila user tidak ingin melakukan login, maka user boleh memilih pilihan "Login as Guest". Tahap ini bertujuan mengambil data diri dari platform media sosial dari user tersebut.

3.3 Home

Menu ini berisi kumpulan dari indeks yang telah dirangkum oleh pengembang aplikasi serta yang sudah didaftarkan oleh pengelola. Pada menu ini, user akan diberikan tampilan tersebut dan memilih sesuai dengan keinginan user tersebut. Pada menu ini juga terdapat beberapa fitur – fitur seperti filter berdasarkan harga, jarak, hingga, fasilitas yang terdapat pada indeks, dan juga terdapat fitur favorit, dimana user dapat memasukkan indeks tertentu sesuai dengan keinginan dari user. Fitur selanjutnya adalah fitur pencarian, pada fitur ini user dapat melakukan pencarian dari nama kost dengan menekan kotak atau tab pencarian yang sudah disediakan pada bagian atas halaman.



Gambar 5. Menu Home, Favorite dan Chat

3.4 Menu Favorite

Menu ini berisi daftar indeks yang sudah ditambahkan oleh user pada menu home.

3.5 Menu Chat

Menu ini berisi tampilan percakapan antara customer dengan pengelola indeks. Pada menu ini customer akan melakukan percakapan kepada pengelola indeks terkait dengan indeks tersebut

4. Kesimpulan

Berdasarkan uraian dan pembahasan sebelumnya, Penelitian ini menggunakan metode prototyping dimana dalam pengembangannya, user akan berperan penuh dalam penentuan desain akhir dari aplikasi ini. Dalam penelitian ini dapat disimpulkan bahwa dengan adanya *e-marketplace* pencarian indekost ini, pengguna dalam mencari dan memesan penginapan kost menjadi lebih mudah. Pengguna yang ingin mencari kost dimudahkan dengan banyaknya pilihan.

Daftar Pustaka

- [1] Marco, R., & Ningrum, B. T. P. (2017). Analisis Sistem Informasi *E-marketplace* Pada Usaha Kecil Menengah (UKM) Kerajinan Bambu Dusun Brajan. *Data Manajemen dan Teknologi Informasi (DASI)*, 18(2), 48-53.
- [2] Rahmadi, Y., Prasetyo, Y. A., & Hasibuan, M. A. (2015). Pengembangan Modul Freemium Aplikasi TEL-US (Telkom University Store) menggunakan Metode Iterative Incremental dan Framework Laravel. *eProceedings of Engineering*, 2(2).
- [3] Sungkono, C. Koleksi Buku 2008 Sistem informasi manajemen: mengelola perusahaan digital buku 1/Kenneth C. Laudon, Jane P. Laudon; Penerjemah: Chriswan Sungkono, Machmudin Eka P.; Editor: Nina Setyaningsih.
- [4] Ratnasari, D., Qur'ani, D. B., & Apriani, A. (2018). Sistem Informasi Pencarian Tempat

- Kos Berbasis Android. Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, 3(1), 32-45.
- [5] Rimate, A. D., Najoan, X., & Sugiarto, B. A. (2017). Rancang Bangun Aplikasi Berbasis Android Untuk Informasi Kegiatan dan Pelayanan Gereja. Jurnal Teknik Elektro dan Komputer, 6(1), 1-6.

Halaman ini sengaja dibiarkan kosong

Prediksi Paket Return Menggunakan Metode Decision Tree Menerapkan Algoritma C4.5 Berbasis Website

Oskar Maha Kasi^{a1}, Agus Muliartara^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹oskaresvada@gmail.com
²muliarta@unud.ac.id

Abstract

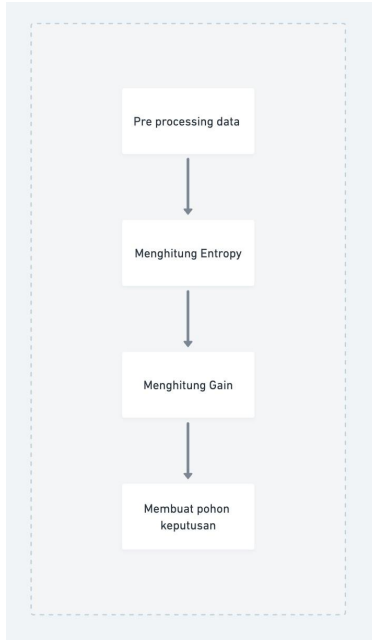
The development of e-commerce technology has provided significant benefits in facilitating online transactions. However, one of the problems that is often faced by e-commerce companies is the management of return packages, where customers return goods that have been purchased. In order to optimize the package return management process, an effective prediction method is needed to estimate the possibility of returning packages by customers. With a website-based package return prediction system, e-commerce companies can take appropriate actions to manage return packages more efficiently. For example, they can adjust their inventory strategy, improve product quality, or provide special promotional offers to customers who have a high probability of returning packages. This can help increase customer satisfaction, reduce the cost of managing return packages, and increase the overall operational efficiency of the company.

Keywords: *Prediksi Paket Return, Decision Tree, Algoritma C4.5*

1. Pendahuluan

Dalam dunia bisnis dan e-commerce, kegiatan pengembalian paket atau return merupakan suatu hal yang lazim terjadi. Pengembalian paket dapat terjadi karena berbagai alasan, seperti barang rusak, ukuran yang tidak sesuai, atau keinginan pelanggan untuk mengembalikan produk yang tidak memenuhi harapan mereka. Mengelola proses pengembalian paket dengan efisien sangat penting bagi perusahaan untuk menjaga kepuasan pelanggan dan meminimalkan kerugian. Dalam menghadapi tantangan tersebut, penggunaan metode prediksi dapat menjadi solusi yang efektif. Salah satu metode yang umum digunakan dalam prediksi paket return adalah metode decision tree (pohon keputusan). Decision tree adalah suatu model prediktif yang menggunakan struktur pohon untuk menggambarkan keputusan dan konsekuensi yang mungkin terjadi. Metode ini mampu memecahkan masalah klasifikasi dan prediksi dengan menggabungkan serangkaian aturan dan kondisi. Dalam penelitian ini, penulis akan menggunakan metode decision tree dengan algoritma C4.5 untuk memprediksi paket return. Algoritma C4.5 adalah salah satu algoritma pembentukan pohon keputusan yang populer dan efisien. Algoritma ini mampu menghasilkan pohon keputusan yang optimal dengan mempertimbangkan faktor-faktor seperti informasi gain dan perhitungan rasio gain. Selain itu, kami juga akan menerapkan prediksi paket return berbasis website. Dengan menggunakan platform berbasis website, perusahaan dapat mengintegrasikan sistem prediksi ini ke dalam infrastruktur mereka, sehingga memungkinkan penggunaan yang mudah dan aksesibilitas yang lebih luas. Tujuan dari penelitian ini adalah untuk mengembangkan suatu sistem prediksi paket return yang akurat dan efisien menggunakan metode decision tree dengan algoritma C4.5. Dengan adanya sistem ini, perusahaan dapat mengidentifikasi dan mengantisipasi paket yang berpotensi untuk mengalami pengembalian, sehingga dapat mengambil tindakan yang diperlukan untuk mengurangi dampak negatif dari pengembalian paket.

2. Metode Penelitian



Gambar 1. Tahapan Penelitian

Pada Gambar 1, dapat dilihat tahapan dari penelitian yang akan dibuat. penelitian diawali dengan melakukan mencari dataset dilanjutkan dengan preprocessing data, menghitung entropy serta gain sehingga dapat menghasilkan urutan node yang tepat sehingga akhirnya akan menjadi pohon keputusan.

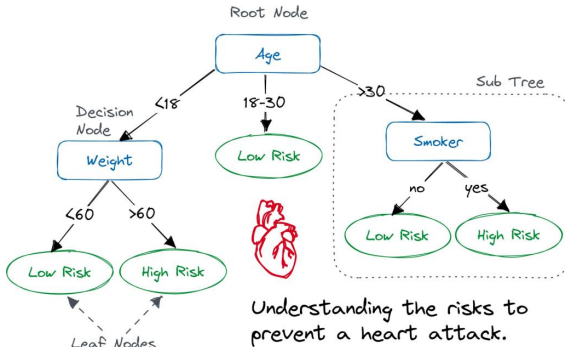
2.1 Pengumpulan Data

Pengumpulan dataset dilakukan dengan cara mengumpulkan data historis pengembalian paket dari data paket pengembalian perusahaan. Data ini harus mencakup berbagai atribut yang relevan, seperti karakteristik produk, informasi pelanggan, alasan pengembalian, dan status pengembalian.

2.2 Pemilihan Algoritma

Algoritma yang akan digunakan pada penelitian ini adalah Decision Tree dengan menggunakan algoritma C4.5 yang akan dibandingkan

a. Decision Tree



Gambar 2. Cara Kerja dari Metode Decision Tree

Decision Tree merupakan sekian dari banyak metode di dalam data mining, dimana data mining adalah proses mengekstrak data dengan menggunakan metode dan algoritma tertentu untuk menghasilkan informasi yang lebih berguna sehingga dapat menjadi dasar dalam pengambilan keputusan (Firmansyah & Yulianto, 2021). Diperkenalkan pertama kali oleh J. Ross Quinlan di awal tahun 1980 dengan nama ID3. Decision tree adalah flowchart yang memiliki struktur pohon dimana masing-masing nodenya memiliki atribut, setiap cabang menggambarkan keluaran dari pengujian dan setiap daun memiliki label (Han et al., 2012). Pada tahun-tahun berikutnya barulah Quinlan memperkenalkan C4.5 sebagai algoritma supervised learning yang lebih baru sebagai penyempurnaan dari ID3.

b. Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma pembentukan pohon keputusan yang dikembangkan oleh Ross Quinlan pada tahun 1993. Algoritma ini merupakan pengembangan dari algoritma sebelumnya yang dikenal sebagai ID3 (Iterative Dichotomiser 3). Algoritma C4.5 memiliki keunggulan dalam mengatasi atribut kontinu, menangani data yang hilang, dan menggunakan gain rasio untuk pemilihan atribut pembagi. Ini membantu dalam menghasilkan pohon keputusan yang lebih akurat dan lebih efisien. Dalam konteks prediksi paket return, algoritma C4.5 dapat digunakan untuk membangun pohon keputusan yang dapat memprediksi kemungkinan pengembalian paket berdasarkan atribut-atribut yang relevan, seperti karakteristik produk, informasi pelanggan, dan alasan.

2.3 Pemodelan Decision Tree

Salah satu tahapan dalam proses penelitian yang sudah dijelaskan pada gambar 1 adalah preprocessing dan pemodelan data. Pada penelitian ini, penulis akan melakukan perancangan model data. Pada tahap metode yang akan digunakan adalah sebagai berikut

a. Perhitungan Gain

MSE merupakan rata rata kuadrat kesalahan yang dihitung dengan menjumlahkan semua kesalahan atau error prediksi yang dihasilkan oleh suatu model kemudian dikuadratkan dan

membaginya dengan jumlah periode prediksi [10]. Berikut merupakan persamaan matematis dari MSE :

$$GAIN(S, A) = Entropy(S) - \sum_{i=1}^n Entropy(S_i) \tag{1}$$

b. Menghitung Nilai Entropy

MAE merupakan selisih absolut antara nilai prediksi dan nilai yang sebenarnya kemudian akan dihitung rata rata dari selisih tersebut. MAE dapat memberikan gambaran rata rata prediksi dengan nilai yang sebenarnya. Berikut merupakan persamaan matematis dari MAE :

$$Entropy(S) = \sum_{i=1}^n P_i * Log_2 P_i \tag{2}$$

3. Hasil dan Diskusi

3.1 Perancangan Model

Pada penelitian ini, jumlah total data yang digunakan adalah sebanyak 50591 baris memuat kolom yang meliputi Id, alamat, provinsi, kabupaten, kecamatan/kota, umur, nilai transaksi, ekspedisi, dll. Dataset yang digunakan merupakan dataset historis penjualan dari suatu perusahaan.



Gambar 3. Dataset yang Digunakan dalam Perancangan Model

Data yang sudah dipersiapkan kemudian akan dicari nilai entropy dan gain sehingga dapat melanjutkan step selanjutnya yaitu pembuatan node. Berikut hasil yang telah didapatkan setelah melakukan perhitungan entropy dan gain.

		Jumlah kasus	Tidak	Ya	Entropy	Info gain	Split info	Gain Ration
Umur	<= 20 tahun	12165	2163	10002	-0,86305143	-26769,91948	0,753200875	-35541,54059
	>= 20 tahun	40418	6910	33508	-0,876016718			
Alamat	<= 80 character	14495	2776	11719	-0,838110367	-22133,22729	0,854115163	-25913,63348
	>= 80 character	36624	6050	30574	-0,887236729			
Ekspedisi	JNE	8384	1878	6506	-0,783998024	-2691,441352	0,864409849	-3113,617174
	JNT	29344	4536	24806	-0,908440394			
	SAD	289	239	50	-0,302688602			
	SICEPAT	434	177	257	-0,569873286			
Nilai COD	<= 200k	96174	5952	21567	-2,275208468	-2691,441352	0,864409849	-3113,617174
	>= 200k	14740	2838	9235	-1,73740876			
Total		50593	8727	41863	-0,873486703			

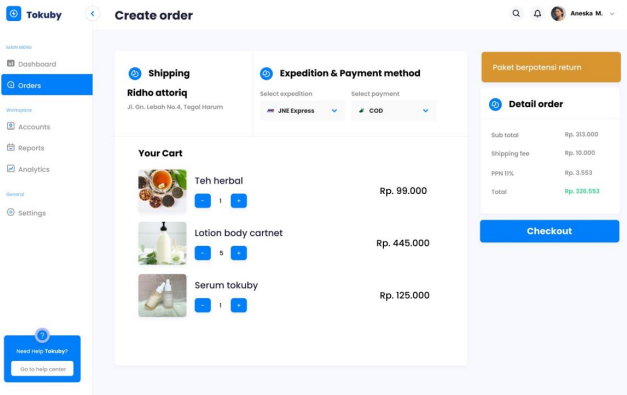
Gambar 4. Hasil Perhitungan Entropy dan Gain

3.2 Integrasi Backend dengan Frontend

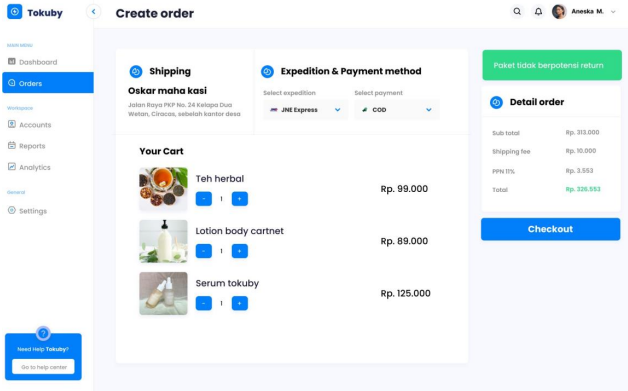
Pada penelitian ini, peneliti menggunakan framework laravel sebagai backend sedangkan Next.Js yang digunakan dalam pembuatan antarmuka sistem sehingga dapat mempermudah seller dalam melakukan prediksi paket return.

Tabel 1. Integration Testing

No	Integration Testing	Status
1	Pengujian pada input order	Succeed
2	Pengujian tombolquantity item tambah dan kurang	Succeed
3	Pengujian pada tombol checkout	Succeed
4	Kesesuaian hasil prediksi dengan prediksi model	Succeed



Gambar 5. Tampilan Antarmuka Sistem Berpotensi Return



Gambar 6. Tampilan Antarmuka Sistem Tidak Berpotensi Return

4. Kesimpulan

Berdasarkan percobaan yang dilakukan pada algoritma Decision Tree dengan menggunakan algoritma C4.5 dapat disimpulkan bahwa metode prediksi paket return menggunakan decision tree dengan algoritma C4.5 mampu memberikan prediksi yang cukup akurat. Model yang dikembangkan berhasil memperoleh tingkat akurasi yang memadai dalam memprediksi apakah suatu paket akan mengalami pengembalian atau tidak. Hal ini dapat memberikan manfaat yang signifikan bagi perusahaan pengiriman dalam mengoptimalkan pengelolaan dan pengendalian proses pengiriman, perlu diperhatikan bahwa untuk mendapatkan prediksi yang akurat dengan menggunakan Decision Tree, prediksi bergantung pada seberapa banyak variasi data yang digunakan dan seberapa baik preprocessing data sebelum data diuji. Secara keseluruhan, penelitian ini menggambarkan bahwa metode decision tree dengan algoritma C4.5 dan penerapannya dalam lingkungan berbasis website dapat menjadi alternatif yang efektif untuk memprediksi paket return.

Daftar Pustaka

- [1] M. D. Kahfi, F. R. Umbara, and H. Ashaury, "Prediksi Pengangguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor," *Informatics and Digital Expert (INDEX)*, vol. 4, no. 2, pp. 75–80, Jan. 2023, doi: 10.36423/index.v4i2.913.
- [2] Y. Dwiyantri, A. Herdiani, and S. Y. Puspitasari, "Memprediksi Status Berlangganan Klien Bank Pada Kampanye Pemasaran Langsung Dengan Menggunakan Metode Klasifikasi Dengan Algoritma C5.0" *eProceedings of Engineering*.
- [3] View of Prediksi Customer Churn dengan Algoritma Decision Tree C4.5 Berdasarkan Segmentasi Pelanggan untuk Mempertahankan Pelanggan pada Perusahaan Retail. <https://jurnal.instiki.ac.id/index.php/jurnalresistor/article/view/219/91>
- [4] A. Nurzahputra, A. R. Safitri, and M. A. Muslim, "Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree" *PRISMA, Prosiding*
- [5] Y. R. Putri, "Prediksi Pola Kecelakaan Kerja Pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 Dan C5.0

Isolation Forest dengan Exploratory Data Analysis pada Anomaly Detection untuk Data Transaksi

I Made Sudarsana Taksa Wibawa^{a1}, Anak Agung Istri Ngurah Eka Karyawati^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹taksawibawa@gmail.com
²eka.karyawati@unud.ac.id

Abstract

Managing value of data is one of the key aspects of presenting analysis for decision making support in various cases. One of such method is by managing detecting anomaly in the data. This research focuses on implementing Isolation Forest result of anomaly detection. This method is used on transaction dataset from Kaggle with about more than 500.000 records. The result this research shows that Isolation Forest used in the dataset have 0.899 in accuracy, 0.00649 in precision, 0.504 in recall, and 0.013 in F1 score.

Keywords: Isolation Forest, iForest, Anomaly Detection

1. Pendahuluan

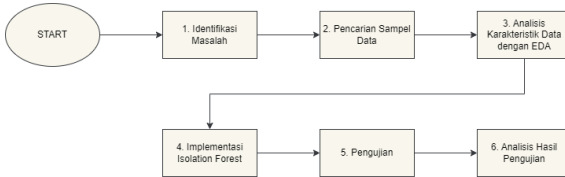
Dalam era Big Data, organisasi menghadapi tantangan baru dalam mengelola dan menganalisis volume data yang besar dan beragam. Kunci dari nilai yang terkandung dalam Big Data adalah kemampuan untuk mengubah data tersebut menjadi wawasan yang bernilai bagi organisasi. Wawasan ini menjadi landasan untuk pengambilan keputusan yang lebih baik, strategi yang lebih efektif, dan keunggulan kompetitif. Karakteristik Big Data sering disebut sebagai 5 V's of Big Data, yaitu volume, variety, velocity, viability, dan value. Dari kelima karakteristik tersebut, value memiliki peran utama dalam menentukan valuasi dari data yang digunakan untuk pengambilan keputusan suatu organisasi. Salah satu tantangan dalam menjaga valuasi dari suatu data adalah adanya anomaly terhadap data tersebut. Anomali data merujuk pada kejadian atau pola yang tidak biasa, tidak terduga, atau tidak sesuai dengan harapan yang ada dalam data. Anomali dapat muncul dalam berbagai bentuk, seperti outlier yang mencolok, kesalahan pengukuran, atau perubahan drastis dalam pola data. Keberadaan anomali dapat menyebabkan bias, ketidaktepatan, atau kesalahan dalam analisis dan pengambilan keputusan. Dalam pengembangan sistem, terdapat beberapa metode yang dapat dilakukan untuk melakukan deteksi anomali data. Salah satu metode yang digunakan adalah machine learning.

Dalam deteksi anomali data, machine learning telah menjadi salah satu metode yang populer dan efektif. Metode ini melibatkan penggunaan algoritma dan teknik pembelajaran mesin untuk mengidentifikasi pola atau perilaku yang tidak biasa dalam data. Salah satu pendekatan dari machine learning yang populer digunakan untuk deteksi anomali adalah unsupervised learning approach. Menurut penelitian yang dilakukan oleh Bakumenko dan rekan mengenai perbandingan deteksi anomali data menggunakan beberapa model machine learning, di dapatkan bahwa pada ranah pendekatan secara unsupervised, Isolation Forest memiliki nilai akurasi tertinggi [2]. Isolation Forest merupakan algoritma unsupervised learning yang didasarkan atas Algoritma pohon keputusan atau decision tree. Secara garis besar cara kerja algoritma ini akan memisahkan anomali dari dataset yang digunakan dengan cara membagi suatu data dengan pohon keputusan sampai anomali benar – benar terpisah [1].

Pada penelitian ini, penulis akan menguji performa dari Isolation Forest untuk deteksi anomali data sehingga dapat digunakan untuk decision making yang lebih baik ke depannya.

2. Metode Penelitian

2.1 Alur Penelitian



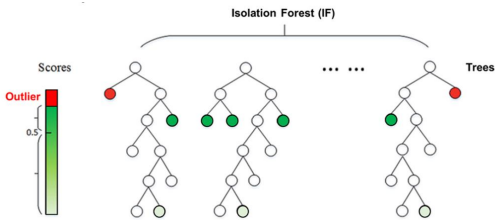
Gambar 1. Alur Penelitian

Alur kerangka penelitian di atas memiliki penjelasan sebagai berikut:

- Identifikasi Masalah: tahapan awal dari penelitian, di mana peneliti mengidentifikasi konsep dari deteksi anomali data melalui studi literatur.
- Pencarian Sampel Data: pada tahapan ini, data dicari dengan menggunakan data sekunder yang berasal dari Kaggle.com. Data yang digunakan merupakan data transaksi.
- Implementasi Isolation Forest: pada tahapan ini, peneliti melakukan pemodelan terhadap data yang telah dianalisis dan dibersihkan dengan algoritma Isolation Forest. Pada tahapan ini akan memanfaatkan bahasa pemrograman Python dan Visual Studio Code untuk implementasinya.
- Pengujian: pada tahapan ini, peneliti akan melakukan training dan testing data terhadap model yang telah dibuat dengan pembagian data yaitu 80% untuk training dan 20% untuk testing model.
- Analisis Hasil Pengujian: pada tahapan ini, peneliti akan menganalisis performa dari Isolation Forest untuk deteksi anomali data. Tolak ukur pengujian yang digunakan meliputi Accuracy, Precision, Recall, F1 Score, dan Average Anomaly Score.

2.2 Isolation Forest

Isolation Forest atau iForest merupakan algoritma unsupervised learning approach, yang didasarkan pada konsep pemisahan atau isolasi data anomali dari data normal. Isolation Forest menggunakan pendekatan berbasis pohon (tree-based approach) untuk membagi data menjadi subgrup yang semakin kecil, dengan tujuan memisahkan data anomali dengan cepat. Pada saat membangun pohon dalam Isolation Forest, data dipecah secara acak dan dipisahkan dengan menggunakan fitur-fitur yang ada. Anomali cenderung membutuhkan jumlah pemisahan yang lebih sedikit dibandingkan dengan data normal. Dengan demikian, ketika suatu data ditempatkan pada posisi yang lebih awal dalam pohon (memiliki jarak lebih pendek dari root), kemungkinan besar data tersebut merupakan anomali. Proses isolasi dan pemisahan berulang-ulang pada pohon-pohon yang dibangun membentuk model deteksi anomali. Kemudian, dengan menggunakan model ini, data baru dapat diuji untuk melihat apakah data tersebut termasuk anomali atau tidak.



Gambar 2. Ilustrasi Isolation Forest Model. Sumber: datrics.ai

3. Hasil dan Diskusi

Berdasarkan alur penelitian dan juga studi literatur telah dijelaskan, maka didapatkan hasil penelitian sebagai berikut.

3.1. Pencarian Sampel Data

Penelitian ini menggunakan data transaksi dari Kaggle.com yang di-publish oleh Edgar Lopez-Rojaz [4]. Format data tersebut dapat dilihat pada **Tabel 1** di bawah.

Tabel 1. Format Dataset

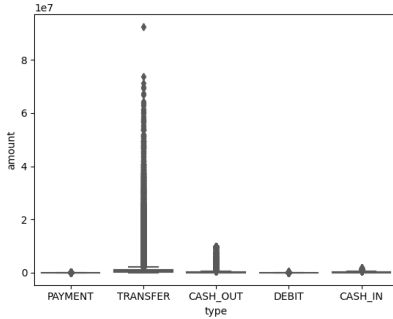
Key	Description
step	Waktu transaksi itu dilakukan, 1 step bernilai 1 jam
type	Tipe transaksi yang dilakukan
amount	Nominal dari transaksi yang dilakukan (CASH_IN, CASH_OUT, DEBIT, TRANSFER, PAYMENT)
nameOrig	Pelanggan yang memulai transaksi
oldbalanceOrg	Saldo awal pelanggan sebelum transaksi
newbalanceOrg	Saldo baru pelanggan setelah transaksi
nameDest	Pelanggan Tujuan transaksi
oldbalanceDest	Saldo awal pelanggan tujuan
newbalanceDest	Saldo akhir pelanggan tujuan
isFraud	Indikator transaksi normal atau tidak
isFlaggedFraud	Indikator normal atau tidak normal jika nominal suatu transaksi melebihi 200.000

3.2. Analisis Karakteristik Data dengan EDA

Proses analisis karakteristik pada dataset yang digunakan dengan EDA akan dilakukan dalam beberapa tahap yaitu sebagai berikut

a. Melihat Distribusi Transaksi Berdasarkan Tipe Transaksi

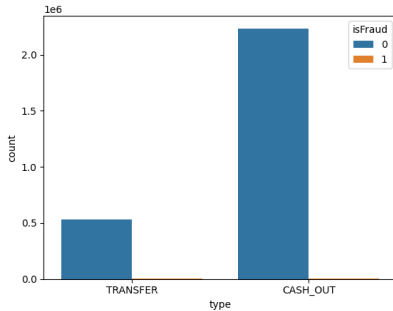
Yang pertama dilakukan adalah melihat persebaran dari data transaksi pada dataset tersebut. Dapat dilihat bahwa ternyata ada beberapa transaksi TRANSFER dan CASH_OUT yang terlihat janggal berdasarkan visualisasi ini.



Gambar 3. Distribusi Transaksi Berdasarkan Tipe Transaksi

b. Melihat Persebaran Anomali pada TRANSFER dan CASH_OUT

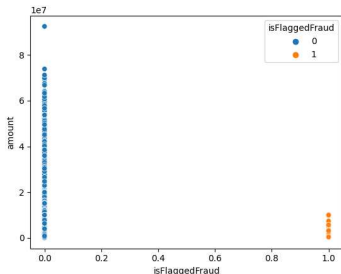
Kemudian dilakukan pengecekan jumlah transaksi yang anomali pada TRANSFER dan CASH_OUT. Ditemukan bahwa sebanyak 4116 CASH_OUT dan 4097 TRANSFER yang dinyatakan anomali.



Gambar 4. Visualisasi Data Normal dan Anomali

c. Analisis Pengaruh Fitur isFlaggedFraud pada Dataset

Setelah mengetahui bahwa persebaran anomali ada pada tipe TRANSFER dan CASH_OUT, perlu dilakukan pengecekan apakah anomali – anomali tersebut memiliki nominal transaksi melebihi 200.000 atau tidak. Setelah dilakukan pengecekan, dinyatakan bahwa fitur tersebut hanya dimiliki oleh transaksi TRANSFER, sehingga perlu dilakukan analisis seberapa pengaruhnya fitur tersebut. Pertama dicek terlebih dahulu apakah seluruh transaksi yang memiliki fitur ini memiliki transaksi lebih dari 200.000 atau tidak.



Gambar 5. Distribusi Transaksi Berdasarkan Fitur isFlaggedFraud

Berdasarkan visualisasi tersebut, dapat dikatakan bahwa ternyata fitur ini tidak berpengaruh pada keanormalan data. Hal ini dikarenakan transaksi yang dilabeli tidak memiliki fitur isFlaggedFraud juga mempunyai nominal transaksi yang melebihi 200.000, sehingga fitur ini dapat dihilangkan dari groundtruth dan hanya isFraud yang akan digunakan.

d. Pembersihan Data (Data Cleaning)

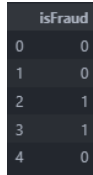
Langkah selanjutnya adalah membersihkan dataset tersebut agar dapat digunakan untuk perhitungan. Mengingat Isolation Forest akan bekerja lebih baik jika menerima data berupa numerik, maka kita perlu menghilangkan beberapa fitur yang tidak relevan. Pada kasus ini, kita akan menghilangkan fitur nameOrig, nameDest, dan isFlaggedFraud karena dianggap tidak berpengaruh signifikan pada perhitungan berdasarkan analisis yang telah dilakukan. Kemudian, kita akan mengubah nilai TRANSFER dan CASH_OUT tadi menjadi numerik yaitu 0 dan 1 untuk mempermudah perhitungan. Selanjutnya, dilakukan pelabelan dari nilai – nilai fitur yang beresla sama - sama bernilai kosong atau 0. Pada kasus ini yang memiliki nilai 0 adalah fitur oldbalanceDest dengan newbalanceDest dan oldbalanceOrg, dengan newbalanceOrg. Sehingga hasil dari pembersihan tadi akan seperti berikut.

step	type	amount	oldbalanceOrg	newbalanceOrg	oldbalanceDest	newbalanceDest	isFraud	
2	1	0	181.00	181.0	0.0	-1.0	-1.00	1
3	1	1	181.00	181.0	0.0	21182.0	0.00	1
15	1	1	229133.94	15325.0	0.0	5083.0	51513.44	0
19	1	0	215310.30	705.0	0.0	22425.0	0.00	0
24	1	0	311685.89	10835.0	0.0	6267.0	2719172.89	0

Gambar 6. Hasil Data Cleaning Terhadap Dataset

e. Pembentukan Ground Truth

Untuk ground truth di sini akan menggunakan fitur isFraud sebagai label untuk nanti digunakan sebagai salah satu parameter untuk melakukan proses prediksi.

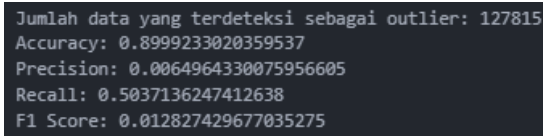


	isFraud
0	0
1	0
2	1
3	1
4	0

Gambar 7. Ground Truth

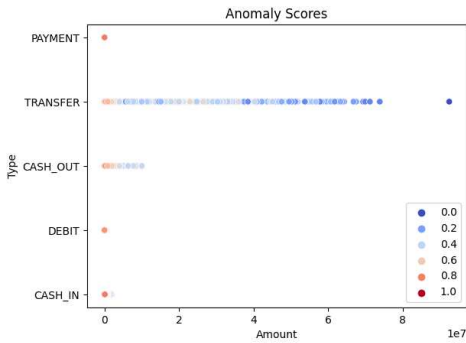
3.3. Implementasi Isolation Forest

Pada tahapan ini, peneliti akan mengimplementasikan Isolation Forest untuk deteksi anomali pada data. Hal pertama yang perlu dilakukan adalah menentukan outlier fraction yang akan digunakan. Nilai ini akan menjadi nilai parameter contamination pada model atau nilai prediksi dari persebaran anomali. Ini dapat dilakukan dengan membagi jumlah transaksi yang dinyatakan dianggap anomali dengan jumlah transaksi normal. Didapatkan nilai contamination sebesar 0.1. Kemudian kita akan membagi dataset tersebut menjadi dua bagian, 80% untuk training dan 20% untuk testing. Kemudian data training tersebut akan dimasukan pada model Isolation Forest. Waktu training ini memakan waktu sebesar 3 menit 52 detik. Berdasarkan pengujian, diperoleh bahwa nilai akurasi dari model yang dibuat mencapai akurasi sebesar 0.899, precision sebesar 0.00649, recall sebesar 0.503, dan F1 score sebesar 0.0128.



```
Jumlah data yang terdeteksi sebagai outlier: 127815  
Accuracy: 0.899233020359537  
Precision: 0.0064964330075956605  
Recall: 0.5037136247412638  
F1 Score: 0.012827429677035275
```

Gambar 8. Hasil Pengujian Model



Gambar 9. Visualisasi Distribusi Skor Anomali

4. Kesimpulan

Berdasarkan penelitian, ditemukan bahwa Isolation Forest memiliki performa yang baik. EDA sangat membantu dalam menganalisis dataset yang digunakan dan mengeliminasi fitur – fitur yang tidak berpengaruh secara signifikan sehingga memudahkan perhitungan pada model. Isolation Forest memiliki performa yang baik dengan total anomali yang ditemukan sebanyak 1646. Algoritma ini juga mampu mencapai anomaly score sebesar 0.47%.

Daftar Pustaka

- [1] P. Patil, "What is exploratory data analysis?," Medium, <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15> (accessed Jun. 11, 2023).
- [2] E. S. Nugroho, "Catatan Belajar Anomaly Detection Menggunakan Algoritma Isolation Forest," Medium, <https://edistywn.medium.com/catatan-belajar-anomaly-detection-menggunakan-algoritma-isolation-forest-4e4d13e61c3d> (accessed Jun. 11, 2023).
- [3] A. Bakumenko and A. Elragal, "Detecting anomalies in financial data using machine learning algorithms," *Systems*, vol. 10, no. 5, p. 130, 2022. doi:10.3390/systems10050130.
- [4] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson. "PaySim: A financial mobile money simulator for fraud detection". In: The 28th European Modeling and Simulation Symposium-EMSS, Larnaca, Cyprus. 2016
- [5] S. V. Daele and G. Janssenswillen, "Identifying the steps in an exploratory data analysis: A process-oriented approach," *Lecture Notes in Business Information Processing*, pp. 526–538, 2023. doi:10.1007/978-3-031-27815-0_38

Halaman ini sengaja dibiarkan kosong

Perancangan Antar Muka Wearable Sistem Bagi Atlet

Ni Wayan Sani Utari Dewi^{a1}, Cokorda Pramatha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹saniutari089@student.unud.ac.id
²cokorda@unud.ac.id

Abstract

Technological developments have had a huge impact on various areas of life, including the world of sports. Today, athletes are leveraging various technologies to improve their performance and gain a better competitive advantage. Athletes must comply with the rules of the game and comply with the rules and regulations that apply to their game. hours of sleep is one of the most important factors in maintaining the stamina of athletes, according to the coach's directions. This study aims to make it easier for coaches to monitor the sleep hours of athletes being mentored. In order to improve the athlete's heart rate monitoring, an application was formed in the form of a prototype. The method used in this study is the mid-fidelity prototype which is the development of a prototype design model. Mid-fidelity prototype is a method for initial design, for purposes after detailed design and usability validation.

Keywords: Athlete, Wearable, Prototype

1. Pendahuluan

Perkembangan teknologi telah memberikan dampak yang sangat besar pada berbagai bidang kehidupan, termasuk dunia olahraga. Saat ini, para atlet memanfaatkan berbagai teknologi untuk meningkatkan performa mereka dan mendapatkan keunggulan kompetitif yang lebih baik. Salah satu inovasi yang muncul adalah sistem wearable yang dirancang khusus untuk atlet.

Atlet adalah orang yang bermain olahraga kompetitif dan memiliki keunggulan dalam olahraga tertentu. Mereka berlatih fisik dan mental secara intensif untuk mencapai performa terbaik dalam kompetisi [1]. Selain dedikasi dan kerja keras dalam latihan, atlet juga perlu menjaga nutrisi yang baik, pemulihan yang baik dan menjaga pola hidup yang seimbang untuk meningkatkan performanya. Mereka juga harus mematuhi peraturan permainan dan mematuhi peraturan dan ketentuan yang berlaku untuk permainan mereka. jam tidur merupakan salah satu factor yang sangat penting dalam mempertahankan stamina dari atlet, sesuai dengan arahan pelatih. Dalam hal pemantauan jam tidur atlet oleh pelatih, penting memilih system pemantau dan pengawasan detak jantung yang tentu mengacu pada jam tidur yang bekerja secara realtime. Melalui aplikasi, pelatih bisa memantau jam tidur atlet.

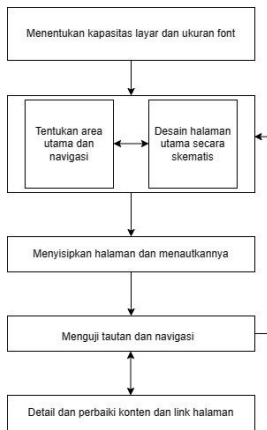
Berdasarkan permasalahan tersebut, user selalu dilibatkan dalam pengembangan sistem ini, metode yang digunakan pada rancangan design antar muka sistem ini yaitu mid-fidelity prototype, Mid-fidelity penjelasan mengacu pada representasi desain atau prototipe yang lebih mendekati tampilan dan fungsionalitas produk akhir, tetapi masih memiliki beberapa kesederhanaan dan keterbatasan. Mid-fidelity sering digunakan sebagai tahap perantara dalam pengembangan produk, di mana desainer dan pengembang dapat mengevaluasi konsep, interaksi, dan fitur-fitur utama sebelum berinvestasi dalam pengembangan yang lebih lanjut. Tahap dari pengembangan sistem ini usability testing, yaitu usability dari sistem perlu diuji agar peneliti bisa menyimpulkan apakah sistem ini dapat digunakan oleh target penggunaanya. Penelitian pengembangan sistem yang menggunakan metode mid-fidelity prototype belum banyak digunakan sebelumnya. Dalam rangka meningkatkan pemantauan detak jantung atlet, dibentuklah sistem yang berupa prototype

ini dapat memanfaatkan teknologi sensor yang terdapat pada smartband untuk melacak dan menganalisis detak jantung atlet secara real-time.

2. Metode Penelitian

2.1 Metode

Metode yang digunakan dalam penelitian ini yaitu mid-fidelity prototype yang merupakan pembangunan model design prototype. Mid-fidelity prototype adalah suatu metode untuk design awal, untuk keperluan setelah desain detail dan validasi kegunaan. Metode ini menyajikan informasi terperinci tentang navigasi, fungsionalitas, konten, dan tata letak, tetapi dalam bentuk skematis ("kerangka gambar").



Gambar 1. Metode Prototype

a. Menentukan Kemampuan Layer dan Ukuran Font

Prototipe harus mensimulasikan resolusi layar dan ukuran layar dimana situs terakhir akan dilihat. Ukuran font mempengaruhi berapa banyak ruang yang dibutuhkan menu, berapa banyak informasi akan masuk ke zona konten, dan berapa banyak item tingkat atas yang bisa masuk ke dalam menu horizontal. Font utama yang akan ditentukan adalah untuk item menu, judul halaman, subtittel, konten tekstual, label bidang, dan tombol.

b. Tentukan Area Utama dan Navigasi

Langkah ini dilakukan secara paralel dengan definisi halaman beranda. Langkah ini terdiri dari menentukan tata letak halaman dan perilaku dasar, khusus untuk unsur-unsur yang tercantum dalam besar tekstual.

c. Desain Tampilan Awal Secara Skematis

langkah ini dilakukan secara paralel dengan menentukan zona utama dan pendekatan navigasi. Ini terdiri dari memasukkan halaman ke dalam prototipe untuk halaman beranda,

dan mendefinisikan halaman secara skematis, khususnya untuk menu dan sistem navigasi.

d. Menyisipkan Halaman dan Menautkannya

Pertama, halaman tujuan dalam prototipe disisipkan untuk setiap link utama di beranda (e.g. menu elements), dan untuk item tertaut di sekunder halaman. Hal terpenting saat ini adalah memberi judul pada halaman sekunder segera sebagai pengingat akan tujuannya. Halaman kosong adalah terutama placeholder untuk menautkan ke halaman dari hyperlink di tempat lain Prototipe. Belum perlu untuk mulai mendesain konten terperinci dari setiap halaman. Kedua, jika tidak dilakukan secara otomatis oleh alat prototyping, maka menu yang ditentukan di halaman beranda dapat ditambahkan ke setiap halaman sekunder, membuka konfigurasi sub-level yang sesuai sesuai kebutuhan. Beberapa halaman sekunder tentu saja mungkin menggunakan menu yang berbeda dari beranda halaman. Ketiga, jika tidak dilakukan secara otomatis oleh alat prototyping, menu item dapat disediakan dengan hyperlink dan tautan lain ke halaman target. Terakhir, konten skematik ditambahkan ke halaman. Pada titik ini bisa cukup untuk menunjukkan zona utama dengan menggambar kotak teks atau persegi Panjang dengan judul deskriptif.

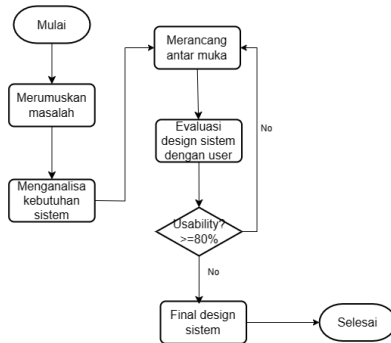
e. Menguji Tautan dan Navigasi

Setelah struktur navigasi dibuat, prototipe diuji informal dengan menavigasi dengan hyperlink melalui jalur yang khas. Ini langkah membantu perancang menentukan apakah navigasi lancar dan efisien. Setelah itu, tautan dan halaman dapat diubah atau ditambahkan jika perlu.

f. Detail dan Sempurnakan Tampilan Halaman

setelah "kerangka" atau kerangka navigasi desain telah dibuat selesai pada langkah sebelumnya, isi halaman dan interaksi adalah Dihilangkan. Dalam formulir dan dialog interaktif, label bidang dan bidang sampel adalah juga ditambahkan. Grafik mungkin juga berguna. Objek dapat disejajarkan kapan saja. Ini adalah tahap menyempurnakan penyesuaian objek dan zona

Pada penelitian ini melakukan *usability testing* untuk menguji kelayakan dari design sistem. Usability kualitas suatu produk merupakan kualitas suatu produk menggunakan efektivitas, efisiensi, dan kepuasan pengguna terhadap fungsinya, dengan cara menggunakan kuisioner efisiensi dan efektivitas untuk menguji sistem [3].



Gambar 2. Alur Penelitian

Tahap pertama dalam penelitian sistem ini adalah merumuskan masalah yang terjadi pada jam tidur atlet. Pada tahap kedua peneliti melakukan Analisa kebutuhan design sistem, Analisa kebutuhan fungsi yang berhubungan dengan pengguna Ketika mengakses sistem. Tahap ketiga yaitu perancangan design antar muka sistem. Perancangan dilakukan dengan mengacu pada Analisa kebutuhan fungsi untuk pengguna. Hasil dari tahapan ini berupa design antar muka yang sudah memiliki beberapa fungsi sesuai dengan tahap ke-2. Selanjutnya pengujian aspek usability, dengan usability testing dari design sistem. Jika nilai usability dari design sistem sudah memenuhi kebutuhan, maka tahap terakhir pada alur penelitian ini adalah finalisasi design antar muka sistem.

2.2 Kebutuhan Fungsional

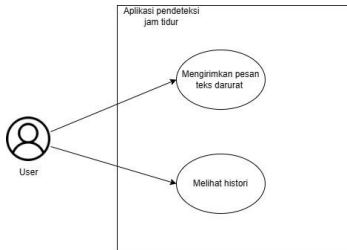
Kebutuhan fungsional adalah persyaratan yang harus dipenuhi oleh suatu sistem atau produk agar dapat berfungsi sesuai dengan tujuan.

Tabel 1. Kebutuhan Fungsional

Komponen	Fungsi
Mengirimkan pesan teks darurat	Berfungsi untuk mengirimkan pesan reminder notifikasi pada perangkat mobile jika terdeteksi jam tidur atlet tidak teratur
Melihat histori	Berfungsi untuk melihat histori pendeteksi jam tidur atlet

2.3 Use Case Diagram

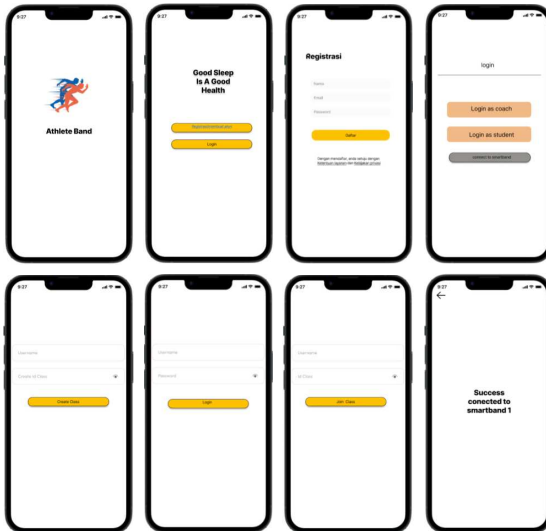
Use case diagram adalah jenis diagram yang digunakan dalam pemodelan perangkat lunak untuk menggambarkan interaksi antara aktor (pengguna atau sistem lain yang berinteraksi dengan sistem yang sedang dianalisis) dan fungsi (use case) yang dapat dilakukan oleh sistem [5]. Berisi penjelasan tentang tahapan-tahapan penelitian yang menggambarkan urutan logis untuk mendapatkan hasil penelitian yang sesuai dengan harapan dan gambaran sistem.

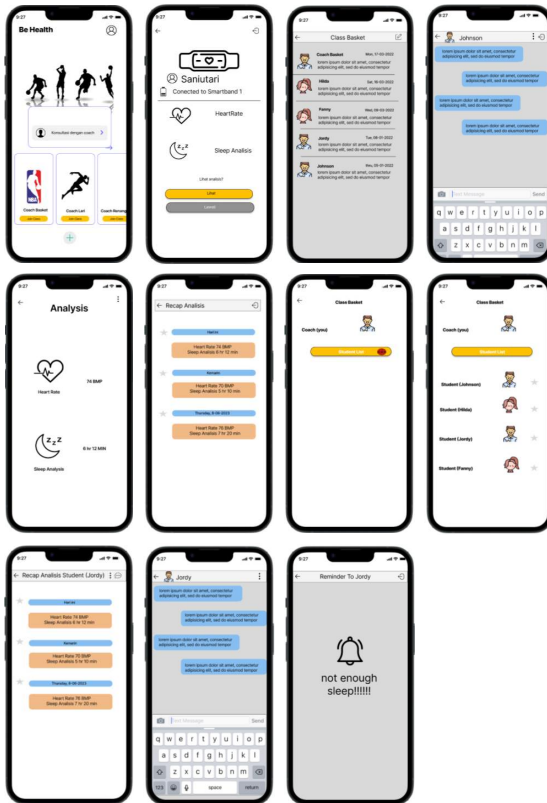


Gambar 3. Use Case Diagram

3 Hasil dan Pembahasan

Hasil penelitian berupa tampilan antar muka sistem deteksi jam tidur atlet, dari hasil uji coba smartband jika sensor yang terdapat pada smartband mendeteksi jam tidur stlet, dan sistem antarmuka yang memberi akses bagi pelatih memantau jam tidur atlet. Pada rancangan design antar muka sistem ini bisa digunakan oleh dua tipe pengguna, yaitu dari pihak atlet dan pelatih, juga memberikan kesempatan bagi pelatih pada saat memberi peringatan saat atlet mengurangi jam tidur. Bagian ini memuat hasil dan pembahasan penelitian dan dapat disajikan dalam bentuk deskripsi, grafik atau gambar.





Gambar 4. Rancangan Desain Antar Muka

3.1 Frame Open

Pada bagian frame ini terdapat icon atlet berlari, memilih icon tersebut karena berlari juga termasuk salah satu cabang olah raga. Pemilihan warna pada icon tersebut karena warna merah dominan [4], berkesan kecepatan dan aksi, menstimulasi detak jantung, nafas, sedangkan warna biru juga tidak kalah menarik karena menurut beberapa orang merasa lebih produktif di dalam ruangan berwarna biru.

3.2 Frame Login

Pada bagian frame ini menggunakan design dan tata letak yang tak jauh berbeda yaitu dominan warna kuning yang warna yang paling menarik perhatian, lebih terang membuat focus teralihkan dari background putih polos, yang berbeda berada pada frame connect smartband yaitu dengan warna yang cenderung lebih gelap dibandingkan dengan simbol login yang lain, bertujuan pada bagian ini lebih penting untuk menentukan login sebagai kelompok user yang berbeda, dan juga connect ke smartband.

3.3 Frame Beranda

Pada bagian ini terdapat beberapa icon olah raga sesuai penamaannya.

3.4 Frame Profile

Pada bagian ini berisi beberapa icon, icon baterai menunjukkan presentase baterai smartband, icon jam menunjukkan bahwa pada bagian frame ini memastikan tersambung jaringan smartband ke smartphone, icon heart rate menunjukkan icon yang biasa digunakan saat menunjukkan heart rate, icon sleep analisis juga menunjukan icon yang sesuai. Pemilihan warna pada simbol 'lihat' dan 'lewati', terdapat perbedaan warna yang berbeda, bertujuan agar user lebih terfokus pada warna kuning terang.

3.5 Frame Notif

Pada bagian recap analisis ini, menggunakan pemilihan warna yang mirip dengan frame lainnya karena menyesuaikan pilihan warna.

3.6 Frame Forum Class

Pada bagian ini berisi beberapa icon yang berwarna, agar terlihat lebih menarik.

3.7 Frame Alarm Reminder

Pada bagian ini tidak berisi warna-warna yang mencolok, bertujuan agar user tidak jenuh memperhatikan warna-warna terang yang terdapat pada frame yang lain.

4 Kesimpulan

Menyesuaikan pembahasam dan hasil, tentang jam tidur atlet yang merupakan salah satu hal penting bagi fisik dan mental secara intensif untuk mencapai performa terbaik dalam kompetisi atlet, dan juga kendala bagi para pelatih bisa ikut memantau memastikan jam tidur yang cukup bagi atlet, maka dibuatkan perancangan antar muka wearable system bagi atlet dalam memberikan upaya mempermudah atlet mematuhi aturan jam tidur. Pada sistem ini memerlukan dua komponen utama yang berupa smartband dan smartphone sebagai sarana menjalankan sistem. Metode yang digunakan pada perancangan sistem ini adalah mid-fidelity prototype, yang merupakan metode yang dapat digunakan metode untuk design awal, untuk keperluan setelah desain detail dan validasi kegunaan.

Daftar Pustaka

- [1] Umar, U., & Fadillah, N. (2019). Pengaruh latihan daya tahan aerobik terhadap kemampuan menembak. *Jurnal Performa Olahraga*, 4(2), 92-100.
- [2] Engelberg, D., & Seffah, A. (2002). A framework for rapid mid-fidelity prototyping of web sites. *Usability: Gaining a Competitive Edge*, 203-215.
- [2] N. I. Kurniati, A. Rahmatulloh, and D. Rahmawati, "Perbandingan Performa Algoritma Koloni Semut Dengan Algoritma Genetika – Tabu Search Dalam Penjadwalan Kuliah,"

- Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 17, 2019, doi: 10.24114/cess.v4i1.11387.
- [3] Sudiarsa, I. W., & Wiraditya, I. G. B. (2020). Analisis Usability Pada Aplikasi Peduli Lindungi Sebagai Aplikasi Informasi Dan Tracking Covid-19 Dengan Heuristic Evaluation. *INTECOMS: Journal of Information Technology and Computer Science*, 3(2), 354-364.
- [4] Monica, M., & Luzar, L. C. (2011). Efek warna dalam dunia desain dan periklanan. *Humaniora*, 2(2), 1084-1096.
- [5] Destriana, R., Kom, M., Husain, S. M., Kom, S., Handayani, N., Kom, M., ... & Kom, S. (2021). *Diagram UML Dalam Membuat Aplikasi Android Firebase" Studi Kasus Aplikasi*

Implementasi Algoritma KNN untuk Memprediksi Performa Siswa Sekolah

I Made Ryan Prana Dhita^{a1}, Gst. Ayu Vida Mastrika Giri^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹ryanprana555@gmail.com
²vida@unud.ac.id

Abstract

One of the factors that influences students graduation rates is their performance in learning. Predicting graduation rates based on student performance has the benefit of analyzing academically underperforming students and providing support to students who face difficulties in the learning process. There are several factors to consider in predicting students' graduation rates, such as academic grades, attitudes, and social factors. However, these factors alone are not sufficient to effectively predict students' performance, and educators also struggle to identify which factors affect students' performance. To predict the performance of school students, the K-Nearest Neighbor (KNN) method is utilized. The K-Nearest Neighbor method is often used in classifying students' performance due to its simplicity and ability to produce significant and competitive results. In this research, the prediction of students' graduation rates is carried out using the KNN method. The results of implementing the prediction of students' performance using the KNN method can serve as a reference for students to improve their achievements and assist educators in considering future teaching materials.

Keywords: KNN, K-Nearest Neighbor, Students Performance, Student

1. Pendahuluan

Saat ini, pendidikan adalah salah satu bagian terpenting dan esensial bagi kehidupan masyarakat. Ini digunakan untuk meningkatkan dan meningkatkan pertumbuhan individu secara akademis dan finansial. Seorang individu yang berpendidikan harus berkontribusi tidak hanya untuk keluarganya tetapi juga untuk masyarakat dan komunitas. Ini semua dicapai melalui pembelajaran yang tepat [1].

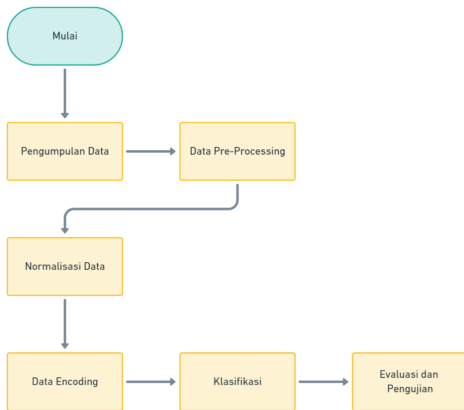
Untuk dapat menghasilkan SDM (sumber daya manusia) yang cakap, berwawasan, kompetitif dan kreatif, lembaga pendidikan diminta untuk menyelenggarakan pendidikan yang bermutu dan berkualitas bagi peserta didiknya [2]. Sehingga untuk mengimplementasikannya dilakukan beberapa perubahan kurikulum guna meningkatkan kualitas pada tingkatan pendidikan tentunya untuk mengetahui berhasil atau tidaknya kurikulum yang telah dibuat diperlukan sebuah evaluasi berupa prediksi hasil belajar siswa di sekolah serta faktor-faktor yang mempengaruhinya.

Maka untuk mewujudkan hal tersebut dibutuhkan suatu alat analisis berupa komputasi cerdas untuk menganalisis bagaimana kinerja siswa, faktor mana yang akan mempengaruhi kinerja mereka, dengan cara apa siswa dapat membuat kemajuan, dan apakah siswa memiliki potensi untuk tampil lebih baik [2].

Pada penelitian ini akan dilakukan klasifikasi performa siswa berdasarkan lingkungan serta hasil akademik siswa dengan K-Nearest Neighbor (KNN) algoritma ini digunakan karena kemampuannya dalam menyederhanakan perhitungan algoritma dan mengoptimalkan waktu. KNN adalah metode yang bekerja dengan mengelompokkan data baru berdasarkan jarak mereka dengan data lainnya. Prediksi akan dilakukan menggunakan algoritma K-Nearest Neighbor (KNN).

2. Metode Penelitian

Penelitian dimulai dari tahapan pengumpulan dataset lalu pemilihan algoritma yaitu algoritma K-Nearest Neighbor, melakukan normalisasi data serta data encoding, klasifikasi lalu evaluasi hasil dari data yang telah diproses, gambar 1. Tahapan penelitian merupakan ilustrasi langkah-langkah metode yang akan dikerjakan.



Gambar 1. Tahapan Penelitian

2.1. Data dan Sumber Data

Penelitian ini menggunakan dataset Student Performance terkait dengan latar belakang serta nilai siswa sekolah yang didapat dari platform kaggle tepatnya pada halaman berikut <https://www.kaggle.com/datasets/rkiattisak/student-performance-in-mathematics>

Dataset Student Performance terkait yang telah diperoleh nantinya pada penelitian ini dibagi menjadi dua yaitu data training sebanyak 85% serta testing 15% dari keseluruhan data yang telah diperoleh

2.2. Analisis Data

Tahap ini bertujuan untuk memastikan integritas data sehingga nantinya tidak menimbulkan masalah pada proses data training atau pelatihan data. data latar belakang siswa sekolah yang digunakan sejumlah 1000 data yang terbagi menjadi 4 kelas yaitu Sangat Kurang, Kurang, Cukup Baik, dan Baik. pada tahap ini dilakukan penghilangan data atau seleksi data dengan menyeleksi atribut apa saja yang diperlukan dalam dataset yang telah didapat terdapat 9 atribut yaitu atribut

gender, race/ethnicity, parental level of education, lunch, test preparation course, math score, reading score, dan average.

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	avarage	Peformance
0	female	group B	bachelor's degree	standard	none	72	72	74	72.666667	Cukup Baik
1	female	group C	some college	standard	completed	69	90	88	82.333333	Baik
2	female	group B	master's degree	standard	none	90	95	93	92.666667	Baik
3	male	group A	associate's degree	free/reduced	none	47	57	44	49.333333	Sangat Kurang
4	male	group C	some college	standard	none	76	78	75	76.333333	Cukup Baik
...

Gambar 2. Sampel Data Student Performance

2.3. Pre-Processing

Pada tahapan ini tujuan utama dari data pre-processing adalah. proses identifikasi, pembenaran, dan/atau penghapusan data yang tidak akurat, tidak lengkap, tidak konsisten, atau tidak relevan dari sebuah dataset. Tujuan utama dari data cleansing adalah memastikan bahwa data yang digunakan dalam analisis atau pemodelan adalah data yang berkualitas tinggi, dapat diandalkan, dan tepat [3].

a. Normalisasi Data

Normalisasi data melibatkan mengubah skala data agar sesuai dengan rentang atau standar yang ditetapkan. Normalisasi dapat melibatkan pemetaan data ke rentang yang spesifik atau transformasi data untuk menghilangkan bias atau asumsi tertentu. Metode min-max normalization digunakan untuk normalisasi data pada penelitian ini dengan mentransformasikan setiap nilai dalam rentang data ke rentang baru antara 0 dan 1 [2]. Nilai terendah dalam data akan menjadi 0, sedangkan nilai tertinggi akan menjadi 1. Berikut merupakan persamaan matematis dari min-max normalization.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Keterangan:

- X = nilai asli dari data yang akan dinormalisasi
- X_{normalized} = nilai yang dinormalisasi dari X
- X_{min} = nilai terendah dalam rentang data
- X_{max} = nilai tertinggi dalam rentang data

b. Data Encoding

Pada tahapan ini proses mengubah variabel kategori atau kualitatif menjadi bentuk numerik agar dapat digunakan dalam analisis dan pemodelan data. Hal ini diperlukan karena sebagian besar algoritma dan model yang digunakan dalam analisis data memerlukan data dalam bentuk numerik. berikut merupakan perubahan data dari masing-kelas dan atribut yang digunakan.

Tabel 1. Contoh Kelas dan Kategori Nilai

Kelas	Nilai Kategori
Kategori 1	0
Kategori 2	1

Kelas	Nilai Kategori
Kategori 3	2
Kategori 4	3

Tabel 2. Contoh Atribut dan Kategori Nilai

Atribut	Nilai Kategori
Parental level	0 - 4
Lunch	0 - 1
Test Preparation	0 - 1

Tabel 1 menunjukkan nilai kelas yang telah melalui proses data encoding pada dataset yang telah di processing dengan data encoding, sedangkan tabel 2 menunjukkan nilai atribut yang telah melalui proses encoding

c. Klasifikasi

Pada tahap ini dataset dipisah menjadi 2 bagian secara acak yaitu: data pelatihan (data training) dan data pengujian (data testing). Data pelatihan (data Training): Data pelatihan digunakan untuk melatih model KNN. pada tahap ini dataset dipisah menjadi 2 bagian secara acak yaitu: data pelatihan (data training) dan data pengujian (data testing). Data pelatihan (data Training): Data pelatihan digunakan untuk melatih model KNN. Data pelatihan terdiri dari contoh-contoh data yang memiliki label kelas yang diketahui. Model KNN akan menggunakan data ini untuk belajar pola dan hubungan antara fitur-fitur yang ada dalam dataset dengan label kelas yang sesuai. sedangkan data pengujian (data Testing) merupakan Data pengujian digunakan untuk menguji kinerja model KNN yang telah dilatih [2][4].

Data pengujian terdiri dari contoh-contoh data yang juga memiliki label kelas yang diketahui, tetapi label kelasnya disembunyikan dari model saat proses pengujian. Model KNN akan memprediksi label kelas untuk data pengujian berdasarkan informasi yang telah dipelajari selama tahap pelatihan. Pemisahan data dilakukan secara acak untuk memastikan bahwa data pelatihan dan data pengujian mewakili dataset secara proporsional. Biasanya, pemisahan data dilakukan dengan membagi dataset secara acak menjadi dua bagian, di mana sebagian besar data digunakan untuk pelatihan (misalnya 70-80% dari dataset) dan sisanya digunakan untuk pengujian (misalnya 20-30% dari dataset). Namun, perbandingan ini dapat bervariasi tergantung pada ukuran dataset dan kebutuhan spesifik [2][5][6]

3. Hasil dan Pembahasan

3.1. Data Encoding

Fitur pada dataset yang memiliki atribut data kualitatif akan diubah kedalam bentuk numerik dengan menggunakan metode one shot encoding pada setiap data sehingga dengan demikian data akan dapat diolah ke dalam perhitungan yang akan dilakukan. Berikut merupakan gambar encoding dataset student performance.

	parental level of education	lunch	test preparation course	math score	reading score	writing score	Performance
0	bachelor's degree	standard	none	72	72	74	3
1	some college	standard	completed	69	90	88	4
2	master's degree	standard	none	90	95	93	4
3	associate's degree	free/reduced	none	47	57	44	1
4	some college	standard	none	76	78	75	3
...
995	master's degree	standard	completed	88	99	95	4
996	high school	free/reduced	none	62	55	55	2
997	high school	free/reduced	completed	59	71	65	2
998	some college	standard	completed	68	78	77	3
999	some college	free/reduced	none	77	86	86	4

Gambar 3. Dataset Sebelum Melakukan Data Encoding

	parental level of education	lunch	test preparation course	math score	reading score	writing score	Performance
0	1	1	1	72	72	74	3
1	4	1	0	69	90	88	4
2	3	1	1	90	95	93	4
3	0	0	1	47	57	44	1
4	4	1	1	76	78	75	3
...
995	3	1	0	88	99	95	4
996	2	0	1	62	55	55	2
997	2	0	0	59	71	65	2
998	4	1	0	68	78	77	3
999	4	0	1	77	86	86	4

Gambar 4. Dataset Sesudah Melakukan Data Encoding

3.2. Normalisasi data

Nilai siswa sekolah bervariasi dapat menyulitkan sistem dalam melakukan perhitungan sehingga normalisasi data perlu dilakukan pada nilai hasil ujian siswa sekolah yang pada penelitian ini metode min-max normalization digunakan melakukan normalisasi pada data. Metode ini dilakukan apa setiap fitur dari data variabel yang memerlukan normalisasi pada data dengan mengubah skala data ke dalam rentang antara. berikut merupakan normalisasi yang dilakukan pada fitur nilai dari dataset.

	parental level of education	lunch	test preparation course	math score	reading score	writing score	Performance
0	1	1	1	2	2	2	3
1	4	1	0	2	2	2	4
2	3	1	1	3	3	3	4
3	0	0	1	1	1	1	1
4	4	1	1	2	2	2	3
...
995	3	1	0	3	3	3	4
996	2	0	1	2	2	2	2
997	2	0	0	2	2	2	2
998	4	1	0	2	2	2	3
999	4	0	1	2	2	2	4

Gambar 5. Mix-Max Normalization

3.3. Pengujian dan Evaluasi

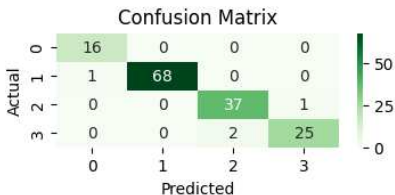
Pada tahap ini, pengujian dan evaluasi dilakukan untuk mengklasifikasikan sejumlah data uji menggunakan model klasifikasi yang telah dibentuk sebelumnya. Pengujian dilakukan dengan jumlah (k-5), di mana k adalah jumlah data uji yang digunakan. Dalam kasus ini, k-5 adalah data uji yang digunakan karena dalam beberapa percobaan yang telah dilakukan k-5 memiliki precision serta recall data yang lebih baik dibandingkan k lainnya dalam pengujian dalam penelitian ini.

```

    Hasil prediksi:
    [2 3 1 2 3 3 1 4 1 2 2 4 4 2 1 2 4 2 4 2 2 3 2 1 2 2 1 2 4 3 4 4 2 2 4 4
    2 3 3 3 2 2 3 3 2 4 2 4 2 1 2 1 2 1 3 2 2 3 2 2 3 2 3 2 2 3 2 2 2 3 4 3 2
    2 2 3 3 1 3 3 4 2 4 2 3 2 4 3 2 3 2 2 2 3 3 2 2 2 2 1 4 2 1 3 3 2 2 2 4 2
    3 1 2 2 3 3 3 4 2 2 3 4 4 3 2 3 3 3 2 2 2 1 4 4 2 2 2 2 4 3 2 2 2 3 1 1 1
    4 4]
    
```

Gambar 6. Hasil Prediksi dengan Nilai K-5

Total data uji yang digunakan adalah 150, yang terbagi menjadi empat kelas berdasarkan distribusinya. Kelas "baik" memiliki 25 data, kelas "cukup baik" memiliki 37 data, kelas "kurang" memiliki 68 data, dan kelas "sangat kurang" memiliki 16 data. Untuk mengukur kinerja model klasifikasi yang telah dibentuk, digunakan metode confusion matrix. Confusion matrix adalah metode yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model dengan nilai sebenarnya dari data uji. Gambar 7 menunjukkan hasil dari confusion matrix yang diperoleh setelah melakukan pengujian terhadap model klasifikasi yang telah dibentuk sebelumnya.



Gambar 7. Mix-Max Normalization

Keterangan :

- 0 = Sangat Kurang
- 1 = Kurang
- 2 = Cukup Baik
- 3 = Baik

Berdasarkan confusion matrix seperti yang terlihat pada Gambar 7, kita dapat mengetahui bahwa pada kelas "Sangat Kurang" dengan 16 sampel data, semua data terprediksi dengan benar sebagai kelas "Sangat Kurang" dan tidak ada data yang salah diprediksi sebagai kelas lain. Pada kelas "Kurang" dengan 69 sampel data, 67 data terprediksi dengan benar sebagai kelas "Kurang", namun terdapat 1 data yang salah diprediksi sebagai kelas "Sangat Kurang" dan 1 data yang diprediksi sebagai kelas "Cukup Baik". Pada kelas "Cukup Baik" dengan 38 sampel data, 37 data terprediksi dengan benar sebagai kelas "Cukup Baik", tetapi terdapat 1 data yang salah diprediksi

sebagai kelas "Baik". Pada kelas "Baik" dengan 27 sampel data, 25 data terprediksi dengan benar sebagai kelas "Baik", namun terdapat 2 data yang salah diprediksi sebagai kelas "Cukup Baik".

Dari confusion matrix tersebut, akurasi model klasifikasi dapat dihitung berdasarkan persentase nilai yang diprediksi dengan benar dibagi dengan total jumlah data uji [4].

$$\text{Average Accuracy} = \frac{\sum_{l=1}^l \frac{tp_l + tn_l}{tp_l + fp_l + fn_l + tn_l}}{l} \quad (2)$$

Setelah melakukan perhitungan tersebut akan di dapatkan data akurasi setiap kelas dalam system yang telah dibuat seperti yang terlihat pada tabel berikut

Tabel 3. Kinerja Model Klasifikasi

Kelas	Accuracy	Precision	Recall
Sangat Kurang	1.0000	1.0000	1.0000
Kurang	0.9933	0.9709	0.9963
Cukup Baik	0.9867	0.9880	0.9855
Baik	0.9733	0.9580	0.9734
Rata-rata	0.9667	0.9569	0.9677

4. Kesimpulan

Penelitian ini menyimpulkan bahwa telah berhasil dibuat sebuah aplikasi untuk melakukan seleksi data penerima beasiswa dengan menggunakan algoritma KNN. Evaluasi algoritma KNN menggunakan metode confusion matrix menunjukkan hasil rata-rata akurasi dari metode KNN ini sebesar 96%. Ini menunjukkan bahwa implementasi algoritma KNN pada prediksi performa siswa sekolah memiliki akurasi yang cukup tinggi dengan demikian klasifikasi performa dengan KNN dapat dijadikan sebagai system pendukung untuk membantu guru serta instansi pendidikan untuk mempertimbangkan langkah-langkah yang dapat dilakukan untuk pemajukan pendidikan di Indonesia kedepannya

Daftar Pustaka

- [1]. Kurniawati, F. N. A., 2022. Meninjau Permasalahan Rendahnya Kualitas Pendidikan Di Indonesia Dan Solusi. *Academy of Education Journal*, Volume 13, p. 13.
- [2]. Daru Prasetyawan, .. R. G., 2022. K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi. *Jurnal Informatika Sunan Kalijaga*, Volume 7, p. 12.
- [3]. Ni Made Rika Padeswari Kusuma, L. G. A., 2022. Implementasi Algoritma K-Nearest Neighbor (K-NN) dalam Deteksi Dini Penyakit Hepatitis C. *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, Volume 1, p. 8. [3]
- [4]. Deti Fusvita1), A. 2., 2021. Penerapan Algoritma KNN (K-Nearest Neighbour) Dalam Klasifikasi Data Pinjaman ANggota Koprasi. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya*, Volume 03, p. 5.
- [5]. Sebastian Raschka, V. M., 2019. *Python Machine Learning*. Third Edition ed. s.l.:Packt.
- [6]. Mutiara Ayu Banjarsari, H. I. B. A. F., 2015. Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4. *jurnal Ilmu Komputer (KLIK)*, Volume 02, p. 15.
- [7]. Lubis, A. R., Lubis, M., & Khowarizmi, A.-. (2020). Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics*, 9(1), 326–338. <https://doi.org/10.11591/eei.v9i1.1464>

Halaman ini sengaja dibiarkan kosong

Penerapan RMS Contrast sebagai Penentu Citra Terbaik Berdasarkan Tingkat Kontras

Shiennyta Florensia Adiriyanto^{a1}, I Gede Arta Wibawa^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹shiennyflorensia@gmail.com
²gede.arta@unud.ac.id

Abstract

This research centers on the objective of identifying the finest image from a burst shot through the utilization of Root Mean Square (RMS) contrast as a guiding metric. The primary focus lies in selecting the visually captivating image by assessing the contrast levels present in each individual shot. By conducting calculations of the RMS contrast values, the optimal image can be determined. The core purpose of this study is to employ RMS contrast as a reliable criterion for selecting the most outstanding image from a burst shot, thereby guaranteeing the production of high-quality visuals, and aiding in the organization and decluttering of smartphone galleries. This research holds significance in addressing the growing need for efficient and effective image selection methods, ensuring that users can effortlessly identify and showcase the best images captured in burst mode. By embracing the power of RMS contrast analysis, individuals can confidently curate their image collections with exceptional visual content while optimizing limited storage space on their smartphones.

Keywords: Burst Shot, RMS contrast, Contrast Level, Best Image, Storage Space.

1. Pendahuluan

Dalam dunia fotografi modern, perkembangan teknologi telah menghadirkan kemudahan dalam pengambilan foto dalam jumlah besar dalam waktu singkat. Fitur burst shot pada kamera digital dan smartphone memungkinkan pengguna untuk menangkap serangkaian foto berurutan hanya dengan menahan tombol rana. Mode ini sangat bermanfaat dalam situasi di mana pengguna ingin menangkap momen yang cepat berubah atau objek yang bergerak dengan cepat. Namun, setelah sesi pemotretan selesai, muncul tantangan baru yaitu memilih gambar terbaik dari sekumpulan foto burst tersebut. Hal ini dapat memakan waktu yang cukup lama, karena pengguna tentu ingin memastikan gambar yang dipilih memiliki kejelasan, ketajaman, dan kontras yang optimal.

Kontras warna merupakan sebuah teknik desain yang digunakan untuk membuat warna pada gambar lebih mencolok. Kontras ditentukan oleh perbedaan dalam warna dan tingkat kecerahan dari objek yang satu dengan yang lainnya dalam jangkauan pandang yang sama [1]. Tingkat kontras warna ini bervariasi dari tinggi hingga rendah, tergantung pada posisi kedua warna pada roda warna. Semakin jauh kedua posisi dari warna tersebut pada roda warna, maka semakin tinggi juga tingkat kontras warnanya. Tidak hanya untuk memperindah foto, kontras warna juga diperlukan terutama bagi orang yang memiliki masalah pada indera penglihatannya untuk dapat membedakan objek dengan lebih mudah [2].

Dengan begitu, salah satu metode yang dapat digunakan dalam mengevaluasi dan memilih gambar terbaik berdasarkan tingkat kontrasnya adalah metrik Root Mean Square (RMS) contrast. RMS contrast adalah metode kuantitatif yang mengukur tingkat kontras dalam sebuah gambar [3]. Metrik ini menghitung variasi intensitas piksel dalam gambar dengan cara menghitung akar kuadrat dari rata-rata selisih kuadrat antara intensitas piksel dengan nilai intensitas rata-rata. Dengan membandingkan nilai RMS contrast dari masing-masing foto dalam burst shot, akan

dapat ditemukan gambar dengan kontras yang optimal, di mana detail-detail gambar terlihat jelas dan tajam.

Selain memberikan hasil yang memuaskan secara visual, penggunaan metrik RMS contrast dalam pemilihan gambar dari burst shot juga membantu mengurangi waktu dan usaha yang diperlukan untuk memilih gambar secara manual. Dengan memanfaatkan metode ini, pengguna dapat dengan mudah menemukan gambar terbaik yang menonjol dan memiliki kualitas visual yang tinggi, sekaligus mengoptimalkan ruang penyimpanan pada galeri smartphone mereka.

2. Metode Penelitian

Metode penelitian memiliki peran yang sangat penting dalam menentukan hasil dari suatu penelitian. Hal ini disebabkan oleh pentingnya pendekatan yang tepat dalam melaksanakan penelitian tersebut. Berikut merupakan urutan dari metode penelitian yang akan dilaksanakan.

2.1. Pengumpulan Data

Pada penelitian ini, akan digunakan data primer berupa foto beberapa objek yang difoto menggunakan burst shot, kemudian disimpan menggunakan format JPG file. Objek yang akan difoto di antaranya adalah tin kotak, buku, boneka, tempat minum dan bunga, dengan sebanyak 5 foto untuk setiap objek tersebut. Data dari pengambilan foto-foto objek dapat dilihat pada **Tabel 1**. Foto akan diberikan nama dengan format "nomor objek'_nomor foto".

2.2. Pre-Processing Data

Data akan diubah menjadi grayscale terlebih dahulu, sehingga penghitungan kontras warna akan digabungkan menjadi satu. Gambar grayscale merupakan gambar yang hanya memiliki warna hitam dan putih dengan skala intensitas 0-255. Angka 0 berarti hitam, 255 berarti putih, dan 1-254 merupakan abu-abu dengan intensitas kecerahan yang berbeda-beda. Semakin kecil angka tersebut, maka semakin gelap pula pikselnya.

Gambar RGB merupakan gambar berwarna yang berdasar dari warna merah, hijau, dan biru dengan skala intensitas 0-255 pada setiap warnanya. Dengan 3 dasar warna tersebut, didapatkan sebanyak 16.777.216 varian campuran warna yang dapat memberatkan proses pengolahan gambar. Dengan begitu, pelaksanaan pre-processing data dengan menggunakan grayscale sangatlah penting [4].

2.3. RMS Contrast


























Root Mean Square (RMS) contrast merupakan salah satu cara yang dapat digunakan dalam menentukan tingkat kontras gambar. Rumus dari RMS contrast untuk gambar $f(x, y)$ dapat dilihat pada persamaan (1) di bawah. Semakin tinggi nilai dari RMS contrast tersebut, maka semakin tinggi pula tingkat kontras pada gambar. Dengan $f(x, y)$ sebagai intensitas piksel dari gambar dua dimensi berukuran X dan Y. Sedangkan, μ_f merupakan rata-rata intensitas dari semua nilai piksel pada gambar. Rumus dari μ_f dapat dilihat pada persamaan (2) [5].

$$f_{rms} = \sqrt{\frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \{f(x, y) - \mu_f\}^2} \quad (1)$$

$$\mu_f = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} f(x, y) \quad (2)$$

Foto yang sudah dilakukan pre-processing dengan grayscale akan dihitung RMS contrast menggunakan rumus tersebut. Foto dengan nilai RMS tertinggi akan menjadi foto dengan tingkat kontras terbaik dan akan ditampilkan.

Tabel 1. Data Foto Objek

Objek	Foto 1	Foto 2	Foto 3	Foto 4	Foto 5
Tin Kotak (1)					
Buku (2)					
Boneka (3)					
Tempat Minum (4)					
Bunga (5)					

* estimasi foto terbaik

3. Hasil dan Diskusi

Berdasarkan metode penelitian yang sudah dipaparkan di atas, berikut merupakan hasil penghitungan dari setiap foto pada masing-masing objek sesuai dengan **Tabel 1**.

```
RMS contrast for 1_1.jpg: 46.32469683818406  
RMS contrast for 1_2.jpg: 45.55963659056988  
RMS contrast for 1_3.jpg: 45.13111158317552  
RMS contrast for 1_4.jpg: 48.53592335230029  
RMS contrast for 1_5.jpg: 49.877362845482416  
  
The best image is: 1_5.jpg
```

Gambar 1. Hasil Penghitungan Objek 1 (Tin Kotak)

Pada objek (1) tin kotak, didapatkan foto 5 sebagai foto terbaik dengan nilai RMS 49.877, dilanjutkan dengan foto 4 dengan nilai RMS 48.536. Foto dengan nilai RMS terburuk adalah foto 3 dengan RMS 45.131. Foto terbaik yang didapatkan, yaitu foto 5 sesuai dengan estimasi foto terbaik,

```
RMS contrast for 2_1.jpg: 55.67153590809478  
RMS contrast for 2_2.jpg: 56.33791575391052  
RMS contrast for 2_3.jpg: 56.42718742325045  
RMS contrast for 2_4.jpg: 56.50383423743579  
RMS contrast for 2_5.jpg: 55.350802154514895  
  
The best image is: 2_4.jpg
```

Gambar 2. Hasil Penghitungan Objek 2 (Buku)

Pada objek (2) buku, didapatkan foto 4 sebagai foto terbaik dengan nilai RMS 56.504, dilanjutkan dengan foto 3 dengan nilai RMS 56.427. Foto dengan nilai RMS terburuk adalah foto 5 dengan RMS 55.351. Foto terbaik yang didapatkan, yaitu foto 4 tidak sesuai dengan estimasi foto terbaik yang merupakan foto 2,

```
RMS contrast for 3_1.jpg: 36.2307858027897  
RMS contrast for 3_2.jpg: 34.76491936142141  
RMS contrast for 3_3.jpg: 34.16883362580635  
RMS contrast for 3_4.jpg: 36.56450304505334  
RMS contrast for 3_5.jpg: 36.2666462737494  
  
The best image is: 3_4.jpg
```

Gambar 3. Hasil Penghitungan Objek 3 (Boneka)

Pada objek (3) boneka, didapatkan foto 4 sebagai foto terbaik dengan nilai RMS 36.564, dilanjutkan dengan foto 5 dengan nilai RMS 36.267. Foto dengan nilai RMS terburuk adalah foto 3 dengan RMS 34.169. Foto terbaik yang didapatkan, yaitu foto 4 sesuai dengan estimasi foto terbaik.

```
RMS contrast for 4_1.jpg: 46.11874252291498
RMS contrast for 4_2.jpg: 45.59398686702662
RMS contrast for 4_3.jpg: 45.45977853431225
RMS contrast for 4_4.jpg: 46.01089614658642
RMS contrast for 4_5.jpg: 45.9550917182774

The best image is: 4_1.jpg
```

Gambar 4. Hasil Penghitungan Objek 4 (Tempat Minum)

Pada objek (4) tempat minum, didapatkan foto 1 sebagai foto terbaik dengan nilai RMS 46.119, dilanjutkan dengan foto 4 dengan nilai RMS 46.011. Foto dengan nilai RMS terburuk adalah foto 3 dengan RMS 45.460. Foto terbaik yang didapatkan, yaitu foto 1 sesuai dengan estimasi foto terbaik.

```
RMS contrast for 5_1.jpg: 47.91176073142965
RMS contrast for 5_2.jpg: 48.10882587074554
RMS contrast for 5_3.jpg: 49.69945322238286
RMS contrast for 5_4.jpg: 49.60001472094993
RMS contrast for 5_5.jpg: 49.57430561006631

The best image is: 5_3.jpg
```

Gambar 5. Hasil Penghitungan Objek 5 (Bunga)

Pada objek (5) bunga, didapatkan foto 3 sebagai foto terbaik dengan nilai RMS 49.699, dilanjutkan dengan foto 4 dengan nilai RMS 49.600. Foto dengan nilai RMS terburuk adalah foto 1 dengan RMS 47.912. Foto terbaik yang didapatkan, yaitu foto 3 sesuai dengan estimasi foto terbaik.

Berdasarkan hasil-hasil tersebut, pada **Tabel 2** diperlihatkan perbandingan antara estimasi foto terbaik dengan foto terbaik berdasarkan penelitian, besertakan perbandingan RMS dari kedua foto tersebut.

Tabel 2. Perbandingan Estimasi dan Hasil Foto Terbaik

Objek	Foto Terbaik		RMS	
	Estimasi	Hasil	Estimasi	Hasil
(1) Tin kotak	5	5	49.877362845482416	49.877362845482416
(2) Buku	2	4	56.33791575391052	56.50383423743579
(3) Boneka	4	4	36.56450304505334	36.56450304505334
(4) Tempat minum	1	1	46.11874252291498	46.11874252291498
(5) Bunga	3	3	49.69945322238286	49.69945322238286

4. Kesimpulan

Penelitian penentu citra terbaik dengan menggunakan penerapan RMS contrast memberikan hasil yang cukup baik. Penelitian ini mencapai hasil akurasi sebesar 80% yang di mana terdapat kesalahan pemilihan foto terbaik pada 1 objek dari 5 objek yang digunakan. Kekurangan yang terdapat pada hasil penelitian ini adalah belum dibuatnya penerapan ini dalam rupa aplikasi. Kedepannya, diharapkan penelitian ini dapat ditingkatkan dengan penambahan fitur maupun fungsi lainnya untuk mendapatkan hasil yang lebih optimal dan akurat.

Daftar Pustaka

- [1] J. Felix, "Penggunaan Kontras Warna dalam Fotografi," *HUMANIORA*, vol. 1, no. 2, 2010.
- [2] R. Almeida and C. Duarte, "Analysis of automated contrast checking tools," *W4A '20: Proceedings of the 17th International Web for All Conference*, no. 18, p. 1, April 2020.
- [3] S. Rahman, M. . M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP Journal on Image and Video Processing*, no. 35, 18 October 2016.
- [4] P. Kaler, "Study of Grayscale imagein Image processing," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 4, no. 11, p. November, 2016.
- [5] A.-A. Bhuiyan and A. R. Khan, "Image Quality Assessment Employing RMS Contrast and Histogram Similarity," *The International Arab Journal of Information Technology*, vol. 15, no. 6, pp. 984-985, November 2018.

Pengujian Fungsionalitas Sistem Pengamanan Digital Watermarking Kartu Indonesia Sehat Menggunakan Algoritma MSB

Ni Putu Intan Cahyani¹, I Gede Santi Astawa²

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia

¹intancahyani006@student.unud.ac.id

²santiastawa@gmail.com

Abstract

The rapid development of technology makes the dissemination of information also develop, and the use of digital data is increasingly used even in the health sector, one of which is when registering at health institutions online so that when registering it is necessary to scan the Healthy Indonesia Card to register. Sometimes there are onkuns who are not responsible for misusing it to do evil things, so based on these problems, the Healthy Indonesia Card (KIS) that will be used should be inserted with a watermark to minimize unwanted events. Watermark is a method of inserting information into digital data that aims to protect data ownership. One technique that can be used is the steganography technique, which uses the MSB (Most Significant Bit) algorithm. The length of time it takes to insert a watermark on a Healthy Indonesia Card (KIS) depends on how much text will be inserted to become a watermark.

Keywords: Watermarking, MSB Algoritma, Watermark KIS, Steganography

1. Pendahuluan

Penyebaran informasi dan data sudah sangat mengalami perkembangan. Semakin maju perkembangan teknologi maka akan semakin mudah dalam mengakses sebuah informasi dan memberikan informasi. Penggunaan data digital berupa teks, audio, gambar dan video juga akan menjadi sangat pesat, hal ini dapat menyebabkan mengakses berbagai data yang ada di internet salah satunya yaitu data digital yang berupa gambar yang menyebabkan penyalahgunaan oleh oknum-oknum yang tidak bertanggung jawab, salah satunya dapat terlihat pada penggunaan Kartu Indonesia Sehat. Semakin berkembang teknologi semakin berkembang pula bidang kesehatan yang menggunakan teknologi, salah satunya dapat terlihat pada pendaftaran online di berbagai instansi kesehatan, dalam melakukan pendaftaran user atau pengguna biasanya diminta untuk mengupload scan Kartu Indonesia Sehat yang dimiliki user, dimasa seperti ini terdapat banyak oknum-oknum yang menggunakan kesempatan itu untuk berbuat jahat terhadap scan Kartu Indonesia Sehat tersebut, oleh sebab itu untuk meminimalisir penyebaran gambar digital memerlukan sebuah identitas yang melekat pada konten agar tidak mudah diklaim oleh pihak yang tidak bertanggung jawab. Identitas yang melekat pada konten digital gambar disebut dengan watermark image

Watermark merupakan sebuah metode yang digunakan untuk menyisipkan suatu informasi pada suatu digital untuk perlindungan kepemilikan dari data tersebut [1]. Data digital yang telah diberikan watermark atau yang telah diberikan tanda perlindungan disebut dengan data bertanda air (watermarked data). Informasi yang disisipkan yaitu informasi yang tidak merusak data digital yang dilindungi. Dalam melakukan proses watermarking digunakanlah sebuah metode dari steganografi dengan menggunakan MSB (Most Significant Bit). Metode Most Significant Bit adalah sebuah metode yang dilakukan dengan cara menjumlahkan bit dengan menggunakan bit dari sebelah kiri dengan nilai bit yang terbesar [2]. Pada tulisan ini akan meningkatkan keamanan watermark pada gambar dengan menggunakan metode Most Significant Bit yang

diimplementasikan pada gambar *.jpeg *.jpg, dan hasil akhir dari tulisan ini dapat digunakan dengan baik untuk menjaga, melindungi dan menghindari penyalahgunaan terhadap Kartu Indonesia Sehat.

2. Metode Penelitian

2.1 Metode Most Significant Bit (MSB)

Metode Most Significant Bit (MSB) adalah sebuah metode yang digunakan dalam Teknik steganografi. Steganografi adalah metode untuk melakukan penyembunyian informasi atau pesan kedalam media lain seperti citra digital, teks, suara atau video. Metoda ini dilakukan untuk menghindari kecurigaan orang lain [3]. Terdapat dua langkah dalam system Steganografi yaitu proses penyembunyian dan recovery data dari berkas penampung. Penyembunyian data dilakukan dengan mengganti bit-bit data di dalam segmen citra dengan bit-bit data rahasia [4]. Pada metode Most Significant Bit (MSB) pesan disisipkan pada bit ke-1 [5]. Pada data digital digit paling kiri merupakan digit yang paling tertinggi, sedangkan digit yang paling kanan merupakan digit terendah.

2.2 Analisis Kebutuhan

a. Kebutuhan Masukan

1. Media yang digunakan untuk penyisipan adalah *.jpeg *.jpg atau *.png
2. Pesan yang disisipkan *.txt

b. Kebutuhan Keluaran

1. File bergambar bertipe *.jpg
2. Chiphertext hasil enkripsi yang berada di dalam gambar.

c. Kebutuhan Proses

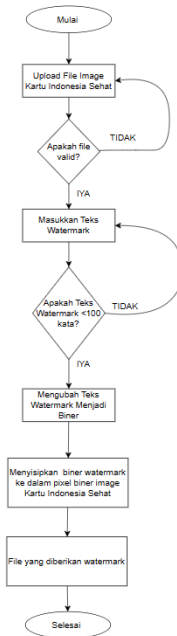
1. Pada sisi pemilih gambar pesan teks di enkripsi untuk mendapatkan ciphertext kemudian ciphertext tersebut disisipkan ke dalam citra gambar secara langsung dan berformat *.JPG dengan metode Most Significant Bit (MSB)

2.3 Gambaran Umum Sistem

Implementasi digital watermarking menggunakan algoritma MSB (Most Significant Bit) dalam tulisan ini yaitu penyisipan watermark dilakukan dengan cara user melakukan upload gambar Kartu Indonesia Sehat menggunakan format gambar yaitu *.jpeg *.jpg dan selanjutnya ada proses pembacaan watermark yaitu user akan mengupload file Kartu Indonesia Sehat dan system akan menampilkan hasil dari watermark.

2.4 Desain Sistem

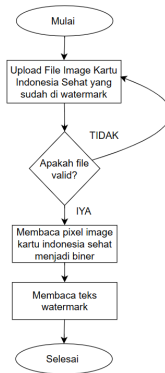
Perancangan sistem dapat dijelaskan dalam flowchart yang digunakan dalam implementasi watermarking dengan menggunakan algoritma MSB (Most Significant Bit)



Gambar 1. Flowchart Sistem Watermark

Penjelasan Flowchart Alur Watermark:

- Input file Kartu Indonesia Sehat yang berekstensi gambar (*.jpg atau *.jpeg)
- Dilakukan pengecekan apakah file yang diupload sudah valid atau tidak
- Membaca file Kartu Indonesia Sehat menjadi biner dan mengubah teks watermark menjadi biner
- Dilanjutkan dengan proses menyisipkan data biner teks watermark ke dalam pixel biner image Kartu Indonesia Sehat dengan menggunakan metode Most Significant Bit
- Menghasilkan output Kartu Indonesia Sehat yang sudah berisi watermark berupa gambar.



Gambar 2. Flowchart Sistem Read Watermark

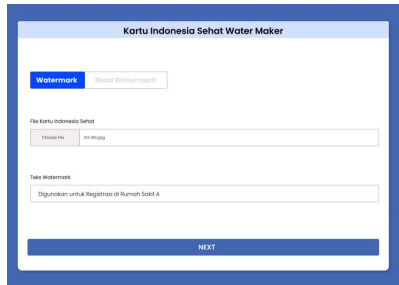
Penjelasan digaram alir pembacaan watermark:

- a. Input berupa file KTP yang berekstensi gambar (*.jpg atau *.jpeg) yang sudah diberi watermark sebelumnya
- b. Proses membaca piksel file KTP menjadi biner kemudian penggabungan biner MSB menjadi teks watermark
- c. Menghasilkan output teks watermark

3. Hasil dan Pembahasan

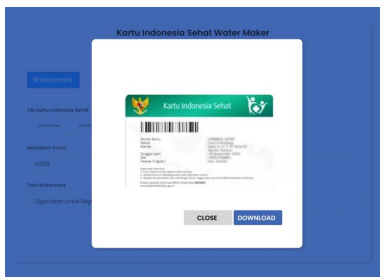
a. Tampilan Awal Sistem

Pada tampilan awal sistem yang telah dibuat terdapat dua halaman yaitu, halaman untuk melakukan watermark pada Kartu Indonesia Sehat, terdapat tiga entri pada halaman watermark ini yaitu masukkan file Kartu Indonesia Sehat yang berekstensi *.jpg, masukkan kunci , dan masukkan pesan rahasia. Sedangkan pada halaman Read Watermark terdapat dua entri yaitu file Kartu Indonesia Sehat dan masukkan kunci



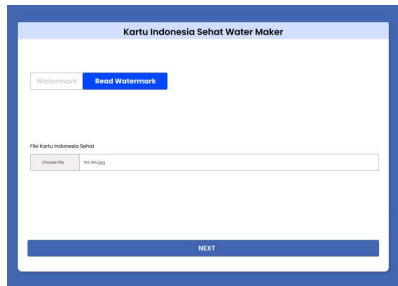
Gambar 3. Antarmuka Watermark

Pada Gambar 3 merupakan tampilan awal pada saat user atau pengguna membuka sistem untuk menambahkan watermark pada Kartu Indonesia Sehat, terdapat entri yang meminta user untuk mengupload gambar Kartu Indonesia Sehat yang berformat (*.jpg atau *.jpeg), kemudian yaitu user menginputkan teks watermark yang akan disisipkan, dengan ketentuan maksimal 99 kata yang dapat disisipkan, jika lebih dari itu sistem akan menolak.



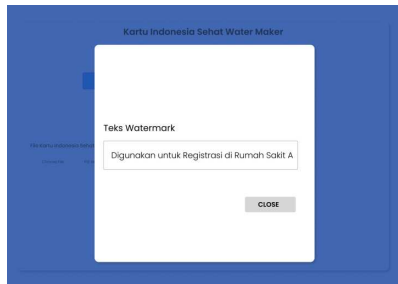
Gambar 4. Antarmuka Hasil Watermark

Pada Gambar 4 merupakan tampilan hasil dari Kartu Indonesia sehat yang sudah diberikan watermark, pada bagian ini user dapat melakukan download pada button download namun apabila user tidak ingin mengunduh user dapat menekan ikon close.



Gambar 5. Antarmuka Read Watermark

Pada Gambar 5 merupakan tampilan dari read watermark, pada bagian ini user diminta untuk mengupload file Kartu Indonesia Sehat yang sebelumnya sudah diberikan watermark dengan format file gambar yaitu (*.jpg atau *.jpeg),kemudian user dapat menekan next untuk mendapatkan hasil dari teks watermark tersebut.



Gambar 6. Antarmuka Hasil Read Watermark

Pada gambar 6 merupakan tampilan antarmuka hasil read watermark, pada tampilan ini sistem akan menampilkan teks yang disisipkan sebelumnya, dan jika user sudah mengetahuinya, user dapat melakukan close dari tampilan diatas.

b. Pengujian Sistem

Metode pengujian sistem menggunakan metode blackbox. Hasil dari pengujian black box adalah sebagai berikut :

Tabel 1. Pengujian Insert Watermark

No	Skenario Pengujian	Test Case	Hasil yang Diharapkan	Hasil Pengujian
1	File Kartu Indonesia Sehat dan teks watermark tidak diisi kemudian klik tombol Save	File Kartu Indonesia Sehat: (kosong) Teks Watermark: (kosong)	Sistem akan menampilkan pesan error, bahwa field tersebut harus di isi	Sesuai Harapan (Valid)
2	Teks Watermark lebih dari 100 karakter	File Kartu Indonesia Sehat: (valid) Teks Watermark: Digunakan untuk melakukan registrasi pendaftaran Rumah Sakit dst.	Sistem akan menampilkan pesan error "Teks Watermark tidak boleh lebih dari 100 karakter"	Sesuai Harapan (Valid)
3	File Kartu Indonesia Sehat diupload sudah sesuai dan jumlah watermark sesuai	File Kartu Indonesia Sehat: (valid file Kartu Indonesia Sehat) Teks watermark: Digunakan untuk registrasi di Rumah Sakit A	Sistem akan menampilkan pop up image Kartu Indonesia Sehat yang watermarknya terlihat	Sesuai Harapan (Valid)
4	Ketika pop up image Kartu Indonesia Sehat tampil dan tombol download di tekan	File Kartu Indonesia Sehat: (valid file Kartu Indonesia Sehat) Teks watermark: Digunakan untuk registrasi di Rumah Sakit A	File Kartu Indonesia Sehat yang muncul akan langsung ter-download di device user	Sesuai Harapan (Valid)

4. Kesimpulan

Dapat disimpulkan bahwa pengujian fungsionalitas sistem pengamanan digital watermarking dengan steganografi menggunakan penyisipan Most Significant Bit (MSB) dapat digunakan untuk menyisipkan pesan dan pesan yang disisipkan tidak tampak pada file Kartu Indonesia Sehat. Hasil dari pengujian ini menunjukkan bahwa fungsionalitas dari sistem pengamanan digital watermarking kartu Indonesia sehat menggunakan algoritma MSB (Most Significant Bit) dapat berjalan dengan baik.

Daftar Pustaka

- [1] Febriani, Shella Rasita. 2016. "Implementasi Digital Watermarking Pada Citra Menggunakan Metode Least Significant Bit". Jurnal Informatika dan Komputer. Vol. 23, No 3.
- [2] Khairani,M. , Nur Wulan. 2018. "Rancang Bangun Metode Most Significant Bit (Msb) Watermarking Video Format Avi". Seminar Nasional dan Prosiding Sains dan Teknik. Vol 1, No 1.

- [3] Rosmala. Dewi, Angga Kusuma K. 2018. "Perbandingan Metode Most Significant Bit dan Least Significant Bit pada Steganografi untuk Keamanan Data Media Digital". MIND Journal. Vol. 3. No 2. Hal 36-46.
- [4] Wahyuni. Meri Sri. 2017. "Implementasi Steganografi Dalam Menyembunyikan Pesan Teks Dengan Metode Msb (Most Significant Bit). Jurnal Nasional Informatika dan Teknologi Jaringan. Vol 1. No 2.
- [5] Ratnasari. Anita Putri, Felix Andika Dwiyanto. 2020. "Metode Steganografi Citra Digital". Sains, Aplikasi, Komputasi dan Teknologi Informasi. Vol 2, No 2

Implementasi Sistem Monitoring pada Bunga Anggrek Menggunakan Arduino dengan Aplikasi Berbasis Website

Gede Krisnawa Sandhya Wandhana^{a1}, I Gusti Agung Gede Arya Kadyanan^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia

^{a1}krisnawa@student.unud.ac.id

^{a2}gungde@unud.ac.id

Abstract

This article will focus on implementing an orchid plant monitoring system using Arduino in a web-based application. The purpose of this system is to monitor important environmental parameters such as temperature, humidity and soil moisture in real time. By collecting data from orchid plants using sensors, the system enables continuous measurement and analysis of these parameters. Collected data is processed and stored by the Arduino, which acts as the central control unit. A web-based application serves as the user interface, allowing remote access and visualization of monitoring data. Implementation of this system provides a practical solution for improving orchid cultivation through accurate and easily accessible monitoring of key parameters. This allows researchers and hobbyists to closely monitor growing conditions and make informed decisions based on the data collected. By maintaining optimal environmental conditions, the system contributes to improved growth, increased yields and successful cultivation of orchid plants. The combination of an Arduino and a web-based application provides an efficient and easy-to-use solution for orchid plant monitoring and management for better results in your orchid cultivation practice.

Keywords : *Orchid Plants, Arduino, Web-Based Application, Real-time Monitoring, Sensors.*

1. Pendahuluan

Dalam beberapa tahun terakhir, perkembangan teknologi telah membawa dampak yang cukup signifikan di berbagai bidang, termasuk di bidang pertanian dan hortikultura. Salah satu aspek penting dalam pertanian yang melibatkan teknologi adalah penggunaan sistem monitoring yang efektif untuk memantau dan mengelola pertumbuhan tanaman secara tepat sesuai dengan parameter yang sudah ditentukan. Dalam masalah ini, bunga anggrek menjadi fokus penelitian, mengingat popularitasnya sebagai tanaman hias yang memiliki nilai ekonomi tinggi [2].

Di lain sisi penggunaan Arduino sebagai platform pengembangan dan aplikasi berbasis website sebagai antarmuka pengguna telah menunjukkan kemajuan yang besar dalam implementasi sistem monitoring pada bunga tumbuhan secara umum [1]. Arduino merupakan mikrokontroler open-source yang dapat diprogram untuk mengendalikan berbagai sensor dan aktuator. Aplikasi berbasis website, di sisi lain, menyediakan antarmuka yang mudah dimengerti dan digunakan oleh pengguna.

Tujuan dari penelitian ini adalah mengimplementasikan sistem monitoring pada bunga anggrek menggunakan Arduino dengan aplikasi berbasis website [1]. Sistem ini dirancang untuk mengukur dan merekam parameter penting seperti suhu, kelembaban udara, dan kelembaban tanah. Data yang dikumpulkan kemudian akan dikirim ke server melalui koneksi internet, dan pengguna dapat mengaksesnya melalui aplikasi berbasis website.

Dalam artikel ilmiah ini, kami akan mempresentasikan rincian implementasi sistem monitoring menggunakan Arduino pada bunga anggrek, termasuk perancangan perangkat keras dan

perangkat lunak. Kami juga akan membahas hasil pengujian dan evaluasi dari sistem yang diimplementasikan [3]. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi monitoring pertanian yang lebih efisien dan terjangkau.

2. Metode Penelitian

2.1 Objek Penelitian Bunga Anggrek

Bunga anggrek merupakan tanaman yang populer dalam bidang hortikultura dan memiliki keunikan dalam morfologi, fisiologi, dan kebutuhan lingkungan [6]. Penelitian ini bertujuan untuk mengembangkan sebuah sistem monitoring yang menggunakan perangkat Arduino sebagai sensor dan pengontrol, serta aplikasi berbasis website sebagai antarmuka pengguna [5].

Sistem monitoring ini akan mengumpulkan data suhu, kelembaban udara, dan kelembaban tanah pada lingkungan tempat bunga anggrek tumbuh [4]. Melalui perangkat Arduino, data akan diambil secara periodik dan dikirim ke aplikasi berbasis website untuk ditampilkan kepada pengguna [5]. Tujuan dari penelitian ini adalah memberikan pemantauan yang akurat dan real-time terhadap parameter lingkungan yang mempengaruhi pertumbuhan dan kesehatan bunga anggrek [7]. Dengan adanya sistem monitoring ini, pengguna dapat dengan mudah memantau kondisi lingkungan bunga anggrek dan mengambil tindakan yang diperlukan untuk menjaga keberhasilan pertumbuhannya [7].

2.2 Perancangan Sistem

Pada tahapan perancangan sistem ini menjelaskan bagaimana sistem menjalankan tugasnya untuk memantau data – data yang didapat dari sensor kelembaban udara dan kelembaban tanah yang didapatkan ketika melakukan observasi dari berbagai sumber didapatkan penjelasan sebagai berikut:

- a. Data yang didapatkan dari sensor suhu, kelembaban udara, dan kelembaban tanah dalam rentang waktu kurang lebih 1 hari kemudian data tersebut akan dikirimkan oleh mikrokontroler yang terhubung ke dalam database yang sudah dibuat melalui protokol HTTP.
- b. Data – data yang sudah terkumpul tersebut akan dibaca oleh sistem dan akan diklasifikasikan menjadi 2 jenis data, yaitu data normal dan data anomali [9].
- c. Sistem akan memberikan informasi berupa data – data yang sudah terkumpul yang berisi keterangan kondisi dan kualitas dari data sensor yang diterima sesuai dengan parameter yang sudah ditentukan sebelumnya.

Sistem ini dapat dikategorikan baik apabila sistem mampu untuk menampilkan kondisi dan kualitas sesuai dengan data yang sudah dikumpulkan oleh sensor dan dapat ditampilkan sesuai dengan keinginan pengguna. Untuk mengetahui lebih jelas kategori dapat melihat tabel dibawah ini.

Tabel 1. Himpunan dan Nilai Keanggotaan

Nama Himpunan	Nilai Keanggotaan
Suhu	Panas
	Normal
	Dingin
Kelembaban Tanah	Kering
	Normal

Nama Himpunan	Nilai Keanggotaan
	Lembab
Kelembaban Udara	Kering
	Normal
	Lembab

2.3 Algoritma Anomaly Detection (One-Class SVM)

a. Proses Algoritma One – Class SVM

Proses algoritma One-Class SVM (Support Vector Machine) dimulai dengan persiapan data pelatihan yang hanya terdiri dari satu kelas data, seperti data normal [11]. Selanjutnya, model One-Class SVM dilatih untuk membangun sebuah batas kelas yang dapat memisahkan data normal dari data anomali [10]. Parameter-parameter seperti kernel, C, dan gamma disesuaikan untuk memperoleh model yang optimal [11]. Setelah model terlatih, algoritma dapat digunakan untuk mendeteksi anomali pada data baru dengan mengklasifikasikan data tersebut berdasarkan posisinya terhadap hyperplane yang telah dipelajari [8]. Algoritma One-Class SVM memiliki kemampuan untuk mengenali data yang tidak umum atau tidak sesuai dengan pola yang telah dipelajari dari data normal.

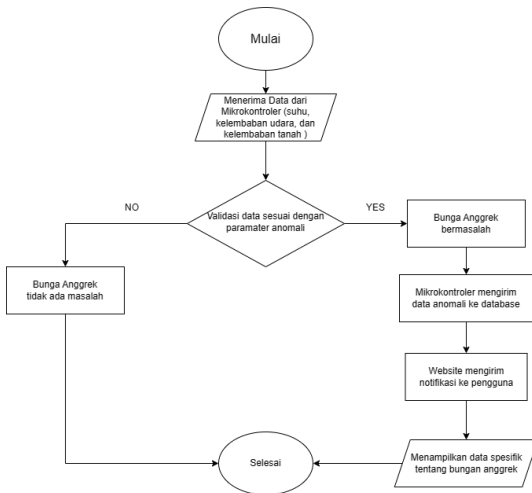
b. Penentuan Support Vectors

Penentuan support vectors adalah proses untuk menemukan dan memilih sejumlah titik data yang merupakan representasi penting dalam menentukan batas keputusan. Support vectors adalah titik-titik data yang berada paling dekat dengan batas keputusan dan berperan penting dalam menentukan posisi serta orientasi dari batas keputusan yang membedakan dua kelas tersebut.

2.4 Desain Sistem

Implementasi dari sistem monitoring bunga anggrek menggunakan alforitma One – Class SVM yang memiliki beberapa fitur utama seperti klasifikasi data berdasarkan sensor suhu, sensor kelembaban tanah, dan kelembaban udara, selain itu sistem akan mengirim notifikasi kepada pengguna ketika tanaman mengalami sebuah masalah atau anomali, untuk lebih jelasnya akan dijelaskan sebagai berikut:

- Fitur data sensor, dimana fitur ini akan menampilkan besaran nilai dari setiap sensor kepada pengguna mulai dari sensor kelembaban udara, kelembaban tanah, dan suhu. Data tersebut akan ditampilkan dalam bentuk tabel sehingga memudahkan pengguna.
- Fitur sensor suhu, dimana fitur ini akan menampilkan informasi rinci dari suhu disekitar tanaman tersebut dengan indikasi suhu panas, normal, ataupun dingin.
- Fitur sensor kelembaban tanah dan udara, yaitu fitur dimana akan menampilkan nilai rinci dari kelembaban tanah dan air untuk menghindari data anomali, jika data tersebut masuk kedalam paramater anomali akan mengirim notifikasi ke pengguna mengenai masalah tersebut.
- Fitur hasil klasifikasi, dimana fitur ini akan menampilkan klasifikasi keseluruhan data yang didapat dari sensor suhu, sensor kelembaban udara dan kelembaban tanah yang ditampilkan dalam bentuk tabel dengan jangkauan sangat baik, baik, dan buruk.



Gambar 1. Flowchart

3. Hasil Pembahasan

3.1 Implementasi Sistem

Sistem monitoring ini menyediakan 5 buah fitur yang keseluruhannya berguna untuk menampilkan data yang didapat dari sensor – sensor yang digunakan. Penjelasan lebih rincinya sebagai berikut, pertama, fitur data sensor, dimana fitur ini akan menampilkan besaran nilai dari setiap sensor kepada pengguna mulai dari sensor kelembaban udara, kelembaban tanah, dan suhu. Data tersebut akan ditampilkan dalam bentuk tabel sehingga memudahkan pengguna. Kedua, fitur sensor suhu, dimana fitur ini akan menampilkan informasi rinci dari suhu disekitar tanaman tersebut dengan indikasi suhu panas, normal, ataupun dingin. Ketiga, fitur sensor kelembaban tanah, yaitu fitur dimana akan menampilkan nilai rinci dari kelembaban tanah dan air untuk menghindari data anomali, jika data tersebut masuk kedalam parameter anomali akan mengirim notifikasi ke pengguna mengenai masalah tersebut. Keempat, fitur sensor kelembaban udara, yang berfungsi untuk melihat nilai dari kelembaban udara yang ditangkap oleh sensor sehingga dapat mengontrol kelembaban pada tumbuhan. Kelima, fitur hasil klasifikasi, dimana fitur ini akan menampilkan klasifikasi keseluruhan data yang didapat dari sensor suhu, sensor kelembaban udara dan kelembaban tanah yang ditampilkan dalam bentuk tabel dengan jangkauan sangat baik, baik, dan buruk.

3.2 Pengujian Sistem

Pengujian sistem ini dilakukan menggunakan metode black box testing yang berfungsi untuk melihat fungsionalitas dari sistem monitoring yang telah dibuat sebelumnya. Pengujian yang akan dilakukan adalah sebagai berikut:

Tabel 2. Pengujian Sistem

Modul yang diujikan	Prosedur Pengujian	Input Atau Masukan	Output Yang Diharapkan	Hasil Yang Pengujian	Kesimpulan
Fitur Data Sensor					
Fitur Sensor Suhu					
Fitur Sensor Kelembaban Udara					
Fitur Sensor Kelembaban Tanah					
Fitur Hasil Klasifikasi					

4. Kesimpulan

Dari penelitian diatas dapat diambil kesimpulan berupa sistem monitoring menggunakan Arduino dan aplikasi berbasis website merupakan pendekatan efektif untuk memantau kondisi pertumbuhan bunga anggrek. Dengan menggunakan sensor suhu, kelembaban udara, dan kelembaban tanah, sistem ini mengumpulkan data real-time yang diolah dan disimpan oleh Arduino. Aplikasi website memberikan antarmuka pengguna untuk pemantauan jarak jauh. Implementasi sistem ini dapat meningkatkan efisiensi dan keberhasilan dalam budidaya bunga anggrek.

Daftar Pustaka

- [1] Arduino. (2021). Arduino - Home. Diakses dari <https://www.arduino.cc/>
- [2] Kodek, M., & Ponomarev, A. (2018). A Review of Arduino Board in Agriculture and Environment Field. 2018 19th International Carpathian Control Conference (ICCC), Zakopane, Poland, hal. 204-209. DOI: 10.1109/CarpathianCC.2018.8446941
- [3] Khan, M. F. R., Al-Zahrani, A. A., & Khan, S. (2020). IoT-Based Plant Monitoring System: A Review. 2020 3rd International Conference on Information and Computer Technologies (ICICT), Riyadh, Saudi Arabia, hal. 186-191. DOI: 10.1109/ICICT48567.2020.9124634
- [4] Rizki, A., & Santoso, B. (2021). Design of an IoT-Based Orchid Flower Monitoring System Using Arduino. Journal of Applied Intelligent Systems, 3(1), 45-54.
- [5] Saputra, A. P., & Hidayatullah, M. F. (2020). Monitoring System for Orchid Flower Growth Using Arduino and IoT. International Journal of Advances in Intelligent Informatics, 6(1), 54-62.
- [6] Halim, A. H., & Lee, R. L. (2018). Wireless Sensor Network for Real-time Monitoring of Orchid Greenhouse. Proceedings of the 3rd International Conference on Electrical Systems, Technology and Information, 1-6.
- [7] Alfiansyah, N., & Gunawan, A. (2017). Design and Implementation of IoT-Based Orchid Flower Monitoring System. Journal of Electrical Engineering and Informatics, 9(1), 97-103.
- [8] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.
- [9] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448-3470.
- [10] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural computation, 13(7), 1443-1471.
- [11] Tax, D. M., & Duin, R. P. (2004). Support vector data description. Machine learning, 54(1), 45-66.

Halaman ini sengaja dibiarkan kosong

Analisis Performa Sistem pada Pemisahan Database Analitik dengan Transaksional

I Made Ari Madya Santosa^{a1}, I Gusti Ngurah Anom Cahyadi Putra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹arimadyasantosa020@student.unud.ac.id
²anom.cp@unud.ac.id

Abstract

The speed and efficiency of a system currently has a very important role. Currently there are several solutions to answer these challenges, one of which is to separate the database on the system. In this study, an analysis was carried out on the separation of databases on a transactional and analytic system. Analysis is carried out with a research flow that begins with problem identification followed by sampling data on technology companies that have transactional systems, then implementing transactional and analytical programs on microservices architecture, conducting tests with Load-Testing, followed by analysis of test results, and then drawing conclusions. After conducting research in accordance with the research flow, it was concluded that the separation of databases on transactional and analytic systems is better for producing faster system performance compared to transactional and analytic systems using the same database.

Keywords: Big Data, Optimization, Transactional, Analytical Systems, Microservices

1. Pendahuluan

Salah satu tantangan dalam sistem informasi saat ini adalah kinerja sistem yang cepat dan efisien. Untuk menyelesaikan masalah ini, pemrosesan data terdistribusi menjadi salah satu solusi yang banyak digunakan [1]. Terdapat dua tipe pemrosesan data dalam sistem distribusi yaitu Transactional Processing (OLTP) dan Analytical Processing (OLAP) yang merupakan cara pemrosesan yang berbeda namun saling melengkapi. OLTP digunakan untuk memproses operasi transaksi bisnis, sedangkan OLAP digunakan untuk analisis dan pengambilan keputusan dengan menggunakan data historis.

Pada saat ini, terdapat beberapa solusi untuk mengintegrasikan OLTP dan OLAP pada sistem yang sama, salah satunya adalah dengan menggunakan pemisahan servis dan database [2]. Pemisahan database memungkinkan pengelolaan data yang lebih efisien, sementara pemisahan servis memungkinkan pengelolaan servis yang lebih efektif. Oleh karena itu, implementasi pemisahan servis dan database analitik dengan transaksional menjadi salah satu solusi yang paling populer saat ini [3].

Namun, penggunaan database terdistribusi juga memiliki beberapa tantangan. Masalah utama dalam penggunaan database terdistribusi adalah konsistensi data dan skalabilitas [4]. Oleh karena itu, metode konsistensi data yang tepat harus digunakan untuk memastikan bahwa data yang tersedia dalam database terdistribusi konsisten dan akurat. Selain itu, sistem database terdistribusi juga harus dapat mengatasi masalah skalabilitas untuk memastikan bahwa sistem dapat menangani volume data yang besar dan meningkat seiring waktu.

Dalam kaitannya dengan performa sistem, pemisahan database dan servis juga dapat membantu dalam mengoptimalkan kinerja sistem. Dalam sebuah penelitian oleh Li et al. (2019), mereka menunjukkan bahwa pemrosesan transaksional dan analitis pada database terdistribusi dapat meningkatkan kinerja sistem, khususnya dalam hal throughput dan latensi [3]. Kim et al. (2019)

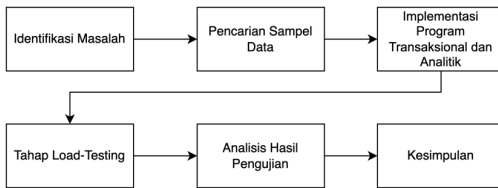
juga melakukan penelitian yang menunjukkan bahwa arsitektur database terdistribusi dapat membantu mengintegrasikan pemrosesan transaksional dan analitis [4].

Dalam rangka memaksimalkan performa sistem, diperlukan juga penggunaan database yang tepat. Widiastuti dan Adhita (2019) mengemukakan bahwa penggunaan database dapat mempengaruhi performa sistem informasi [5]. Oleh karena itu, pemilihan database yang tepat menjadi kunci penting dalam mengoptimalkan performa sistem.

2. Metode Penelitian

2.1. Alur Penelitian

Penulis mengikuti alur dan juga tahapan yang ditunjukkan pada **Gambar 1** untuk memastikan kelancaran dari penelitian ini



Gambar 1. Alur penelitian

Penjelasan dari alur penelitian di atas sebagai berikut.

a. Identifikasi Masalah

Pada awal penelitian, penulis melakukan identifikasi permasalahan di lapangan dan juga melakukan penelitian dengan menggunakan literatur yang dapat membantu penelitian. Kemudian, masalah yang ditemukan oleh penulis adalah lambatnya pengambilan data pada database dan tingginya penggunaan CPU dan RAM ketika melakukan transaksional dan juga analitik terhadap suatu layanan.

b. Pencarian Sampel Data

Pada tahap ini, sampel data dicari melalui survey terhadap perusahaan teknologi yang memiliki produk berupa sistem transaksional. Dilakukan observasi dan wawancara untuk mengetahui bagaimana bentuk data dari sistem transaksional yang dimiliki oleh perusahaan terkait.

c. Implementasi Program Transaksional dan Analitik Menggunakan Arsitektur Microservices

Pada tahap ini, penulis membuat sebuah sistem untuk melakukan transaksional dan analitik. Terdapat satu servis untuk melakukan transaksional dan satu servis analitik serta penulis juga membuat dua database dengan data yang sama.

d. Tahap Load-Testing

Pengujian dilakukan dengan metode *Load-Testing* menggunakan K6 dan memperhatikan beberapa metris.

e. Analisis Hasil Pengujian

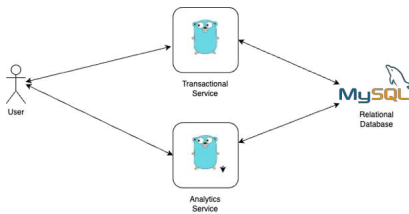
Peneliti melakukan analisis terhadap beberapa metris dan hasil database monitoring untuk menarik sebuah kesimpulan.

f. Kesimpulan

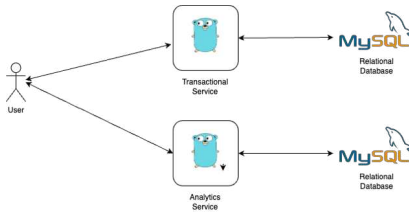
Pada tahap ini, penulis menarik kesimpulan dari beberapa metris dan hasil database monitoring untuk menarik kesimpulan dari hasil penelitian ini.

2.2. Arsitektur Microservices

Penelitian ini menggunakan arsitektur *microservices* untuk melihat dampak dari layanan *single-point-of-failure*, atau layanan yang kegagalannya mengganggu pengoperasian sistem. Ini adalah masalah fatal dalam arsitektur layanan mikro. *Microservices* adalah model arsitektur [6]. Ide dasar di balik jenis arsitektur ini adalah membagi sistem yang lebih besar menjadi sistem yang lebih kecil. Hal ini dapat mengurangi kompleksitas sistem [7], membuat sistem yang lebih kecil menjadi lebih mandiri, dan membuat hubungan antara sistem atau layanan menjadi kurang rumit. Selain itu, sistem yang mandiri membutuhkan proses pengembangan dan implementasi yang lebih mandiri. Keuntungan lain dari *microservices* adalah jika suatu layanan gagal, kegagalan tersebut hanya diisolasi dan tidak mengganggu layanan lain, kecuali jika layanan tersebut merupakan kegagalan tunggal. **Gambar 2** dan **Gambar 3** menunjukkan arsitektur yang penulis gunakan dalam penelitian ini,



Gambar 2. Arsitektur Microservices 1



Gambar 3. Arsitektur Microservices 2

2.3. Load-Testing

Pengujian yang dilakukan pada penelitian ini menggunakan metode *Load-Testing* dengan bantuan K6. Metode *Load-Testing* akan mensimulasikan pengguna yang mengirim permintaan sesuai dengan parameter yang diberikan ke program [8]. K6 merupakan alat bantu untuk melakukan *Load-Testing* yang bersifat *open-source*. Dengan ini kita bisa menguji performa dan keandalan sistem untuk mencari tahu masalah lebih awal. K6 akan mengembalikan hasil berupa beberapa metris yang berkaitan tentang permintaan yang telah dikirim. Metris ini akan membantu penulis untuk mengukur kecepatan penerimaan data pada penelitian ini. Beberapa metris yang digunakan pada penelitian dapat dilihat pada **Tabel 1**

Tabel 1. Metris Load-Testing

Key	Deskripsi
<i>http_req_duration</i>	Metris yang merepresentasikan durasi dari permintaan hingga menerima data pada HTTP Request (<i>http_req_sending</i> + <i>http_req_waiting</i> + <i>http_req_receiving</i>)

3. Hasil dan Diskusi

Berdasarkan alur penelitian dan juga studi literatur yang telah dijelaskan pada bab sebelumnya. Maka hasil penelitian yang peneliti lakukan dapat dijabarkan sebagai berikut.

3.1. Pencarian Sampel Data

Penelitian ini melakukan pencarian sampel data ke perusahaan teknologi yang memiliki sistem transaksional. Sampel data dicari dengan cara melakukan survey dan observasi terhadap perusahaan terkait. Didapatkan format data pada tabel transaksi yang digunakan oleh perusahaan dalam sistem transaksional yang dapat dilihat pada **Tabel 2**

Tabel 2. Format Data Tabel Transaksi

Column	Type
<i>transaction_no</i>	Int
<i>date</i>	Varchar
<i>product_no</i>	Int
<i>product_name</i>	Varchar
<i>price</i>	Double
<i>quantity</i>	Int
<i>customer_no</i>	Int
<i>country</i>	varchar

3.2. Implementasi Program Transaksional dan Analitik Menggunakan Arsitektur *Microservices*

Proses implementasi program transaksional dan analitik dilakukan pada basis kode atau *codebase* berbeda dengan menggunakan bahasa pemrograman Go. Potongan kode inti dari program transaksional dapat dilihat pada **Tabel 3** dan analitik pada **Tabel 4**.

Tabel 3. Potongan Kode Inti Transaksional

Baris	Kode
67	func createTransaksi(c *gin.Context) {
68	var transaksi Transaksi
69	
70	if err := c.ShouldBindJSON(&transaksi); err != nil {
71	c.JSON(http.StatusBadRequest, gin.H{"error":
72	err.Error()})
	return
73	}
74	
75	if err := db.Create(&transaksi).Error; err != nil {
76	c.JSON(http.StatusInternalServerError,
	gin.H{"error": err.Error()})
77	return
78	}
79	
80	c.JSON(http.StatusOK, transaksi)
81	}
82	
83	func getTransaksi(c *gin.Context) {
84	id := c.Param("id")
85	var transaksi Transaksi
86	
87	if err := db.First(&transaksi, id).Error; err != nil {
88	if errors.Is(err, gorm.ErrRecordNotFound) {
89	c.JSON(http.StatusNotFound, gin.H{"error":
	"record not found"})
90	} else {
91	c.JSON(http.StatusInternalServerError,
	gin.H{"error": err.Error()})
92	}
93	return
94	}
95	
96	c.JSON(http.StatusOK, transaksi)
97	}
98	
99	func updateTransaksi(c *gin.Context) {
100	id := c.Param("id")
101	var transaksi Transaksi
102	
103	if err := db.First(&transaksi, id).Error; err != nil {
104	if errors.Is(err, gorm.ErrRecordNotFound) {
	c.JSON(http.StatusNotFound, gin.H{"error":
105	"record not found"})
	} else {
106	c.JSON(http.StatusInternalServerError,
107	gin.H{"error": err.Error()})
	}
108	return
109	}
110	
111	if err := c.ShouldBindJSON(&transaksi); err != nil {
112	c.JSON(http.StatusBadRequest, gin.H{"error":
113	err.Error()})
	return

Baris	Kode
114	}
115	
116	if err := db.Save(&transaksi).Error; err != nil {
117	c.JSON(http.StatusInternalServerError,
118	gin.H{"error": err.Error()})
	return
119	}
120	
121	c.JSON(http.StatusOK, transaksi)
122	}
123	
124	func deleteTransaksi(c *gin.Context) {
125	db := c.MustGet("db").(*gorm.DB)
126	
127	// Get the transaksi record with the specified ID
128	var transaksi Transaksi
129	if err := db.Where("id = ?",
130	c.Param("id")).First(&transaksi).Error; err != nil {
131	c.JSON(http.StatusNotFound, gin.H{"error": "Record not
	found!"})
132	return
133	}
134	
135	// Delete the transaksi record from the database
136	db.Delete(&transaksi)
137	
138	// Return a success message as JSON
139	c.JSON(http.StatusOK, gin.H{
140	"message": "Record deleted successfully!",
141	})
142	}

Tabel 4. Potongan Kode Inti Analitik

Baris	Kode
70	func getRevenueByProduct(c *gin.Context) {
71	var revenueByProduct []RevenueByProduct
72	
73	if err :=
	db.Table("transaksis").Select("product_no, sum(price *
	quantity) as
74	revenue").Group("product_no").Find(&revenueByProduct).Err
	or; err != nil {
75	c.JSON(http.StatusInternalServerError,
76	gin.H{"error": err.Error()})
77	return
78	}
79	
80	c.JSON(http.StatusOK, revenueByProduct)
81	}
82	
83	
84	func getRevenueByDate(c *gin.Context) {
	var revenueByDate []struct {

Baris	Kode
85	Date string `json:"date"``
86	Revenue float64 `json:"revenue"``
87	}
	if err := db.Table("transaksis").Select("date,
88	sum(price * quantity) as
	revenue").Group("date").Find(&revenueByDate).Error; err
89	!= nil {
90	c.JSON(http.StatusInternalServerError,
91	gin.H{"error": err.Error()})
92	return
93	}
	c.JSON(http.StatusOK, revenueByDate)
	}

Pada **Tabel 3** merupakan kode program untuk melakukan transaksi pada umumnya seperti membuat transaksi, mengubah transaksi, melihat transaksi, dan juga menghapus transaksi. Kemudian pada kode analitik di **Tabel 4**, analitik yang dilakukan adalah menghitung pendapatan berdasarkan tanggal dan berdasarkan produk.

3.3. Tahap Load-Testing

Pengujian dari sistem ini adalah dengan melakukan *Load-Testing* menggunakan K6. Program untuk melakukan *Load-Testing* ditulis dengan menggunakan bahasa pemrograman *JavaScript*. Kode pengujian menggunakan K6 dapat dilihat pada kode di **Tabel 5**.

Table 5. Kode Load-Testing

Baris	Kode
1	import http from 'k6/http';
2	import { check, sleep } from 'k6';
3	
4	export let options = {
5	stages: [
6	{ duration: '1m', target: 20 }
7]
8	};
9	
10	export function testCreateTransaksi() {
11	let payload = {
12	country: 'Indonesia',
13	customer_no: 1,
14	date: '2023-04-24',
15	price: 100.0,
16	product_name: 'Product A',
17	product_no: 1,
18	quantity: 1,
19	transaction_no: 1
20	};
21	
22	let headers = { 'Content-Type': 'application/json' };
23	
24	

Baris	Kode
25	let res = http.post('http://localhost:8081/transaksi',
26	JSON.stringify(payload), { headers: headers });
27	check(res, { 'status is 201': (r) => r.status === 201 });
28	}
29	
30	export function testGetProductRevenue() {
	let res =
	http.get('http://localhost:8080/transaksi/revenue-by-
31	product');
32	
33	check(res, { 'status is 200': (r) => r.status === 200 });
34	}
35	
36	export function testUpdateTransaksi() {
37	let payload = {
38	country: 'Indonesia',
39	customer_no: 1,
40	date: '2023-04-24',
41	price: 200.0,
42	product_name: 'Product A',
43	product_no: 1,
44	quantity: 2,
45	transaction_no: 1
46	};
47	
48	let headers = { 'Content-Type': 'application/json' };
49	
	let res = http.put('http://localhost:8081/transaksi/1',
50	JSON.stringify(payload), { headers: headers });
51	
52	check(res, { 'status is 200': (r) => r.status === 200 });
53	}
54	
55	export function testDeleteTransaksi() {
56	let res = http.del('http://localhost:8081/transaksi/1');
57	
58	check(res, { 'status is 200': (r) => r.status === 200 });
59	}
60	
61	export default function () {
62	// Create a new transaksi record
63	testCreateTransaksi();
64	
65	// Get the total revenue by product
66	testGetProductRevenue();
67	
68	// Update an existing transaksi record
69	testUpdateTransaksi();
70	
71	// Delete an existing transaksi record
72	testDeleteTransaksi();
73	
74	// Wait for 1 second before repeating the cycle
75	sleep(1);
	}

Implementasi dari *Load-Testing* dapat dilihat pada penggalan kode di atas. Pada kode tersebut terdapat beberapa fungsi yang menjalankan HTTP request ke beberapa endpoint untuk melakukan testing fungsionalitas dari API. Terdapat fungsi untuk membuat transaksi baru, mengambil total revenue dari produk, mengubah transaksi yang sudah ada, dan menghapus transaksi yang sudah ada. Selain itu, terdapat fungsi default yang menjalankan seluruh fungsi di atas secara berurutan dan akan diulang setiap 1 detik hingga durasi yang diatur pada `options.stages` habis. Pada `options.stages`, terdapat satu tahap (*stage*) pengujian yang mengatur bahwa selama 1 menit, akan ada 20 koneksi yang dibuat ke aplikasi. Dalam setiap HTTP request yang dilakukan, juga terdapat pengecekan (*check*) apakah status response dari API sesuai dengan yang diharapkan, misalnya status 201 untuk *create* atau 200 untuk *read/update/delete*. Sehingga didapatkan hasil dari *Load-Testing* seperti pada **Gambar 4** dan juga **Gambar 5** di bawah.

```
checks.....: 25.00% x 1040 / 3120
data_received.....: 48 MB 797 kB/s
data_sent.....: 810 kB 13 kB/s
http_req_blocked.....: avg=8.16µs min=0s med=2µs max=1.22ms p(90)=5µs p(95)=7µs
http_req_connecting.....: avg=3.22µs min=0s med=0s max=52µs p(90)=0s p(95)=0s
http_req_duration.....: avg=39.79ms min=193µs med=1.99ms max=316.77ms p(90)=146.83ms p(95)=186.51ms
{ expected_response:true }...: avg=78.15ms min=684µs med=59.62ms max=316.77ms p(90)=186.58ms p(95)=219.96ms
http_req_failed.....: 50.00% x 2000 / 2000
http_req_receiving.....: avg=11.46µs min=5µs med=30µs max=1.2ms p(90)=84µs p(95)=108µs
http_req_sending.....: avg=14.54µs min=2µs med=11µs max=534µs p(90)=27µs p(95)=36µs
http_req_tls_handshaking.....: avg=0s min=0s med=0s max=0s p(90)=0s p(95)=0s
http_req_waiting.....: avg=39.73ms min=177µs med=1.92ms max=316.65ms p(90)=146.7ms p(95)=186.44ms
http_req.....: 4168 68.047317/s
iteration_duration.....: avg=1.16s min=1.05s med=1.14s max=1.32s p(90)=1.23s p(95)=1.26s
iterations.....: 1040 17.211895/s
vus.....: 20 min=20 max=20
vus_max.....: 20 min=20 max=20

running (1m00.4s), 00/20 VUs, 1040 complete and 0 interrupted iterations
default / ..... 20 VUs 1m0s
```

Gambar 4. Hasil Testing Dengan Satu Database

```
checks.....: 25.00% x 1060 / 3180
data_received.....: 49 MB 818 kB/s
data_sent.....: 826 kB 14 kB/s
http_req_blocked.....: avg=10.07µs min=0s med=2µs max=1.97ms p(90)=4µs p(95)=5µs
http_req_connecting.....: avg=5.57µs min=0s med=0s max=719µs p(90)=0s p(95)=0s
http_req_duration.....: avg=54.74ms min=180µs med=1.13ms max=353.53ms p(90)=135.45ms p(95)=166.55ms
{ expected_response:true }...: avg=68.62ms min=585µs med=46.65ms max=353.53ms p(90)=166.55ms p(95)=197.69ms
http_req_failed.....: 50.00% x 2120 / 2120
http_req_receiving.....: avg=14.21µs min=4µs med=22µs max=2.41ms p(90)=62µs p(95)=80µs
http_req_sending.....: avg=11.02µs min=2µs med=9µs max=356µs p(90)=21µs p(95)=29µs
http_req_tls_handshaking.....: avg=0s min=0s med=0s max=0s p(90)=0s p(95)=0s
http_req_waiting.....: avg=34.7ms min=176µs med=1.1ms max=353.38ms p(90)=135.37ms p(95)=166.47ms
http_req.....: 4240 59.99297/s
iteration_duration.....: avg=1.13s min=1.05s med=1.13s max=1.37s p(90)=1.2s p(95)=1.24s
iterations.....: 1060 17.498247/s
vus.....: 20 min=20 max=20
vus_max.....: 20 min=20 max=20

running (1m00.6s), 00/20 VUs, 1060 complete and 0 interrupted iterations
default / ..... 20 VUs 1m0s
```

Gambar 5. Hasil Testing Dengan Database Berbeda

3.4. Analisis Hasil Pengujian

Dalam pengukuran kinerja sistem transaksional dan analitik dengan database yang sama dan berbeda pada penelitian ini, dapat menggunakan perbandingan hasil dari *Load-Testing* pada kedua metode tersebut. Perbandingan dapat dilihat dengan mudah pada **Tabel 6**.

Tabel 6. Hasil metris Load-Testing

Metris	Sistem Dengan Satu Database	Sistem Dengan Database Berbeda	Performa Tertinggi
http_req_duration	186.51 milidetik	166.55 milidetik	Sistem Dengan Database Berbeda

Bedasarkan tabel diatas dapat dilihat metris *http_req_duration* menggunakan p (95) atau persentase *request* pada kecepatan yang jatuh pada persentase ke-95 agar lebih tepat dan menghindari *outlier* pada kedua hasil *Load-Testing*, Sistem transaksional dan analitik dengan menggunakan *database* yang berbeda mendapatkan hasil yang lebih baik dengan jangka waktu dari *request* hingga menerima data lebih cepat dibandingkan dengan satu *database*.

4. Kesimpulan

Bedasarkan metris *http_req_duration* menggunakan persentase *request* pada kecepatan yang jatuh pada persentase ke-95 atau p (95) dari hasil pengujian penelitian yang dilakukan yang dapat dilihat pada **Tabel 6**, ditemukan bahwa pemisahan *database* untuk sistem transaksional dan analitik memiliki kecepatan sistem yang lebih cepat dibanding sistem transaksional dan analitik hanya menggunakan satu *database*. Sehingga dapat direkomendasikan untuk pemisahan *database* pada sistem transaksional dan analitik, namun diperlukan juga sinkronisasi *database* agar data yang digunakan untuk transaksional dan analitik sama.

Daftar Pustaka

- [1] Jafarzadeh, A., & Alizadeh, A. (2018). Design and implementation of a high-performance software-defined storage system for big data applications. *IEEE Access*, 6, 29934-29948.
- [2] Chen, Y., Zhang, X., Hu, F., Guo, M., & Yan, X. (2019). An Automatic Data Separation Method for Analytical and Transactional Services. *IEEE Access*
- [3] Li, J., Li, C., Li, L., Xu, X., & Feng, Y. (2019). Performance evaluation of transactional and analytical processing on distributed column-oriented database. *IEEE Access*, 7, 20934-20944.
- [4] Kim, T., Lee, Y., Lee, S., Lee, S., & Song, J. (2019). A distributed database architecture for integrating transactional and analytical data processing. *Future Generation Computer Systems*, 99, 205-215.
- [5] Widiastuti, I., & Adhita, D. (2019). Pengaruh Penggunaan Database Terhadap Performa Sistem Informasi. *Journal of Information Technology and Computer Science*, 4(1), 1-7.
- [6] F. Rademacher, S. Sachweh, and A. Zundorf, "Differences between model-driven development of service-oriented and microservice architecture," *Proc. - 2017 IEEE Int. Conf. Softw. Archit. Work. ICSAW 2017 Side Track Proc.*, pp. 38-45, 2017, doi: 10.1109/ICSAW.2017.32.
- [7] M. Shahin, M. Ali Babar, and L. Zhu, "Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices," *IEEE Access*, vol. 5, no. March, pp. 3909-3943, 2017, doi: 10.1109/ACCESS.2017.2685629.
- [8] H. Schulz, T. Angerstein, and A. Van Hoorn, "Towards automating representative load testing in continuous software engineering," *ICPE 2018 - Companion 2018 ACM/SPEC Int. Conf. Perform. Eng.*, vol. 2018-January, pp. 123-126, 2018, doi: 10.1145/3185768.3186288.

Analisis Performa Write/Read Kompresi Delta Encoding pada Data Logging Menggunakan Go Benchmark

I Putu Gede Mahardika Adi Putra^{a1}, I Ketut Gede Suhartana^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana

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

¹gedemahardika@student.unud.ac.id

²kg.suhartana@unud.ac.id (Corresponding Author)

Abstract

Compression for data logging in environmental health monitoring is a serious concern. Recording of environmental health history is carried out by monitoring the fluctuations of the physical parameters. The record stored in a warehouse from the logging system will increase over time. So, that's the reason why compression on time-series data logging is needed. But, the addition of compression algorithm like delta encoding allows for increased latency. Therefore, the performance of write/read of delta encoding must be analyzed. One method to analyze is the Go Benchmark. The test parameter used in this paper is the number of iterations and times per iteration taken from the Go Benchmark's output. The other test parameter is compression ratio and storage saving taken from the size of data before and after compression. There are 4 tests case used: single data write, single data read, multiple data write, and multiple data read. As the result, single data write/read and multiple data read work optimally in delta compression with the similar test result with baseline. But multiple data write not working optimally with times per iteration 10,254 times higher than baseline.

Keywords: Data Compression, Go Benchmark, Delta Encoding, Data Logging, Performance.

1. Pendahuluan

Kompresi pada data logging pemantauan kesehatan lingkungan menjadi hal krusial yang perlu perhatian serius. Pemantauan kesehatan lingkungan melalui parameter kondisi fisik lingkungan menghasilkan data time-series dalam jumlah yang besar. Parameter fisik yang dapat dijadikan acuan untuk mengetahui kesehatan lingkungan adalah kandungan gas di udara, kelembapan, suhu, dan ada atau tidaknya cahaya. Kandungan gas di udara dapat memberikan informasi keberadaan gas beracun untuk manusia. Kelembapan dan suhu memberikan informasi kecenderungan berkembangnya bakteri [1]. Data hasil pemantauan dari beberapa parameter dapat memberikan informasi yang utuh mengenai kondisi lingkungan terkini. Status lingkungan layak ditinggali, diperlukan perbaikan, atau tidak layak ditinggali dapat diketahui dengan lebih cepat sehingga mendukung keputusan yang akan dilakukan selanjutnya [2]. Selain status kesehatan lingkungan pada satu waktu, pemantauan juga berfungsi untuk mencatat riwayat kesehatan lingkungan. Riwayat ini berfungsi untuk melihat fluktuasi parameter-parameter fisik lingkungan serta hubungan antarparameter yang dipantau. Karena pentingnya fungsi data pemantauan kesehatan lingkungan, maka diperlukan sebuah metode kompresi yang dapat digunakan untuk mereduksi ukuran data time series yang didapatkan dari pemantauan.

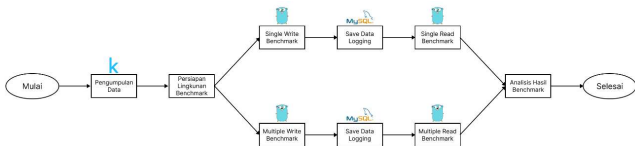
Perangkat Internet of Things (IoT) yang mengirimkan data time series termasuk ke dalam perangkat data logger. Sistem data logger adalah sistem yang dapat memonitor parameter fisik seperti suhu dan radiasi dalam jumlah yang masif [3]. Data logging yang dihasilkan oleh data logger selanjutnya disimpan ke sistem basis data. Data logging digunakan untuk tujuan analisis dan sistem pendukung keputusan dalam proses pemantauan keadaan lingkungan. Data yang disimpan di sistem basis data akan membentuk sebuah gudang data (warehouse) yang menggambarkan keadaan lingkungan dari waktu ke waktu. Semakin lama waktu pemantauan,

semakin banyak data time series yang didapatkan. Besarnya porsi data dan terus bertambahnya ukuran data memicu pembengkakan ukuran data yang disimpan pada basis data.

Metode kompresi dapat digunakan untuk mengatasi pembengkakan ukuran data. Mengecilnya ukuran data memungkinkan pengurangan biaya jaringan dan penyimpanan data di server [4]. Salah satu metode kompresi pada data time series adalah delta encoding. Metode ini termasuk ke dalam lossless compression, yaitu metode kompresi yang menjaga agar tidak ada data yang hilang pada proses dekompresi data yang terkompresi [5]. Namun, penambahan algoritma kompresi terkadang dapat menambah latensi dan penggunaan memori pada server [4]. Oleh karena itu, perlu dianalisis perbedaan performa write dan read data dengan dan tanpa menggunakan delta encoding. Salah satu metode yang dapat digunakan untuk mengetahui perbedaan performanya, terutama pada server adalah dengan menggunakan Go Benchmark.

2. Metode Penelitian

2.1 Alur Penelitian



Gambar 1. Alur Penelitian

Alur penelitian digambarkan pada Gambar 1. Penelitian dimulai dari pengumpulan data pemantauan keadaan lingkungan. Data didapatkan dari penyedia dataset, yaitu Kaggle. Data timestamp dan identitas perangkat pengirim yang tersedia pada dataset dapat digunakan untuk menyimulasikan data yang dikirim oleh beberapa data logger. Sebelum menuju tahap benchmark, disiapkan lingkungan benchmark terlebih dahulu. Lingkungan benchmark yang disiapkan berupa sistem operasi dan text editor atau integrated development environment (IDE) untuk menjalankan benchmark. Selanjutnya, tahap pengujian dilakukan dengan menggunakan Go Benchmark dengan database management system (DBMS) MySQL. Pengujian terjadi dalam dua tahap, yang mana pada masing-masing tahap terjadi tahap write dan read data. Tahap pertama adalah pengujian single write/read data dan tahap kedua adalah pengujian multiple write/read data.

Tahap pertama digunakan untuk mengetahui konsistensi performa write dan read data untuk satu data yang sama. Satu data tersebut secara terus menerus di-write ke basis data berdasarkan fungsi benchmark, baik dengan tambahan proses kompresi maupun tidak. Selanjutnya, proses read merupakan kebalikan proses write-nya. Satu data yang sama, baik data terkompresi maupun tidak, dibaca terus menerus dengan menggunakan fungsi benchmark.

Tahap kedua digunakan untuk mengetahui performa read dan write untuk multiple data, di mana sebanyak 10.000 data ditulis ke basis data menggunakan fungsi benchmark. Proses tulis data ini dilakukan secara terus menerus, baik untuk yang disertai proses kompresi maupun tidak. Selanjutnya, proses read juga dilakukan untuk 10.000 data, baik untuk data terkompresi maupun tidak.

2.2 Pengumpulan Data

Data logging pemantauan keadaan lingkungan yang digunakan pada penelitian ini adalah data sekunder yang didapatkan dari Kaggle. Sebagai komunitas, Kaggle mengizinkan penggunaannya untuk mendapatkan dataset yang telah dibagikan oleh pengguna lain [6]. Parameter-parameter fisik lingkungan yang tersedia pada dataset adalah suhu, kelembapan udara, kandungan gas

karbon monoksida dan liquid petroleum gas (LPG) di udara, kandungan asap di udara, dan ketersediaan cahaya. Data logging ini dikumpulkan selama satu minggu oleh pemilik, mulai dari 12 Juli 2020 pukul 00.00 UTC hingga 19 Juli 2020 pukul 23.59 UTC. Pemantauan dilakukan dengan menggunakan tiga buah perangkat IoT yang sejenis, dilengkapi dengan alamat masing-masing perangkat. Total data logging yang disediakan sebanyak 405.184 data [7].

2.3 Go Benchmark

Bahasa pemrograman Go memiliki fungsionalitas bawaan untuk melakukan benchmark yang berada pada package testing. Benchmark pada Go menggunakan aksioma yang dimulai dengan kata "Benchmark". Aksioma ini menandakan fungsi yang dibuat adalah sebuah fungsi benchmark yang menerapkan fungsionalitas benchmark. Eksekusi fungsi benchmark menggunakan perintah "go test" dengan penanda "-bench" [8]. Fungsi benchmark menjalankan kode target sebanyak b.N kali. Selama fungsi benchmark dijalankan, b.N disesuaikan hingga fungsi benchmark berlangsung cukup lama untuk diukur waktunya. Secara bawaan, benchmark pada Go dijalankan selama satu detik untuk mendapatkan hasil yang signifikan secara statistik [8], [9]. Benchmark akan dijalankan sampai fungsi menghasilkan return, fail, atau skip.

2.4 Delta Encoding

Delta encoding (juga disebut delta compression) merupakan salah satu algoritma yang termasuk ke dalam lossless compression. Algoritma ini dapat digunakan untuk mengurangi biaya penyimpanan dan mengurangi bandwidth untuk mengirimkan data. Ide utama pada algoritma ini adalah untuk meningkatkan kesamaan data sehingga rentang dari dataset akan lebih kecil daripada data asli [10]. Berikut merupakan rumusan delta encoding.

$$\sigma = x_i - x_{i-1} \tag{1}$$

2.5 Perangkat Pengujian

Benchmark proses write/read penggunaan metode kompresi delta encoding dilakukan dengan menggunakan Go Benchmark. Proses pengujian dijalankan dengan metode Go Benchmark pada DBMS MySQL. Pada saat pengujian dilangsungkan, beban kerja perangkat disamakan dengan tidak membuka aplikasi selain Visual Studio Code dan LAMP yang menyediakan paket MySQL server. Spesifikasi perangkat pengujian dapat dilihat pada Tabel 1 di bawah.

Tabel 1. Spesifikasi Perangkat Pengujian

Bagian	Nama dan Spesifikasi
Sistem Operasi	Kali Linux 2023.1
Arsitektur	64-bit
Prosesor	Intel Core i3-2350M
RAM	8 GB

2.6 Parameter Pengujian

Parameter yang digunakan untuk sebagai pembandingan dalam proses benchmark adalah number of iteration dan time per iteration. Number of iteration menunjukkan banyaknya iterasi proses write/read yang dijalankan oleh fungsi benchmark, sedangkan time per iteration menunjukkan waktu yang dibutuhkan untuk melakukan sekali iterasi. Semakin tinggi nilai number of iteration dan semakin rendah nilai time per iteration menunjukkan kasus yang lebih baik. Ukuran data sebelum dan setelah kompresi juga digunakan sebagai parameter pengujian. Parameter ukuran data digunakan untuk mengetahui rasio kompresi dan penghematan ruang yang menunjukkan kinerja algoritma kompresi. Rasio kompresi adalah perbandingan ukuran sebelum kompresi dengan ukuran data setelah kompresi. Rasio ini adalah perhitungan pengurangan ukuran data

relatif oleh algoritma kompresi. Selanjutnya, penghematan ruang menentukan pemotongan ukuran data relatif terhadap ukuran data yang tidak terkompresi. Penghematan ruang biasanya dinotasikan dalam bentuk persentase [11]. Berikut formula rasio kompresi dan penghematan ruang.

$$\text{Compression ratio} = \frac{\text{Uncompressed file size}}{\text{Compressed file size}} \quad (2)$$

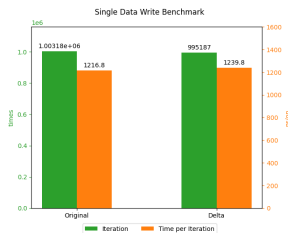
$$\text{Space saving} = 1 - \frac{\text{Compressed file size}}{\text{Uncompressed file size}} \quad (3)$$

3. Hasil dan Pembahasan

```
1- $ go test -bench BenchmarkInsertDeltaData
goos: linux
goarch: amd64
pkg: go-delta
cpu: Intel(R) Core(TM) i3-2350M CPU @ 2.30GHz
BenchmarkInsertDeltaData-4      998538      1208 ns/op
PASS
ok      go-delta      2.041s
```

Gambar 2. Proses Go Benchmark

Terdapat beberapa informasi yang didapatkan saat menjalankan Go Benchmark. Pada Gambar 2 terdapat informasi sistem operasi yang digunakan, arsitektur sistem operasi, package yang sedang menjalankan benchmark, prosesor perangkat, fungsi benchmark yang dijalankan, jumlah core prosesor yang digunakan, jumlah iterasi pada proses benchmark, waktu untuk menjalankan satu iterasi, status benchmark "pass" atau "failed", dan total waktu yang diperlukan untuk menjalankan benchmark. Pengujian menggunakan Go Benchmark dijalankan menggunakan beberapa kasus. Kasus-kasus yang diuji adalah single data write, single data read, multiple data write, dan multiple data read. Setiap kasus benchmark dijalankan sebanyak lima kali untuk mendapatkan nilai benchmark yang lebih presisi dengan menggunakan nilai rata-rata dari kelima pengujian. Setiap kasus juga menerapkan write/read untuk proses tanpa kompresi sebagai baseline dan proses kompresi menggunakan delta encoding.



Gambar 3. Grafik Perbandingan Single Data Write pada Baseline dan Delta Encoding

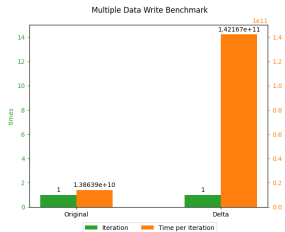
Kasus single data write pada grafik di Gambar 3 menunjukkan proses write data asli (baseline) dan delta encoding memiliki nilai rata-rata iterasi yang tinggi. Saat dijalankan Go Benchmark, rata-rata nilai iterasi baseline adalah 1.003.184 kali dan delta encoding dengan rata-rata nilai iterasi 995.187 kali. Meskipun nilai iterasi pada delta encoding lebih rendah, nilainya hanya terpaud 7.997 kali iterasi. Selanjutnya, pada parameter time per iteration, baseline dapat menyelesaikan satu iterasi dengan waktu 1216,8 nanodetik, sedangkan metode delta encoding memiliki nilai yang lebih tinggi, yaitu 1239,8 nanodetik, terpaud 23 nanodetik. Pada kasus ini, rata-

rata nilai iterasi metode delta encoding saat benchmark adalah 0,992 kali lebih sedikit daripada baseline dan dengan rata-rata waktu per iterasi lebih tinggi 1,019 kali daripada baseline. Perbedaan perbandingan metode delta encoding yang tidak terpaut jauh dari baseline menunjukkan delta encoding efektif diterapkan untuk kasus single data write.



Gambar 4. Grafik Perbandingan Single Data Read pada Baseline dan Delta Encoding

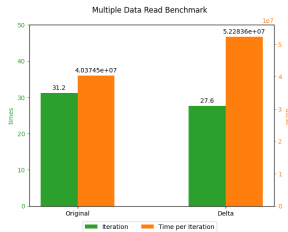
Kasus single data read merupakan kebalikan dari single data write. Satu data, baik data asli maupun yang telah di-delta encoding disimpan ke MySQL lalu dibaca dan di-benchmark. Grafik pada Gambar 4 merupakan hasil benchmark single data read. Pembacaan data asli (baseline) dapat menghasilkan iterasi rata-rata sebanyak 1.786,2 kali, sedangkan pembacaan data yang telah di-delta encoding menghasilkan iterasi rata-rata sebanyak 1.760,2 kali, terpaut 26 kali. Pembacaan data asli memerlukan waktu rata-rata 625.119 nanodetik untuk menjalankan satu kali iterasi, sedangkan pembacaan data yang telah di-delta encoding memerlukan waktu rata-rata 637.736 nanodetik untuk satu kali iterasi. Waktu ini terpaut 12.617 nanodetik. Pada kasus ini, nilai iterasi metode delta encoding saat benchmark adalah 0,985 kali lebih rendah dari baseline. Sedangkan, waktu per iterasi metode delta encoding lebih tinggi 1,02 kali daripada baseline. Perbedaan perbandingan parameter iteration dan time per iteration yang kecil (tidak mencapai 2 kali) menunjukkan proses decode data yang telah di-delta encoding tidak jauh lebih berat daripada membaca data asli yang tidak dikenai proses kompresi.



Gambar 5. Grafik Perbandingan Multiple Data Write pada Baseline dan Delta Encoding

Kasus selanjutnya adalah multiple data write. Pada kasus ini, benchmark dilakukan dengan menuliskan sebanyak 10.000 data, baik data asli (baseline) maupun yang telah di-delta encoding. Grafik pada Gambar 5 menunjukkan proses penulisan 10.000 data ke database merupakan proses yang berat. Benchmark pada baseline maupun delta encoding untuk kasus ini hanya terjadi 1 kali iterasi. Namun, time per operation menunjukkan baseline dan delta encoding memiliki waktu yang terpaut jauh. Baseline memerlukan waktu rata-rata 13.863.918.358,2

nanodetik (± 13 detik) untuk satu iterasi, sedangkan delta encoding memerlukan waktu rata-rata 142.167.422.453,4 nanodetik (± 142 detik) untuk satu iterasi. Terpaut ± 129 detik, di mana delta encoding memerlukan waktu untuk satu iterasi 10,254 kali lebih tinggi daripada baseline. Tingginya waktu untuk satu iterasi ini disebabkan karena proses encode pada data mengharuskan untuk mengambil jumlah nilai delta pada masing-masing field terlebih dahulu dari basis data. Jumlah nilai delta pada data-data sebelumnya digunakan untuk mencari nilai delta untuk data yang akan dimasukkan ke basis data. Perbedaan time per iteration pada proses delta encoding menunjukkan proses kompresi data dalam jumlah yang banyak sekaligus tidak efektif untuk dilakukan dengan menggunakan delta encoding.



Gambar 6. Grafik Perbandingan Multiple Data Read pada Baseline dan Delta Encoding

Kasus multiple data read merupakan proses pembacaan pada data sensor yang telah dilakukan delta encoding. Grafik pada Gambar 6 menggambarkan benchmark untuk multiple data read. Proses pembacaan data asli (baseline) saat di-benchmark dapat dilakukan iterasi rata-rata sebanyak 31,2 kali, sedangkan untuk data yang telah di-delta encoding rata-rata sebanyak 27,6 kali. Proses pembacaan data dengan delta encoding memiliki rata-rata iterasi yang lebih kecil, yaitu lebih sedikit 3,6 kali dibandingkan dengan baseline. Waktu satu iterasi yang diperlukan untuk membaca 10.000 data pada baseline adalah 40.374.522,6 nanodetik (± 40 milidetik), sedangkan pada data dengan delta encoding adalah 52.283.589,4 nanodetik (± 52 milidetik). Waktu satu iterasi ini terpaut 11.909.066,8 nanodetik (± 11 milidetik). Pada kasus ini, nilai iterasi pada metode delta encoding lebih rendah 0,88 kali dibandingkan baseline. Sedangkan waktu rata-rata per iterasi pada metode delta encoding lebih tinggi 1,294 kali dibandingkan baseline. Mirip dengan kasus multiple data write, waktu rata-rata per iterasi yang lebih tinggi pada metode delta encoding disebabkan karena adanya proses decode dengan mencari jumlah data pada setiap nilai delta untuk mendapatkan data asli.

Pada parameter ukuran data, data logging tanpa proses kompresi memiliki ukuran 1.293.040 byte, sedangkan data logging yang telah dikompresi dengan menggunakan delta encoding memiliki ukuran 722.873 byte. Rasio kompresi pada 10.000 data logging sensor adalah sebesar 1,789 (1,789:1) dengan penghematan ruang sebesar 44,096%. Nilai penghematan ruang ini menunjukkan delta encoding dapat digunakan untuk mengompresi data dengan baik, yaitu hampir setengah dari ukuran data sebelum dikompresi. Namun, perlu diperhatikan bahwa waktu yang diperlukan akan semakin bertambah untuk mengompresi data dengan jumlah yang semakin banyak pula.

4. Kesimpulan

Penelitian yang dilakukan bertujuan untuk menganalisis performa write/read algoritma delta encoding terhadap 10.000 data logging sensor pendeteksi kesehatan lingkungan. Berdasarkan parameter pengujian, yaitu number of iteration, time per iteration, rasio kompresi, dan penghematan ruang, delta encoding menunjukkan performa yang optimal, terutama pada kasus single data write, single data read, dan multiple data read. Performa optimal ini ditunjukkan oleh

nilai pada masing-masing parameter pengujian yang tidak terpaut jauh dengan baseline. Namun perlu diperhatikan saat menggunakan delta encoding untuk mengompresi beberapa data dalam jumlah yang besar sekaligus. Waktu yang diperlukan akan jauh lebih tinggi dibandingkan dengan baseline, yaitu mencapai 10,254 kali lebih tinggi. Oleh karena itu, delta encoding akan lebih baik digunakan untuk proses write satu data pemantauan (single data write) sehingga beban kerja server akan lebih kecil jika dibandingkan dengan menuliskan beberapa data pemantauan sekaligus, terlebih lagi data dengan jumlah yang besar. Selanjutnya, jika diperlukan untuk write banyak data pemantauan sekaligus (multiple data write), maka perlu dipertimbangkan untuk membagi data menjadi beberapa kluster untuk mengurangi waktu write data untuk satu kali eksekusi.

Daftar Pustaka

- [1] O. R. Pinontoan dan O. J. Sumampouw, Dasar Kesehatan Lingkungan. Sleman: Deepublish Publisher, 2019. Diakses: 29 April 2023. [Daring]. Tersedia pada: https://books.google.co.id/books?hl=en&lr=&id=ki3HDwAAQBAJ&oi=fnd&pg=PR5&dq=parameter+kesehatan+lingkungan&ots=rPxKCO0Nxl&sig=Z7B4cieoBi9Vvc4NfktA6fpzjw4&redir_esc=y#v=onepage&q&f=false
- [2] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, dan B. Qureshi, "An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques," *Sensors*, vol. 20, no. 21, hal. 6076, Okt 2020, doi: 10.3390/S20216076.
- [3] O. A. Ahmad, H. Sayed, K. A. Jalal, D. Y. Mahmood, W. H. Habeeb, dan O. A. Ahmed, "Design and implementation of an indoor solar emulator based low-cost autonomous data logger for PV system monitoring," *Int. J. Power Electron. Drive Syst.*, vol. 10, no. 3, hal. 1645–1654, 2019, doi: 10.11591/ijpeds.v10.i3.pp1645-1654.
- [4] D. Blalock, S. Madden, dan J. Guttag, "Sprintz: Time Series Compression for the Internet," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, hal. 1–23, Sep 2018, doi: 10.1145/3264903.
- [5] H. Devi Kotha, M. Tummanapally, dan V. K. Upadhyay, "Review on Lossless Compression Techniques," *J. Phys. Conf. Ser.*, vol. 1228, no. 1, hal. 012007, Mei 2019, doi: 10.1088/1742-6596/1228/1/012007.
- [6] M. H. Mahmoud, "What is a Kaggle?," 1 Juni 2022. <https://www.kaggle.com/general/328265> (diakses 1 Mei 2023).
- [7] G. A. Stafford, "Environmental Sensor Telemetry Data," 21 Juli 2020. <https://www.kaggle.com/datasets/garystafford/environmental-sensor-data-132k> (diakses 1 Mei 2023).
- [8] B. Strecansky, *Hands-On High Performance with Go*. Birmingham: Packt Publishing Ltd., 2020. Diakses: 3 Mei 2023. [Daring]. Tersedia pada: https://books.google.co.id/books?hl=id&lr=&id=5C7ZDwAAQBAJ&oi=fnd&pg=PP1&dq=golang+benchmark&ots=IFrriR9L-z&sig=gJymaEU5rPZ924sGVVntzn23Mdk&redir_esc=y#v=onepage&q=golangbenchmark&f=false
- [9] Go, "testing package," 2 Mei 2023. <https://pkg.go.dev/testing> (diakses 3 Mei 2023).
- [10] A. Saidani, X. Jianwen, dan D. Mansouri, "A Lossless Compression Approach Based on Delta Encoding and T-RLE in WSNs," *Wirel. Commun. Mob. Comput.*, vol. 2020, hal. 1–10, Sep 2020, doi: 10.1155/2020/8824954.
- [11] A. Gopinath dan M. Ravisankar, "Comparison of Lossless Data Compression Techniques," in 2020 International Conference on Inventive Computation Technologies (ICICT), Feb 2020, hal. 628–633. doi: 10.1109/ICICT48043.2020.9112516.

Halaman ini sengaja dibiarkan kosong

Purwarupa Perangkat Monitoring Ketinggian Air Tandon Menggunakan TinkerCAD

I Kadek Widiarthawan^{a1}, I Komang Ari Mogi^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana

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

¹kadekwidiarthawan092@Student.unud.ac.id

²arimogi@unud.ac.id

Abstract

Along with population growth and urban development, the demand for clean water is increasing. Water reservoirs are an effective solution to this problem, by storing large amounts of water in specially designed tanks. Water reservoirs are one of the important components in the household water supply system. This study aims to make it easier for users to monitor the condition of the water in the reservoir. The tools used are Arduino uno R3, green LED, yellow LED, red LED, ultrasonic distance sensor, DC motor, piezo, push button, and h-bridge motor driver. This system is designed using the prototype method, so that users can easily see real-time water levels via mobile devices.

Keywords: Water reservoir, Arduino uno R3, Prototype.

1. Pendahuluan

Air merupakan sumber daya yang sangat berharga bagi kehidupan manusia dan ekosistem di seluruh dunia [1]. Namun, sayangnya, sering kali terjadi pemborosan air yang tidak perlu, baik itu dalam rumah tangga, perkantoran, atau bahkan di skala yang lebih besar seperti perkotaan. Pemborosan air dapat terjadi karena berbagai alasan, mulai dari kebiasaan yang kurang bijaksana dalam menggunakan air hingga kebocoran yang tak terdeteksi dalam infrastruktur air.

Untuk mengatasi masalah ini, berbagai upaya telah dilakukan untuk mengurangi pemborosan air, salah satunya adalah dengan menciptakan sensor yang terpasang di tandon atau wadah penyimpanan air. Sensor ini berfungsi untuk mendeteksi secara akurat level air yang ada di dalam tandon dan memberikan informasi yang diperlukan untuk mengelola penggunaan air dengan lebih efisien [2].

Sensor yang ada di tandon air dapat mengukur tingkat air yang tersedia dan memberikan peringatan saat air mencapai tingkat minimum atau maksimum yang telah ditentukan. Dengan adanya sensor ini, pengguna dapat memantau konsumsi air mereka secara real-time dan mengambil langkah-langkah untuk mengurangi pemborosan air yang tidak perlu [3].

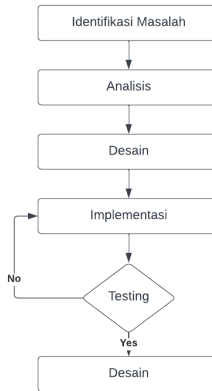
Penerapan sensor di tandon air juga dapat membantu dalam pengelolaan air secara keseluruhan di sebuah bangunan atau komunitas. Dengan informasi yang akurat tentang konsumsi air. Dalam era yang semakin terhubung secara digital, sensor-sensor ini juga dapat terhubung dengan sistem pemantauan yang lebih luas, seperti Internet of Things (IoT). Hal ini memungkinkan pengguna untuk mengakses informasi tentang penggunaan air secara langsung melalui perangkat pintar mereka, bahkan dari jarak jauh [4].

Sensor inovatif yang dirancang khusus untuk mengukur ketinggian tandon air dengan akurasi tinggi dan kemudahan penggunaan. Sensor tersebut akan diimplementasikan ke dalam prototipe sebagai bagian dari sistem atau perangkat keras yang dikembangkan. Sensor tersebut akan terhubung dengan mikrokontroler atau perangkat elektronik lainnya dalam prototipe untuk mengumpulkan data [5].

2. Metode Penelitian

2.1 Metode

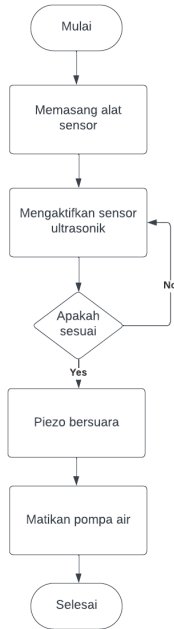
Metode rapid prototyping adalah pendekatan yang digunakan dalam pengembangan prototipe dengan tujuan menghasilkan prototipe fungsional secara cepat. Metode ini memanfaatkan teknologi canggih, seperti pencetakan 3D, pencetakan cepat, atau teknik manufaktur lainnya untuk membuat prototipe dengan presisi dan kecepatan tinggi [6].



Gambar 1. Metode Rapid Prototype

- a. Identifikasi masalah
- b. Identifikasi masalah-masalah dalam pelaksanaan proses mengukur ketinggian air dan evaluasi kinerja berjalan dengan menekankan pada aspek kelayakan pengembangan sistem.
- c. Analisis
- d. Pada tahap analisis, penelitian melakukan beberapa langkah seperti:
- e. Melakukan klasifikasi masalah dan solusi yang diterapkan pada sistem
- f. Desain
- g. Pada tahapan desain dilakukan analisa data dan pemodelan sistem.
- h. Implementasi
- i. Pada tahap implementasi peneliti melakukan perangkaian pada sensor ultrasonik serta Bahasa C++ untuk menjalankan rangkaian tersebut.
- j. Testing
- k. Pada tahapan pengujian peneliti melakukan percobaan yang telah selesai.

2.2 Alur Sistem

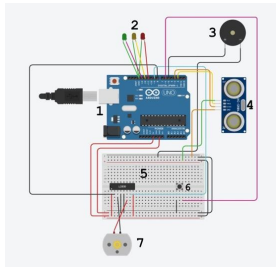


Gambar 2. Alur Sistem

3. Hasil dan Diskusi

Pada penelitian ini menggunakan metode rapid prototyping untuk membuat prototyping pada penelitian ini menggunakan tinkercad. Tinkercad merupakan aplikasi web dengan kemudahannya dalam mengakses dan menggunakan alat ini dari berbagai perangkat yang terhubung ke internet tanpa batasan perangkat keras atau sistem operasi tertentu. Dalam memprogram arduino di tinkercad menggunakan bahasa pemrograman C++.

3.1 Rangkaian Arduino



Gambar 4. Rangkaian Arduino

Deskripsi:

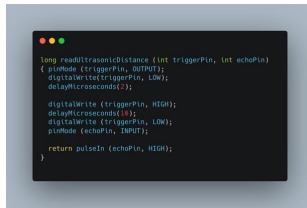
1. Arduino Uno R3
2. Lampu LED
3. Piezo
4. Ultrasonic Distance Sensor
5. H-bridge Motor Driver
6. Pushbutton
7. DC Motor

Tabel 1. Fungsi Komponen

Komponen	Fungsi
Arduino Uno R3	Untuk membaca data yang diterima dari sensor
LED	Memberikan simbol
Ultrasonic Distance Sensor	Untuk mengukur jarak antara sensor dan permukaan air
DC Motor	Untuk menggerakkan mekanisme yang memonitor tinggi rendahnya air
Piezo	Untuk mengukur perubahan tekanan yang terjadi akibat perubahan ketinggian air (berupa suara).
Pushbutton	Untuk mengontrol atau memicu tindakan tertentu
H-bridge Motor Driver	Untuk mengendalikan arah putaran motor DC

a. Bahasa C++

Bahasa C++ pada Arduino adalah penggunaan bahasa C++ dalam pengembangan perangkat keras dengan menggunakan platform Arduino. Arduino adalah platform open-source yang menggunakan mikrokontroler sebagai dasar pengembangan, dan bahasa C++ digunakan sebagai bahasa pemrograman utama untuk menulis kode yang mengontrol dan berinteraksi dengan perangkat keras.



```
long readUltrasonicDistance (int triggerPin, int echoPin)
{
  pinMode (triggerPin, OUTPUT);
  digitalWrite(triggerPin, LOW);
  delayMicroseconds(2);

  digitalWrite (triggerPin, HIGH);
  delayMicroseconds(10);
  digitalWrite (triggerPin, LOW);
  pinMode (echoPin, INPUT);

  return pulseIn (echoPin, HIGH);
}
```

Gambar 5. Code

Kode tersebut merupakan fungsi dalam bahasa C++ untuk mengukur jarak menggunakan sensor ultrasonik pada Arduino. Fungsi tersebut menggunakan pin trigger dan pin echo yang telah ditentukan sebagai argumen fungsi.



```
void setup() {
  Serial.begin(9600);

  pinMode (1, OUTPUT);
  pinMode (3, OUTPUT);
  pinMode (7, OUTPUT);
  pinMode (LED3, OUTPUT);
  pinMode (LED2, OUTPUT);
  pinMode (LED1, OUTPUT);
  pinMode (PUSH, INPUT);
  digitalWrite (PUSH, HIGH);
}
```

Gambar 6. Code

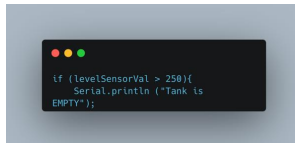
Fungsi setup () digunakan untuk melakukan inisialisasi awal saat Arduino dinyalakan.



```
void loop(){
  echoPin = 3;
  triggerPin = 2;
  levelSensorVal = 0.01723 * readUltrasonicDistance(2, 3);
  Serial.println("Level Control System");
}
```

Gambar 7. Code

Fungsi loop () dieksekusi secara berulang terus-menerus setelah fungsi setup () selesai dieksekusi.



```
if (levelSensorVal > 250){  
  Serial.println ("Tank is  
  EMPTY");  
}
```

Gambar 8. Code

Pernyataan if digunakan untuk melakukan pengecekan kondisi. Jika kondisi di dalam tanda kurung () bernilai benar (nilai levelSensorVal lebih besar dari 250), maka blok kode di dalam kurung kurawal {} akan dieksekusi.

Dalam hal ini, jika nilai levelSensorVal lebih besar dari 250, maka pesan "Tank is EMPTY" akan ditampilkan pada Serial Monitor menggunakan perintah Serial.println().



```
else if (levelSensorVal <= 400 && levelSensorVal > 50)  
{  
  digitalWrite (8, HIGH);  
  digitalWrite (9, HIGH);  
  digitalWrite (LED1, LOW);  
  digitalWrite (LED3, LOW);  
  digitalWrite (LED2, HIGH);  
}
```

Gambar 9. Code

Kode tersebut merupakan bagian dari sebuah pernyataan kondisional yang memeriksa nilai variabel bernama "levelSensorVal" dan melakukan beberapa tindakan berdasarkan kondisi tersebut. Berdasarkan kode yang diberikan, jika nilai "levelSensorVal" kurang dari atau sama dengan 400 dan lebih dari 50, tindakan-tindakan berikut akan dilakukan:

- Pin 8 akan diatur ke keadaan HIGH.
- Pin 9 akan diatur ke keadaan HIGH.
- Pin "LED1" akan diatur ke keadaan LOW.
- Pin "LED3" akan diatur ke keadaan LOW.
- Pin "LED2" akan diatur ke keadaan HIGH.



```
else {  
  Serial.println("Tank is  
  FULL");  
}
```

Gambar 10. Code

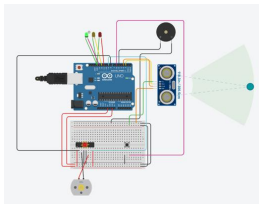
Jika semua kondisi tidak terpenuhi, Dalam hal ini, pesan "Tank is FULL" akan ditampilkan pada Serial Monitor menggunakan perintah Serial.println().



Gambar 11. Code

Jika tombol pada pin PUSH ditekan atau saklar dalam keadaan tertutup, pesan "Alarm Bunyi" akan ditampilkan pada Serial Monitor dan bunyi dengan frekuensi 19 Hz akan dihasilkan pada pin 5 selama 2 detik. Dapat menambahkan perintah atau blok kode lain di dalam blok if tersebut untuk menjalankan aksi tambahan ketika kondisi terpenuhi.

b. Rangkaian Arduino LED Hijau



Gambar 12. Rangkaian Arduino

Pada gambar diatas lampu LED warna hijau menyala yang menandakan air pada tandon sedang diisi air/tidak isi air. Lampu LED hijau menyala dari rentang 250 - 300 cm, pada monitor akan menampilkan "Tank is EMPTY" atau pada tandon sedang tidak ada air/sedang diisi air.

4. Kesimpulan

Dari uraian dan hasil analisis di atas dapat disimpulkan bahwa data atau informasi dapat dihasilkan melalui aktivitas sehari-hari, misalnya saat memeriksa tinggi air di dalam tandon air. Data ini dapat dihasilkan oleh sensor yang terpasang pada tandon air. Metode ini juga dapat dikembangkan untuk penelitian selanjutnya yang cakupannya lebih luas. Data ini dikumpulkan dari tandon air yang di pasang menggunakan arduino uno R3. Sistem berguna untuk menginformasikan kepada pemilik tentang ketinggian air di tandon melalui sistem. Perancangan tandon air ini melibatkan penggunaan sensor ultrasonik berbasis Arduino Uno. Sensor akan memantulkan sinyal ultrasonik dari penampung air dan kembali ke sensor. Setelah volume air di dalam tandon mencapai batas 50 cm dibawah sensor, maka fizeo akan memberi peringatan berupa suara untuk mengingatkan mematikan pompa air.

Daftar Pustaka

- [1] Suntari, L. (2017). *Dinamika Pengelolaan Sumber Daya Air di Desa Pudak Kulon Kecamatan Pudak Kabupaten Ponorogo* (Doctoral dissertation, Universitas Muhammadiyah Ponorogo).
- [2] Darmawan, I. D. M. B. A., Mogi, I. K. A., & Santiyasa, I. W. (2017). Sistem Instalasi Air Rumah Terkomputerisasi Berbasis Mikrokontroler Dengan Perintah Sms. *JST (Jurnal Sains dan Teknologi)*, 6(1).
- [3] Susantok, M., & Ramadhan, T. (2021). Manajemen Ketersediaan dan Penggunaan Air pada Rumah Tangga Berbasis IoT. *Jurnal Elektro dan Mesin Terapan*, 7(1), 1-10.
- [4] Dewanto, E., Yoseph, J., & Rifâ, M. (2018). Tandon Air Otomatis Dengan Sistem Monitoring Melalui Android Berbasis Arduino Uno. *Autocracy: Jurnal Otomasi, Kendali, dan Aplikasi Industri*, 5(01), 8-16.
- [5] Yuliaminuddin, V., & Bintoro, J. (2020). Prototipe Sistem Kontrol Dan Monitoring Pada Tangki Air Berbasis Internet of Things. *Autocracy: Jurnal Otomasi, Kendali, dan Aplikasi Industri*, 7(1), 27-34.
- [6] Kusumah, H., & Kurniawan, F. (2018, July). Perancangan Prototipe Smart Chair Berbasis Arduino. In *Proceeding Seminar Nasional Sistem Informasi dan Teknologi Informasi* (Vol. 1, No. 1, pp. 281-286).

Halaman ini sengaja dibiarkan kosong

Pengenkripsian dan Dekripsi Gambar Menggunakan Algoritma AES dengan MAC untuk Peningkatan Keamanan

Ni Wayan Amanda Putri Astawa^{a1}, I Made Widiartha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia

¹putrimanda985@gmail.com

²madewidiartha@unud.ac.id

Abstract

With the increasing importance of data security in digital image transmission and storage, this research presents an implementation of an image encryption and decryption program using the Advanced Encryption Standard (AES) algorithm combined with Message Authentication Code (MAC) for enhanced security. The program utilizes AES in Cipher Block Chaining (CBC) mode to ensure confidentiality and integrity of the image data. The unique key and initialization vector (IV) enhance the security of the encryption process. Additionally, the inclusion of MAC ensures data integrity and prevents unauthorized modifications during transmission or storage. The program offers a user-friendly web-based interface for easy usability. The implemented solution provides a high level of security for image data and can be applied in various applications requiring secure image transmission and storage. The effectiveness and reliability of the program are demonstrated through experimental results and evaluation.

Keywords: *Advanced Encryption Standard, Message Authentication Code, Hash, Encryption, Decryption*

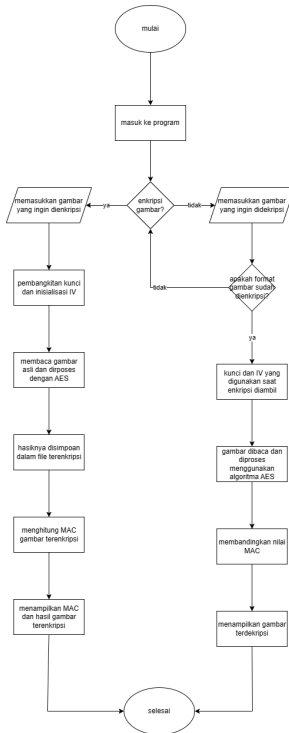
1. Pendahuluan

Dalam era digital yang terus berkembang, pertukaran informasi secara konvensional sudah mulai ditinggalkan dan digantikan dengan dunia maya atau digital. Yang dimana segala informasi dapat diakses melalui internet secara bebas. Salah satu contoh nyatanya adalah penggunaan gambar sebagai media komunikasi dan pertukaran informasi telah menjadi sangat umum untuk dilakukan [1]. Namun, dengan adanya perkembangan teknologi, kebutuhan untuk melindungi keamanan dan integritas gambar juga turut meningkat. Ancaman keamanan seperti peretasan, pencurian data, dan modifikasi yang tidak sah dapat mengancam kerahasiaan dan integritas gambar yang sensitif. Untuk dapat mengantisipasi terjadinya ancaman keamanan tersebut, diperlukan metode keamanan yang efektif dan juga kuat dalam menjaga kerahasiaan dan juga integritas gambar.

Algoritma Advanced Encryption Standard (AES) dapat menjadi salah satu solusi yang cukup efektif dalam mengantisipasi terjadinya ancaman dalam citra gambar. Metode AES akan digunakan untuk mengenkripsi gambar, sehingga hanya pihak yang memiliki kunci enkripsi yang tepat dapat mendekripsikannya. Selanjutnya, digunakan pula algoritma Message Authentication Code (MAC) yang akan memastikan integritas data dengan menandatangani gambar yang telah dienkripsi, sehingga perubahan yang tidak sah dapat terdeteksi. Dalam penelitian ini, penulis juga menggunakan fungsi Hash sebagai mekanisme verifikasi integritas dengan menyimpan nilai hash dari gambar yang telah dienkripsi dan ditandatangani.

2. Metode Penelitian

2.1 Desain Sistem



Gambar 1. Flowchart Sistem

Pertama, pengguna diminta untuk memilih gambar yang akan dienkripsi. Setelah gambar asli dipilih, program membuka gambar tersebut dan melakukan proses enkripsi. Gambar terenkripsi kemudian disimpan ke dalam file. Selain itu, program juga menghitung MAC (Message Authentication Code) dari gambar asli sebagai tanda verifikasi. MAC tersebut disimpan dalam file terpisah yang akan digunakan nanti.

Ketika pengguna ingin mendekripsi dan memverifikasi gambar terenkripsi, mereka memilih gambar terenkripsi yang akan diolah. Program membuka gambar terenkripsi dan memulai proses verifikasi MAC. Verifikasi ini penting untuk memastikan bahwa gambar terenkripsi tidak

mengalami perubahan atau kerusakan yang tidak sah. Jika verifikasi MAC berhasil, program melanjutkan dengan proses dekripsi. Gambar terenkripsi didekripsi dan hasilnya disimpan ke dalam file terpisah. Akhirnya, gambar terdekripsi ditampilkan kepada pengguna untuk dilihat.

2.2 Pemrosesan Awal

Pada tahap ini, gambar yang diunggah oleh pengguna diambil sebagai input. Langkah pertama yang dilakukan adalah memvalidasi format gambar untuk memastikan bahwa gambar yang diunggah memiliki format yang valid, seperti JPEG, PNG, atau format gambar lainnya. Validasi ini penting untuk memastikan bahwa gambar dapat diterima dan diproses dengan benar. Setelah validasi format, gambar yang valid kemudian akan dienkripsi menggunakan algoritma enkripsi yang dipilih, dalam kasus ini yaitu Advanced Encryption Standard (AES). Tujuan dari enkripsi ini adalah untuk melindungi kerahasiaan gambar dengan mengubah kontennya menjadi bentuk yang tidak dapat dibaca oleh pihak yang tidak berwenang.

Selanjutnya, dalam rangka memverifikasi integritas gambar, dilakukan pembangkitan Message Authentication Code (MAC) menggunakan algoritma hash seperti SHA-256. MAC ini berguna untuk memastikan bahwa gambar tidak mengalami perubahan atau kerusakan pada saat ditransmisikan atau disimpan. MAC akan menjadi referensi untuk memverifikasi apakah gambar telah mengalami perubahan yang tidak sah. Setelah gambar dienkripsi dan MAC dihasilkan, gambar terenkripsi beserta MAC-nya disimpan dalam penyimpanan yang relevan, seperti file atau database. Hal ini bertujuan untuk menjaga keamanan gambar yang terenkripsi dan memfasilitasi proses pemulihan saat diperlukan.

Dalam keseluruhan pemrosesan awal ini, langkah-langkah tersebut memiliki peran penting dalam menjaga keamanan dan integritas gambar pada aplikasi berbasis web. Validasi format gambar memastikan bahwa gambar yang diunggah sesuai dengan harapan, sementara enkripsi dan MAC membantu melindungi kerahasiaan dan memverifikasi integritas gambar.

2.3 Proses Enkripsi

Proses enkripsi dalam program ini dilakukan menggunakan algoritma AES (Advanced Encryption Standard) dalam mode CBC (Cipher Block Chaining), yang memberikan tingkat keamanan yang tinggi. Pertama-tama, pengguna memilih gambar yang akan dienkripsi melalui antarmuka aplikasi berbasis web. Setelah gambar dipilih, program membuka gambar asli dan mempersiapkan kunci dan vektor inisialisasi (IV) yang diperlukan untuk proses enkripsi. Selanjutnya, gambar asli dienkripsi dengan menggunakan algoritma AES dalam mode CBC. Pada mode CBC, setiap blok gambar dienkripsi secara berurutan dengan blok sebelumnya sebagai vektor inisialisasi. Hal ini membantu dalam menciptakan perubahan yang signifikan pada setiap blok enkripsi, sehingga menjaga kerahasiaan dan keamanan data.

Selain enkripsi gambar, program juga menerapkan algoritma MAC (Message Authentication Code) untuk memastikan integritas dan otentikasi data. Algoritma MAC yang digunakan dalam program ini adalah HMAC-SHA256 (Hash-based Message Authentication Code dengan fungsi hash SHA256). MAC dihasilkan dari gambar terenkripsi dan kunci yang sama yang digunakan dalam proses enkripsi. Fungsi MAC memastikan bahwa gambar terenkripsi tidak mengalami perubahan atau manipulasi selama proses penyimpanan atau transfer. Dengan menggabungkan enkripsi AES dalam mode CBC dan algoritma MAC, program ini memberikan keamanan yang kuat terhadap gambar yang dienkripsi. Enkripsi melindungi kerahasiaan gambar asli, sementara MAC memastikan integritas dan otentikasi data pada tahap dekripsi. Kombinasi kedua teknik ini memberikan lapisan keamanan yang kokoh dan melindungi gambar terenkripsi dari serangan dan manipulasi yang tidak sah.

2.4 Proses Dekripsi

Proses dekripsi dalam program ini dilakukan setelah pengguna memilih gambar terenkripsi melalui antarmuka aplikasi berbasis web. Setelah gambar terenkripsi dipilih, program membuka file gambar terenkripsi dan memulai proses dekripsi. Pertama, program membaca isi file gambar

terenkripsi dan mempersiapkan kunci dan vektor inisialisasi (IV) yang sama yang digunakan dalam proses enkripsi. Kunci dan IV ini penting untuk mendekripsi gambar dengan benar. Selanjutnya, program menggunakan algoritma AES dalam mode CBC untuk mendekripsi gambar terenkripsi. Dengan menggunakan kunci dan IV yang tepat, setiap blok gambar terenkripsi didekripsi secara berurutan. Proses dekripsi ini mengembalikan gambar ke bentuk aslinya.

Selama proses dekripsi, program juga memverifikasi integritas dan otentikasi gambar dengan menggunakan algoritma MAC (Message Authentication Code). Program menghasilkan MAC baru dari gambar terdekripsi dan menggunakan kunci yang sama yang digunakan dalam proses enkripsi. MAC ini kemudian dibandingkan dengan MAC asli yang disimpan bersama gambar terenkripsi. Jika kedua MAC cocok, itu menunjukkan bahwa gambar tidak mengalami perubahan selama proses penyimpanan atau transfer. Setelah proses dekripsi selesai, gambar terdekripsi ditampilkan ke pengguna melalui antarmuka aplikasi. Pengguna dapat melihat gambar dalam bentuk aslinya, seperti sebelum dienkripsi. Proses dekripsi ini memberikan kemampuan untuk mengembalikan gambar ke bentuk aslinya setelah melalui proses enkripsi. Algoritma AES dalam mode CBC dan verifikasi MAC memastikan bahwa gambar terdekripsi akurat dan tidak mengalami modifikasi yang tidak sah.

3. Hasil dan Pembahasan

3.1 Uji Coba Sistem

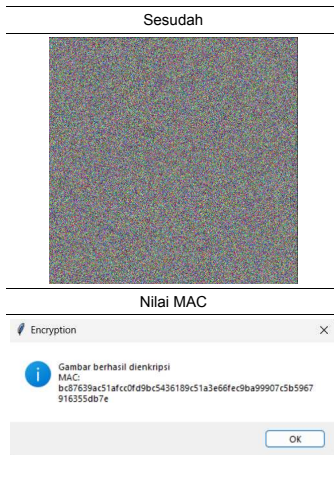
a. Proses Enkripsi

Pengujian ini dilakukan untuk melihat apakah algoritma AES dapat mengenkripsi gambar dengan langkah-langkah sebagai berikut:

1. Memilih gambar yang akan dienkripsi.
2. Menghasilkan kunci acak dan inisialisasi vektor (IV) untuk AES.
3. Membuka gambar asli.
4. Mengenkripsi gambar menggunakan AES dalam mode CBC dengan kunci dan IV yang dihasilkan.
5. Menghasilkan MAC (Message Authentication Code) dari gambar terenkripsi untuk integritas dan autentikasi.
6. Menyimpan gambar terenkripsi dan MAC ke dalam file.
7. Menghasilkan output berupa gambar terenkripsi dan MAC.

Tabel 1. Proses Enkripsi



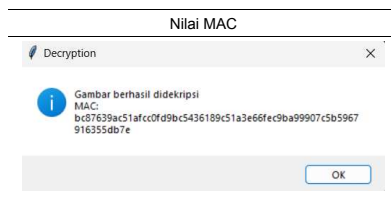


b. Proses Dekripsi

Pengujian ini dilakukan untuk melihat apakah algoritma AES dapat mendekripsi gambar dengan langkah-langkah sebagai berikut:

1. Memilih gambar terenkripsi yang akan didekripsi.
2. Membuka gambar terenkripsi dan MAC dari file.
3. Memverifikasi integritas dan autentikasi gambar terenkripsi menggunakan MAC.
4. Mendapatkan kunci dan IV yang digunakan saat proses enkripsi.
5. Mendekripsi gambar menggunakan AES dalam mode CBC dengan kunci dan IV yang sesuai.
6. Menghilangkan padding yang ditambahkan saat enkripsi.
7. Menyimpan gambar terdekripsi ke dalam file.
8. Menghasilkan output berupa gambar terdekripsi.

Tabel 2. Proses Dekripsi



Dapat kita lihat bahwa nilai MAC dari enkripsi dan dekripsi sama. Hal itu menunjukkan bahwa gambar tidak mengalami perubahan selama proses penyimpanan atau transfer.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan program enkripsi dan dekripsi gambar dengan menggunakan algoritma AES dan MAC. Program ini memberikan tingkat keamanan yang tinggi terhadap data gambar dengan menggunakan kunci dan IV yang unik, serta mode CBC untuk melindungi kerahasiaan data. Adanya algoritma MAC juga menjaga integritas data. Implementasi ini dapat digunakan dalam berbagai aplikasi yang membutuhkan keamanan data gambar.

Daftar Pustaka

- [1] B. S. A. Priandana and I. M. Widiartha, "Pengembangan Aplikasi Berbasis Mobile Untuk Pengamanan Teks Menggunakan Metode Advanced Encryption Standard dan Least Significant Bit," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*
- [2] G. D. M. Zulma, H. B. Seta and T. Yuniati, "Implementasi Algoritma AES Dan Bcrypt untuk Pengamanan File Dokumen," *Jurnal Informatik*, 2022.
- [3] G. Y. Pangestu, A. I. Hadiana and P. N. Sabrina, "Kriptografi Untuk Enkripsi Ganda Pada Gambar Menggunakan Algoritma AES (Advanced Encryption Standard) Dan RC5 (Rivest Code 5)," *Informatics And Digital Expert (Index)*, vol. IV, 2022.
- [4] A. T. Hlaing and M. T. M. Win, "Secure Image Steganography using Canny Magic LSB Substitution Method and HMAC Algorithm," in *Universal Academic Cluster International July Conference*, Bangkok, 2019.
- [5] R. N. Sihombin and Y. Hasan, "Implementasi Algoritma Advanced Encryption Standard (Aes) Dalam Mengamankan File Citra Dokumen," in *Seminar Nasional Inovasi dan Teknologi Informasi SNITI*, Medan, 2016.

Analisa UI dalam Perancangan Aplikasi Ekspor Hasil Laut “FishGo!” dengan Metode SUS

Ketut Agus Cahyadi Nanda^{a1}, I Putu Gede Hendra Suputra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹gusnanda1170@student.unud.ac.id
²hendra.suputra@unud.ac.id

Abstract

This research discusses the development of the Fish Go! application, which aims to facilitate fishermen in finding market opportunities, particularly within the business sector. The author identifies several challenges faced by fishermen in terms of marketing and selling their catch. Additionally, there are opportunities to meet the demand for marine products in the tourism sector in Indonesia. The Fish Go! application seeks to integrate fishermen and the business sector into a single digital platform, with the goal of expanding the fishermen's market reach and providing convenience for buyers in obtaining fresh and high-quality seafood. This research involves a literature study to gain a deeper understanding of application interface analysis and usability in general. The User Journey Map method is used to identify the problems faced by potential users of the Fish Go! application. The usability testing phase is conducted using the System Usability Scale (SUS) method, involving the distribution of questionnaires to 20 respondents. The quantitative data obtained from the SUS scores indicate that the Fish Go! application has a good level of usability and is suitable for further development, with an average SUS score of 84, exceeding the average threshold for usability testing. The results of this research show that the Fish Go! application has the potential to support partnerships between fishermen and the business sector in utilizing marine resources in Indonesia. By leveraging technology and digital platforms, it is expected to improve efficiency and provide broader market opportunities for fishermen, while also meeting the demand of the business sector for fresh and high-quality seafood.

Keywords: Fishermen, Market, Usability, Partnerships, Challenges

1. Pendahuluan

Indonesia sebagai salah satu negara maritim terbesar di dunia namun masih belum memanfaatkan hasil laut dengan optimal. Berbagai peraturan dan kebijakan telah dikeluarkan untuk mendukung peningkatan hasil produksi perikanan di Indonesia. Wilayah perairan di Indonesia memiliki luasan yang lebih besar dibanding luas daratan di mana antara perairan yang satu dengan yang lainnya memiliki karakteristik yang berbeda-beda. Misi Indonesia untuk menjadi poros maritim dunia tidak dapat dilepaskan dari peran nelayan dalam produktivitasnya untuk menghasilkan hasil laut seperti ikan, kerang, lobster dan sebagainya.

Nelayan kecil hingga menengah sangat bergantung terhadap industri perikanan yang merupakan sumber pencaharian utama mereka, namun tak jarang nelayan menghadapi tantangan dalam memasarkan dan menjual hasil tangkapannya, dan ketika dijual kepada tengkulak harga jual yang ditawarkan tidak wajar [1]. Di sisi lain nelayan lokal biasanya terbatas pada pasar lokal yang tidak mampu menampung semua hasil tangkapan mereka [2].

Di balik permasalahan tersebut, terdapat peluang untuk memenuhi permintaan hasil laut dari sektor-sektor bisnis pariwisata, seperti hotel, resort, rumah makan, maupun jasa boga yang menggunakan hasil laut sebagai menu makanannya. Terdapat lebih dari 29.742 lini bisnis di

Indonesia yang memanfaatkan hasil laut untuk usahanya [3]. Hal ini menjadi peluang yang besar untuk pasar Business to business di Indonesia.

Berdasarkan hal-hal di atas, maka penulis berinovasi dengan menghadirkan aplikasi Fish Go! yang bertujuan untuk memudahkan nelayan mencari pangsa pasar khususnya dengan sektor bisnis, sehingga dapat menjalin kemitraan jangka panjang. Selain itu, nelayan dapat memperluas jangkauan pasarnya dan mengurangi pemborosan pendapatan yang tidak terjual. Bagi pihak pembeli hal ini dapat memberikan kemudahan untuk memperoleh hasil tangkapan yang segar dan berkualitas. Dengan adanya Fish Go! penulis berharap dapat membantu mengintegrasikan nelayan (supply) dan sektor bisnis (demand) dalam satu platform digital. Sejalan dengan SDGs poin 14 yakni menjaga dan menggunakan secara berkelanjutan sumber daya laut.

1.1. User Journey Map

Journey Map pada dasarnya adalah sebuah visualisasi dari proses atau langkah yang seorang pengguna akan lalui dalam rangka meraih tujuan. Langkah-langkah yang dimaksud adalah berbagai skenario yang mungkin terkait bagaimana pengguna itu berinteraksi dengan produk yang dirancang. Melalui User Journey Map inilah pengguna dapat membayangkan peristiwa dan interaksi utama selama proses penggunaan aplikasi [4].

1.2. Wireframe

Wireframe pada dasarnya adalah sebuah kerangka awal atau sketsa sebelum antarmuka akhir dari sebuah aplikasi (high fidelity design) didesain. Wireframe menjadi salah satu tahapan penting dalam sebuah desain produk aplikasi yang tentunya harus dipahami dengan baik. Tahapan ini penting sebelum stakeholder atau pihak terkait (dalam hal ini bisa saja calon penggunanya) menyetujui tata letak informasi pada aplikasi dalam menyelesaikan tugas atau masalah di dalamnya sebelum desain akhir antarmuka aplikasi atau UI dibuat [5].

1.3. Wireflows

Wireflows pada dasarnya adalah sebuah kombinasi antara Wireframe dengan flowchart. Wireflows adalah sebuah sketsa rancangan dari interaksi pengguna yang mungkin pada aplikasi untuk mengetahui bagaimana alur interaksi dari sketsa Wireframe tadi berjalan [6].

1.4. High-Fidelity Design

High fidelity pada dasarnya adalah sebuah rancangan akhir atau rancangan pasti dari tampilan antarmuka pengguna (user interface) aplikasi yang nantinya akan digunakan dan diimplementasikan menjadi aplikasi oleh tim pengembang untuk didistribusikan ke pengguna. Pada high fidelity design ini sudah ada beberapa komponen desain yang terisi, seperti ilustrasi, warna, dan icon. Perancangan ini bertujuan untuk menunjukkan visualisasi antarmuka yang semirip mungkin dengan produk akhir sehingga ini akan berefek dalam memudahkan penentuan komponen spesifik dari desain untuk diuji oleh pengguna. Pada tahap ini, rancangan sketsa Wireframe yang telah dirancang sebelumnya akan dibuat lebih mendetail dan nyata [7].

1.5. System Usability Scale (SUS)

System Usability Scale pada dasarnya adalah kuesioner yang dapat digunakan untuk mengukur usability dari sebuah aplikasi dari sudut pandang subjektif pengguna. SUS berisi 10 pertanyaan dengan 5 buah opsi yang bisa dipilih oleh pengguna menurut tingkat kepuasan mereka [8]. Setiap pertanyaan memiliki poin likert dari 0 – 4 sebagai tanggapan. Suatu produk dianggap memiliki usability yang baik jika skor akhir dari SUS sama dengan atau berada di atas 68. Berikut adalah rumus perhitungan dari skor SUS:

- a. Setiap pertanyaan dengan nomor ganjil, skor setiap pertanyaan yang didapat dari skor responden dikurangkan 1.

- b. Setiap pertanyaan dengan nomor genap, skor akhir yang didapat dari nilai 5 akan dikurangkan dengan skor pertanyaan yang dipilih responden.
- c. Skor akhir dari SUS didapat dari hasil penjumlahan kontribusi skor poin (skor pertanyaan ganjil ditambah skor pertanyaan genap) dikalikan 2.5 sehingga skor yang aslinya memiliki rentang 0-40, akan menjadi 0-100.

Setelah mendapat skor akhir sesuai perhitungan tadi, akan didapatkan hasil skor rata-rata SUS dari semua responden yang kemudian akan bisa didapatkan hasil kepuasan pengguna secara menyeluruh terkait aplikasi.

1.6. Figma

Figma adalah salah satu software atau perangkat lunak yang membantu desainer dalam merancang desain suatu aplikasi, baik dalam bentuk platform mobile ataupun web. Figma dikenal dengan penggunaannya yang gratis, bahkan bisa dengan mudah mengajukan Figma Education Plan untuk mendapatkan beberapa keuntungan tambahan, serta penggunaannya yang tidak memerlukan untuk memasang aplikasi di perangkat, tetapi cukup dengan hanya membuka di browser. Figma juga memiliki basis komunitas besar yang sangat membantu para desainer dalam mencari aset desain, sebagai alat bantu desain, dan lainnya.

1.7. Maze

Maze adalah salah satu alat bantu untuk melakukan pengujian usability. Maze mampu dalam melakukan semua jenis metode pengujian yang ada pada usability dan memberikan sebuah hasil yang berupa data kuantitatif dari tingkat kepuasan pengguna yang bisa digunakan dan dihitung untuk membantu proses pengembangan dari sebuah rancangan antarmuka aplikasi.

2. Metode Penelitian

Pada penelitian ini, penulis menggunakan metode System Usability Scale atau SUS yang merupakan salah satu metode usability yang sering digunakan dalam proses pengujian usability dari sebuah aplikasi. Metode ini digunakan untuk mengukur tingkat kepuasan pengguna terkait antarmuka dari aplikasi yang telah dirancang yang nantinya tingkat kepuasan ini akan menghasilkan data kuantitatif yang terukur melalui perhitungan skor SUS. Adapun beberapa tahapan di dalam penelitian ini adalah sebagai berikut.

2.1. Studi Literatur

Pada penelitian ini, penulis memulai penelitian dengan melakukan studi literatur terhadap materi yang berkaitan dengan analisis antarmuka aplikasi, pengujian usability testing dengan metode System Usability Scale (SUS), serta materi terkait usability secara umum. Studi literatur didapat dari banyak sumber lain seperti artikel, jurnal, dan lainnya yang terpublikasi agar analisis yang dilakukan nanti dapat berjalan dengan terukur dan sesuai kaidah.

2.2. Identifikasi Masalah

Tahap berikutnya dari penelitian ini, penulis melakukan identifikasi masalah untuk mencari celah atau kendala-kendala yang dirasakan oleh calon pengguna dari aplikasi fishgo yang dalam hal ini adalah nelayan atau pihak yang ingin melakukan ekspor hasil laut mereka. Tujuan dari identifikasi masalah ini agar penulis dapat menyusun sketsa kebutuhan dasar dalam mendesain tampilan aplikasi berdasarkan data permasalahan yang didapat. Identifikasi masalah yang penulis lakukan di penelitian ini menggunakan User Journey Map. Metode ini memiliki kemampuan menggali masalah lebih baik pada suatu proses yang melibatkan pengguna [9].

2.3. Perancangan Wireframe dan Wireflow

Tahap selanjutnya, penulis akan melakukan perancangan solusi berupa sketsa desain untuk tampilan aplikasi dari "FishGo!" ini berdasarkan identifikasi permasalahan pada pengguna.

Bentuk dari sketsa desain ini berupa Wireframe dan Wireflow. Perancangan dari kedua bentuk sketsa desain tersebut juga dirancang dengan tools bernama Figma.

2.4. Pembuatan Tampilan High-Fidelity

Tahap berikutnya, penulis akan membuat desain berdasarkan data yang telah terkumpul dan terkonversi menjadi Wireframe. Melalui Wireframe, desain kemudian akan dikembangkan menuju bentuk high fidelity (rancangan desain akhir) yang kemudian akan diuji kelayakannya oleh pengguna melalui kuesioner usability SUS (System Usability Scale). Pembuatan tampilan high fidelity design akan menggunakan tools yang sama, yaitu Figma.

2.5. Pengujian Usability dengan System Usability Scale

Pada tahap terakhir penelitian ini adalah pengujian usability yang merupakan tahap penentu apakah sebuah rancangan desain aplikasi yang telah dibuat dan dibangun memiliki tingkat kemudahan pemakaian yang baik dan layak untuk dikembangkan lebih jauh. Penulis akan melakukan sebuah pengujian usability dengan metode SUS atau System Usability Scale berupa 10 pertanyaan SUS melalui penyebaran kuesioner kepada 20 responden yang akan menghasilkan data kuantitatif. Tetapi sebelum itu, responden akan diberikan task scenario dalam rangka untuk menguji usability dari prototype aplikasi dengan tools Maze, setelah itu responden akan diarahkan untuk mengisi kuesioner 10 pertanyaan SUS tadi berdasarkan kesan responden dalam menyelesaikan task scenario. Data kuantitatif dari skor SUS dihasilkan dari 5 buah opsi respons dari kategori sangat tidak setuju (STS), tidak setuju (TS), netral (N), setuju (S), sangat setuju (SS) dengan skala likert 0-4. Pertanyaan-pertanyaan tersebut adalah:

3. Hasil dan Diskusi

Berdasarkan alur penelitian dan juga studi literatur yang telah dijelaskan, didapatkan hasil penelitian sebagai berikut.

3.1. Rancangan User Journey Map

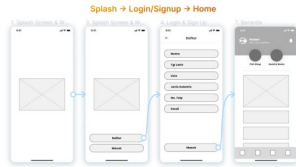


Gambar 1. User Journey Map

User Journey Map di atas memberikan gambaran perjalanan dari pengguna sebelum dan sesudah menggunakan produk hingga selesai menggunakan. Di dalam User Journey Map tidak hanya ada perjalanan pengguna tapi juga berisi grafik tingkat kenyamanan pengguna, deskripsi lengkap terkait pengalaman pengguna dan juga grafik tingkat opportunity yang bisa dijadikan sebuah pendapat untuk meningkatkan kualitas produk.

3.2. Wireframe & Wireflow

a. Alur dari Splash Screen menuju Halaman Utama



Gambar 2. Alur dari Splash Screen Menuju Halaman Utama

Pada Gambar 2, terlihat alur dari rancangan dasar desain tampilan antarmuka aplikasi serta interaksi yang akan terjadi oleh pengguna ketika baru memulai aplikasi. Aplikasi akan menampilkan halaman splash yang berisi logo, kemudian proses masuk ataupun daftar, hingga masuk ke halaman beranda atau halaman utama.

b. Alur dari Beranda menuju Fish Shop



Gambar 3. Alur dari Beranda Menuju Fish Shop

Berikutnya adalah alur yang menunjukkan rancangan dasar desain tampilan antarmuka aplikasi dari halaman utama atau beranda menuju halaman Fish Shop yang merupakan salah satu fitur utama aplikasi seperti yang terlihat pada Gambar 3. Pengguna hanya perlu menekan tombol Fish Shop pada halaman beranda untuk bisa menuju ke halaman Fish Shop.

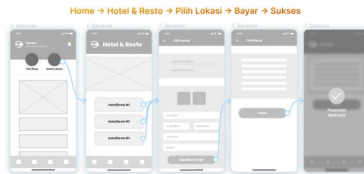
c. Alur dari Beranda menuju Hotel & Resto



Gambar 4. Alur dari Beranda Menuju Hotel & Resto

Berikutnya alur yang menunjukkan rancangan dasar desain tampilan antarmuka aplikasi dari halaman beranda menuju halaman hotel & resto yang juga merupakan salah satu fitur utama lain dari aplikasi seperti yang terlihat pada Gambar 4. Seperti halnya pada Gambar 3, pengguna hanya perlu menekan tombol Hotel & Resto pada halaman beranda untuk bisa masuk ke halaman Hotel & Resto.

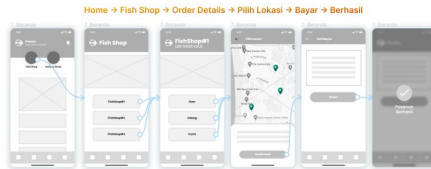
d. Alur dari Nelayan Mengirimkan Supply



Gambar 5. Alur dari Nelayan Ketika akan Mengirimkan Supply Hasil Laut

Selanjutnya, pada Gambar 5 ditunjukkan rancangan alur dari Wireframe antarmuka aplikasi dari beranda hingga pembayaran pengiriman supply. Terlihat interaksi ketika pengguna memilih Hotel atau Resto yang memerlukan pasokan, kemudian pengguna memasukkan data pengiriman berupa kategori barang, kemasan, dimensi kemasan, hingga beratnya, sampai pada perkiraan tarif bayar dan konfirmasi pembayaran berhasil.

e. Alur Restoran Melakukan Pemesanan

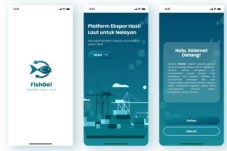


Gambar 6. Alur Restoran dalam Melakukan Pemesanan

Gambar 6 menunjukkan alur dari pengguna (dalam hal ini pihak konsumen) ketika melakukan pemesanan pasokan hasil laut ke kelompok pemasok. Pengguna menekan tombol Fish Shop, memilih kelompok pemasok, memilih jenis hasil laut dan jumlah hasil laut (dalam kwintal) yang diperlukan, hingga memilih tempat atau lokasi pengiriman serta melakukan pembayaran.

3.3. Implementasi Desain

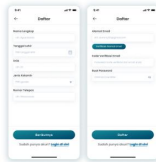
a. Tampilan Splash Screen dan Onboarding Screen



Gambar 7. Tampilan Splash Screen dan Onboarding Screen

Terlihat pada Gambar 7 desain tampilan Splash Screen dan Onboarding Screen yang berisi logo, kata sambutan serta tombol untuk melakukan Daftar/Sign up dan Masuk/Login.

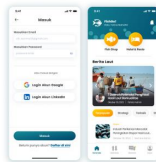
b. Tampilan Antarmuka Signup



Gambar 8. Tampilan Antarmuka Signup

Terlihat pada Gambar 8 tampilan dari antarmuka ketika pengguna melakukan Sign up atau daftar akun pada aplikasi.

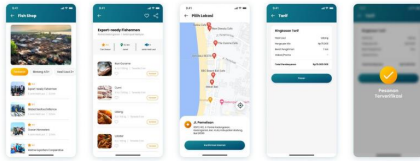
c. Tampilan Antarmuka Login dan Beranda



Gambar 9. Tampilan Antarmuka Login dan Beranda

Terlihat pada Gambar 9 tampilan dari antarmuka ketika pengguna melakukan Login atau masuk akun pada aplikasi. Pengguna bisa masuk dengan akun email yang didaftarkan sebelumnya pada aplikasi, atau bisa juga masuk dengan memasukkan akun google atau linkedin. Ketika pengguna sudah berhasil masuk, maka akan tampil menu beranda atau halaman utama aplikasi.

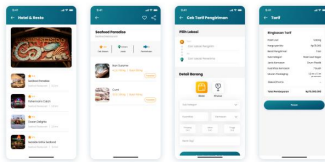
d. Tampilan Fitur Fish Shop, Detail Kelompok Nelayan, dan Konfirmasi Pemesanan



Gambar 10. Tampilan Fitur Fish Shop, Detail Kelompok Nelayan, dan Konfirmasi Pemesanan

Gambar 10 memperlihatkan fitur inti pertama dari aplikasi ini, yaitu fitur Fish Shop, yaitu fitur yang diperlukan oleh para stakeholder seafood seperti Resto dan Hotel untuk bisa memesan pasokan hasil laut kepada kelompok pemasok. Pengguna dapat memilih kelompok pemasok, kemudian memilih jenis hasil laut dengan satuan berat kwintal, memilih lokasi pengiriman, serta konfirmasi pembayaran.

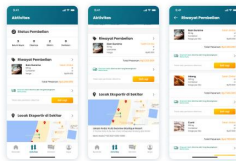
e. Tampilan Fitur Hotel & Resto, Detail, dan Konfirmasi Pengiriman



Gambar 11. Tampilan Fitur Hotel & Resto, Detail, dan Konfirmasi Pengiriman

Gambar 11 memperlihatkan fitur inti yang kedua dari aplikasi ini, yaitu fitur Hotel & Resto yang merupakan fitur untuk para kelompok pemasok hasil laut. Pengguna dapat memilih stakeholder seafood seperti hotel & resto yang membutuhkan pasokan hasil laut, memberikan penawaran jenis hasil laut berdasarkan kebutuhan stakeholder tersebut, memasukkan data pengemasan barang/shipment, hingga melakukan konfirmasi pengiriman.

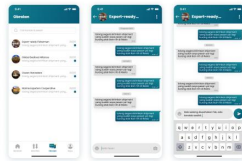
f. Tampilan Menu Aktivitas



Gambar 12. Tampilan Menu Aktivitas

Gambar 12 merupakan tampilan dari menu kedua dari aplikasi, yaitu menu aktivitas yang berisi status pemesanan/pengiriman, riwayat pembelian/pengiriman pengguna, serta melacak lokasi terdekat dari stakeholder seafood/kelompok nelayan yang bersangkutan.

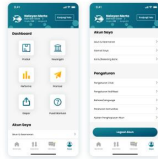
g. Tampilan Menu Obrolan



Gambar 13. Tampilan Menu Obrolan

Gambar 13 merupakan tampilan menu obrolan yang merupakan menu ketiga dari aplikasi yang berisi riwayat obrolan masuk terhadap pihak stakeholder/kelompok pemasok hasil laut.

h. Tampilan Menu Akun



Gambar 14. Tampilan Menu Akun

Terakhir adalah menu Akun sebagai menu keempat pada aplikasi yang merupakan tempat dashboard serta pengaturan dasar aplikasi. Di menu akun ini juga pengguna bisa melakukan logout dari aplikasi.

3.4. Hasil Pengujian Metode System Usability Scale

Dari hasil pengujian menggunakan metode System Usability Scale, diperoleh hasil sebagai berikut:

Tabel 2. Hasil Survei Sebelum Diolah

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

Skor SUS										
Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
4	5	1	4	2	4	2	3	3	4	2
5	4	2	5	2	5	2	5	1	3	2
6	5	1	4	1	4	3	5	2	4	2
7	5	2	5	1	5	2	4	3	4	1
8	3	2	4	1	4	2	4	2	5	1
9	5	2	5	2	5	2	4	1	4	1
10	4	1	5	2	5	2	5	2	5	2
11	5	1	5	2	4	3	5	2	4	2
12	4	2	4	1	4	2	4	1	5	2
13	4	2	5	2	5	2	5	2	5	2
14	5	1	5	3	5	1	4	1	5	1
15	5	2	4	1	4	2	5	2	5	1
16	5	2	5	2	5	2	4	2	5	1
17	4	1	5	2	4	2	5	1	4	2
18	3	3	4	2	5	3	5	1	4	2
19	4	1	5	1	5	2	4	2	4	1
20	5	2	5	2	4	2	5	2	5	2

Setelah itu, skor diproses dengan aturan SUS sehingga mendapatkan hasil seperti berikut:

Tabel 3. Hasil Survei Setelah Diolah

Skor SUS												
Resp.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Sum	x2.5
1	4	2	4	2	4	2	4	3	4	4	33	82,5
2	4	3	4	3	4	4	4	2	4	3	35	87,5
3	3	3	3	4	3	2	3	3	4	4	32	80
4	4	4	3	3	3	3	2	2	3	3	30	75
5	3	3	4	3	4	3	4	4	2	3	33	82,5
6	4	4	3	4	3	2	4	3	3	3	33	82,5
7	4	3	4	4	4	3	3	2	3	4	34	85
8	2	3	3	4	3	3	3	3	4	4	32	80
9	4	3	4	3	4	3	3	4	3	4	35	87,5
10	3	4	4	3	4	3	4	3	4	3	35	87,5
11	4	4	4	3	3	2	4	3	3	3	33	82,5
12	3	3	3	4	3	3	3	4	4	3	33	82,5

Skor SUS												
Resp.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Sum	x2.5
13	3	3	4	3	4	3	4	3	4	3	34	85
14	4	4	4	2	4	4	3	4	4	4	37	92,5
15	4	3	3	4	3	3	4	3	4	4	35	87,5
16	4	3	4	3	4	3	3	3	4	4	35	87,5
17	3	4	4	3	3	3	4	4	3	3	34	85
18	2	2	3	3	4	2	4	4	3	3	30	75
19	3	4	4	4	4	3	3	3	3	4	35	87,5
20	4	3	4	3	3	3	4	3	4	3	34	85
Total Keseluruhan											1680	
Rata-rata												84

Setelah mendapat sekian banyak data dari 20 responden, diperoleh hasil rata-rata 84 yang berada di atas batas nilai kelayakan rata-rata pada pengujian SUS, yakni di atas 68.

4. Kesimpulan

Berdasarkan pengujian yang telah dilaksanakan, dapat disimpulkan bahwa rancangan desain tampilan aplikasi "FishGo" yang diuji dengan metode System Usability Scale (SUS) mendapatkan rata-rata nilai 84 dari total 20 responden yang telah mencoba prototype dan mengisi kuesioner SUS. Ini menyimpulkan bahwa "FishGo" memiliki tampilan yang efektif dan mudah digunakan serta layak untuk diimplementasikan menjadi sebuah aplikasi di pengembangan selanjutnya.

Daftar Pustaka

- [1] Lubis, Ernani, dkk. (2012). Besaran Kerugian Nelayan dalam Pemasaran Hasil Tangkapan: Kasus Pelabuhan Perikanan Nusantara Palabuhanratu. *Maspari Journal*; 4(2), 159-167.
- [2] Retnowati, Herlina, dkk. (2014). Strategi Peningkatan Kinerja Nelayan dalam Rantai Pasok Ikan Layur melalui Pengembangan Modal Insani di Pelabuhanratu. *Manajemen IKM*. ISSN 2085-8418.
- [3] B. Pusat Statistik, "Statistik Hotel Dan Akomodasi Lainnya Di Indonesia Hotel And Other Accommodation Statistics In Indonesia."
- [4] R. F. A. Aziza, "Analisis Kebutuhan Pengguna Aplikasi Menggunakan User Persona Dan User Journey" *INFOS*, vol. 3, no. 2, p. 7, 2020.
- [5] M. S. Hartawan, "Penerapan User Centered Design (Ucd) Pada Wireframe Desain User Interface Dan User Experience Aplikasi Sinopsis Film" *JEIS*, vol. 2, no. 1, p. 44, 2022.
- [6] F. E. Permana, et al., "Perancangan User Experience Sistem Informasi Manajemen Magang pada Jurusan Sistem Informasi menggunakan Pendekatan Human-Centered Design (HCD)" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 9, p. 2864, 2022.
- [7] A. R. Perdanakusuma, et al., "User Interface Design Development on PT SAMUDERA INDONESIA Website Using Human Centered Design Method Perancangan User Interface E-Learning Training Program Kemaritiman dan Logistik PT. Samudera Indonesia dengan Metode User Centered Design (UCD)" *JICTE*, vol. 4, no. 1, p. 4, 2021.
- [8] A. Sidik, "Penggunaan System Usability Scale (SUS) Sebagai Evaluasi Website Berita Mobile" *Technology*, vol. 9, no. 2, p. 5, 2018, doi: <http://dx.doi.org/10.31602/tji.v9i2.1371>
- [9] A. Nurfitri, et al., "Pemetaan User Journey untuk Sistem Informasi Praktik Kerja Lapangan Fakultas Ilmu Komputer Universitas Brawijaya" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 8, p. 7543, 2019.

Halaman ini sengaja dibiarkan kosong

Sistem Informasi Prediksi Penjualan E-Commerce Menggunakan Analisis Data Historis dan Algoritma MLR

I Gusti Putu Wisnu Wardhana^{a1}, I Wayan Santiyasa^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹igpwisnuw@gmail.com
²santiyasa@unud.ac.id

Abstract

Accurate sales to improve business planning and decision making. This study aims to design an information system that utilizes historical data and multiple linear regression algorithms to predict e-commerce sales. This study addresses the current challenges in forecasting uncertain sales by analyzing historical sales data and identifying relevant independent variables, such as marketing efforts, economic factors, and customer behavior. Through the implementation of a multiple linear regression algorithm, the system calculates the relationship between these variables and sales, enabling accurate predictions. The proposed information system provides valuable insights for businesses to optimize inventory management, marketing strategy and resource allocation. The experimental results show the effectiveness of the system in forecasting e-commerce sales, resulting in increased operational efficiency and revenue. This research contributes to the field of e-commerce analytics and assists businesses in making data-driven decisions for sustainable growth.

Keywords: e-commerce, sales prediction, historical data, multiple linear regression, forecasting

1. Pendahuluan

Perkembangan pesat industri e-commerce dalam beberapa tahun terakhir telah menciptakan tantangan baru dalam menganalisa dan memprediksi penjualan. Dalam rangka mengoptimalkan efisiensi operasional dan meningkatkan pengambilan keputusan bisnis, prediksi penjualan yang akurat menjadi krusial. Oleh karena itu, penelitian ini bertujuan untuk merancang sistem informasi yang memanfaatkan data historis dan algoritma regresi linier berganda untuk melakukan prediksi penjualan e-commerce dengan tingkat akurasi yang tinggi.

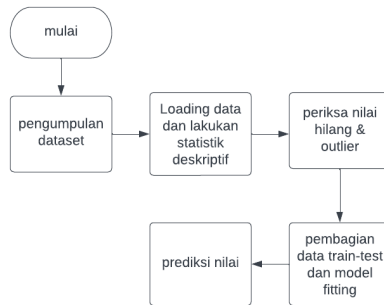
Untuk mencapai tujuan tersebut, penelitian ini menggunakan dataset "Linear Regression E-commerce" [1]. Dataset ini berisi data pelanggan yang membeli pakaian online di sebuah toko. Analisis data historis dari dataset ini bertujuan untuk mengidentifikasi pola dan tren penjualan masa lalu guna memprediksi penjualan masa depan. Dalam analisis ini, diterapkan algoritma regresi linier berganda yang mempertimbangkan faktor-faktor seperti karakteristik pelanggan, detail pembelian, dan informasi waktu. Dataset baru ini dapat memberikan wawasan relevan terkait faktor-faktor yang mempengaruhi penjualan

Beberapa penelitian sebelumnya memberikan kontribusi penting dalam pengembangan sistem informasi prediksi penjualan e-commerce. Tolstoy (2016) [2] menunjukkan bahwa algoritma regresi linier berganda dapat digunakan sebagai metode untuk memprediksi penjualan online. Penelitian lain yang sejenis dilakukan oleh Khan dan Uddin (2021) [3] yang menemukan bahwa regresi linier berganda dapat menjadi pilihan utama dalam prediksi penjualan di industri e-commerce. Sebagaimana dijelaskan oleh Basyir (2020) [4], dalam rancang bangun sistem informasi, penting untuk memperhatikan preferensi pengguna, sehingga sistem yang dibangun dapat menjadi user-friendly dan dapat diakses dengan mudah.

Ma et al. (2017) [5] mencatat bahwa data historis penjualan adalah input yang penting dalam memprediksi penjualan. Mereka menambahkan bahwa analisis pola dan tren penjualan masa lalu dapat memberikan wawasan berharga dalam memprediksi penjualan masa depan. Hasil dari analisis pola tersebut dapat mengarah pada pengembangan model prediksi yang tepat sehingga dapat membantu menjaga stok produk yang tepat dengan memaksimalkan penjualan dengan waktu yang tepat. Penelitian oleh Poon dan Zhong (2016) [6] mempelajari pentingnya mempertimbangkan interaksi antara faktor-faktor yang mempengaruhi penjualan e-commerce dalam model prediksi, karena hal ini dapat menghasilkan ramalan yang lebih akurat dan membantu perusahaan membuat keputusan yang lebih baik dalam pengelolaan persediaan dan optimasi penjualan.

Dalam proyek ini, akan dilatih model machine learning menggunakan regresi linier berganda dengan memanfaatkan dataset "Online Retail II". Selanjutnya, akan dikembangkannya sebuah antarmuka sederhana menggunakan framework Streamlit yang memungkinkan pengguna untuk memilih produk dari dataset, memilih jangka waktu tertentu, dan mendapatkan prediksi jumlah penjualan di masa mendatang. Dengan demikian, sistem informasi yang dihasilkan diharapkan dapat membantu perusahaan e-commerce dalam mengoptimalkan keputusan bisnis berdasarkan prediksi penjualan yang akurat.

2. Metode Penelitian



Gambar 1. Tahapan Penelitian

Penelitian ini dimulai dengan mengumpulkan dataset yang diperlukan untuk studi. Dataset dimuat menggunakan library pandas, dan statistik deskriptif dihitung untuk memperoleh pemahaman tentang data tersebut. Nilai yang hilang diperiksa dan ditangani dengan tepat. Outlier diidentifikasi menggunakan visualisasi dan dihapus dari dataset. Data kemudian dibagi menjadi set pelatihan dan pengujian menggunakan pembagian pelatihan-pengujian. Model regresi linear disesuaikan dengan data pelatihan untuk mempelajari pola dan hubungan yang ada. Model yang telah dilatih digunakan untuk membuat prediksi pada data pengujian. Akhirnya, nilai yang diprediksi dievaluasi dan dianalisis untuk menilai kinerja dan mengambil kesimpulan dari penelitian tersebut.

2.1 Pengumpulan Data

Data untuk penelitian ini akan dikumpulkan melalui pengunduhan dataset dari platform Kaggle. Dataset ini akan berisi informasi historis tentang penjualan e-commerce, termasuk variabel seperti jumlah transaksi, waktu transaksi, kategori produk, harga produk, dan informasi pelanggan. Dataset ini akan memberikan kerangka kerja yang diperlukan untuk melakukan analisis dan prediksi penjualan e-commerce. Penelitian Sari dan Ibrani (2017) [7] yang membahas prediksi menggunakan regresi linier memvariasikan distribusi rasio data training: data testing menjadi 70:30, 80:20, dan 90:10 untuk menentukan ukuran training data yang optimal. Maka dari itu, dilakukan perbandingan tiga variasi tersebut

2.2 Pemodelan Regresi Linier Berganda

Model regresi linear berganda berguna ketika mempelajari hubungan antara beberapa variabel independen dan satu variabel dependen. Untuk membangun model-model tersebut, para peneliti pertama-tama mengidentifikasi variabel-variabel yang kemungkinan besar mempengaruhi variabel dependen. Hal ini dapat dicapai dengan menguji variabel-variabel tersebut untuk korelasi signifikan dengan variabel dependen menggunakan tingkat signifikansi, yang sering kali ditetapkan pada $p < 0,20$ [8]. Variabel-variabel yang memenuhi kriteria ini kemudian dimasukkan ke dalam model, yang dibangun menggunakan algoritma seperti metode bertahap [8], kuadrat terkecil biasa (OLS), atau penurunan gradien [9].

2.3 Evaluasi dan Validasi

Setelah model regresi linier berganda dibangun, akan dilakukan evaluasi dan validasi model. Langkah-langkah yang dilakukan meliputi:

a. MSE (Mean Squared Root)

MSE merupakan rata-rata kuadrat kesalahan yang dihitung dengan menjumlahkan semua kesalahan atau error prediksi yang dihasilkan oleh suatu model kemudian dikuadratkan dan membaginya dengan jumlah periode prediksi [12]. Berikut merupakan persamaan matematis dari MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - F_i)^2 \quad (1)$$

Keterangan:

- X_i = Data aktual pada periode ke- i
- F_i = Nilai hasil prediksi atau prediksi pada period ke- i
- n = Banyaknya sampel

b. MAE (Mean Absolute Error)

MAE metrik yang digunakan untuk mengukur selisih rata-rata absolut antara nilai aktual dan nilai yang diprediksi dalam suatu model regresi. MAE dihitung dengan mengambil rata-rata dari selisih absolut antara nilai yang diprediksi dan nilai sebenarnya [13]. Berikut merupakan persamaan matematis dari MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Keterangan:

- n = Banyaknya sampel

y_i = nilai aktual pada sampel ke i
 \hat{y}_i = nilai prediksi pada sampel ke i

c. R2 Score

R2 Score digunakan untuk mengukur sejauh mana variabilitas suatu model statistic atau machine learning dapat menjelaskan variasi data yang diamati. Range nilai dari R2 Score adalah antara 0 hingga 1. Berikut merupakan persamaan matematis dari R2 Score:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \tag{3}$$

Keterangan:

y_i = Nilai aktual dari variabel dependen
 \hat{y}_i = Nilai prediksi dari variabel dependen

d. Integrasi Model dengan Sistem

Langkah terakhir dalam metode penelitian ini adalah integrasi model prediksi dengan sistem yang relevan, seperti sistem manajemen penjualan e-commerce. Model yang telah dibangun akan digunakan untuk memberikan prediksi penjualan e-commerce berdasarkan input yang diberikan oleh sistem. Dalam hal ini, output dari model dapat digunakan sebagai panduan dalam pengambilan keputusan bisnis dan perencanaan strategi penjualan.

3. Hasil dan Diskusi

3.1. Perancangan Model

Data yang digunakan pada penelitian ini adalah dataset customer ecommerce yang terdiri dari email, alamat, avatar, waktu sesi, waktu kunjungan, jangka langganan, serta jumlah yang dibayarkan per tahun. Adapun dataset yang digunakan dapat diakses pada link berikut:

<https://www.kaggle.com/datasets/ivadavvaibhav/ecommerce-customer-device-usage>

Metode yang digunakan adalah regresi linear berganda.

	Avg. Session Length	Time on App	Time on Website	Length of Membership
0	34.497268	12.655651	39.577668	4.082621
1	31.926272	11.109461	37.268959	2.664034
2	33.000915	11.330278	37.110597	4.104543
3	34.305557	13.717514	36.721283	3.120179
4	33.330673	12.795189	37.536653	4.446308
...
495	33.237660	13.566160	36.417985	3.746573
496	34.702529	11.695736	37.190268	3.576526
497	32.646777	11.499409	38.332576	4.958264
498	33.322501	12.391423	36.840086	2.336485
499	33.715981	12.418808	35.771016	2.735160

Gambar 2. Dataset yang digunakan dalam perancangan model

Data yang didapatkan dari dataset kemudian akan dibagi menjadi data latih dan data uji. Pada penelitian ini, rasio data latih: data uji didistribusikan menjadi 70:30, 80:20, dan 90:10. Pada gambar 2, rasio ini direpresentasikan dalam bentuk desimal yang menunjukkan jumlah data latih.

	Test Ratio	MSE	MAE	R2 Score
0	0.3	93.434699	7.583105	0.981536
1	0.2	89.729365	7.358401	0.979355
2	0.1	71.811240	6.612771	0.982737

Gambar 3. Performa Algoritma Regresi Linier Berganda

Berdasarkan hasil yang diperoleh dari model regresi, dapat kita lihat bahwa rasio pelatihan dan pengujian yang berbeda digunakan untuk mengevaluasi kinerja. Model ini dilatih dengan menggunakan 70% dari data pada skenario pertama, 80% pada skenario kedua, dan 90% pada skenario ketiga, sedangkan data yang tersisa digunakan untuk pengujian.

Metrik evaluasi yang digunakan untuk mengukur kinerja model adalah Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R2 Score. Metrik-metrik ini memberikan wawasan tentang sejauh mana model cocok dengan data dan dapat memprediksi penjualan produk e-commerce.

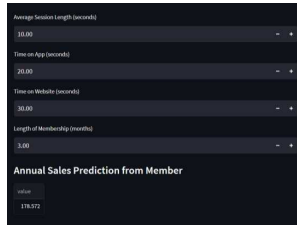
Dari hasil yang diperoleh, dapat diamati bahwa seiring dengan peningkatan ukuran set pelatihan, kinerja model juga meningkat. Hal ini terlihat dari nilai yang semakin mengecil pada MSE dan MAE, yang menunjukkan bahwa model cocok dengan data dengan baik. Dengan rasio tes 0.3, 0.2, dan 0.1, dapat dilihat bahwa Nilai MSE mengecil dari 93.4, 89.7, lalu 71.8. Untuk rasio tes yang sama, nilai MAE juga semakin berkurang, 7.59 untuk rasio 0.3, 7.36 untuk rasio 0.2, lalu 6.61 untuk rasio 0.1. Nilai MAE dan MSE yang lebih rendah menunjukkan tingkat akurasi model regresi lebih tinggi.

Selain itu, R2 Score, yang menggambarkan proporsi variansi dalam variabel target yang dijelaskan oleh model, juga meningkat seiring dengan peningkatan ukuran set pelatihan. Dengan nilai R2 sekitar 0,98, dapat disimpulkan bahwa model ini mampu menjelaskan sekitar 98% variansi dalam data penjualan. Hal ini menunjukkan bahwa model regresi linier berganda dengan menggunakan data historis dan fitur-fitur yang dipilih sangat efektif dalam memprediksi penjualan e-commerce.

Nilai MSE dan MAE rendah menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang kecil. Selain itu, jika nilai R2 mendekati 1, maka model mampu menjelaskan sebagian besar variansi yang ada dalam data penjualan E-commerce. Dengan mempertimbangkan ketiga metrik ini, dapat disimpulkan bahwa model regresi linier berganda dalam proyek ini memiliki performa yang baik dalam memprediksi penjualan E-commerce berdasarkan analisis data historis.

Secara keseluruhan, hasil ini menunjukkan potensi penggunaan data historis dan algoritma regresi linier berganda dalam membangun sistem prediksi penjualan e-commerce. Analisis dan optimasi lebih lanjut dapat dilakukan untuk meningkatkan kinerja model dan mengeksplorasi fitur-fitur tambahan yang dapat berkontribusi pada prediksi penjualan yang akurat.

3.2. Integrasi Backend dengan Frontend



Gambar 4. Antarmuka aplikasi berbasis streamlit

Digunakan library streamlit untuk mengimplementasi antarmuka interaktif untuk mempermudah melakukan prediksi pendapatan e-commerce masing-masing member berdasarkan data yang dimiliki terkait member tersebut. Library streamlit dipilih karena penggunaannya yang sederhana sehingga mempercepat proses perancangan sistem.

4. Kesimpulan

Berdasarkan hasil penelitian dan analisis yang dilakukan, dapat diambil kesimpulan sebagai berikut:

- Model regresi linier berganda yang dikembangkan dalam penelitian ini mampu memberikan prediksi penjualan E-commerce yang akurat dan mendekati nilai sebenarnya.
- Evaluasi model menggunakan metrik MSE (Mean Squared Error), MAE (Mean Absolute Error), dan R2 (R-squared) menunjukkan hasil yang baik, dengan nilai-nilai sebagai berikut:
 - Pada uji ratio 0.3: MSE = 93.434699, MAE = 7.583105, R2 Score = 0.981536.
 - Pada uji ratio 0.2: MSE = 89.729365, MAE = 7.358401, R2 Score = 0.979355.
 - Pada uji ratio 0.1: MSE = 71.811240, MAE = 6.612771, R2 Score = 0.982737.
- Hasil pengujian dengan menggunakan tiga rasio uji yang berbeda menunjukkan konsistensi kinerja model dalam menghasilkan prediksi yang akurat.
- MSE yang rendah pada setiap rasio uji menunjukkan tingkat kesalahan prediksi yang minim.
- MAE yang rendah mengindikasikan bahwa model memiliki tingkat kesalahan prediksi yang kecil secara keseluruhan.
- R2 Score yang mendekati 1 menunjukkan bahwa model dapat menjelaskan sebagian besar variasi dalam data penjualan E-commerce.
- Hasil ini menunjukkan bahwa model regresi linier berganda yang dikembangkan memiliki kemampuan yang baik dalam memprediksi dan menjelaskan penjualan E-commerce.
- Dengan demikian, sistem informasi prediksi penjualan yang dikembangkan dalam penelitian ini dapat digunakan sebagai alat yang efektif dalam membantu perusahaan E-commerce dalam mengoptimalkan penjualan.

Daftar Pustaka

- [1] [1] S. Kolawale, "Linear Regression E-Commerce Dataset," [Online]. Available: <https://www.kaggle.com/datasets/kolawale/focusing-on-mobile-app-or-website>.
- [2] D. Tolstoy, A. Jonsson, and D. D. Sharma, "The influence of a retail firm's geographic scope of operations on its international online sales," International Journal of Electronic

- Commerce, vol. 20, no. 3, pp. 293–318, 2016. doi:10.1080/10864415.2016.1121760
- [3] A. H. Khan and G. S. Uddin, "Predicting e-commerce sales using machine learning techniques," *Journal of Retailing and Consumer Services*, vol. 59, 2021. doi: 10.1016/j.jretconser.2020.102332.
- [4] A. Basyir, M. Maskur, and I. Nuryasin, "Rancang Bangun Sistem Informasi Pertandingan Pencak Silat Berbasis Website Menggunakan Metode User Centered Design (Ucd)," *JR*, vol. 12, no. 2, pp. 1663–1670, 2020. doi:10.22219/repositor.v2i12.571
- [5] J. Ma, Y. Liang, G. Liu, and P. Zhou, "Forecasting Fashion Retail Sales with a Hybrid Method Combining Singular Spectrum Analysis and Back-Propagation Neural Network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 8, 2017. doi:10.1142/S0218001417560101.
- [6] Poon and Zhong, "A Forecasting Model of Online Retail Sales incorporating Traffic and Advertising Efforts," *Journal of Business Research*, vol. 69, no. 11, pp. 4958–4964, 2016. doi: 10.1016/j.jbusres.2016.05.032.
- [7] Sari and Ibrani, "Prediksi Harga Saham Menggunakan Regresi Linear Berganda Berbasis Machine Learning," *Jurnal Rekayasa Sistem dan Teknologi Informasi (RESTI)*, vol. 5, no. 2, pp. 168–174, 2021. doi:10.29207/resti.v5i2.2067
- [8] D. Castilho, A. Silva, F. Gimenes, R. Nunes, A. Pires, and C. Bernardes, "Factors Related To the Patient Safety Climate In An Emergency Hospital," *Rev. Latino-Am. Enfermagem*, vol. 28, 2020. doi:10.1590/1518-8345.3353.3273
- [9] N. Hirschall, S. Norby, M. Weber, S. Maedel, S. Amir-Asgari, and O. Findl, "Using Continuous Intraoperative Optical Coherence Tomography Measurements Of the Aphakic Eye For Intraocular Lens Power Calculation," *Br J Ophthalmol*, vol. 99, no. 1, pp. 7–10, 2014. doi:10.1136/bjophthalmol-2013-304731
- [10] F. Gouzi et al., "Blunted Muscle Angiogenic Training-response In Copd Patientsversusedentary Controls," *Eur Respir J*, vol. 4, no. 41, pp. 806–814, 2012. doi:10.1183/09031936.00053512
- [11] G. Khalaf and M. Iguemane, "Multicollinearity and A Ridge Parameter Estimation Approach," *J. Mod. App. Stat. Meth.*, vol. 2, no. 15, pp. 400–410, 2016. doi:10.22237/jmasm/1478002980
- [12] H. Setyawan, S. H. Fitriasih, and R. T. Vuldari, "Prediksi Tingkat Produksi Buah Kelapa Sawit dengan Metode Single Moving Average," *J. TIKomSiN*, vol. 9, no. 2, pp. 1–10, 2021. ISSN: 2338-4018 <https://doi.org/10.30646/tikomsin.v9i2.53>
- [13] C. Willmott and K. Matsuura, "Advantages of the Mean Absolute Error (Mae) Over The Root Mean Square Error (Rmse) In Assessing Average Model Performance," *Clim. Res.*, no. 30, pp. 79–82, 2005. doi:10.3354/cr030079

Halaman ini sengaja dibiarkan kosong

Analisa Sistem Rekomendasi Konten Youtube Berdasarkan Durasi Menonton Menggunakan Content-Based Filtering

I Gede Ngurah Wahyu Ananta^{a1}, I Wayan Supriana^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
wahyuananta5@gmail.com
wayan.supriana@unud.ac.id

Abstract

In today's era, the internet is a facility in social life that causes phobias, complete and necessary information is difficult to obtain again. However, how YouTube provides consistently using an algorithm designed for YouTube content recommendations which is an online video media can witness important moments instantly to individuals who are not on television media so that all users can get useful information and entertainment from website media. For some reason Youtube is used as social media with the highest user level from Instagram. Therefore, we make an experiment to categorize the right content to be a crucial factor in producing accurate and meaningful recommendations. In a system analysis, it recommends content on Youtube based on individual categories using the basic concept of the content based filtering algorithm and how it is implemented in the context of YouTube. The model training is carried out using the cosine similarity method which aims to compare the similarity between the contents of these representations. Evaluation of the model can provide insight into how effective the algorithm is in producing relevant recommendations. The steps in the recommendation system analysis are literature study, data collection, model training, and model evaluation by increasing understanding of content-based filtering algorithms.

Keywords: YouTube, content recommendations, content-based filtering, cosine similarity

1. Pendahuluan

Studi pendahuluan dapat memfokus tentang bagaimana perilaku menonton konten di YouTube secara konsisten yang mempengaruhi algoritma untuk rekomendasi konten YouTube yang merupakan media video online dapat menyaksikan momen untuk mencari acara kesukaan secara instan kepada individu yang tidak ada di media televisi yang menjadi semua pengguna bisa didapatkan suatu informasi yang bermanfaat serta hiburan dari media situs web yang berupa video secara bebas tanpa batasan secara langsung. Menurutny dari survei pengguna internet ini bahwa aplikasi yang paling aktif untuk aktivitas berbagai mendominasi adalah Youtube diperoleh sebesar 88% dari jumlah penduduk Indonesia yang aktif menggunakan media sosial. Namun dengan alasan mengapa Youtube yang dijadikan sebagai aplikasi media social yang paling banyak dipakai adalah karena Youtube yang bisa memberikan informasi dan hiburan yang kepada setiap individu dalam bentuk video sehingga Youtube yang menjadi kategori TV dan video kedua yang paling banyak ditelusuri di situs web pada tahun 2020, dengan sebanyak 651 juta pencarian per bulan. Ini mengalahkan Instagram dan kawan-kawan. Dengan demikian, data yang menunjukkan bahwa banyaknya suatu situs yang ditelusuri sehingga data tersebut bisa mendukung konteks media social [1]. Namun ada beberapa alasan yang kebanyakan kalangan senang menikmati sarana Youtube adalah mengenai durasi iklan yang singkat dan sedikit dari televisi, bisa menonton secara on the go, banyaknya pilihan menonton, bisa mengetahui trending, bebas menonton tidak perlu akses akun ataupun berlangganan yang tanpa perlu menunggu atau takut ketinggalan acara, dan bisa berkarya dengan segala kondisi ekonomi. Untuk membangun sebuah topik yang diangkat maka adapun perlu dibahas didalam suatu sistem alur yang bekerja

adalah metode Content-based filtering dalam merekomendasikan konten kemiripan dengan apa yang sering diklik maupun disukai oleh pengguna dengan topik yang sama. Kemudian, data riwayat tersebut otomatis merekam aktivitas user di Youtube yang menyebabkan suatu peristiwa disaat dia sedang menonton dan konten apa yang mereka dinikmati salah satunya adalah melalui mesin proses YouTube Recommendation Algorithm yang bertujuan untuk mengukur parameter di dalam waktu menjalankan platform tersebut artinya jika seseorang seberapa sering menyaksikan atau menikmati yang berbagai topik yang sedang tren lalu dicari atau sekedar langsung diputar saja kemudian adapun faktor dapat dipertimbangkan oleh algoritma tersebut adalah pola kebiasaan yang dilakukan oleh pengguna serupa. Tetapi dengan catatan bahwa platform Youtube tidak bisa secara terbuka yang mengungkapkan seluruh detail mengenai pola algoritma mereka, dan algoritma ini terus berkembang seiring waktunya. YouTube juga mengambil langkah-langkah untuk memastikan bahwa algoritma itu mematuhi kebijakan platform, termasuk kebijakan tentang konten yang aman dan relevan. Berikut adalah beberapa rekomendasi konten YouTube yang berasal dari algoritma content-based filtering adalah seseorang yang berpengalaman sebagai contoh jika seseorang suka menonton video tentang teknologi komputer, maka yang diajarkan oleh algoritma content-based filtering yang akan merekomendasikan video tentang teknologi lainnya, seperti video tentang gadget terbaru atau tutorial penggunaan software yang tertentu melalui tata letak pada aplikasi tersebut baik di menu beranda maupun scrolling dibawahnya namun tetapi jika dia sering menonton tentang musik, maka algoritma content-based filtering yang bekerja didalam suatu pemrosesan yang akan merekomendasikan video tentang musik lainnya, seperti video klip musik atau konser musik. Metode content-based filtering yang mempertimbangkan fitur-fitur dari konten yang disukai oleh pengguna, seperti tag, deskripsi, dan judul, namun untuk menentukan seberapa jawaban yang muncul dapat dikatakan konten untuk ditonton berikutnya dengan isi yang relevan berdasarkan pola mengklik video yang secara terus menerus. Dengan demikian, metode ini memiliki kelebihan dalam merekomendasikan konten yang spesifik dan personal ada juga memiliki kelemahan dalam merekomendasikan konten yang berbeda dari minat pengguna.

Algoritma content-based filtering pada YouTube menentukan rekomendasi video dengan mempertimbangkan fitur-fitur dari konten yang disukai oleh pengguna, seperti tag, deskripsi, dan judul. Algoritma ini kemudian mencari video lain yang memiliki fitur-fitur serupa dengan konten yang disukai oleh pengguna menggunakan metode cosine similarity. Metode cosine similarity digunakan untuk mengukur seberapa mirip dua vektor dalam ruang fitur yang diterapkan pada aplikasi berbasis mobile dengan menggunakan persentase atau data yang terukur terhadap kemiripan produk dengan hasil kemiripan yang tertinggi [3]. Semakin tinggi nilai cosine similarity, semakin mirip kedua vektor tersebut. Dengan demikian, semakin tinggi nilai cosine, semakin mirip konten kedua video tersebut dan semakin besar kemungkinan video tersebut direkomendasikan kepada pengguna. Algoritma content-based filtering pada YouTube juga mempertimbangkan riwayat tontonan dan perilaku pengguna dalam menonton video sebelumnya, seperti rewind dan rewatch, untuk merekomendasikan video yang sesuai dengan minat dan riwayat pengguna.

2. Metode Penelitian

Pada bagian ini yang dilakukan beberapa penelitian sebelumnya yaitu analisa e-commerce namun dilakukan dengan studi kasus berbeda terkait analisa konten di Youtube adalah sebagai berikut :

2.1 Studi Literatur

Untuk suatu pada algoritma content-based filtering di Youtube bahwa alur pada rekomendasi konten yang melibatkan pengguna suatu kumpulan informasi dalam konten youtube meliputi tanggal, judul, channel, kategori, atau durasi yang bisa membuat rekomendasi terbaik yang personal. Algoritma ini memiliki mengukur kemiripan konten yang bernama cosine similarity [3]. Hasil penelitian nantinya bahwa bisa menunjukkan pada Content-Based Filtering yang efektif dari durasi menonton dalam menghasilkan rekomendasi relevan berdasarkan preferensi pengguna dan karakteristik video.

2.2. Pengumpulan Data

Dalam suatu sistem ini yang dapat mengumpulkan topik video yang ada di Youtube tentu bisa digunakan untuk mengetahui informasi tentang merekomendasi konten. Kemudian data ini harus mencakup pada informasi seperti judul video, nama channel, tag, kategori, dan informasi usia terkait video tersebut. mengidentifikasi saat simulasi aplikasi Youtube berjalan dengan pencarian atau mengklik langsung dari tampilan beranda dan menyimpan data history tersebut secara sembunyi khusus untuk mode samaran dalam percobaan langsung.

2.3. Pelatihan Model

Setelah melakukan pemodelan, maka data dapat dilatih oleh model machine learning video berdasarkan pencarian judul video tersebut yang diimbangi dengan kategori ke suatu algoritma yang dapat bekerja dalam pembelajaran mesin seperti collaborative filtering, content-based filtering, atau hybrid filtering yang bertujuan untuk menghasilkan suatu rekomendasi. Dalam suatu implementasinya maka proses ini akan membandingkan minat dengan pengguna produk termasuk platform digital lalu untuk mengoptimalkan parameter model dalam proses mencapai kinerja terbaik yang menggunakan tools cosine similarity.

2.4. Pengujian Model

Setelah kita pelatihan berdasarkan judul video maka akan diuji dari beberapa kumpulan konten tersebut yang dipakai dalam mengolah suatu rekomendasi konten. Dari suatu pengujian sistem yang akan membantu dalam memvalidasi yang terkait dengan keefektifan algoritma, penyesuaian yang dilakukan, serta meningkatkan pengalaman pengguna didalam menemukan suatu konten yang relevan dari Youtube.

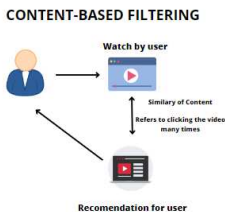
2.5 Evaluasi Model

Dengan adanya evaluasi model ini maka diperlukan menguji dalam suatu kinerja pemodelan. Kemudian, ini dapat membagikan data konten Youtube yang menjadi antara set pelatih dan pengujian. Adapun yang dalam evaluasi model ini berdasarkan waktu menonton yang dapat diberikan sinyalir kepada suatu sistem rekomendasi mengacu ke pengguna. Untuk memastikan bahwa penonton dapat dikatakan puas dengan konten yang akan ditonton, sistem tersebut dapat mengukur tiap konten video tersebut [4]. untuk mengevaluasi model ini, saya menggunakan metrik evaluasi precision dengan rumus dasarnya adalah:

$$\text{recommender system precision: } P = \frac{\# \text{ of our recommendations that are relevant}}{\# \text{ of items we recommended}} \quad (1)$$

2.6 Perancangan Sistem Rekomendasi

Penelitian ini telah melakukan tahapan-tahapan dalam pengolahan data dimana sebuah sistem rekomendasi ini yang menggunakan algoritma Content-Based Filtering berdasarkan kata kunci di judul video tersebut serta durasi menontonnya. Maksud dari sistem rekomendasi ini untuk memberikan saran dalam proses pengambilan keputusan tiap item. Berikut gambar secara umum dari algoritma Content-Based Filtering

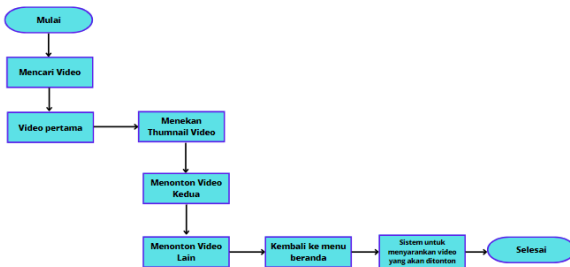


Gambar 1. Alur Content Based Filtering

Lalu, disini pada gambar diatas mengenai content based filtering ini yang membahas bagaimana suatu alur secara umum pada gambar diatas ini yang dirancang dalam suatu rekomendasi konten berupa media video online di Youtube. Implementasi algoritma ini melibatkan perbandingan minat pengguna sesuai data riwayat tontonan yang melalui faktor-faktor diperhitungkan yang berasal dari user sendiri yaitu seberapa sering mengklik suatu video dan menambahkan ke daftar favorite apabila akun tersebut telah dilogin. Namun, sebelum membangun sistem ini dapat diutamakan adalah mengumpulkan perferensi suatu pengguna terhadap konten kesenangan mereka. Profil pengguna yang dibangun berdasarkan pilihan konten, preferensi, dan selera mereka, yang kemudian ini digunakan untuk lakukan peringkat pengguna. Jadi, algoritma tersebut pada dasarnya mencari kesamaan antara konten-konten yang ada. Kemiripan dan kedekatan ini antara konten-konten diukur berdasarkan kesamaan isi, seperti genre, kategori, dan aspek-aspek lainnya yang dalam konten tersebut yang diharapkan bisa memberikan saran secara singkat sesuai dengan minat dan preferensi para pengguna. Dalam konteks tersebut ini bisa membantu pengguna menemukan konten yang relevan serta menarik berdasarkan suatu kesamaan isi dari konten yang telah mereka dinikmati.

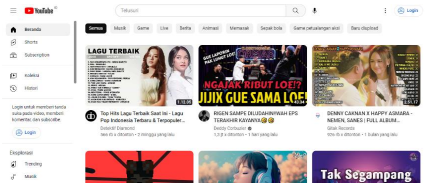
3. Hasil dan Pembahasan

Berikut adalah sebuah gambaran flowchart yang digunakan dalam menganalisa perancangan sistem untuk user melihat beberapa video yang ingin ditonton, saat ditonton, dan seterusnya maka yang akan keluar adalah iklan atau rekomendasi video berikutnya akan ditonton disaat nanti pada menu beranda :



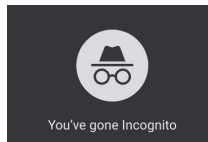
Gambar 2. Flowchart

Kemudian, untuk membuktikan suatu algoritma yang bekerja di platform digital Youtube maka yang kami akan lakukan dengan percobaan dalam mensimulasikan yang menggunakan mode samaran. Ini tentu bertujuan untuk mensetting awal layaknya akun baru pada tampilan beranda yang bersifat sementara. Berikut adalah hasil dokumentasinya :



Gambar 2. Flowchart

- a. User dapat membuka aplikasi atau website Youtube yang tersedia. Kemudian, ketika diproses oleh sistem sampai di tampilan beranda dan pastikan dalam keadaan mode samaran. Untuk pada tampilan awal atau beranda anda dapat melihat daftar konten.



Gambar 2. Flowchart

- b. Mode samaran ini berfungsi untuk menampilkan gambar miniature video atau judul-judul video secara singkat pada halaman beranda. Selain itu, mode ini dapat membantu dalam mengeksplorasi dan menemukan konten video yang menarik tanpa perlu membuka setiap video secara penuh.



Gambar 2. Flowchart

- c. Untuk mencari video yang spesifik, maka user dengan mengklik video simbol pencarian yang terletak di sebelah kanan atas layar. Setelah ini anda cukup mengetikkan topik yang ingin dicari setelah dikonfirmasi maka akan muncul daftar video dengan topik yang ingin menelusuri.



Gambar 2. Flowchart

- d. Setelah menemukan video yang diinginkan dari hasil mesin penelusuran terhadap suatu konten maka user dapat mengklik langsung ke suatu video untuk dapat diputar. Jika setelah itu, anda dapat menonton sebagian selama 3 menit sesuai kebutuhan dari total keseluruhan video yang dimiliki oleh satu judul video. Setelah itu, anda dapat memilih satu video lagi untuk melanjutkan menonton video berikutnya atau tetap pada video yang sama.

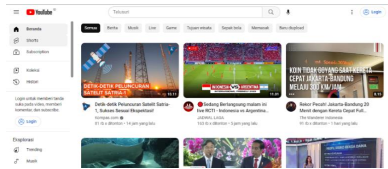


Gambar 2. Flowchart

- e. Lalu, jika user dapat menggantikan video berikutnya dengan topik yang sama maupun lainnya. Anda dapat mencari video baru atau menggunakan fitur rekomendasi secara otomatis pada saat menikmati konten yang diinginkan di samping yang sedang diputar.
- f. Setelah menonton 5 video dengan durasi bervariasi dalam berselang waktu sekitar 10 menit, maka pengguna bisa kembali untuk beralih ke tampilan beranda. Hal ini dapat memberikan pengalaman total waktu menonton yang dapat mempengaruhi rekomendasi konten video dari sistem tersebut.



Gambar 2. Flowchart



Gambar 2. Flowchart

- g. Setelah 10 menit dijeda, maka sistem bisa menunjukan rekomendasi konten yang serupa dan relevan dengan topik yang telah anda tonton sebelumnya di dalam preview tersebut. Semakin lama menonton serta berinteraksi dengan konten di Youtube maka yang muncul terjadi adalah banyak rekomendasi mengenai konten berhubungan dengan topik-topik tertentu di halaman beranda. Dengan mengikuti langkah ini dimana web/aplikasi Youtube ini melakukan simulasikan pengalaman dengan mencari, menonton, dan menjelajahi berbagai video yang sesuai dengan minat dan kebutuhan anda.

Adapun sistem rekomendasi dari konten Youtube menggunakan parameter dan algoritma untuk bekerja secara otomatis. Lalu beberapa parameter yang digunakan dalam sistem rekomendasi bekerja konten di Youtube dengan mempertimbangkan beberapa faktor yaitu Riwayat menonton, preferensi pengguna, interaksi pengguna, konteks video, kesesuaian demografis, popularitas trending, pengelompokan konten, durasi tontonan, dan fitur rekomendasi tambahan. Namun, perlu untuk diperhatikan bahwa Youtube terus mengembangkan dan memperbaiki algoritma rekomendasi mereka serta parameter yang dijadikan faktor digunakan dalam sistem rekomendasi dapat berubah seiring dengan waktu.

Kemudian, dapat dipastikan dengan percobaan cosine simily yang dapat dikatakan untuk mengukur kemiripan. Pada sebelumnya pengukuran ini dengan biasa digunakan dalam penambangan teks, pencarian informasi, dan aplikasi pembelajaran mesin seperti klasifikasi dokumen, sistem rekomendasi, dan pengelompokan lalu berikut pada tabelnya:

Title	Duration (Minutes)	Total Similarity
Spesifikasi Serta Manfaat Satelit Satria-1	4	1. 0.40824829 0.32732684 0.36927447 0.35355339
Layanan Internet Hingga Pelosok, Satelit Satria-1 Siap Mengorbit	6	0.40824829 1. 0.4454354 0.40201513 0.38490018
Satelit Satria-1 Diluncurkan Ke Orbit, Dapat Percepat Akses Layanan Internet Merata Se-Indonesia	1	0.32732684 0.4454354 1. 0.40291148 0.46291005
Peluncuran Satelit Satria-1 Sukses, Internet Di Indonesia Akan Semakin Cepat	3	0.36927447 0.40201513 0.40291148 1. 0.34815531
Satelit Satria-1 Diluncurkan, Kini Desa-Desa Bisa Dapat Internet	3	0.35355339 0.38490018 0.46291005 0.34815531 1.

Dari data tersebut melalui tahap pelatihan permodelan yang menyatakan bahwa terdapat penyesuaian yang dilakukan untuk pengukuran dengan menggunakan metode cosine similarity yang memberikan hasil dari pengklasifikasi konten dari masing-masing judul video yang berbeda maupun kata-kata yang sama.

4. Kesimpulan

Pada simulasi yang menjalankan aplikasi Youtube secara langsung dalam analisa sebuah sistem yang berjalan yang mengetahui umpan balik pada pilihan video tersebut di menu beranda telah diterapkan untuk memastikan privasi pengguna dan memberi pengalaman disaat menonton. Dengan menganalisis durasi dan judul video yang ditonton oleh pengguna, arsitektur program ini menggunakan fungsi tools/library seperti CountVectorizer dan cosine_similarity dari perpustakaan sklearn untuk menghitung skor kemiripan yang mengetahui bagaimana cara kerja merekomendasikan konten yang relevan berdasarkan preferensi pengguna. Sistem yang diterapkan menggunakan content-based filtering untuk merekomendasikan video YouTube. Perpaduan antara judul dan lama menontonnya maka yang disebut atribut dalam analisis pengklasifikasian konten disaat berjalan. Dengan memasukkan umpan balik pengguna, pembuatan profil pengguna, pemfilteran kolaboratif, dan pembaruan waktu nyata, sistem dapat lebih ditingkatkan dalam jangka waktu yang belum pasti untuk memberikan rekomendasi yang lebih akurat dan dipersonalisasi dilakukan pengukuran antara judul video dengan durasi menonton untuk meningkatkan pengalaman pengguna secara keseluruhan di platform YouTube.

Daftar Pustaka

- [1] Jati, Nala Prasasta; ABIDIN, Zainal; Oxcygentri, Mickey. Pengaruh Sosial Media Youtube Terhadap Perilaku Literasi Siswa SMA Utama. *Linimasa: Jurnal Ilmu Komunikasi*, 2022, 5.1: 67-75.
- [2] Rahmadi, Ridho, et al. Analisis Sentimen Politik Berdasarkan Big Data dari Media Sosial Youtube: Sebuah Tinjauan Literatur. *Automata*, 2021, 2.1.
- [3] Sujasman, M. Bagus; Diana, Diana; SYAZILI, Ahmad. Implementasi Metode Cosine Similarity Untuk Rekomendasi Produk Pada Aplikasi Penjualan Berbasis Mobile. In: *Bina Darma Conference on Computer Science (BDCCS)*. 2020. p. 162-171.
- [4] Cristos Goodrow. "indonesia.googleblog.com" : Memahami cara kerja sistem rekomendasi Youtube. 18 10 2021. [Online]. Available: <https://indonesia.googleblog.com/2021/10/memahami-cara-kerja-sistem-rekomendasi.html?m=1>. [Accessed 9 06 2023].
- [5] Fitrianti, Betharia Sri; Fachurrozi, Muhammad; Yusliani, Novi. Sistem Rekomendasi Artikel Ilmiah Berbasis Web Menggunakan Content-based Learning dan Collaborative Filtering. *Generic*, 2018, 10.1: 17-23.

Perhitungan Nilai Besaran Fisis Mammografi Jenis Histopatologi IDC dan ILC

Anak Agung Ngurah Frady Cakranegara^{a1}, Ida Ayu Gde Suwiprabayanti Putra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹gungfrady04@gmail.com
²agsuwiprabayantiputra@unud.ac.id

Abstract

In this study, the main objective was to calculate the range of physical values contained in mammography X-ray images and determine the physical quantities that are significant in differentiating between the histopathological types of ILC (Invasive Lobular Carcinoma) and IDC (Invasive Ductal Carcinoma). The research method involved collecting data from 152 mammograms consisting of 7 ILCs and 145 IDCs from doctor Sutomo Surabaya's radiology database. The range of physical values such as entropy, contrast, second angular moment, differential invest moment, mean, deviation, entropy of Hdiff, angular moment of Hdiff, and mean of Hdiff are calculated and compared between ILC and IDC using the Anova statistical test. The results showed that there were differences in the range of physical quantity values between ILC and IDC. Significant parameters in differentiating the two types of histopathology are mean1, mean2, mean3, and mean4. In conclusion, IDC has a higher peak than ILC, and the range of ILC physical quantities is higher than IDC.

Keywords: X-ray, IDC, ILC, Mammografi, Anova

1. Pendahuluan

Kanker payudara merupakan penyebab kematian terbanyak setelah kanker rahim, banyak penelitian yang dilakukan untuk mendeteksi dini kanker payudara, seperti : texture coding [1] , edge detection [2] ,adaptive k-mean clustering [3] , self similar fractal [4], fractal feature [5], neural network [6], kekre's [7], SVM classifier [8], texture resemblance marker [9], extraction [10], accurate method (M. Rizzi,at.al.,2010), contour description [11], bilateral asymmetry S. K. Bandyopadhyay (2010), orthogonal polynomials model [12], dual tree complex [13], gabor features [14], fuzzy clustering [15], k-means and fuzzy c-means (N. Singh,at.al.,2011), vector quantization technique (H.B. Kekre,at.al.,2009), kohonen network SOM and LVQ [16], T salis entropy & a type II fuzzy (Mohanalin,at.al.,2010), foveal method (Oh Whi-Vin,at.al,2009), wavelet [17]. Belum ada yang menggunakan besaran fisis untuk mengkalsifikasi histopatologi kanker payudara. Pada penelitian ini bertujuan untuk menghitung range nilai besaran fisis yang terdapat pada hasil foto sinar-X mammografi dan menentukan besaran fisis apa saja yang benar-benar signifikan mampu membedakan ILC dan IDC. Pada makalah ini diorganisasi sebagai berikut. Bagian 2 bahan dan metoda, bagian 3 hasil, Bagian 4 pembahasan dan kesimpulan dibahas pada bagian 5.

2. Metode Penelitian

Metode penelitian adalah suatu langkah ilmiah yang digunakan untuk memecahkan suatu masalah guna mencapai tujuan tertentu. Penelitian ini bertujuan untuk membangun sebuah ontologi yang dapat menjadi basis komputerisasi di bidang perfilman untuk pengembangan sistem rekomendasi pemilihan film. Dalam membangun sebuah ontologi, diperlukan sebuah metode yang disebut Methontology. Methontology adalah salah satu yang memberikan keuntungan dalam kegiatan konseptualisasi rinci pada setiap tahap dan juga memiliki

kemampuan untuk mengatur ulang ontologi. Langkah-langkah metodenya adalah Spesifikasi, akuisisi, konsep, integrasi, implementasi, evaluasi, dokumentasi

a. Mammogram

Pertama-tama kami mengambil data record pasien dari ruang radiologi yang sudah pasti nilai histopatologinya dan sudah diperiksa oleh dokter ahli onkologi yang sudah berpengalaman lebih dari 20 tahun. Setelah mendapatkan data record pasien kemudian kami lanjut mengambil gambar mammogram dari data base ruang radiologi sesuai data record pasien. Tidak semua data record pasien ada gambar mamogramnya, karena tidak semua pasien melakukan pemeriksaan mamografi dirumah sakit tersebut. Data mammogram merupakan data sekunder yang diambil dari data base rumah sakit Dokter Soetomo Surabaya mulai bulan Januari 2023 sampai Mei 2023, mammogram yang memenuhi kriteria inklusi berikut dimasukkan dalam penelitian ini. 1) mammogram dengan lesi payudara yang mencurigakan terdeteksi dan data recordnya ada di ruang onkologi. 2) Pasien tidak menjalani biopsi, kemoterapi atau intervensi lainnya sebelum pemeriksaan. 3) Diameter lesi payudara adalah lebih besar dari 1cm. 4) Karakteristik lesi adalah dikonfirmasi oleh patologi. Dari 200 data hanya 152 data memenuhi kriteria inklusi. 152 mammogram yang terdiri dari 7 ILC dan 145 IDC.

b. Akuisisi Gambar

Gambar mammogram diambil dari alat mamografi merek Kodak tipe dryview 6800 laser imager dengan seting KV= 30, MAS = 25, brightness = 7, latitude = 11, kontras = -4, ukuran film = 18x24 cm.

c. Analisis Gambar

Seorang dokter ahli radiologi dengan pengalaman lebih dari 20 tahun dalam diagnosis lesi payudara, menganalisis gambar mammogram. ROI dipilih di area lesi yang paling jelas dengan ukuran 2 x 2 cm. kemudian dilakukan perbaikan kontras mammogram. Parameter fisika yang ada pada mammogram dihitung secara otomatis menggunakan persamaan (1 s/d 10) [18].

$$Entropy = - \sum_{y_q=y_1}^{y_t} \sum_{y_r=y_1}^{y_t} [H(y_q, y_r, d)] \log [H(y_q, y_r, d)] \tag{1}$$

$$Contrast = \sum_{y_q=y_1}^{y_t} \sum_{y_r=y_1}^{y_t} (y_q - y_r)^2 H(y_q, y_r, d) \tag{2}$$

$$Moment\ Anguler\ kedua = \sum_{y_q=y_1}^{y_t} \sum_{y_r=y_1}^{y_t} [H(y_q, y_r, d)]^2 \tag{3}$$

$$Momen\ Differensial\ Invers = \sum_{y_q=y_1}^{y_t} \sum_{y_r=y_1}^{y_t} \left[\frac{H(y_q, y_r, d)}{1 + (y_q - y_r)^2} \right] \tag{4}$$

$$Mean = \sum_{y_q=y_1}^{y_t} y_q H_m(y_q, d) \tag{5}$$

$$Deviation = \sqrt{\sum_{y_q=y_1}^{y_t} [y_q - \sum_{y_p=y_1}^{y_t} y_p H_m(y_p, d)]^2 H_m(y_q, d)} \tag{6}$$

$$H_{diff}(i, d) = \sum_{y_q=|y_q-y_r|=i}^{y_t} \sum_{y_r=y_1}^{y_t} H(y_q, y_r, d) \tag{7}$$

$$Entropy\ dari\ H_{diff} = - \sum_{i=i_1}^{i_t} H_{diff}(i, d) \log H_{diff}(i, d) \tag{8}$$

$$ASM\ dari\ H_{diff}(i, d) = \sum_{i=i_1}^{i_t} [H_{diff}(i, d)]^2 \tag{9}$$

$$Mean\ H_{diff} = \sum_{i=i_1}^{i_t} i H_{diff}(i, d) \tag{10}$$

dengan y_0, y_1, y_2, y_3 berturut-turut adalah nilai *gray-level pixel* kesatu, nilai *gray-level pixel* kedua, nilai awal nol, nilai akhir 255, jarak antar dua *pixel*. Hasil perhitungan tersebut kemudian ditabulasi menggunakan excel, nilai entropy ILC dan ILC dikumpulkan jadi satu file, begitu juga sepuluh parameter fisika yang lain. Kemudian dilakukan Analisa statistic Anova table 1. Kami juga membuat visual pemisahan background dengan objek massa yang mencurigakan seperti terlihat pada gambar 1. Kami akan melihat apakah ada perbedaan visual ILC dan IDC.

d. Transformasi Data

Pada penelitian ini kami menggunakan 3 jenis data, yaitu data asli (tidak ditransformasi), data ditransformasi biner yang nilainya dari 0 sampai 1, data ditransformasi bipolar yang nilainya dari -1 sampai 1. Persamaan untuk normalisasi data sebagai berikut:

$$\text{Data Biner} = 0.8 * (X - \text{Min}) / (\text{Max} - \text{Min}) + 0.1 \tag{11}$$

$$\text{Data Bipolar} = 0.8*(X - \text{Min}) / (\text{Max} - \text{Min}) + 0.1 + (X - \text{Min}) / (\text{Max} - \text{Min}) - 1 \tag{12}$$

Dimana:

- X = nilai intensitas piksel data asli
- Min = nilai minimum dari keseluruhan data
- Max = nilai maksimum dari keseluruhan data.

e. Analisis Statistic

Setelah parameter fisik dihitung menggunakan persamaan (1) sampai (10), kemudian hasil perhitungan parameter fisik tersebut ditransformasi ke Biner dan Bipolar menggunakan persamaan (11) dan (12), kemudian dianalisa statistik menggunakan Uji Anova untuk mencari parameter fisika yang benar-benar mampu membedakan ILC dan IDC seperti terlihat pada table 1.

Table 1. Hasil Uji Statistik Anova.

No	Parameter	Jenis	Transformasi											
			Tidak				Biner				Bipolar			
			Mean	Median	Variance	P	Mean	Median	Variance	P	Mean	Median	Variance	P
1	Env1	ILC	3.8171667	3.64627	0.007	0.907	0.0706962	0.8911345	0.097	0.371	0.184117	0.4305027	0.463	0.371
		IDC	3.8194336	3.64668	0.023	0.6287115	0.6695326	0.026	0.2891109	0.3747138	0.3747138	0.126	0.26	
2	Env2	ILC	3.891179	3.69801	0.006	0.933	0.0462005	0.8453933	0.083	0.25	0.109012	0.3271137	0.479	0.25
		IDC	3.896448	3.69217	0.024	0.8239689	0.6607593	0.028	0.278117	0.3811084	0.132	0.28		
3	Env3	ILC	3.9813429	3.71025	0.008	0.952	0.0414639	0.8118671	0.078	0.2	0.0933072	0.251701	0.385	0.2
		IDC	3.877772	3.71489	0.024	0.6289765	0.6657443	0.027	0.288472	0.3792047	0.138	0.28		
4	Env4	ILC	3.8888114	3.70141	0.008	0.956	0.0500551	0.5951444	0.08	0.208	0.1145394	0.2163248	0.405	0.208
		IDC	3.888545	3.72287	0.024	0.6330853	0.6718775	0.027	0.303887	0.3887243	0.139	0.28		
5	Env5	ILC	3.8952887	3.7021	0.008	0.963	0.0537113	0.5818088	0.08	0.19	0.1210304	0.1840717	0.405	0.19
		IDC	3.894768	3.70259	0.024	0.8112384	0.6749559	0.027	0.3718859	0.4046638	0.138	0.28		
6	Env6	ILC	3.8086771	3.7003	0.008	0.934	0.0562452	0.5678019	0.082	0.236	0.1400507	0.1525543	0.475	0.236
		IDC	3.8949498	3.73294	0.024	0.6460628	0.6881328	0.028	0.3305309	0.4232682	0.144	0.28		
7	Env7	ILC	3.70003	3.69404	0.008	0.975	0.0633577	0.5432091	0.084	0.183	0.1420548	0.0974239	0.423	0.183
		IDC	3.8662033	3.7345	0.023	0.6532126	0.6831435	0.028	0.3447283	0.43485728	0.141	0.28		
8	Env8	ILC	3.8977043	3.69274	0.008	0.984	0.0791215	0.5801387	0.089	0.237	0.1790234	0.1303701	0.449	0.237
		IDC	3.8959895	3.73428	0.023	0.659988	0.7017668	0.028	0.3592279	0.4539752	0.143	0.28		
9	Env9	ILC	3.8996874	3.69113	0.008	0.986	0.0847186	0.5864521	0.091	0.236	0.1602523	0.1314678	0.448	0.236
		IDC	3.8841883	3.73761	0.022	0.6848231	0.7134161	0.028	0.370852	0.4881728	0.141	0.28		
10	Env10	ILC	3.8919357	3.67994	0.005	0.99	0.0818865	0.5343939	0.092	0.237	0.1842871	0.0752331	0.468	0.237
		IDC	3.8905294	3.73155	0.022	0.6817626	0.7073611	0.028	0.363663	0.4665624	0.14	0.28		
11	Cont1	ILC	339.7289071	255.32723	63295.793	0.726	0.3026988	0.2158745	0.078	0.802	-0.4291488	-0.6397325	0.394	0.802
		IDC	315.138488	276.21954	31977.801	0.2039475	0.2559134	0.023	-0.4638182	-0.5491949	0.115	0.28		
12	Cont2	ILC	682.7721929	372.39403	220317.26	0.858	0.0794803	0.1734698	0.08	0.704	-0.4962038	-0.7349265	0.483	0.704
		IDC	499.8705874	432.54899	30561.08	0.084	0.2894186	0.2955383	0.027	-0.4208416	-0.5275389	0.171	0.28	
13	Cont3	ILC	888.0785014	481.8838	235495.235	0.937	0.2927707	0.1848808	0.079	0.93	-0.4862588	-0.7094688	0.482	0.93
		IDC	686.788411	581.12857	186338.435	0.2978804	0.2834479	0.02	-0.454814	-0.522421	0.102	0.28		
14	Cont4	ILC	778.0520114	584.98815	283056.082	0.904	0.3032494	0.1964218	0.08	0.987	-0.4428389	-0.680515	0.404	0.987
		IDC	801.8032399	687.85765	234957.878	0.3008317	0.2816762	0.02	-0.4481287	-0.5362285	0.102	0.28		
15	Cont5	ILC	888.5953447	678.8894	398194.888	0.813	0.3094878	0.200431	0.081	0.971	-0.43138	-0.6740322	0.408	0.971
		IDC	927.9188498	815.85116	328188.927	0.3045389	0.2739698	0.021	-0.444197	-0.5091622	0.104	0.28		
16	Cont6	ILC	888.7688614	768.3899	389675.922	0.768	0.3088763	0.2041325	0.081	0.981	-0.4277687	-0.6887016	0.412	0.981
		IDC	888.7688614	768.3899	389675.922	0.768	0.3088763	0.2041325	0.081	0.981	-0.4277687	-0.6887016	0.412	0.981

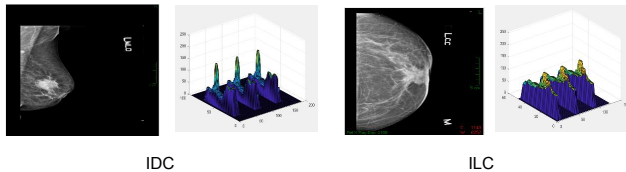
Table 2. Parameter yang Signifikan

Jenis Data	Parameter signifikan
Tidak ditransformasi	Mean 1, mean 2, mean 3 dan mean 4
Ditarnsformasi Biner	MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10
Ditransformasi Bipolar	MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, DEV 7, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10

3. Hasil dan Diskusi

3.1 Hasil

Hasil visualisasi IDC dan ILC dengan cara memisahkan background dengan massa yang mencurigakan ternyata memiliki visual yang berbeda seperti gambar 1.



Gambar 1. IDC dan ILC

Hasil perhitungan parameter fisis menggunakan persamaan (1) sampai (10) diperoleh ring nilai parameter fisis seperti table 3.

Table 3. Range Nilai Besaran Fisis Film ILC dan IDC untuk Data Tidak Dinormalisasi

No	Besaran Fisis	ILC	IDC
1	Entropy	3.49541- 3.79207	3.1047 - 3.93558
2	Contrast	151.01428- 3024.44113	86.08063- 4736.43024
3	Moment anguler kedua	0.00019- 0.00073	0.00013- 0.0009
4	Moment Differensial Invers.	0.02213- 0.06482	0.0113- 0.392
5	Mean.	127.05144- 182.39332	71.16284- 210.21205
6	Deviation.	19.39466- 48.20695	11.06751- 93.33126
7	Entropy of H_{diff} .	1.41346- 2.00655	1.29424- 2.1479
8	<i>moment anguler of H_{diff}</i>	0.01272- 0.04542	0.00806 - 0.06171
9	Mean of H_{diff}	9.72898- 38.68645	7.27355- 55.92737

3.2 Diskusi

Dari gambar 1 terlihat ada perbedaan visual IDC dengan ILC dilihat dari puncak grafiknya, IDC mempunyai puncak lebih tinggi dari ILC. Dari table 3 terlihat range nilai parameter fisika ILC lebih

tinggi dari IDC. Dari table 1 terlihat besaran fisis yang berpengaruh untuk membedakan jenis histopatologi *ILC* dan *IDC* untuk data tidak ditransformasi adalah mean1, mean2, mean3 dan mean4, sedangkan yang ditransformasi Biner adalah MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10, sedangkan untuk Bipolar adalah MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, DEV 7, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10 Untuk penelitian berikutnya kami akan menggunakan parameter fisik yang signifikan ini sebagai parameter masukan dari metode JST Backpropagation untuk TA kami, dan kami akan menguji kinerja dari metode JST Backpropagation, dengan cara menghitung nilai Accuracy, Sensitivity, Specificity dan precision.

4. Kesimpulan

Dari visul grafik IDC mempunyai puncak lebih tinggi dari ILC, range nilai parameter fisika ILC lebih tinggi dari IDC, besaran fisis yang berpengaruh untuk membedakan jenis histopatologi *ILC* dan *IDC* adalah mean1, mean2, mean3 dan mean4, sedangkan yang ditransformasi Biner adalah MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10, sedangkan untuk Bipolar adalah MA 2, MA 3, MA 4, MA 5, MA 6, MA 7, MA 8, MA 9, MA 10, MD 2, MD 4, DEV 7, EntrHd 9, EntrHd 10, MAHd 6, MaHd 7, MAHd 8, MAHd 9, MAHd 10.

Daftar Pustaka

- [1] F. Eddaoudi and F. Regragui, "Microcalcifications Detection in Mammographic Images Using Texture Coding," 2011.
- [2] B. Senthikumar and G. Umamaheswari, "A Novel Edge Detection Algorithm for the Detection of Breast Cancer," 2011. [Online]. Available: <http://www.eurojournals.com/ejsr.htm>
- [3] B. Charan Patel, "An Adaptive K-means Clustering Algorithm for Breast Image Segmentation," 2010.
- [4] B. Charan Patel, "Early Detection of Breast Cancer using Self Similar Fractal Method," 2010.
- [5] D. Sankar and T. Thomas, "Fractal Features based on Differential Box Counting Method for the Categorization of Digital Mammograms." [Online]. Available: <http://www.mirlabs.org/ijcisim>
- [6] J. June June, "Detection of Microcalcification Clusters in Mammograms using Neural Network," 2010.
- [7] "Dr.H.B.Kekre_print".
- [8] F. Eddaoudi, F. Regragui, A. Mahmoudi, and N. Lamouri, "Masses Detection Using SVM Classifier Based on Textures Analysis," 2011.
- [9] M. Nielsen *et al.*, "A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer," *Cancer Epidemiol*, vol. 35, no. 4, pp. 381–387, Aug. 2011, doi: 10.1016/j.canep.2010.10.011.
- [10] M. Vasantha, D. V Subbiah Bharathi, and R. Dhamodharan, "Medical Image Feature, Extraction, Selection and Classification," 2010.
- [11] P. H. Tsui, Y. Y. Liao, C. C. Chang, W. H. Kuo, K. J. Chang, and C. K. Yeh, "Classification of benign and malignant breast tumors by 2-D analysis based on contour description and scatterer characterization," *IEEE Trans Med Imaging*, vol. 29, no. 2, pp. 513–522, Feb. 2010, doi: 10.1109/TMI.2009.2037147.
- [12] R. Krishnamoorthy, N. Amudhavalli, and M. K. Sivakolundu, "Identification of Microcalcifications with Orthogonal Polynomials Model," 2010.
- [13] K. Sujatha and V. C. Sumitha, *Dual Tree Complex with Modified Complex Ridgelets for Image Denoising in Digital Mammographic Images*. 2007.
- [14] Y. Zheng, "Breast cancer detection with gabor features from digital Mammograms," *Algorithms*, vol. 3, no. 1, pp. 44–62, Mar. 2010, doi: 10.3390/a3010044.
- [15] S. Saheb Basha and D. K. Satya Prasad, "Automatic Detection of Breast Cancer Mass In Mammograms Using Morphological Operators And Fuzzy C-Means Clustering," 2005. [Online]. Available: www.jatit.org

- [16] "classification of the medical images by the kohonen network SOM and LVQ".
- [17] S. Bouyahia, J. Mbainabeye, and N. Ellouze, "Wavelet Based Microcalcifications Detection in Digitized Mammograms." [Online]. Available: <http://www.enit.rnu.tn>
- [18] A. A. N. Gunawan, S. Poniman, and I. W. Supardi, "Classification of breast cancer grades using physical parameters and K-nearest neighbor method," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 2, pp. 637–644, Apr. 2019, doi: 10.12928/Telkomnika.V17I2.9797.

Uji Performansi Algoritma LR dan RFR pada Implementasi Sistem Prediksi Harga Rumah

I Putu Teddy Dharma Wijaya^{a1}, Ida Bagus Dwidasmara^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹ dharmawijaya077@student.unud.ac.id
² dwidasmara@unud.ac.id (Corresponding Author)

Abstract

Currently the house has become one of the needs that must be met. The price of a house is the main parameter that determines whether a person or organization buys or invests. In general, house prices are influenced by several factors, including building area, land area, number of bedrooms, number of bathrooms and number of garages. Currently, there are many websites devoted to providing information about buying and selling houses. This of course makes it easier for someone when looking for a house with the desired specifications without the need to come directly to the location. However, the house buying and selling platform does not provide a house price prediction feature that is in accordance with user specifications. This means someone who is planning to buy a house does not get an initial idea of the costs that must be spent to own the desired home. Therefore, in this study, researchers will design a web app-based house price prediction system that can make it easier for users to get predictions of the desired house price. In this study the prediction algorithms to be used are linear regression and random forest. Both algorithms will be analyzed for their performance and then the algorithm with the best level of accuracy will be applied as a predictive model which will be integrated with the user interface display.

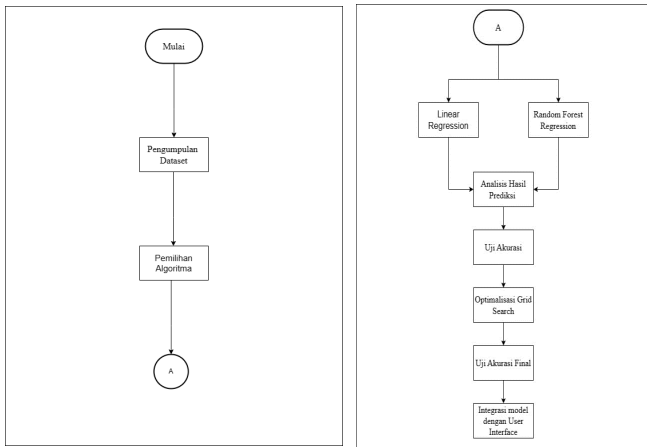
Keywords: *Harga Rumah, Linear Regression, Random Forest Regression*

1. Pendahuluan

Secara umum, rumah merupakan bangunan yang digunakan sebagai tempat tinggal dalam jangka waktu tertentu. Saat ini, rumah sudah menjadi kebutuhan primer bahkan sudah dijadikan alat untuk berinvestasi di masa yang akan datang dikarenakan harga yang cenderung berubah tiap waktunya serta banyaknya proses transaksi jual beli. Seiring berjalannya waktu, kebutuhan fisiologis manusia akan semakin bertambah salah satunya adalah kebutuhan dalam membeli rumah. Pengusaha properti akan berlomba-lomba membeli lahan perumahan yang kemudian akan dijual kembali kepada pembeli. Hal inilah yang menjadi salah satu faktor kuat yang menyebabkan harga rumah cenderung naik setiap tahunnya. Namun hal ini juga yang akan membuat masyarakat ragu dalam membeli rumah yang diinginkan. Harga yang terus berubah dan tidak bisa diprediksi membuat investor atau pembeli rumah memerlukan sebuah sistem yang dapat melakukan prediksi harga rumah berdasarkan beberapa parameter. Saat ini, salah satu metode machine learning yaitu prediksi sudah banyak digunakan sebagai bahan pertimbangan dalam pengambilan keputusan pembelian [1]. Hingga kini sudah banyak jurnal penelitian yang membahas prediksi harga rumah diantaranya, penelitian dengan judul "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression" yang dilakukan oleh M.Mu'tashm diperoleh hasil akurasi dari Multiple Linear Regression dalam prediksi harga rumah sebesar 66% [2]. Penelitian serupa yang dilakukan oleh Sefto P. (2016) dengan judul penelitian "Prediksi Harga Tanah menggunakan Algoritma Linear Regression" memberikan hasil akurasi yang tinggi dan RMSE terendah dengan data training 70% dan sisanya digunakan sebagai data testing yaitu sebesar 30% [3]. Berdasarkan penelitian yang sudah dilakukan sebelumnya dengan topik yang sama, algoritma Linear Regression merupakan algoritma yang paling banyak digunakan dalam penelitian topik prediksi harga rumah. Pada penelitian ini, selain algoritma

Linear Regression yang digunakan untuk prediksi harga rumah, penulis juga akan menggunakan algoritma Random Forest Regression sebagai pembandingan dari algoritma Linear Regression. Random Forest Regression merupakan algoritma terawasi yang menggunakan metode pembelajaran ensemble untuk regresi sebagai teknik dengan melakukan penggabungan prediksi dari beberapa prosedur pemecahan dalam machine learning untuk menghasilkan hasil yang akurat pada model tunggal yang dirancang [4][5]. Terdapat beberapa penelitian mengenai penerapan dari algoritma Random Forest dalam kasus prediksi, misalnya pada penelitian yang dilakukan oleh Agri P., dkk (2019) dengan judul "Prediksi Pergerakan Harga Saham dengan Metode Random Forest Menggunakan Trend Deterministic Data Preparation". Dari penelitian tersebut dihasilkan bahwa akurasi model yang dibangun dengan algoritma Random Forest pada data Trend deterministic adalah sebesar 61,40% dan pada data non-Trend deterministic adalah sebesar 75,76% [6]. Penelitian yang dilakukan oleh Egas S. dan Ida A. (2022) dengan judul "Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat." Dari penelitian tersebut diperoleh hasil precision sebesar 78%, recall 100% dan akurasi sebesar 80%. Pada penelitian kali ini, penulis akan melakukan komparasi performa algoritma Linear Regression dengan Random Forest Regression. Komparasi performa meliputi MSE (Mean Squared Root), MAE (Mean Absolute Error) dan R2 Score dari kedua model tersebut. Penulis juga akan melakukan optimalisasi kinerja dari kedua algoritma tersebut dengan metode grid search serta mencari kombinasi yang terbaik dari pembagian data train dan data test. Algoritma yang memiliki performa terbaik akan digunakan sebagai bagian dari model yang akan diintegrasikan dengan user interface pada sistem prediksi.

2. Metode Penelitian



Gambar 1. Tahapan Penelitian

Pada Gambar 1, dapat dilihat tahapan dari penelitian yang akan dibuat. Terdapat tujuh proses dalam penelitian ini. Pada tahap pertama proses dimulai dari pengumpulan dataset, pemilihan algoritma yaitu Linear Regression dan Random Forest, analisis hasil prediksi, uji akurasi,

optimalisasi dengan metode grid search, uji akurasi final kemudian integrasi model dengan user interface:

2.1 Pengumpulan Data

Pengumpulan dataset dilakukan dengan mengunduh dataset harga rumah yang didapatkan dari platform Kaagle pada laman "https://www.kaggle.com/datasets/gustiosamba/datarumahjaksel". Berdasarkan data yang didapatkan, dapat dilihat bahwa dataset terdiri dari 4 kolom yaitu harga rumah, luas bangunan, luas tanah, jumlah kamar mandi, jumlah kamar tidur dan jumlah garasi dengan total baris sebanyak 1010. Kolom luas bangunan, luas tanah, jumlah kamar mandi, jumlah kamar tidur dan jumlah garasi akan dijadikan features sedangkan kolom harga akan menjadi target features. Pada proses train, jumlah data yang digunakan akan divariasikan sebanyak 70%, 80% dan 90%. Hal ini dilakukan dengan tujuan untuk mengetahui jumlah data train optimal yang dapat digunakan.

2.2 Pemilihan Algoritma

Algoritma yang akan digunakan pada penelitian ini adalah Random Forest Regressoin dan Linear Regression. Kedua algoritma ini kemudian akan dibandingkan akurasi yang dihasilkan dengan metode MSE (Mean Squared Root), MAE (Mean Absolute Error) dan R2 Score

a. Linear Regression

Linear regression merupakan metode statistik yang digunakan untuk mengetahui pengaruh beberapa variabel terhadap satu buah variabel yang menjadi target [7]. Secara matematis Linear Regression akan mencari hubungan variabel dependen (y) dengan variabel independen (x) [8]. Berikut merupakan persamaan matematis dari metode Linear Regression [9]:

$$y = a + wx \quad (1)$$

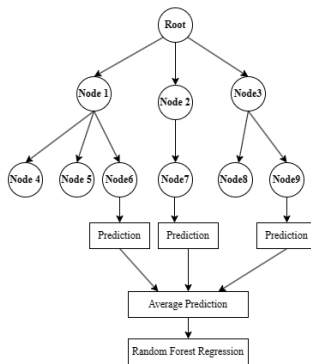
$$a = y - wx \quad (2)$$

$$w = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} \quad (3)$$

Dengan **w** merupakan vektor bobot sedangkan **a** adalah intercept yang merupakan sebuah nilai skalar yang di dalam bidang machine learning disebut dengan istilah bias. Sehingga, secara umum model linear regression dapat diartikan sebagai penggunaan suatu fungsi lurus atau datar. Jadi pembelajaran (learning) pada model linear adalah untuk menentukan dua parameter yaitu **a** dan **w**.

b. Random Forest Regression

Random Forest Regression merupakan kombinasi pohon keputusan sedemikian sehingga setiap pohon akan bergantung pada nilai nilai dari vektor acak yang disamping secara independen dan dengan distribusi yang sama untuk semua pohon dalam hutan tersebut. Keunggulan dari Random Forest Regression dibandingkan dengan Linear Regression adalah pada seleksi acak untuk memilih setiap simpul (node) yang mampu menghasilkan tingkat kesalahan yang rendah. Proses pelatihan dari Random Forest Regression melibatkan pembangunan banyak pohon keputusan secara paralel dan independen. Ketika melakukan prediksi, setiap pohon akan memberikan prediksi berdasarkan fitur yang ada serta hasil prediksi setiap pohon yang sudah dibangun. Random Forest Regression dapat bekerja optimal pada data yang tidak seimbang dan memiliki kemampuan untuk mengevaluasi kepentingan fitur dalam prediksi. Dalam algoritma Random Forest Regression parameter jumlah pohon (**n_estimators**), kedalaman pohon (**max_depth**) serta jumlah fitur yang digunakan dalam setiap pohon (**max_features**) akan mempengaruhi kinerja dari algoritma Random Forest Regression.



Gambar 2. Cara kerja dari algoritma Random Forest Regression

2.3 Uji Akurasi

Salah satu tahapan dalam proses penelitian yang sudah dijelaskan pada gambar 1 adalah uji akurasi. Pada penelitian ini, penulis akan melakukan uji akurasi sebanyak dua kali yaitu uji akurasi model tanpa optimalisasi Grid Search serta dengan optimalisasi Grid Search. Pada tahap uji akurasi, metode yang akan digunakan adalah sebagai berikut

a. MSE (Mean Squared Root)

MSE merupakan rata rata kuadrat kesalahan yang dihitung dengan menjumlahkan semua kesalahan atau eror prediksi yang dihasilkan oleh suatu model kemudian dikuadratkan dan membaginya dengan jumlah periode prediksi [10]. Berikut merupakan persamaan matematis dari MSE:

$$MSE = \frac{1}{n} \sum_i^n (X_i - F_i)^2 \quad (4)$$

Keterangan:

- X_i = Data aktual pada periode ke-i
- F_i = Nilai hasil prediksi atau prediksi pada period ke-i
- n = Banyaknya sampel

b. MAE (Mean Absolute Error)

MAE merupakan selisih absolut antara nilai prediksi dan nilai yang sebenarnya kemudian akan dihitung rata rata dari selisih tersebut. MAE dapat memberikan gambaran rata rata prediksi dengan nilai yang sebenarnya. Berikut merupakan persamaan matematis dari MAE:

$$MAE = \frac{1}{n} \sum_i^n |y_i - \hat{y}| \quad (5)$$

Keterangan:

- n = Banyaknya sampel
- yi = nilai aktual pada sampel ke i
- \hat{y} = nilai prediksi pada sampel ke i

c. R2 Score

R2 Score digunakan untuk mengukur sejauh mana variabilitas suatu model statistic atau machine learning dapat menjelaskan variasi data yang diamati. Range nilai dari R2 Score adalah antara 0 hingga 1. Berikut merupakan persamaan matematis dari R2 Score:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{6}$$

Keterangan:

- yi = Nilai aktual dari variabel dependen
- \hat{y}_i = Nilai aktual dari variabel dependen
- \bar{y} = Nilai rata rata dari variabel dependen

3. Hasil dan Diskusi

3.1 Perancangan Model

Pada penelitian ini, jumlah total data yang digunakan adalah sebanyak 1010 baris dengan 9 buah kolom yang meliputi No, Nama Rumah, Harga, Luas Bangunan (LB), Luas Tanah (LT), Kamar Mandi (KM), Kamar Tidur (KT), Garasi (GRS). Dataset yang digunakan merupakan dataset harga rumah yang terdapat pada daerah Jakarta Selatan. Informasi lebih jelas dari dataset yang digunakan dapat diakses pada <https://www.kaggle.com/datasets/gustiosamba/datarumahjaksel>.

NO	NAMA RUMAH	HARGA	LB	LT	KT	KM	GRS
0	1 Rumah Murah Hook Tebet Timur, Tebet, Jakarta S...	3800000000	220	220	3	3	0
1	2 Rumah Modern di Tebet dekat Stasiun, Tebet, Ja...	4600000000	180	137	4	3	2
2	3 Rumah Mewah 2 Lantai Hanya 3 Menit Ke Tebet, T...	3000000000	267	250	4	4	4
3	4 Rumah Baru Tebet, Tebet, Jakarta Selatan	4300000000	40	25	2	2	0
4	5 Rumah Bagus Tebet komp Gudang Peluru It 350m, ...	9000000000	400	355	6	5	3

Gambar 3. Dataset yang Digunakan Dalam Perancangan Model

Data yang sudah dipersiapkan kemudian akan dibagi menjadi data latih dan juga data uji. Pada penelitian ini data latih akan divariasikan jumlahnya yaitu sebanyak 70%, 80% dan 90%. Hasil yang didapatkan dari algoritma Random Forest Regression adalah sebagai berikut

Tabel 1. Performa Algoritma Random Forest Regression

Jumlah Dataset	MSE	MAE	R2 SCORE
70%	24119623.92	1979.89	0.683
80%	24262964.70	1825.28	0.667
90%	1184164.61	1708.59	0.83

Dengan konfigurasi yang sama pada algoritma Linear Regression maka diperoleh hasil sebagai berikut:

Tabel 2. Performa Algoritma Linear Regression

Jumlah Dataset	MSE	MAE	R2 SCORE
70%	23723834.90	2102.96	0.69
80%	26499374.32	2103.67	0.64
90%	14911490.56	2091.32	0.75

Untuk mendapatkan performa terbaik dari kedua algoritma tersebut, maka langkah selanjutnya adalah melakukan optimalisasi menggunakan metode Grid Search. Berikut merupakan konfigurasi pada Grid Search pada kedua model:

Tabel 3. Parameter yang Digunakan pada Grid Search

Algoritma	Parameter 1	Parameter 2	Parameter 3	Parameter 4
Random Forest Regression	n_estimators = [100, 200, 300]	max_depth = [5, 10, 15]	min_samples_split = [2,5,10]	min_samples_leaf = [1,2,3,4]
Linear Regression	positive = [True, False]	fit_intercept = [True, False]	copy_X = [True, False]	-

Tabel 4. Nilai parameter hasil optimisasi Grid Search pada Algoritma Random Forest Regression

Jumlah Dataset	Max_depth	Min_samples_leaf	Min_sample_leaf	N_estimators
70%	15	3	2	300
80%	20	2	2	200
90%	20	3	5	300

Tabel 5. Performa Algoritma Random Forest Regression hasil optimisasi Grid Search

Jumlah Dataset	MSE	MAE	R2 SCORE
70%	25280743.26	2029.53	0.67
80%	23733368.41	1791.13	0.67
90%	10700807.45	1795.69	0.84

Tabel 6. Nilai Parameter hasil Optimisasi Grid Search pada Algoritma Linear Regression

Jumlah Dataset	Positive	Fit_intercept	Copy_X
70%	False	False	True
80%	False	True	True
90%	False	False	True

Tabel 7. Performa Algoritma Linear Regression Hasil Optimisasi Grid Search

Jumlah Dataset	MSE	MAE	R2 SCORE
70%	15039552.95	2143.05	0.75

Jumlah Dataset	MSE	MAE	R2 SCORE
80%	26499374.32	2103.67	0.63
90%	15039552.95	2143.05	0.75

Berdasarkan hasil percobaan, baik sebelum dan sesudah dilakukan optimisasi dengan metode grid search diperoleh hasil bahwa Random Forest Regression memiliki rata-rata performa yang lebih baik dibandingkan dengan Linear Regression terutama pada data latih diatas dari 70%. Ketika data training dibuat menjadi sebanyak 70% dan dengan teknik optimisasi grid search, algoritma Linear Regression menunjukkan performa yang jauh lebih baik dibandingkan dengan algoritma Random Forest Regression terutama pada nilai MSE dan R2 Score. Hal ini membuktikan bahwa meskipun algoritma Random Forest Regression membentuk n estimators pohon namun tidak menjamin memiliki performa yang selalu lebih baik dibandingkan dengan algoritma Linear Regression. Berdasarkan data tabel 5, model terbaik yang diperoleh adalah model yang menggunakan Algoritma Random Forest Regression dengan jumlah dataset sebanyak 90%, dan dengan parameter yang sesuai pada tabel 4.

3.2 Integrasi Model dengan User Interface

Pada penelitian ini, peneliti menggunakan library streamlit yang digunakan dalam pembuatan antarmuka sistem sehingga dapat mempermudah pengguna dalam melakukan prediksi harga rumah. Library streamlit dipilih karena penggunaannya yang mudah sehingga mempercepat proses deployment sistem yang dirancang. Dalam proses integrasi antara model dengan User Interface diperlukan sebuah tes yang dinamakan dengan integration testing. Berikut merupakan integration testing yang sudah dilakukan pada sistem prediksi harga rumah yang sudah dirancang.

Tabel 8. ...

No	Integration Testing	Status
1.	Pengujian pada text input	Succeded
2.	Pengujian tombol tambah pada number input	Succeded
3.	Pengujian pada tombol submit	Succeded
4.	Sistem menampilkan hasil prediksi berdasarkan model	Succeded



Gambar 4. Tampilan Antarmuka Sistem yang Dirancang

4. Kesimpulan

Berdasarkan percobaan yang dilakukan pada algoritma Linear Regression dengan Random Forest Regression diperoleh hasil akhir bahwa algoritma Random Forest Regression memiliki performa yang lebih baik dibandingkan dengan algoritma Linear Regression. Terutama pada persentase jumlah data train sebanyak 90% dan sudah dilakukan optimisasi dengan teknik Grid Search. Dimana pada persentase jumlah data train tersebut Algoritma Random Forest lebih baik hampir 10% jika dibandingkan dengan algoritma Linear Regression. Hal ini disebabkan pada algoritma Random Forest menggunakan lebih dari satu model pohon ($n_estimators$) yang kemudian dicari nilai tengah atau rata-rata dari masing-masing model pohon yang sudah dibangun. Namun pada persentase data train 70% dan 80% baik Random Forest Regression maupun Linear Regression tidak menunjukkan perbedaan yang signifikan. Oleh karenanya jumlah dari data train berperan penting dalam perancangan model ini.

Daftar Pustaka

- [1] G. N. A. d. D. Fitriah, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *J. Telemat*, vol. 14, p. 79–86, 2019.
- [2] F. N. R. F. N. R. Evi Febrion Rahayuningtyas, "Prediksi Harga Rumah Menggunakan General Regression Neural Network," *JURNAL INFORMATIKA*, vol. 8, pp. 59-66, 2021.
- [3] S. Pratama, "PREDIKSI HARGA TANAH MENGGUNAKAN ALGORITMA LINEAR REGRESSION," *Technologia*, vol. 7, 2016.
- [4] T. P. J. A. Yani, "Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest," *J. Tek. Elektro*, vol. 9, p. 24–29, 2017.
- [5] "S. Dutalia, A. K. Lalo, P. Batarius, Y. Carmeneja, H. Siki," *Implementasi Algoritma C4 . 5 Untuk Klasifikasi Penjualan*, vol. 06, p. 1–12, 2021.
- [6] F. R. Setiawan, "Prediksi Pergerakan Harga Saham dengan Metode Support Vector Machine (SVM) Menggunakan Trend Deterministic Data Preparation(Studi Kasus Saham Perusahaan PT Astra International Tbk, PT Garuda Indonesia Tbk, dan PT Indosat Tbk)," *e-Proceeding of Engineering*, vol. 5, p. 8356, 2018.
- [7] T. W. P. I. Heru Wahyu, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 4, p. 364, 2019.
- [8] I. K. G. S. Indra Permana Putra, "Perbandingan Akurasi Algoritma Regresi Linier, Regresi Polinomial, dan Support Vector Regression Pada Model Sistem Prediksi Harga Rumah," *JNATIA*, vol. 1, 2022.
- [9] T. K. a. R. Nindyasari, "Forecasting Dengan Metode Regresi Linier Pada Sistem Penunjang Keputusan Untuk Memprediksi Jumlah Penjualan Batik (Studi Kasus Kub Sarwo Endah Batik Tulis Lasem)," *J. Mantik Penusa*, vol. 1, p. 71–92, 2017.
- [10] S. H. F. R. T. V. Heri Setyawan, "Prediksi Tingkat Produksi Buah Kelapa Sawit dengan Metode Single Moving Average," *J.TIKomSiN*, vol. 9, pp. 1-10, 2021.

Perancangan Ontologi Semantik : Representasi Digital Oleh - Oleh Khas Bali

Ngakan Made Alit Wiradhanta^{a1}, Ida Bagus Made Mahendra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹alitwiradhanta08@gmail.com
²ibm.mahendra@unud.ac.id

Abstract

Bali Island, also known as the Island of the Gods, is a popular tourist destination for those who want to enjoy its breathtaking natural beauty. However, many tourists face difficulties in finding authentic Balinese souvenirs to bring back home. To address this issue, a system is needed to assist tourists in easily and quickly finding information about unique Balinese souvenirs using ontology, a semantic web knowledge base. The semantic web is a website development approach aimed at automation and understanding the meaning and relationships of data, resulting in a web experience tailored to visitors' preferences. This research aims to develop an ontology model to be used in the domain of Balinese souvenirs. The ontology design is developed using the Protégé application, where the ontology model is structured with hierarchical classes, slots, properties, and other elements. It is expected that this ontology model will provide information related to semantic web-based application metadata pertaining to Balinese souvenirs.

Keywords: Ontologi, Oleh - Oleh Khas Bali, Web Semantik.

1. Pendahuluan

Dalam era digital yang semakin maju, teknologi informasi dan internet telah mengubah cara kita mengakses dan berinteraksi dengan informasi. Pariwisata menjadi salah satu sektor yang terpengaruh oleh perubahan ini. Para wisatawan saat ini bergantung pada platform digital untuk mencari informasi tentang destinasi wisata yang mereka kunjungi, termasuk oleh-oleh khas yang menjadi salah satu aspek penting dalam pengalaman wisata.

Bali, dengan keindahan alamnya yang memesona dan kekayaan budaya yang unik, merupakan destinasi wisata yang sangat populer. Pulau ini terkenal dengan oleh-oleh khasnya yang beragam, seperti kerajinan tangan, pakaian tradisional, lukisan, perhiasan, dan makanan khas. Oleh-oleh ini tidak hanya menjadi kenang-kenangan bagi para wisatawan, tetapi juga menjadi lambang dari budaya dan identitas Bali.

Namun, seringkali para wisatawan menghadapi kesulitan dalam mencari informasi yang lengkap dan dapat dipercaya tentang oleh-oleh khas Bali. Informasi yang ada terkadang tersebar di berbagai sumber yang tidak terstruktur, dan sulit untuk menemukan informasi yang relevan dan terbaru. Untuk mengatasi masalah ini, perancangan ontologi semantik menjadi penting dalam membentuk representasi digital yang sistematis dan konsisten tentang oleh-oleh khas Bali.

Ontologi semantik adalah struktur pengetahuan yang memungkinkan pengorganisasian dan pemahaman data secara lebih baik. Dalam perancangan ontologi semantik untuk oleh-oleh khas Bali, akan dilakukan identifikasi dan pengklasifikasian berbagai jenis oleh-oleh, termasuk atribut-atribut yang relevan seperti bahan, teknik pembuatan, asal-usul, dan nilai budaya. Dengan ontologi semantik yang dirancang dengan baik, para wisatawan akan dapat dengan mudah menemukan informasi yang mereka cari, melihat hubungan antarobyek, dan memperoleh wawasan yang lebih dalam tentang oleh-oleh khas Bali.

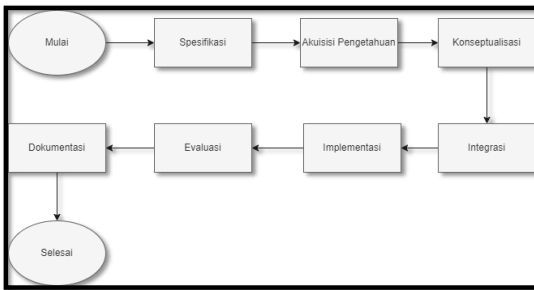
Penelitian ini akan menggunakan aplikasi Protégé, yang merupakan alat yang handal untuk membangun ontologi. Dalam proses perancangan, akan dibangun struktur hierarkis kelas, slot, properti, dan elemen lainnya untuk mencerminkan kompleksitas dan relasi antarobyek oleh-oleh khas Bali. Model ontologi ini akan diimplementasikan dalam sistem yang dapat diakses oleh para wisatawan melalui platform digital.

Dengan adanya ontologi semantik yang representatif dan komprehensif tentang oleh-oleh khas Bali, diharapkan para wisatawan akan memiliki akses yang lebih baik dan pengalaman yang lebih memuaskan dalam mencari, memahami, dan memilih oleh-oleh yang sesuai dengan minat dan preferensi mereka. Selain itu, ontologi semantik ini juga akan menjadi sumber informasi yang berharga bagi peneliti, pelaku industri, dan pecinta budaya Bali dalam mempelajari dan mempromosikan warisan budaya pulau ini.

Secara kesimpulan, perancangan ontologi semantik untuk representasi digital oleh-oleh khas Bali merupakan langkah penting dalam mengatasi tantangan dalam mencari informasi tentang oleh-oleh khas secara digital. Dengan memanfaatkan teknologi web semantik, ontologi ini akan memfasilitasi penemuan, pemahaman, dan apresiasi terhadap kekayaan budaya Bali melalui pengalaman wisata yang lebih mendalam dan berarti.

2. Metodologi Penelitian

Penelitian ini dimulai dengan melakukan tinjauan pustaka dan merumuskan permasalahan yang akan mendukung pelaksanaan penelitian. Setelah merumuskan permasalahan, langkah selanjutnya adalah mencari solusi dan merancang antarmuka pengguna dari sistem berbasis mobile yang akan diuji. Desain yang telah selesai dirancang akan diuji menggunakan metode System Usability Scale (SUS). Metode ini melibatkan distribusi kuesioner kepada sejumlah responden, dan setelah mencapai target yang ditentukan, akan dilakukan perhitungan nilai SUS. Selanjutnya, dilakukan perbaikan desain dengan mengkaji kembali tinjauan pustaka dan perumusan masalah yang telah dibuat. Apabila desain yang telah diperbaiki sesuai dengan harapan, maka akan diberikan hasil dan saran terkait penelitian ini.



Gambar 1. Alur Metode Penelitian

2.1. Ontologi

Ontologi merupakan bentuk representasi resmi mengenai konsep, entitas, hubungan, dan properti dalam suatu bidang pengetahuan. Tujuannya adalah untuk mengatur pengetahuan dengan cara yang terstruktur dan memiliki arti sehingga komputer dapat memahami dan memproses informasi dengan lebih efisien. Ontologi menjadi landasan dalam pengembangan sistem cerdas, penambahan data, dan integrasi informasi. Dengan menggunakan ontologi,

sistem komputer mampu menafsirkan informasi, mengenali hubungan antara entitas, dan menyediakan konteks yang kaya dalam analisis dan pengambilan keputusan.

2.2. Semantik Web

Web semantik merupakan sebuah pendekatan yang digunakan untuk meningkatkan pemahaman dan pengolahan informasi di dalam lingkungan web. Web semantik memiliki peran yang krusial. Web semantik memungkinkan penyusunan ontologi yang terstruktur dengan baik, memahami makna dan hubungan antar konsep, serta mengorganisir informasi secara sistematis. Dengan menerapkan prinsip-prinsip web semantik, ontologi semantik yang dirancang akan memiliki kemampuan untuk mengorganisir dan menyajikan informasi yang terperinci. Web semantik memfasilitasi penautan data yang terstruktur dan bermakna, sehingga memungkinkan komputer untuk memahami konteks dan hubungan antar informasi secara lebih efisien.

2.3. Methontology

Methontology merupakan sebuah pendekatan yang sangat relevan dalam proses perancangan ontologi semantik. Pendekatan ini menawarkan langkah-langkah yang terstruktur dan sistematis untuk merancang ontologi dengan efisiensi yang tinggi. Salah satu komponen penting dari Methontology adalah kemampuannya untuk menentukan kebutuhan informasi yang spesifik, mengidentifikasi konsep dan hubungan yang relevan, serta merancang struktur ontologi yang konsisten dan akurat. Melalui penerapan Methontology, ontologi semantik yang dihasilkan dapat memiliki representasi yang bermakna dan konsisten dalam domain pengetahuan yang dituju. Dalam perancangan ontologi semantik, Methontology menjadi alat yang sangat berharga karena memungkinkan pengelolaan pengembangan ontologi dengan pendekatan yang terstruktur dan berbasis pada prinsip-prinsip formal. Metodologi ini memastikan bahwa ontologi yang dibangun mematuhi standar yang baik dan mampu memenuhi kebutuhan informasi yang spesifik yang diinginkan oleh pengguna.

2.4. SPARQL

SPARQL (SPARQL Protocol and RDF Query Language) adalah sebuah bahasa query yang memiliki fungsi utama untuk melakukan ekstraksi dan manipulasi data yang disimpan dalam format RDF (Resource Description Framework). Dengan menggunakan SPARQL, pengguna dapat mengakses dan memanipulasi informasi yang terdapat dalam ontologi secara efektif. Dalam konteks perancangan ontologi semantik, SPARQL menjadi alat yang sangat penting dalam memperoleh data yang relevan dari ontologi yang telah dibangun. Melalui penggunaan sintaksis SPARQL, pengguna dapat menentukan pola pencarian yang spesifik untuk mengekstrak informasi yang diinginkan dari ontologi. SPARQL juga menyediakan berbagai fitur dan fungsionalitas yang berguna, seperti penggabungan data, pengurutan, pembatasan hasil, dan agregasi, yang memungkinkan untuk melakukan analisis yang lebih kompleks terhadap data semantik. Dengan bantuan SPARQL, pengguna dapat melakukan query terhadap ontologi semantik untuk menjelajahi dan mendapatkan informasi yang terkait dengan konsep, entitas, relasi, dan properti yang ada dalam ontologi. Hasil dari query SPARQL dapat digunakan untuk mendukung pengambilan keputusan yang berdasarkan data, memahami keterkaitan antar konsep, serta mendapatkan wawasan yang lebih dalam tentang data semantik yang terdapat dalam ontologi.

3. Hasil dan Pembahasan

Bagian ini mencakup temuan dan analisis dari penelitian, yang bisa disampaikan dalam bentuk deskripsi yang rinci atau diilustrasikan melalui bagan atau gambar.

3.1. Spesifikasi

- a. Domain : Oleh – Oleh khas Bali
- b. Tanggal :
- c. Dirancang oleh : Ngakan Made Alit Wiradhanta

- d. Diimplementasikan oleh : Ngakan Made Alit Wiradhanta
- e. Level Formalitas : Formal
- f. Ruang Lingkup : Oleh – Oleh khas Bali
- g. Sumber Pengetahuan : Internet, Jurnal

3.2. Akuisisi Pengetahuan

Proses akuisisi pengetahuan melibatkan pengumpulan pengetahuan dari satu atau beberapa sumber. Dalam penelitian ini, tahap akuisisi pengetahuan terdiri dari:

- a. Melakukan identifikasi pengetahuan yang diperoleh melalui jurnal , sumber internet terpercaya.
- b. Melakukan studi literatur melalui situs web resmi pemilik usaha oleh-oleh di Bali.

3.3. Konseptualisasi

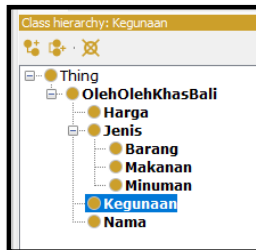
Konsep dari ontologi mempunyai tujuan untuk mengatur dan mengelola sebuah pengetahuan yang didapat selama proses akuisisi. Setelah sebuah model konseptual dirancang, metodologi yang diusulkan hanya berfokus pada transformasi model konseptual menjadi sebuah model formal yang dapat diimplementasikan kedalam bahasa ontologi yang sudah ditentukan sebelumnya.

3.4. Integrasi

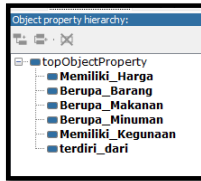
Tahap integrasi adalah tahap dimana dilakukan penggabungan terhadap sebuah ontologi yang dirancang sebelumnya sehingga dapat disesuaikan dengan domain yang diteliti yaitu Oleh-Oleh Khas Bali .Dengan pemilihan ontologi yang sesuai dengan desain yang telah dirancang dapat mempermudah dalam mencapai hasil yang diharapkan. Dengan menggunakan ontologi yang tepat, dapat meningkatkan efisiensi dan akurasi dalam pengelolaan dan pemanfaatan pengetahuan yang telah diperoleh.

3.5. Implementasi

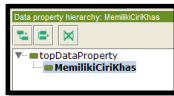
Tahap konseptualisasi memiliki tujuan untuk mengorganisir pengetahuan dalam domain menjadi bentuk konseptual serta menjaga dan mengelola pengetahuan yang diperoleh selama proses akuisisi pengetahuan. Setelah model konseptual telah dibangun, metode ini akan berubah untuk mentransformasikan model konseptual tersebut menjadi model formal yang akan diimplementasikan dalam bahasa ontologi. Ontologi ini akan dikembangkan untuk domain Busana Tradisional dan akan diatur dalam bentuk kelas dan sub-kelas seperti yang ditunjukkan pada gambar di bawah ini



Gambar 2. Class dari Ontologi Oleh-Oleh Khas Bali



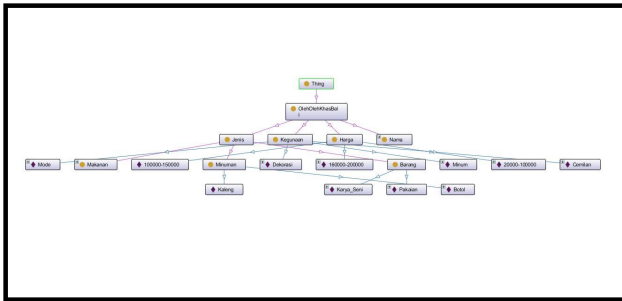
Gambar 3. Object Property dari Ontologi Oleh-Oleh Khas Bali



Gambar 4. Data Property dari Ontologi Oleh-Oleh Khas Bali



Gambar 5. Individu dari Ontologi Oleh-Oleh Khas Bali



Gambar 6. Ontograf dari Ontologi Oleh-Oleh Khas Bali

3.6. Evaluasi

Tahap evaluasi dilakukan dengan pengujian menggunakan SPARQL yang terdapat dalam aplikasi Protégé 4.2. Serta pertanyaan yang akan diajukan dapat diubah menjadi query SPARQL, maka dapat menampilkan hasil ontologi yang telah dibuat.

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX oleholeh: <http://www.semanticweb.org/alitw/ontologies/2023/5/untitled-ontology-18#>

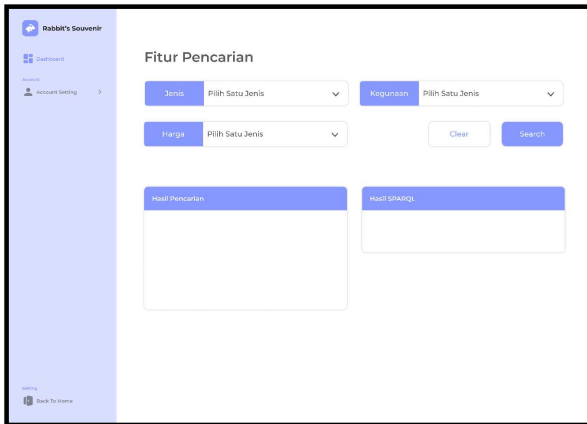
SELECT * WHERE {
  ?OlehOlehKhasBali oleholeh:Berupa_Barang oleholeh:Pakaian.
  ?OlehOlehKhasBali oleholeh:Memiliki_Harga oleholeh:20000-100000
}
```

OlehOlehKhasBali
Udeng
Baju_Barong

Gambar 7. Hasil Query dari Ontologi Oleh-Oleh Khas Bali

3.7. Dokumentasi

Hasil dokumentasi dari penelitian ontologi dari ontologi oleh-oleh khas Bali ini berupa prototype dalam laporan ini



Gambar 8. Prototype dari Ontologi Oleh-Oleh Khas Bali

4. Kesimpulan

Setelah melakukan analisis dan diskusi, kami berhasil menciptakan sebuah ontologi yang berkaitan dengan domain Oleh-oleh Khas Bali. Pengembangan ontologi ini dilakukan menggunakan aplikasi Protégé 4.2 dengan menerapkan metode Methontology. Ontologi ini terdiri dari 8 kelas, 6 properti objek, 1 properti data, dan 22 individu yang merepresentasikan entitas dan hubungan dalam domain Oleh-oleh Khas Bali.

Selanjutnya, kami melakukan evaluasi dan pengujian terhadap model ontologi yang diusulkan. Kami menggunakan serangkaian pertanyaan yang sering diajukan oleh pengguna untuk mencari informasi tentang Oleh-oleh Khas Bali. Dengan melakukan pengujian ini, kami bertujuan untuk memastikan bahwa ontologi yang telah dibuat dapat memberikan informasi yang relevan dan berguna kepada pengguna terkait Oleh-oleh Khas Bali di Indonesia.

Referensi

- [1]. Fajri, R.N., Manurung, R., & Adriani, M. (2015). Ontologi Web Semantik untuk Penyediaan Informasi Pariwisata di Indonesia. *Jurnal Ilmu Komputer dan Informasi*, 8(1), 1-8. .
- [2]. Fajri, R.N. (2016). Pengembangan Ontologi Web Semantik untuk Penyediaan Informasi Pariwisata di Indonesia. Tesis. Universitas Indonesia.
- [3]. Badron, Y.F., Agus, F., & Hatta, H.R. (2017). Studi Tentang Pemodelan Ontologi Web Semantik dan Prospek Penerapan pada Bibliografi Artikel Jurnal Ilmiah. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 2(1), 1-6.
- [4]. Satria, A., Herdiani, A., & Effendy, V. (2016). Analisis Keterhubungan Ontology Pada Web Semantik Menggunakan Semantic-based Ontology Matching. *eProceedings of Engineering*, 3(3), 3241-3248.
- [5]. Pratiwi, D. (2017). Pengembangan Ontologi Web Semantik pada Domain Pengetahuan Kesehatan. Tesis. Universitas Gadjah Mada.

Halaman ini sengaja dibiarkan kosong

Klasifikasi Genre Musik Menggunakan Support Vector Machine Berdasarkan Spectral Features

I Gusti Agung Ngurah Diputra Wiraguna^{a1}, Luh Arida Ayu Rahning Putri^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹ngurahdiputra075@student.unud.ac.id
rahningputri@unud.ac.id

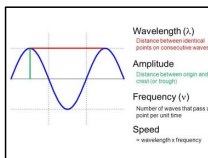
Abstract

This research focuses on music genre classification based on spectral features and Support Vector Machine (SVM). Features such as Spectral Centroid, Spectral Rolloff, Spectral Flux, and Spectral Bandwidth are extracted from MP3 music audio. The dataset comprising 4 music genres is utilized for training and testing the system. The extracted spectral features are fed into the SVM classifier to predict the genre of test samples. Python and machine learning are both used in developing the system while the experimental results demonstrate the effectiveness of SVM in accurately classifying music genres based on the current extracted features. The proposed approach contributes to automated music genre classification systems, facilitating music organization, recommendation, and retrieval. This research promotes advancements in music information retrieval and enhances user experience in music-related applications.

Keywords: Music Feature Extraction, MP3, Music, Spectral Features, SVM

1. Pendahuluan

Musik merupakan salah satu bentuk seni yang melibatkan indra pendengaran manusia. Konsep musik sama halnya dengan gelombang bunyi dimana terjadi perubahan tekanan udara dan gelombang akan terus bertransmisi hingga mencapai telinga manusia. Gelombang bunyi tergolong gelombang longitudinal sehingga sudut antara arah getaran dan arah rambatan gelombangnya adalah nol. Selama perambatannya gelombang longitudinal akan membentuk pola rapatan dan regangan yang menjadi pemklasifikasian satu gelombang longitudinal. Gelombang bunyi juga memiliki komponen gelombang pada umumnya seperti amplitudo, frekuensi, periode, fase, panjang gelombang, dan cepat rambat gelombang [11].



Gambar 1. Gelombang Bunyi dan Komponenya

Perbedaan antara musik dengan gelombang bunyi pada umumnya terletak pada karakteristik gelombang tersebut. Dalam musik dikenal beberapa istilah seperti melodi, ritme, harmoni, dan lainnya yang memiliki karakteristik berbeda dengan gelombang bunyi pada umumnya sehingga terkesan enak didengar. Karakteristik setiap gelombang bunyi berbeda beda bergantung pada frekuensi gelombang tersebut. Dalam mendengarkan musik sering disadari bahwa musik keluaran sekarang terdengar berbeda dengan musik beberapa tahun lalu. Bahkan bisa saja

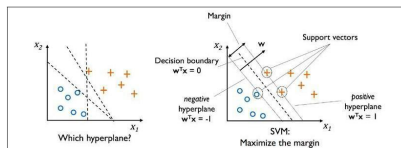
musik - musik yang dirilis pada tahun yang sama terdengar berbeda. Perbedaan dari musik tersebut terletak dari gaya pembawaan musik. Perbedaan gaya musik biasa disebut sebagai genre musik. Setiap genre musik berbeda - beda sesuai dengan musisinya. Contoh musik klasik yang memiliki gaya musik yang tenang dan volume suara tergolong rendah. Sedangkan musik rock memiliki volume suara yang keras dan tempo yang lebih cepat. Semakin berkembangnya teknologi, musik tidak harus didengarkan secara langsung dari sumber suara namun juga dapat didengarkan jarak jauh. Musik dapat disimpan dalam sebuah file dengan berbagai jenis format seperti WAV dan MP3 yang masing - masing memiliki kelebihan tersendiri. Terdapat banyak platform musik yang memungkinkan kita untuk mendengarkan musik secara gratis bahkan mengunduh file musik tersebut untuk disimpan dan didengarkan kapan saja kita mau. Sebuah platform musik digital biasanya memberi pengguna fitur untuk mencari jenis musik yang mereka inginkan. Selain itu juga terdapat fitur rekomendasi musik yang menampilkan beberapa musik yang sekiranya mungkin dinikmati oleh pengguna. Sistem rekomendasi membagi musik ke dalam beberapa kategori salah satunya genre. Platform tersebut mampu memberi rekomendasi dengan genre yang sesuai karena sistem mencatat riwayat pengguna seperti musik apa yang dicari dan musik apa yang didengarkan. Setelah mengakses musik pada riwayat pengguna sistem pada platform tersebut akan mengklasifikasikan musik berdasarkan metadata musik sistem. Sebuah sistem rekomendasi berbasis kecerdasan buatan dapat dijabarkan berdasarkan sumber pengetahuannya. Beberapa sumber tersebut meliputi social, individual, dan content. Social berupa sumber secara umum, individual berupa sumber yang lebih spesifik ke tiap perorangan, dan content merupakan sumber yang berdasarkan musik itu sendiri mulai dari fitur yang sederhana hingga yang cukup rumit [6]. Dalam mengklasifikasikan musik berdasarkan genre maupun aspek lainnya. Perlu dilakukan ekstraksi fitur dari musik tersebut. Fitur sama dengan karakteristik yang dibahas sebelumnya dan membuat setiap musik/suara memiliki perbedaan masing - masing. Salah satu ekstraksi fitur pada musik adalah fitur frequency domain. Ekstraksi fitur dari sebuah musik penting dilakukan untuk mengubah nilai musik tersebut kedalam bentuk numerik karena komputer hanya dapat memproses input bentuk numerik. Selain itu ekstraksi fitur juga berfungsi sebagai ruang fitur dalam metode klasifikasi yang akan menjadi pembeda setiap kelas.

1.1 SVM (Support Vector Machine)

Salah satu metode klasifikasi yang dapat digunakan adalah Support Vector Machine. Support Vector Machine merupakan sebuah metode klasifikasi yang merupakan penjabaran dari PLA (Perceptron Learning Algorithm). Jika pada PLA kita dapat mengklasifikasikan data dalam bentuk linear regresi, maka SVM akan memaksimalkan margin antara kelas data. Margin pada SVM merujuk pada jarak antara hyperplane pembagi dengan setiap sampel data latih pada tiap kelas data dengan jarak terdekat ke hyperplane [10]. Margin dapat dibedakan menjadi positif dan negatif berdasarkan letak kelas pada hyperplane yang dapat dirumuskan sebagai berikut [10].

$$w\theta + wt x_{\text{positive}} = 1 \tag{1}$$

$$w\theta + wt x_{\text{negative}} = -1 \tag{2}$$



Gambar 2. Hyperplane pada 2 Kelas dan Hyperplane Optimal yang Diperoleh [10]

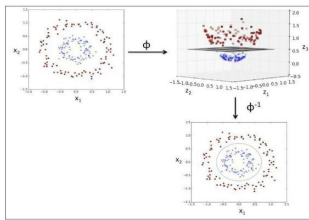
Support Vector Machine dibagi menjadi beberapa bentuk berupa jenis kernel yang digunakan. Kernel tersebut antara lain linear, RBF (Radial Basis Function), Sigmoid, dan Polinomial.

Penggunaan kernel pada SVM berfungsi dalam klasifikasi data dengan bentuk non-linier. Data non-linier yang dimaksud berupa kelas data yang memiliki banyak dimensi nilai sehingga sulit direpresentasikan secara linier. Penggunaan kernel dalam menangani kelas non-linier yaitu membuat kombinasi non-linier dari fitur asal untuk diproyeksikan ke dalam dimensi yang lebih tinggi melalui mapping (ϕ) [10]. Sebagai contoh kelas data berdimensi 2 akan dibentuk menjadi dimensi 3 yang nantinya dapat dipisahkan secara linear dengan rumus berikut [10].

$$\phi(x_1, x_2) = (z_1, z_2, z_3) = (x_1, x_2, x_1^2 + x_2^2) \tag{3}$$

Dimana:

- a. x_1, x_2 = fitur awal pada dimensi 2
- b. z_1, z_2, z_3 = fitur hasil konversi ke bentuk dimensi 3

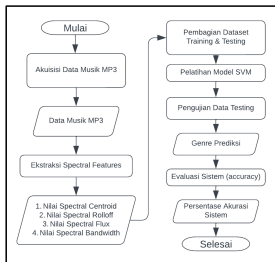


Gambar 3. SVM dengan Fitur Kelas 2 dan 3 Dimensi [10]

Setiap file musik tentu memiliki nilai fitur yang berbeda - beda. Dengan dilakukannya klasifikasi sebuah musik dapat dikategorikan ke dalam label tertentu dan dapat dijadikan sebagai acuan dalam mengklasifikasikan musik lainnya. Dengan dilakukannya penelitian ini diharapkan mampu mengklasifikasikan musik berdasarkan genre sekaligus menguji metode SVM dalam mengklasifikasikan genre musik sesuai data yang diberikan.

2. Metode Penelitian

Terdapat beberapa langkah dalam sistem mulai dari persiapan data hingga proses klasifikasi dan evaluasi.



Gambar 4. Flowchart Sistem

2.1 Akuisisi Data

Data yang digunakan berupa data sekunder yang diperoleh dari website www3.projects.science.uu.nl yang merujuk pada penelitian "Studying emotion induced by music through a crowdsourcing game" [3]. Data musik berformat MP3 yang merupakan data musik terkompresi. Data disimpan pada folder setiap genre berbeda yang dijadikan satu directory berupa folder "source".

Tabel 1. Jumlah dan Pembagian Data Musik

Genre	Jumlah
Classical	100
Electronics	100
Pop	100
Rock	100

Audio selanjutnya akan melalui Python modules librosa berupa "librosa.load()". Beberapa variabel seperti y dan sr akan muncul dimana y merupakan bentuk denotasi sinyal audio (data musik) berbentuk array numpy dan sr merupakan sampling rate yang melambangkan frekuensi dari y [8]. Penggunaan modul ini berfungsi untuk memperoleh nilai amplitudo (y) dan sampling rate (sr) dari setiap file musik yang sudah diakuisisi.

2.2 Ekstraksi Fitur Musik

Ekstraksi fitur yang dilakukan pada data musik berfokus pada fitur frekuensi (spectral features). Proses ekstraksi menggunakan bantuan modules Python yaitu librosa dan numpy. Beberapa fitur yang diekstrak antara lain spectral centroid, spectral rolloff, spectral flux, dan spectral bandwidth. Karena setiap data musik memiliki durasi yang berbeda maka nilai dari spectral features dapat memiliki jumlah yang berbeda saat proses ekstraksi. Sehingga digunakan fungsi "mean. ()" untuk mencari rata - rata dari setiap nilai spectral features tersebut.

Tabel 2. Contoh Hasil Ekstraksi Spectral Features

File	Label	Centroid	Rolloff	Flux	Bandwidth
1.mp3	Classical	1308.785	2067.708	0.953	1428.417
2.mp3	Classical	667.434	865.712	0.861	906.712
3.mp3	Electronic	2385.439	4882.837	1.194	2528.466
4.mp3	Electronic	1242.056	2518.873	1.408	1622.783
5.mp3	Pop	2516.364	5698.397	1.342	2766.53
6.mp3	Pop	2590.617	5800.017	1.273	2737.064
7.mp3	Rock	1566.794	3348.488	1.507	2086.703
8.mp3	Rock	1875.557	4144.43	1.16	2153.481

2.3 Pembagian Dataset Training dan Data Testing

Data musik pada folder setiap genre kemudian akan dipilih secara acak untuk dijadikan data uji dan data latih. Pembagian data dibantu oleh Python modules berupa "sklearn.model_selection" dengan ukuran data tes sebesar 0,2. Hal ini berarti sebanyak 20% dari data akan dialokasikan sebagai data uji dan 80% sisanya sebagai data latih. Data uji dipilih secara acak dari setiap genre [9].

2.4 Pelatihan Model SVM

Pelatihan model SVM menggunakan Python modules "sklearn.svm" untuk memanggil modul "SVC". Secara default jenis SVM yang digunakan adalah kernel RBF. Pada dasarnya rumus perhitungan kernel RBF berupa [12].

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4)$$

Nilai γ secara default dapat didefinisikan:

$$\gamma = \frac{1}{n \text{ features} * \sigma^2} \quad (5)$$

Dimana:

- $\|x - x'\|^2$ menyatakan pangkat dari jarak Euclidean data antar fitur.
- γ menyatakan skalar pengaruh setiap data kelas dalam mencapai decision boundary model.

Setelah pemanggilan model SVM perlu digunakan fungsi "fit. ()" untuk memasukkan data latih yang sudah dibagi sebelumnya ke dalam model SVM. terdapat sebanyak 320 sampe musik sebagai data latih yang terpilih secara acak.

2.5 Pengujian

Pengujian sistem dilakukan dengan menguji data uji dengan model SVM yang sudah dilatih. Sebanyak 80 sampel musik acak dijadikan data uji saat pembagian data latih dengan data uji. Pengujian data testing menggunakan fungsi "predictions ()" pada SVC. Selain pengujian menggunakan data uji juga dilakukan evaluasi berupa akurasi dari sistem dalam menguji data uji. Hasil dari perhitungan akurasi nantinya akan dikonversikan kedalam bentuk persen.

$$\text{accuracy} = \frac{\sum \text{TP}}{\sum \text{Prediction}} \times 100\% \quad (6)$$

Dimana:

- TP (True Positive) merupakan berapa banyak prediksi genre yang sesuai dengan genre asalnya.
- Prediction merupakan jumlah total pengujian.

3. Hasil dan Diskusi

Hasil Pengujian berupa hasil data uji dan evaluasi sistem berupa perhitungan akurasi.

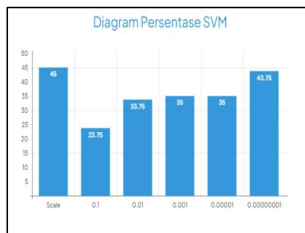
Tabel 3. Perbandingan Hasil Prediksi Data Uji

File (MP3)	Label	Prediksi	File (MP3)	Label	Prediksi
1.mp3	pop	classical	45.mp3	rock	pop
10.mp3	pop	classical	46.mp3	classical	classical
100.mp3	classical	pop	47.mp3	classical	rock
11.mp3	pop	rock	48.mp3	classical	pop
12.mp3	classical	classical	49.mp3	classical	pop
13.mp3	classical	rock	5.mp3	rock	rock
14.mp3	rock	rock	50.mp3	electronic	classical
15.mp3	classical	classical	51.mp3	electronic	electronic
16.mp3	pop	rock	52.mp3	classical	pop
17.mp3	electronic	pop	53.mp3	electronic	pop
18.mp3	classical	classical	54.mp3	classical	pop
19.mp3	rock	rock	55.mp3	rock	rock
2.mp3	classical	classical	56.mp3	rock	pop
20.mp3	classical	pop	57.mp3	classical	pop
21.mp3	electronic	classical	58.mp3	rock	rock
22.mp3	classical	classical	59.mp3	classical	rock
23.mp3	pop	pop	6.mp3	pop	pop
24.mp3	rock	rock	60.mp3	electronic	classical
25.mp3	pop	rock	61.mp3	classical	classical
26.mp3	rock	rock	62.mp3	electronic	pop
27.mp3	classical	classical	63.mp3	electronic	rock
28.mp3	classical	classical	64.mp3	pop	rock
29.mp3	rock	rock	65.mp3	pop	pop

3.mp3	pop	pop	66.mp3	classical	classical
30.mp3	classical	classical	67.mp3	electronic	classical
31.mp3	rock	rock	68.mp3	classical	pop
32.mp3	electronic	rock	69.mp3	rock	rock
33.mp3	pop	rock	7.mp3	pop	classical
34.mp3	pop	pop	70.mp3	rock	electronic
35.mp3	electronic	rock	71.mp3	classical	classical
36.mp3	rock	rock	72.mp3	rock	rock
37.mp3	electronic	rock	73.mp3	electronic	classical
38.mp3	pop	pop	74.mp3	classical	classical
39.mp3	classical	classical	75.mp3	classical	classical
4.mp3	pop	rock	76.mp3	rock	rock
40.mp3	electronic	rock	77.mp3	pop	rock
41.mp3	electronic	pop	78.mp3	pop	classical
42.mp3	rock	electronic	79.mp3	rock	rock
43.mp3	electronic	classical	8.mp3	electronic	pop
44.mp3	electronic	classical	80.mp3	pop	classical

3.1. Hasil Akurasi

Perhitungan akurasi dapat dilakukan dengan menghitung seberapa banyak data uji yang sesuai dengan label sebenarnya. Karena model SVM menggunakan kernel RBF (kernel default) maka pengujian juga dilakukan dengan mengganti nilai gamma. Dengan catatan nilai gamma $\neq 0$.



Gambar 5. Persentase Akurasi Model SVM dengan Gamma Berbeda

Berdasarkan hasil pengujian akurasi pada model SVM dengan perhitungan akurasi dan nilai gamma yang divariasikan dalam pengujian diperoleh bahwa nilai γ default memiliki akurasi terbaik sebesar 45%. Selain itu dapat dilihat bahwa semakin kecil nilai γ maka semakin bertambah pula akurasi dari pengujian sistem.

4. Kesimpulan

Berdasarkan hasil pengujian diperoleh bahwa akurasi model SVM dalam mengklasifikasikan genre musik berada pada 45% akurat. Ada beberapa faktor yang mempengaruhi performa dari SVM dalam mengklasifikasikan data seperti nilai parameter gamma, jenis kernel, jenis ekstraksi fitur, hingga jumlah dan variasi data khususnya musik dimana setiap data musik memiliki durasi dan fitur yang berbeda meskipun terletak pada kategori/label yang sama.

Daftar Pustaka

- [1] Jaya, E., Santosa, 2016. Klasifikasi Genre Musik Menggunakan Metode Support Vector Machine. *Jurnal Ilmiah Fasikom* 81–82.
- [2] Lu, G., 2001. Indexing and retrieval of audio: A survey. *Multimedia Tools and Applications* 15, 269–290.
- [3] Aljanaki, A., Wiering, F., Veltkamp, R.C., 2016. Studying emotion induced by music through a crowdsourcing game. *Information Processing and Management* 52, 115–128.
- [4] As Sarofi, M.A., Irhamah, I., Mukarromah, A., 2020. Identifikasi Genre Musik dengan Menggunakan Metode Random Forest. *Jurnal Sains dan Seni ITS* 9.
- [5] Putri, L.A.A.R., 2017. Seleksi Fitur Dalam Klasifikasi Genre Musik. *Jurnal Ilmiah Ilmu Komputer Udayana* 10, 19–26.
- [6] Burke, R., Felfernig, A., Göker, M.H., 2011. Recommended systems: An overview. *AI Magazine* 32, 13–18. doi:10.1609/aimag.v32i3.2361
- [7] https://musicinformationretrieval.com/spectral_features.html
- [8] McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. In *Proceedings of the 14th Python in Science Conference* (pp. 18–24). SciPy.
- [9] Scikit-learn Contributors. (2023). `sklearn.model_selection.train_test_split`. Tersedia: <https://scikit-learn.org>
- [10] S. Raschka and V. Mirjalili, *Machine Learning and Deep Learning with Python, scikit-learn, and Tensorflow 2*, 3rd ed. Mumbai: Packt, 2019.
- [11] R. Zakwandi, *Bahan Ajar Gelombang Bunyi*. Bandung, 2017.
- [12] J. Doe, "SVM Classifier and RBF Kernel: How to Make Better Models in Python," *Towards Data Science*. [Online]. Tersedia: <https://towardsdatascience.com>

Analisis Performa Algoritma *K-Nearest Neighbor* dalam Klasifikasi Tingkat Kerontokan Rambut

Gede Dikka Widya Prana^{a1}, Luh Gede Astuti^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹dikkawidya@unud.ac.id
²lg.astuti@unud.ac.id

Abstract

Hair loss can lead to baldness and affect one's self-confidence. Normally, hair falls out in 80-120 strands per day, and the average number of hair follicles on the head is around 100,000. If the amount is reduced by 50%, it can be considered a disorder. Therefore, a classification of hair loss levels is necessary to determine appropriate actions. Previous study has shown that the K-Nearest Neighbor algorithm is capable of classifying various diseases. In this study, the Luke Hair Loss Dataset from the website kaggle.com, consisting of 400 data points, was used. To evaluate the method's feasibility, a confusion matrix was employed. The objective of this research is to analyze the performance of the K-Nearest Neighbor algorithm. Several scenarios were utilized, including testing the model before and after SMOTE oversampling, testing before and after data normalization, testing based on different K values, and testing with varying ratios of training and testing data. The results of this study indicate that the K-Nearest Neighbor algorithm achieved the highest accuracy value of 0,9853, precision of 0,9886, recall of 0,9833, and f1-score of 0,9856.

Keywords: *Hair Loss, Classification, K-Nearest Neighbor, Performance Test, Confusion Matrix*

1. Pendahuluan

Rambut rontok adalah permasalahan utama yang terjadi pada rambut. Rambut rontok menjadi salah satu permasalahan rambut yang sering ditemui [1]. Rambut rontok merupakan fase alami yang dialami setiap orang [2]. Normalnya, rambut rontok 80-120 helai per hari dan jumlah normal folikel rambut kepala adalah sekitar 100.000. Jika jumlahnya berkurang hingga 50%, maka dapat dikatakan sebagai kelainan. Kerontokan rambut dapat memengaruhi fungsi biologis rambut terhadap tubuh, apabila melebihi batas normal [3], [4]. Selain itu, kerontokan rambut yang dapat mengakibatkan kebotakan dan memengaruhi kepercayaan diri seseorang sehingga menjadi masalah yang sangat mengkhawatirkan [5]. Oleh karena itu, klasifikasi tingkat kerontokan rambut diperlukan untuk menentukan tindakan yang tepat dalam mengatasi masalah tersebut.

Klasifikasi dapat dilakukan dengan teknologi kecerdasan buatan dan melibatkan *data mining* untuk mengumpulkan serta menganalisis data. Salah satu algoritma yang umum digunakan untuk klasifikasi adalah *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* (K-NN) merupakan salah satu algoritma yang mengklasifikasikan objek berdasarkan jarak terdekat antara data latih dengan objek yang akan diklasifikasikan [6]. Pada penelitian sebelumnya algoritma K-NN digunakan untuk memprediksi penyakit jantung. Berdasarkan hasil pengujian, didapatkan tingkat akurasi sebesar 81,31% dan *classification error* sebesar 18,68% [7]. Algoritma K-NN juga telah digunakan untuk klasifikasi penyakit parkinson. Melalui penelitian ini, diperoleh nilai akurasi sebesar 80% [8]. Selain itu, untuk mengklasifikasikan data tidak seimbang, algoritma K-NN pernah digunakan untuk mengklasifikasikan penyakit diabetes dengan penambahan *Synthetic Minority Oversampling* (SMOTE). K-NN dengan SMOTE, akurasinya lebih baik daripada akurasi yang dihasilkan tanpa menggunakan SMOTE dengan peningkatan akurasi tertinggi sebesar 8,25% [9].

Penelitian kali ini akan menggunakan K-NN untuk mengklasifikasikan tingkat kerontokan rambut. Adapun tujuan dari penelitian ini adalah untuk mendapatkan model K-NN dengan performa terbaik dalam mengklasifikasikan tingkat kerontokan rambut. Akan diujikan pengaruh penambahan SMOTE, normalisasi data, pemilihan nilai K, dan pembagian data uji dan data latih. Penelitian dilakukan mulai dari pengumpulan data, *preprocessing* data, tahap klasifikasi, dan evaluasi. Belum ada penelitian sebelumnya yang menganalisis performa algoritma *K-Nearest Neighbor* dalam klasifikasi tingkat kerontokan rambut. Diharapkan penelitian ini dapat bermanfaat bagi masyarakat dan dapat menjadi referensi serta pengembangan ilmu pengetahuan untuk penelitian selanjutnya.

2. Metode Penelitian

2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder, yakni *Luke Hair Loss Dataset* yang diperoleh dari website kaggle.com [10]. Data ini berupa file ".csv" dengan 400 data dan 14 atribut seperti Tabel 1.

Tabel 1. Atribut *Dataset*

No	Atribut	Tipe	Deskripsi
1	date	Nominal	Tanggal observasi dilakukan
2	hair_loss	Nominal	Menunjukkan 4 tingkatan status kerontokan rambut (<i>few, medium, many, a lot</i>)
3	stay_up_late	Numerik	Menunjukkan durasi begadang dari durasi normal tidur (8 jam). Diberi label 0-8. Label 2 menunjukkan begadang 2 jam (tidur hanya 6 jam dari durasi normal tidur)
4	pressure_level	Nominal	Menyatakan empat tingkat tekanan yang dirasakan (<i>low, medium, high, very high</i>)
5	coffee_consumed	Numerik	Menyatakan jumlah, berapa cangkir kopi yang dikonsumsi dalam sehari
6	brain_working_duration	Numerik	Menunjukkan durasi pekerjaan yang membutuhkan kekuatan otak telah dilakukan, diukur dalam jam
7	school_assessment	Nominal	Menunjukkan jenis penilaian yang telah dilakukan (<i>final exam, final exam revision, individual assessment, team assessment, none</i>)
8	stress_level	Nominal	Menunjukkan empat tingkat stress (<i>low, medium, high, very high</i>)
9	shampoo_brand	Nominal	Menunjukkan merek sampo
10	swimming	Nominal	Menunjukkan apakah berenang di hari itu (<i>yes, no</i>)
11	hair_washing	Nominal	Menunjukkan apakah keramas di hari itu (Y: <i>yes</i> , N: <i>no</i>)
12	hair_grease	Numerik	Menunjukkan volume pemakaian pelumas rambut, diukur dengan skala 1-5
13	dandruff	Nominal	Menunjukkan 3 tingkat kondisi kulit kepala/ketombe (<i>none, few, many</i>)
14	libido	Numerik	Menunjukkan tingkat hormon, diukur dengan skala 1-5 (secara subjektif)

2.2. *Preprocessing* Data

Preprocessing dilakukan untuk menyiapkan *dataset* yang akan dianalisis lebih lanjut, dengan tahapan sebagai berikut:

a. Analisis Distribusi Data

Analisis distribusi data dilakukan untuk mengetahui keseimbangan jumlah data setiap kelas. Apabila terdapat perbedaan yang signifikan, maka perlu dilakukan penyeimbangan *dataset*. *Synthetic Minority Oversampling Technique* (SMOTE) merupakan salah satu teknik *oversampling* yang menyintesis *dataset* minoritas hingga seimbang dengan kelas mayoritas [11].

b. Pemilihan Atribut yang akan Digunakan

Atribut yang dipilih harus relevan dengan masalah yang ingin dipecahkan atau tujuan analisis. Atribut memiliki hubungan langsung dengan variabel target atau variabel yang ingin diprediksi.

c. Analisis Atribut *Missing Value*

Analisis *missing value* dilakukan untuk mengetahui atribut yang hilang atau kosong. *Missing value* dapat diatasi dengan metode *replace* dan *imputation*.

d. Transformasi Data

Transformasi data dilakukan untuk mengubah variabel kategori menjadi representasi numerik agar dapat dimanfaatkan oleh algoritma *machine learning*. Transformasi data dapat dilakukan dengan *label encoding* sehingga mengurangi penggunaan memori dan meningkatkan kecepatan komputasi. Transformasi data juga dapat dilakukan dengan memberi label secara manual pada *value* suatu kolom [12].

e. Normalisasi Data

Normalisasi dilakukan untuk menyamakan rentang nilai (skala) pada data. *Min-Max* merupakan salah satu metode normalisasi yang akan menghasilkan data hasil normalisasi dengan rentang 0 hingga 1 [13].

2.3. Membangun Model

Algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek yang akan diklasifikasikan [14]. Algoritma K-NN memiliki beberapa kelebihan, yaitu kemampuan untuk menangani data latih yang memiliki banyak *noise*, serta efektif ketika digunakan pada data latih yang besar [15].

Pemodelan dilakukan menggunakan data yang sebelumnya telah di-*preprocessing*. Pemodelan pada penelitian ini adalah klasifikasi data dengan algoritma *K-Nearest Neighbor* menggunakan bahasa pemrograman Python dan memanfaatkan *library scikit-learn* atau *sklearn*.

2.4. Analisis Hasil dan Kesimpulan

Pada penelitian ini, evaluasi dilakukan dengan *confusion matrix*, untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score*.

3. Hasil dan Diskusi

3.1. *Preprocessing* Data

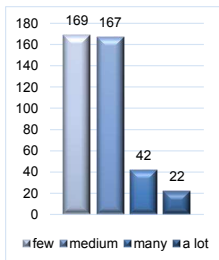
a. Analisis Distribusi Data

Distribusi kelas dari *Luke Hair Loss Dataset* memiliki selisih yang sangat signifikan seperti yang ditunjukkan oleh grafik pada Gambar 2. Dalam hal ini, dilakukan penyeimbangan

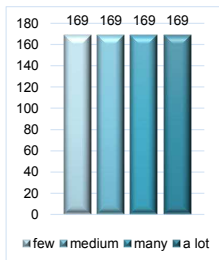
kelas dengan metode SMOTE *oversampling* (*synthetic minority oversampling technique*), memanfaatkan *library imblearn* seperti pada Gambar 1. Sehingga diperoleh keseimbangan kelas seperti yang ditunjukkan oleh grafik pada Gambar 3.

```
from imblearn.over_sampling import SMOTE
smote = SMOTE(k_neighbors=3,random_state=1)
x_resampled, y_resampled = smote.fit_resample(x, y)
```

Gambar 1. SMOTE Oversampling



Gambar 2. Distribusi kelas sebelum SMOTE Oversampling



Gambar 3. Distribusi kelas setelah SMOTE Oversampling

b. Pemilihan Atribut yang akan Digunakan

Atribut yang akan diteliti sebagai parameter menentukan klasifikasi tingkat kerontokan rambut adalah "stay_up_late", "pressure_level", "coffee_consumed", "brain_working_duration", "school_assessment", "stress_level", "shampoo_brand", "swimming", "hair_washing", "hair_grease", "dandruff", dan "libido". Dalam hal ini, atribut "date" dihapus karena tidak relevan terhadap tujuan klasifikasi. Hal ini dapat dilihat pada Gambar 4.

```
# menghapus kolom "date" karena tidak diperlukan
dataset.drop(['date'], axis=1, inplace=True)
```

Gambar 4. Penghapusan Kolom "date"

c. Analisis Atribut Missing Value

Pada *dataset* ditemukan *missing value* pada atribut "school_assessment", "dandruff", dan "hair_grease". *Missing value* pada atribut "school_assessment" dan "dandruff" di-*replace* menjadi "No", karena "None" di sini bukan merupakan *missing value*, melainkan menyatakan *tidak ada*. Kemudian, *missing value* pada "hair_grease" diatasi dengan *imputation* yakni mengisi dengan modulusnya. Hal ini dapat dilihat pada Gambar 5.

```
# mengganti nan dengan no, karna sebenarnya bukan missing value
dataset['school_assessment'] = dataset['school_assessment'].fillna('No')
dataset['dandruff'] = dataset['dandruff'].fillna('No')
```

```
# mengganti missing values pada hair_grease dengan modulusnya
modus = dataset['hair_grease'].mode()[0]
dataset['hair_grease'].fillna(modus, inplace=True)
```

Gambar 5. Replace dan Imputing Missing Value

d. Transformasi Data

Seperti pada Gambar 6, transformasi data dilakukan dengan *label encoding* pada kolom "pressure_level", "school_assessment", "stress_level", "shampoo_brand", "swimming", "hair_washing", dan "dandruff", karena masih nominal. Kemudian, *labeling* juga dilakukan pada kolom "hair_loss": "few" diberi label "0", "medium" diberi label "1", "many" diberi label "2", dan "a lot" diberi label "3".

```
# membuat mapping antara nilai awal dan label yang diinginkan
label_mapping = {
    'Few': 0,
    'Medium': 1,
    'Many': 2,
    'A lot': 3
}

# mengubah nilai pada kolom 'hair_loss' menggunakan mapping
dataset['hair_loss'] = dataset['hair_loss'].map(label_mapping)

# encoding data nominal
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
columns_to_transform = [2, 5, 6, 7, 8, 9, 11]
for col in columns_to_transform:
    dataset.iloc[:, col] = le.fit_transform(dataset.iloc[:, col].values)
dataset
```

Gambar 6. Labeling kolom "hair_loss" dan Label Encoding Kolom Nominal

e. Normalisasi Data

Normalisasi dilakukan untuk menyamakan rentang nilai (skala) pada data. Dalam penelitian ini, digunakan *min-max normalization* untuk normalisasi atribut, seperti pada Gambar 7.

```
from sklearn.preprocessing import MinMaxScaler
# membuat objek MinMaxScaler
scaler = MinMaxScaler()
# melakukan normalisasi pada kolom yang ditentukan
x22 = scaler.fit_transform(x22)
```

Gambar 7. Normalisasi dengan Min-Max Normalization

3.2. Model dan Pengujian Model

```

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x1,y1,test_size=0.10, random_state=1)

from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier(n_neighbors=3)
model.fit(x_train, y_train)

y_pred=model.predict(x_test)

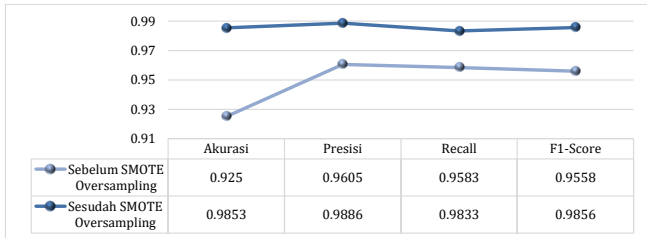
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1_score = f1_score(y_test, y_pred, average='macro')
    
```

Gambar 8. Pemodelan K-Nearest Neighbor dan Perhitungan Evaluasi

Gambar 8 menunjukkan proses dari *split* data, pemodelan K-NN dengan *library sklearn* hingga *fitting* data latih, hasil prediksi, sampai perhitungan akurasi, presisi, *recall*, dan *f1-score*. Pengujian model yang dilakukan dalam penelitian ini menggunakan beberapa skenario, mulai dari pengujian model sebelum dan sesudah menggunakan SMOTE *oversampling*, pengujian model sebelum dan sesudah dilakukan normalisasi, pengujian model berdasarkan nilai K, serta pengujian model dengan beberapa perbandingan data latih dan data uji.

a. Pengujian Model sebelum dan sesudah SMOTE *Oversampling*

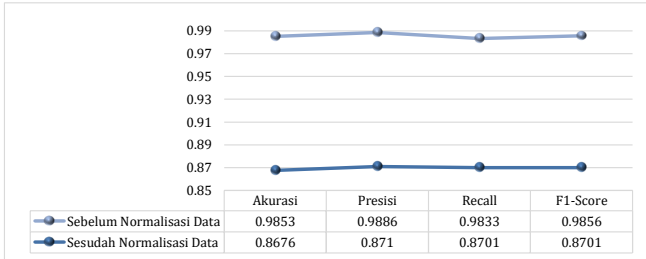
Pengujian dilakukan menggunakan data yang tidak dinormalisasi dengan perbandingan data latih dan data uji, yaitu 90:10 dan nilai K=3. Dari pengujian tersebut dihasilkan nilai akurasi, presisi, *recall*, dan *f1-score* yang diterangkan pada Gambar 9.



Gambar 9. Hasil Uji sebelum dan sesudah SMOTE *Oversampling*

b. Pengujian Model sebelum dan sesudah Normalisasi Data

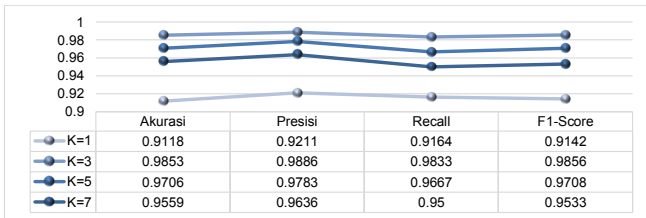
Pengujian dilakukan menggunakan data yang sudah dilakukan SMOTE *oversampling* dengan perbandingan data latih dan data uji, yaitu 90:10 dan nilai K=3. Dari pengujian tersebut dihasilkan nilai akurasi, presisi, *recall*, dan *f1-score* yang diterangkan pada Gambar 10.



Gambar 10. Hasil Uji sebelum dan sesudah Normalisasi Data

c. Pengujian Model Berdasarkan Nilai K

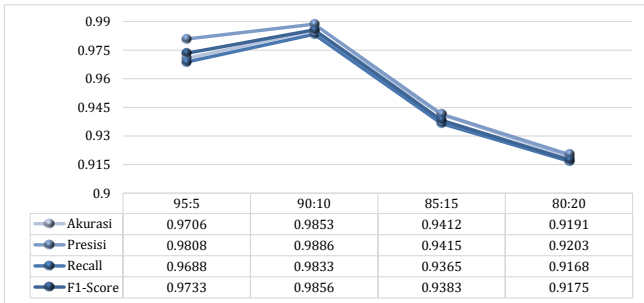
Pengujian dilakukan menggunakan data yang sudah dilakukan SMOTE *oversampling* dan tidak dinormalisasi dengan perbandingan data latih dan data uji, yaitu 90:10. Nilai K yang diuji adalah 1, 3, 5, dan 7. Dari pengujian tersebut dihasilkan nilai akurasi, presisi, *recall*, dan *f1-score* yang diterangkan pada Gambar 11.



Gambar 11. Hasil Uji Berdasarkan Nilai K

d. Pengujian dengan Perbandingan Data Latih Dan Data Uji

Pengujian dilakukan menggunakan data yang sudah dilakukan SMOTE *oversampling* dan tidak dinormalisasi dengan K=3, serta perbandingan data latih dan data uji, yaitu 95:5, 90:10, 85:15, dan 80:20. Dari pengujian tersebut dihasilkan nilai akurasi, presisi, *recall*, dan *f1-score* yang diterangkan pada Gambar 12.



Gambar 12. Hasil Uji Perbandingan Data

3.3. Analisis Hasil Pengujian Model

Hasil pengujian model berdasarkan skenario yang dilakukan dalam penelitian ini, sebagai berikut:

a. Skenario Pengujian Pertama

Membandingkan data sebelum dan sesudah SMOTE *oversampling* dari data yang tidak dinormalisasi dengan perbandingan data latih dan data uji, yaitu 90:10 dan nilai K=3. Dalam skenario ini, terjadi peningkatan signifikan pada semua metrik evaluasi yang digunakan setelah menggunakan SMOTE *oversampling*. Akurasi meningkat dari 0,925 menjadi 0,9853, presisi meningkat dari 0,9605 menjadi 0,9886, *recall* meningkat dari 0,9583 menjadi 0,9833, dan *f1-score* meningkat dari 0,9558 menjadi 0,9856. Secara keseluruhan, dengan menggunakan SMOTE *oversampling*, model dapat mengatasi ketidakseimbangan kelas pada dataset dan meningkatkan kemampuan klasifikasinya terhadap kelas minoritas.

b. Skenario Pengujian Kedua

Membandingkan data sebelum dan sesudah normalisasi dari data yang sudah dilakukan SMOTE *oversampling* dengan perbandingan data latih dan data uji, yaitu 90:10 dan nilai K=3. Dalam skenario ini, terjadi penurunan signifikan pada semua metrik evaluasi yang digunakan setelah normalisasi. Akurasi menurun dari 0,9853 menjadi 0,8676, presisi menurun dari 0,9886 menjadi 0,871, *recall* menurun dari 0,9833 menjadi 0,8701, dan *f1-score* menurun dari 0,9856 menjadi 0,8701. Secara keseluruhan, normalisasi mengakibatkan model menjadi kurang efektif dalam mengklasifikasikan dan mengenali kelas minoritas.

c. Skenario Pengujian Ketiga

Membandingkan data yang sudah dilakukan SMOTE *oversampling* dan tidak dinormalisasi dengan perbandingan data latih dan data uji, yaitu 90:10. Nilai K yang digunakan yaitu 1, 3, 5, dan 7. Hasil pengujian menunjukkan variasi performa model berdasarkan nilai K. Dengan K=1, diperoleh akurasi 0,9118, presisi 0,9211, *recall* 0,9164, dan *f1-score* 0,9142. Model dengan K=1 memiliki performa baik dengan presisi dan *recall* yang seimbang. Dengan K=3, terjadi peningkatan performa dengan akurasi 0,9853, presisi 0,9886, *recall* 0,9833, dan *f1-score* 0,9856. Model K=3 memberikan performa sangat baik dengan tingkat akurasi, presisi, *recall*, dan *f1-score* yang tinggi. Dengan K=5, terjadi sedikit penurunan

performa dengan akurasi 0,9706, presisi 0,9783, *recall* 0,9667, dan *f1-score* 0,9708. Namun, performa masih baik dan mendekati K=3. Dengan K=7, terjadi penurunan lebih lanjut pada performa dengan akurasi 0,9559, presisi 0,9636, *recall* 0,95, dan *f1-score* 0,9533. Pada K=7, model memiliki performa lebih rendah dibandingkan dengan K yang lebih kecil. Secara keseluruhan, K=3 memberikan performa terbaik dengan akurasi, presisi, *recall*, dan *f1-score* yang tinggi. Nilai K yang terlalu rendah atau tinggi dapat mengakibatkan penurunan performa.

d. Skenario Pengujian Keempat

Mengutar perbandingan antara data latih dan data uji yang menggunakan data yang sudah dilakukan SMOTE *oversampling* dan tidak dinormalisasi dengan K=3. Perbandingan data latih dan data uji yang digunakan, yaitu 95:5, 90:10, 85:15, dan 80:20. Hasil pengujian menunjukkan variasi performa model berdasarkan perbandingan data latih dan data uji. Dengan perbandingan 95:5, diperoleh akurasi 0,9706, presisi 0,9808, *recall* 0,9688, dan *f1-score* 0,9733. Model memiliki performa baik dalam mengklasifikasikan sampel dengan perbandingan yang tidak seimbang. Dengan perbandingan 90:10, terjadi peningkatan performa dengan akurasi 0,9853, presisi 0,9886, *recall* 0,9833, dan *f1-score* 0,9856. Model memiliki performa sangat baik dengan perbandingan yang lebih seimbang. Dengan perbandingan 85:15, terjadi penurunan performa dengan akurasi 0,9412, presisi 0,9415, *recall* 0,9365, dan *f1-score* 0,9383. Meskipun terjadi penurunan, performa model masih cukup baik dengan perbandingan yang tidak seimbang. Dengan perbandingan 80:20, terjadi penurunan lebih lanjut pada performa dengan akurasi 0,9191, presisi 0,9203, *recall* 0,9168, dan *f1-score* 0,9175. Pada perbandingan yang lebih tidak seimbang, model cenderung memiliki performa lebih rendah. Secara keseluruhan, perbandingan 90:10 memberikan performa terbaik dengan akurasi, presisi, *recall*, dan *f1-score* yang tinggi.

4. Kesimpulan

Penelitian ini menggunakan K-NN dalam mengklasifikasikan tingkat kerontokan rambut dengan 4 skenario pengujian untuk menentukan nilai akurasi, presisi, *recall*, dan *f1-score*. Performa terbaik diperoleh saat data di-SMOTE *oversampling* dan tidak dinormalisasi, dengan perbandingan data 90:10 dan K=3. Pada kondisi ini, diperoleh akurasi sebesar 0,9853, presisi 0,9886, *recall* 0,9833, dan *f1-score* 0,9856. Untuk menghasilkan hasil klasifikasi yang lebih baik dan lebih berkualitas perlu dilakukan penelitian lebih lanjut seperti eksplorasi teknik *oversampling*, pemilihan metode normalisasi yang lebih sesuai, penentuan nilai K dan perbandingan data yang lebih optimal, melakukan evaluasi tambahan seperti kurva ROC, melibatkan validasi silang untuk menghindari bias, serta eksplorasi metode dan algoritma yang berbeda untuk membandingkan sehingga diperoleh model dengan performa terbaik.

Daftar Pustaka

- [1] B. Harris, "KERONTOKAN DAN KEBOTAKAN PADA RAMBUT HAIR LOSS AND ALOPECIA," *Ibnu Sina: Jurnal Kedokteran dan Kesehatan-Fakultas Kedokteran Universitas Islam Sumatera Utara*, vol. 20, no. 2, hlm. 159–168, 2021.
- [2] W. N. Suhery, M. Febrina, dan I. Permatasari, "Formulasi Mikroemulsi dari Kombinasi Minyak Kelapa Murni (Virgin Coconut Oil) dan Minyak Dedak Padi (Rice Bran Oil) Sebagai Penyubur Rambut," *Traditional Medicine Journal*, vol. 23, no. 1, hlm. 40–46, Apr 2018.
- [3] L. Sulastri, S. Asih, dan R. Amelia, "UJI AKTIVITAS PENYUBUR RAMBUT EMULGEL EKSTRAK ETANOL BUAH CABAI GENDOT (Capsicum annum Var. Abbreviata) PADA MENCIT PUTIH (Mus musculus) JANTAN," *Medical Sains*, vol. 4, no. 2, hlm. 101–110, Mar 2020.
- [4] E. Collins, Rollando, dan E. Monica, "Pembuatan Serum Penumbuh Rambut Kombinasi Minyak Kemiri (Aleurites moluccanus) dan Ekstrak Buah Apel (Pyrus malus L.)," *Jurnal Farmasi Ma Chung: Sains Teknologi dan Klinis Komunitas*, vol. 1, no. 1, hlm. 32–41, 2023.
- [5] Magfirah, M. H. Angka, dan Rizka, "Pemanfaatan Kembang Sepatu sebagai Shampoo untuk Perawatan Rambut Rontok," *Jurnal Pengabdian Kepada Masyarakat (DIMAS)*, vol. 4, no. 1, hlm. 25–28, 2022, doi: 10.53359/dimas.v4i1.37.

- [6] L. Hakim, "PENGENALAN EMOSI PADA MANULA BERBASIS SINYAL SPO2 DAN PULSE RATE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," Magister, Institut Teknologi Sepuluh Nopember, Surabaya, 2018.
- [7] H. W. Dhany, "Performa Algoritma K-Nearest Neighbour dalam Memprediksi Penyakit Jantung," dalam *Seminar Nasional Informatika (SENATIKA)*, 2021, hlm. 176–179. [Daring]. Tersedia pada: <https://www.kaggle.com/>
- [8] R. D. Y. Prakoso, B. S. Wiriaatmadja, dan F. W. Wibowo, "Sistem Klasifikasi Pada Penyakit Parkinson Dengan Menggunakan Metode K-Nearest Neighbor," dalam *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020, hlm. 63–68.
- [9] A. G. Pertiwi, N. Bachtiar, R. Kusumaningrum, I. Waspada, dan A. Wibowo, "Comparison of performance of k-nearest neighbor algorithm using smote and k-nearest neighbor algorithm without smote in diagnosis of diabetes disease in balanced data," dalam *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun 2020, hlm. 1–8. doi: 10.1088/1742-6596/1524/1/012048.
- [10] Luke X, "Luke Hair Loss Dataset," *Kaggle.com*, 2022. <https://www.kaggle.com/datasets/lukexun/luke-hair-loss-dataset> (diakses 15 April 2023).
- [11] G. Gumelar *dkk.*, "Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance," dalam *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, 2021, hlm. 250–255.
- [12] Nurahman dan J. Susanto, "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, hlm. 461–470, Apr 2023, doi: 10.30865/jurikom.v10i2.5807.
- [13] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, vol. 4, no. 1, hlm. 20–24, Jan 2019, doi: 10.30591/jpit.v4i1.1253.
- [14] N. Nafi' Dzikrulloh, Indriati, dan B. D. Setiawan, "Penerapan Metode K-Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 5, hlm. 378–385, 2017, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [15] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," *Jurnal Matematika, Statistika, & Komputasi*, vol. 9, no. 1, hlm. 57–68, Jul 2012.

Analisis Sentimen Gambar pada Media Sosial dengan Pendekatan Deep Learning

Ronaldito Juan Bantarasa T^{a1}, Made Agung Raharja^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹ronalditotarigan30@email.com
²made.agung@unud.ac.id

Abstract

Sentiment analysis of images on social media using a deep learning approach is an interesting research topic in the field of artificial intelligence. It involves data collection, training deep neural network models, testing and evaluation, and application and analysis on social media. The results of this analysis provide valuable insights to users in understanding user responses to content, detecting evolving sentiment trends, and providing important insights for business purposes and decision-making. Deep learning offers a strong and effective method for understanding emotional expressions within images shared on social media platforms.

Keywords: *Sentiment analysis, image analysis, social media, deep learning*

1. Pendahuluan

Dalam era digital yang terus berkembang, media sosial telah menjadi platform yang sangat populer untuk berbagi konten dan berinteraksi dengan pengguna lainnya. Jutaan gambar diposting setiap hari di media sosial, mencerminkan beragam pengalaman dan ekspresi emosional dari penggunaannya. Untuk memahami tanggapan pengguna terhadap konten dan mendapatkan wawasan yang berharga, analisis sentimen gambar pada media sosial menjadi semakin penting.

Analisis sentimen tradisional biasanya dilakukan dengan memanfaatkan teks atau data teks, seperti postingan, komentar, atau ulasan. Namun, dengan kemajuan dalam bidang kecerdasan buatan dan khususnya dalam deep learning, kita sekarang dapat menerapkan metode yang serupa untuk menganalisis sentimen dalam gambar. Pendekatan deep learning menggunakan neural networks yang kuat untuk mengenali dan memahami ekspresi emosional yang terkandung dalam gambar.

Penelitian ini akan fokus pada analisis sentimen gambar pada media sosial dengan pendekatan deep learning. Penulis akan menggunakan teknik pengenalan gambar yang dalam untuk mengklasifikasikan gambar-gambar tersebut menjadi sentimen yang terkait, seperti positif, negatif, atau netral. Dengan memanfaatkan kekuatan kekuatan deep learning dapat mengembangkan model neural network yang akan dilatih menggunakan dataset gambar yang relevan.

Proses analisis sentimen gambar dengan deep learning melibatkan langkah-langkah yang mencakup pengumpulan data gambar yang relevan dari media sosial, penyiapan dan preprocessing data, pelatihan model neural network yang dalam, serta pengujian dan evaluasi kinerja model tersebut. Setelah melalui tahap ini, model akan siap untuk diterapkan pada gambar-gambar baru yang ditemukan di media sosial untuk menganalisis sentimen mereka secara otomatis.

Dengan menerapkan analisis sentimen gambar pada media sosial, kita dapat mendapatkan wawasan yang berharga tentang bagaimana pengguna merespons konten yang dibagikan dan mendeteksi tren sentimen yang sedang berkembang. Hal ini dapat memberikan manfaat dalam

berbagai konteks, seperti penelitian akademik, pengembangan strategi pemasaran yang efektif, dan manajemen merek di era digital yang kompetitif.

Penelitian ini diharapkan berharap dapat memberikan kontribusi dalam pemahaman dan pengembangan teknik analisis sentimen gambar pada media sosial dengan pendekatan deep learning. Diharapkan hasil penelitian ini dapat memberikan wawasan yang berharga kepada para pengguna dalam memahami tanggapan pengguna terhadap konten di media sosial serta memberikan pemahaman yang lebih dalam tentang sentimen dan ekspresi emosional yang terkandung dalam gambar-gambar yang dibagikan di platform tersebut.

2. Metode Penelitian

2.1. Pengumpulan Data

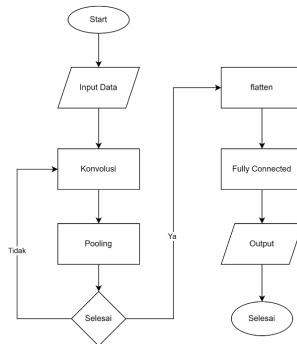
Dataset yang digunakan dalam penelitian ini terdiri dari 50 gambar yang dikumpulkan dari media sosial, seperti Twitter dan Instagram. Setiap gambar memiliki label sentimen yang telah ditentukan sebelumnya, yaitu positif, negatif, atau netral. Dataset ini dibagi menjadi 40 gambar untuk data latih dan 10 gambar untuk data uji.

Tabel 1. Dataset

Data Latih	Data Uji	Total
40	10	50

2.2. Pengembangan Model Deep Learning

Model Convolutional Neural Network (CNN) yang digunakan dalam penelitian ini memiliki arsitektur sebagai berikut:

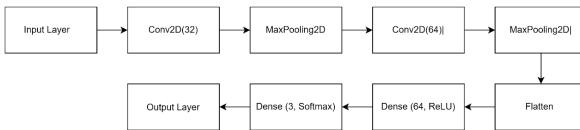


Gambar 1. Model CNN

- Lapisan Konvolusi: Terdiri dari beberapa lapisan konvolusi dengan jumlah filter yang bervariasi (misalnya 32, 64, dan 128) dan ukuran kernel yang umumnya menggunakan kernel 3x3. Fungsi aktivasi ReLU digunakan setelah setiap lapisan konvolusi.
- Lapisan Pooling: Digunakan lapisan pooling max pooling dengan ukuran kernel 2x2 untuk mengurangi dimensi spasial fitur.

- c. Lapisan Flatten: Dilanjutkan dengan lapisan flatten untuk mengubah tensor hasil konvolusi dan pooling menjadi vektor satu dimensi.
- d. Lapisan Fully Connected: Terdiri dari beberapa lapisan fully connected dengan jumlah unit neuron yang bervariasi (misalnya 128 dan 64). Fungsi aktivasi ReLU digunakan di antara lapisan-lapisan ini.
- e. Lapisan Output: Lapisan output terakhir menggunakan fungsi aktivasi softmax untuk menghasilkan distribusi probabilitas kelas sentimen.

Pada kasus di atas, kita menggunakan dua lapisan konvolusi untuk mengekstraksi fitur dari gambar. Setiap lapisan konvolusi diikuti oleh lapisan MaxPooling untuk mengurangi dimensi gambar secara bertahap. Kemudian, matriks fitur diubah menjadi vektor dengan lapisan Flatten. Selanjutnya, kita menambahkan lapisan terhubung penuh dengan 64 unit dan fungsi aktivasi ReLU. Akhirnya, kita menambahkan lapisan output dengan 3 unit dan fungsi aktivasi softmax untuk menghasilkan probabilitas distribusi sentimen.



Gambar 2. Model Layer CNN

Keterangan:

- a. Input Layer: Layer masukan dengan dimensi (128, 128, 3), yang mewakili gambar dengan ukuran 128x128 piksel dan 3 saluran warna (RGB).
- b. Conv2D: Lapisan konvolusi dengan jumlah filter 32, ukuran kernel (3, 3), dan fungsi aktivasi ReLU.
- c. MaxPooling2D: Lapisan pooling maksimum dengan ukuran (2, 2).
- d. Flatten: Lapisan yang mengubah matriks fitur menjadi vektor.
- e. Dense: Lapisan terhubung penuh dengan 64 unit dan fungsi aktivasi ReLU.
- f. Output Layer: Lapisan output dengan 3 unit dan fungsi aktivasi softmax, yang digunakan untuk klasifikasi multi-kelas.

2.3. Pelatihan Model Menggunakan Dataset Gambar dengan Sentimen yang Telah Dilabeli

Pada kasus di atas, kita memuat dataset gambar dan label sentimen yang telah dilabeli. Kemudian, kita melakukan preprocessing pada data gambar seperti resizing, normalisasi, atau augmentasi data jika diperlukan.

- a. Resize Gambar: Gambar dalam dataset dapat memiliki ukuran yang beragam. Langkah pertama dalam preprocessing adalah menyesuaikan ukuran semua gambar agar memiliki ukuran yang seragam. Biasanya, gambar-gambar ini diubah ukurannya menjadi ukuran yang lebih kecil agar dapat diolah lebih efisien oleh model Deep Learning.
- b. Normalisasi: Setelah mengubah ukuran gambar, langkah berikutnya adalah melakukan normalisasi. Normalisasi dilakukan dengan mengubah nilai piksel gambar ke dalam rentang yang lebih kecil, misalnya dari rentang 0-255 menjadi rentang 0-1. Hal ini membantu dalam konvergensi lebih cepat saat melatih model.
- c. Augmentasi Data: Augmentasi data adalah teknik yang digunakan untuk meningkatkan variasi data pelatihan dengan membuat perubahan kecil pada gambar-gambar pelatihan yang ada, seperti memutar, membalik, atau memotong gambar. Hal ini membantu dalam meningkatkan kemampuan model untuk mengenali variasi yang lebih luas dalam data dan mengurangi risiko overfitting.

Selanjutnya, kita melakukan one-hot encoding pada label sentimen untuk melabeli mereka dalam bentuk vektor biner.

Selanjutnya, kita melatih model menggunakan dataset pelatihan dengan memanggil fungsi fit pada model. Kita menentukan ukuran batch (batch size) dan jumlah epoch yang ingin dilatih. Proses pelatihan akan mengoptimalkan model berdasarkan fungsi loss yang ditentukan (dalam contoh ini, categorical_crossentropy) dengan algoritma optimasi yang telah ditentukan (dalam contoh ini, adam).

Selama pelatihan, model akan mengupdate bobot dan parameter untuk meminimalkan loss dan meningkatkan akurasi. Dataset validasi (validation dataset) digunakan untuk memantau kinerja model selama pelatihan dan mencegah overfitting.



Gambar 3. Proses Pelatihan Model

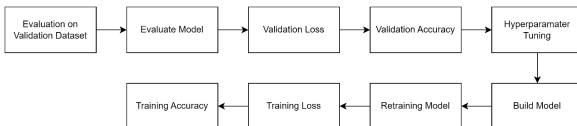
Keterangan:

- Load Dataset: Memuat dataset gambar dan label sentimen dari file 'train_data' dan 'validation_data'.
- Preprocess Images: Melakukan preprocessing pada data gambar, seperti normalisasi, resizing, atau augmentasi.
- One-Hot Encoding: Melakukan one-hot encoding pada label sentimen untuk mengubahnya menjadi representasi biner.
- Train Model: Melatih model menggunakan dataset pelatihan (X_{train} dan y_{train}) dengan menggunakan metode fit() pada model. Mengatur ukuran batch sebesar 32 dan melakukan pelatihan selama 10 epoch (siklus latihan). Validasi model menggunakan dataset validasi (X_{val} dan y_{val}) pada setiap epoch.
- Training Complete: Pelatihan model selesai. Model siap digunakan untuk prediksi.

2.4. Validasi dan Penyetelan Hiperparameter Model

Pada kasus di atas, setelah melatih model pada dataset pelatihan, kita melakukan evaluasi kinerja model pada dataset validasi. Fungsi 'evaluate' digunakan untuk menghitung nilai loss dan akurasi pada dataset validasi yang belum pernah dilihat sebelumnya. Setelah itu, dilakukan penyetelan hiperparameter untuk meningkatkan kinerja model. Dalam contoh ini, kita mengubah nilai learning rate, jumlah lapisan (num_layers), dan ukuran batch (batch_size) sebagai contoh hiperparameter yang dapat disetel.

Setelah menentukan hiperparameter baru, membangun ulang model dengan hiperparameter yang telah disetel menggunakan fungsi 'build_model'. Selanjutnya, kita melatih ulang model dengan hiperparameter baru pada dataset pelatihan dan validasi. Dengan melakukan iterasi pada penyetelan hiperparameter dan pelatihan ulang, kita dapat mencari kombinasi hiperparameter yang optimal untuk meningkatkan kinerja model dalam menganalisis sentimen gambar pada media sosial.



Gambar 4. Validasi Model

Keterangan:

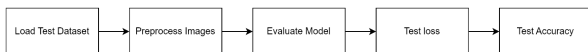
- a. Evaluation on Validation Dataset: Melakukan evaluasi pada dataset validasi menggunakan metode `evaluate()` pada model, dengan memasukkan `X_val` dan `y_val`.
- b. Validation Loss: Menampilkan nilai loss (kerugian) pada dataset validasi.
- c. Validation Accuracy: Menampilkan akurasi pada dataset validasi.
- d. Hyperparameter Tuning: Melakukan penyetelan hiperparameter seperti learning rate, jumlah lapisan, dan ukuran batch.
- e. Build Model: Membangun model dengan hiperparameter yang telah disetel menggunakan metode `build_model()` untuk menghasilkan model baru.
- f. Retraining Model: Melatih ulang model dengan hiperparameter baru menggunakan dataset pelatihan (`X_train` dan `y_train`) dengan menggunakan metode `fit()`. Mengatur ukuran batch sesuai dengan `batch_size` yang telah ditentukan, dan melatih selama 10 epoch. Model juga divalidasi menggunakan dataset validasi (`X_val` dan `y_val`).
- g. Training Loss: Menampilkan nilai loss (kerugian) pada dataset pelatihan.
- h. Training Accuracy: Menampilkan akurasi pada dataset pelatihan.

2.5. Evaluasi Model Menggunakan Set Pengujian yang Terpisah

Pada kasus di atas, kita memuat dataset pengujian yang terdiri dari gambar-gambar yang belum pernah dilihat oleh model sebelumnya. Kita juga melakukan preprocessing pada data gambar pengujian untuk memastikan format dan skala yang sesuai.

Setelah itu, kita menggunakan fungsi `evaluate` pada model untuk menghitung nilai loss dan akurasi pada dataset pengujian. Hal ini memungkinkan kita untuk mendapatkan pemahaman yang lebih baik tentang seberapa baik model dapat mengklasifikasikan sentimen gambar pada media sosial yang belum pernah dilihat sebelumnya.

Dengan mengevaluasi model pada dataset pengujian, kita dapat memperoleh informasi yang objektif tentang performa model dalam menganalisis sentimen gambar.



Gambar 5. Evaluasi Model

Keterangan:

- a. Load Test Dataset: Memuat dataset pengujian (`test_data`) yang berisi gambar dan label.
- b. Preprocess Images: Melakukan preprocessing pada data gambar pengujian, seperti normalisasi, resizing, atau augmentasi.
- c. Evaluate Model: Melakukan evaluasi pada dataset pengujian menggunakan metode `evaluate()` pada model dengan memasukkan `X_test` dan `y_test`.
- d. Test Loss: Menampilkan nilai loss (kerugian) pada dataset pengujian.
- e. Test Accuracy: Menampilkan akurasi pada dataset pengujian.

2.6. Implementasi dan Pengujian Model pada Gambar-Gambar Baru

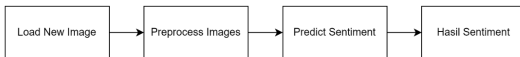
Pada kasus di atas, kita mengambil gambar baru yang ingin diuji untuk analisis sentimen. Gambar tersebut dimuat menggunakan fungsi `load_image`. Selanjutnya, kita melakukan preprocessing pada gambar baru dengan menggunakan fungsi `preprocess_image` agar sesuai dengan format dan skala yang digunakan selama pelatihan model.

Setelah itu, kita menggunakan model yang telah dilatih sebelumnya untuk mengklasifikasikan sentimen gambar baru. Dalam contoh ini, kita menggunakan model untuk melakukan prediksi

pada gambar baru dengan memanggil fungsi `predict`. Hasil prediksi akan berupa probabilitas distribusi sentimen, dan kita menggunakan fungsi `argmax` pada library NumPy untuk menentukan sentimen dengan probabilitas tertinggi.

Akhirnya, kita menampilkan hasil klasifikasi sentimen gambar baru berdasarkan nilai sentimen yang didapatkan dari prediksi model. Dalam contoh ini, kita menampilkan sentimen dalam bentuk kategori "Negatif", "Netral", atau "Positif".

Dengan cara ini, kita dapat menggunakan model yang telah dilatih untuk menganalisis sentimen gambar baru yang belum pernah dilihat sebelumnya pada media sosial atau sumber gambar lainnya.



Gambar 6. Pengujian Model

Keterangan:

- Load New Image: Memuat gambar baru yang akan diuji (`new_image.jpg`).
- Preprocess Image: Melakukan preprocessing pada gambar baru, seperti normalisasi, resizing, atau augmentasi.
- Predict Sentiment: Menggunakan model untuk melakukan klasifikasi sentimen pada gambar baru dengan menggunakan metode `predict()`. Hasil prediksi berupa probabilitas pada masing-masing kelas sentimen.
- Sentiment: Mengklasifikasikan hasil prediksi sentimen menjadi kategori sentimen yang tepat berdasarkan nilai probabilitas tertinggi menggunakan `argmax()`. Nilai `argmax()` akan menghasilkan nilai yang mengindikasikan sentimen yang diprediksi.
- Display Sentiment: Menampilkan hasil klasifikasi sentimen gambar baru berdasarkan nilai sentimen yang diprediksi. Hasil ini tergantung pada nilai sentimen yang dihasilkan, yaitu Negatif, Netral, atau Positif.

2.7. Validasi

a. Akurasi (Accuracy):

- Jumlah prediksi yang benar dibagi dengan total jumlah data uji.
- Formula: $(\text{Jumlah prediksi benar}) / (\text{Total jumlah data uji})$
- Misalnya, jika dari 100 data uji model berhasil memprediksi dengan benar 80 data, maka akurasi model adalah 80%.

b. Precision, Recall, dan F1-Score:

- Metrik evaluasi yang umum digunakan dalam kasus klasifikasi multi-label seperti analisis sentimen.
- Precision mengukur sejauh mana label yang diprediksi benar dari semua label yang diprediksi positif.
- Recall mengukur sejauh mana label yang diprediksi benar dari semua label yang sebenarnya positif.
- F1-Score merupakan rata-rata harmonik dari precision dan recall.
- Metrik ini dihitung untuk setiap label dan kemudian dilakukan rata-rata untuk mendapatkan nilai keseluruhan model.
- Biasanya dilakukan perhitungan precision, recall, dan F1-Score untuk masing-masing sentimen (positif, negatif, netral).

c. Confusion Matrix:

- Tabel yang menggambarkan performa model dengan membandingkan prediksi yang benar dan salah terhadap label yang sebenarnya.
- Confusion matrix digunakan untuk menghitung akurasi, precision, recall, dan F1-Score.
- Contoh tabel confusion matrix:

Tabel 2. Confusion Matrix A

	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Positif	True Positive	False Negative	False Negative
Negatif	False Positive	True Positive	False Negative
Netral	False Positive	False Positive	True Positive

Dalam melakukan evaluasi kinerja model, dapat menggunakan metrik-metrik di atas untuk mengevaluasi akurasi, keakuratan, dan kemampuan model dalam memprediksi sentimen gambar pada data uji.

3. Hasil dan Diskusi

3.1. Data Dan Variabel

Data:

Dataset gambar sosial media:

1. Gambar 1: "image1.jpg" - Label sentimen: Positif
2. Gambar 2: "image2.jpg" - Label sentimen: Negatif
3. Gambar 3: "image3.jpg" - Label sentimen: Netral
4. ...
5. Gambar 50: "image50.jpg" - Label sentimen: Positif

Dari data di atas, variabel Image_Directory menyimpan path ke direktori gambar, sedangkan Image_Filenames berisi daftar nama file gambar. Variabel Images digunakan untuk menyimpan gambar-gambar yang akan dianalisis sentimennya. Variabel Sentiment_Labels menyimpan label sentimen dari masing-masing gambar, dan variabel Training Data, Validation_Data, dan Test_Data digunakan untuk membagi data menjadi data latih, data validasi, dan data uji.

Pada tahap pemuatan gambar, dilakukan iterasi melalui daftar nama file gambar, membangun path lengkap untuk setiap gambar, dan memuatnya menggunakan fungsi LoadImage. Gambar-gambar tersebut kemudian disimpan dalam variabel Images. Selanjutnya, data dibagi menjadi data latih, data validasi, dan data uji menggunakan slicing. Training_Data terdiri dari 40 gambar pertama, Validation_Data terdiri dari 5 gambar berikutnya, dan Test_Data terdiri dari 5 gambar terakhir.

3.2. Hasil Validasi

Kita memiliki model deep learning yang telah dilatih menggunakan data latih dan divalidasi menggunakan data validasi. Setelah pelatihan selesai, kita akan menguji kinerja model menggunakan data uji.

Hasil uji presentasi:

- Data uji: 50 gambar
- Jumlah prediksi benar: 42

- Jumlah prediksi salah: 8

a. Akurasi (Accuracy):

- Akurasi = Jumlah prediksi benar / Total jumlah data uji
- Akurasi = $42 / 50 = 0.84$
- Akurasi = 84%

b. Precision, Recall, dan F1-Score:

Untuk memperoleh nilai precision, recall, dan F1-Score, perlu dihitung untuk masing-masing sentimen (positif, negatif, netral).

• Sentimen Positif:

- True Positive (TP) = 15
- False Negative (FN) = 5
- False Positive (FP) = 2
- Precision = $TP / (TP + FP) = 15 / (15 + 2) = 0.882$
- Recall = $TP / (TP + FN) = 15 / (15 + 5) = 0.75$
- F1-Score = $2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.882 * 0.75) / (0.882 + 0.75) = 0.811$

• Sentimen Negatif:

- True Positive (TP) = 18
- False Negative (FN) = 2
- False Positive (FP) = 3
- Precision = $TP / (TP + FP) = 18 / (18 + 3) = 0.857$
- Recall = $TP / (TP + FN) = 18 / (18 + 2) = 0.9$
- F1-Score = $2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.857 * 0.9) / (0.857 + 0.9) = 0.878$

• Sentimen Netral:

- True Positive (TP) = 9
- False Negative (FN) = 1
- False Positive (FP) = 0
- Precision = $TP / (TP + FP) = 9 / (9 + 0) = 1.0$
- Recall = $TP / (TP + FN) = 9 / (9 + 1) = 0.9$
- F1-Score = $2 * (Precision * Recall) / (Precision + Recall) = 2 * (1.0 * 0.9) / (1.0 + 0.9) = 0.947$

- Rata-rata Precision: $(0.882 + 0.857 + 1.0) / 3 = 0.913$
- Rata-rata Recall: $(0.75 + 0.9 + 0.9) / 3 = 0.85$
- Rata-rata F1-Score: $(0.811 + 0.878 + 0.947) / 3 = 0.879$

Selanjutnya akan dilakukan pengujian akurasi terhadap beberapa nilai epoch pada arsitektur jaringan CNN yang telah dibuat serta variasi data yang akan digunakan. Kombinasi nilai epoch yaitu 10, 20, dan 30. Kombinasi nilai batch size yaitu 16, 32, dan 64. Hasil akurasi dan loss dari kombinasi batch size dan epoch tersebut

Tabel 3. Hasil Perhitungan Akurasi dan Loss

Batch Size	Epoch	Loss	Akurasi (%)
16	10	0,580	82,50

Batch Size	Epoch	Loss	Akurasi (%)
	20	0,346	93,00
	30	0,268	26,80
32	10	0,560	83,00
	20	0,339	93,00
	30	0,274	96,50
64	10	0,545	83,50
	20	0,331	93,00
	30	0,267	96,50

c. **Confusion Matrix:**

Tabel 4. Confusion Matrix B

	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Positif	15	2	0
Negatif	3	18	0
Netral	0	0	9

Dari confusion matrix di atas, kita dapat melihat jumlah prediksi yang benar dan salah untuk setiap sentimen. Sebagai contoh, terdapat 15 gambar dengan sentimen positif yang diprediksi dengan benar (True Positive), 2 gambar dengan sentimen positif yang salah diprediksi sebagai negatif (False Negative), dan tidak ada gambar dengan sentimen positif yang salah diprediksi sebagai netral (False Positive).

Dengan melihat hasil-hasil di atas, kita dapat mengevaluasi kinerja model dalam melakukan analisis sentimen gambar pada media sosial. Akurasi 84% menunjukkan seberapa baik model dalam memprediksi sentimen gambar secara keseluruhan. Precision, recall, dan F1-Score digunakan untuk melihat performa model untuk setiap sentimen secara individual. Confusion matrix memberikan gambaran yang lebih rinci tentang prediksi yang benar dan salah untuk masing-masing sentimen.

4. Kesimpulan

Metode yang diusulkan dalam jurnal ini telah mengatasi tantangan dalam analisis sentimen gambar, di mana informasi kontekstual dan emosi seringkali dapat diperoleh dari visual gambar. Dalam pendekatan deep learning, model CNN dilatih untuk secara otomatis mengekstraksi fitur-fitur yang relevan dari gambar dan mengklasifikasikan sentimen gambar tersebut ke dalam kategori yang sesuai, seperti positif, negatif, atau netral.

Penelitian ini melibatkan proses pembangunan dan pelatihan model CNN menggunakan dataset gambar yang telah dianotasi dengan label sentimen. Model CNN kemudian dievaluasi menggunakan metrik evaluasi yang tepat, seperti akurasi, presisi, dan recall.

Hasil eksperimen menunjukkan bahwa pendekatan deep learning dengan menggunakan model CNN mampu memberikan klasifikasi sentimen gambar pada media sosial dengan tingkat akurasi yang cukup tinggi. Hal ini menunjukkan potensi penggunaan deep learning dalam analisis sentimen gambar, yang dapat memberikan wawasan yang berharga dalam pemahaman dan interpretasi konten visual di media sosial.

Daftar Pustaka

[1] Saif M. Mohammad, "Imagisaurus: An Interactive Visualizer of Valence and Emotion in the Roget's Thesaurus," *National Research Council Canada*.

- [2] A. Mondaref Jon, dan I. V. Papatungan, "ANALISIS SENTIMEN PADA MEDIA SOSIAL INSTAGRAM KLUB PERSIJA JAKARTA MENGGUNAKAN METODE NAIVE BAYES," *i Informatika – Program Sarjana Universitas Islam Indonesia*.
- [3] Ahmad Fauzi, M. F. Akbar, dan Y. F. A. Asmawan "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *i Informatika – Program Sarjana Universitas Islam Indonesia. JURNAL INFORMATIKA, Vol.6 No.1 April 2019, pp. 77-83*

Pengembangan Model Ontologi Cerita Rakyat Bali

Ni Made Julia Budiantari¹, Ngurah Agus Sanjaya ER²

¹Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹juliabudiantari004@student.unud.ac.id
²agus_sanjaya@unud.ac.id

Abstract

This research aims to develop an ontology related to the domain of Balinese folklore. Balinese folklore is an important part of Bali's cultural heritage and has moral values, as well as being an inspiration in art and performance. Ontology is a method to organize and categorize information in a structured way, and has been widely applied in various fields. However, research on the ontology of Balinese folklore is still limited. This research will propose and develop a comprehensive and structured ontology model for Balinese folklore. The steps to be taken include analysis of relevant ontologies, development of an ontology schema, organization and classification of information, and testing and validation of the ontology model built. The method used in this research is methontology, which includes the stages of specification, knowledge acquisition, conceptualization, integration, evaluation, and documentation. The result of this research is a structured Balinese folklore ontology that has been evaluated using SPARQL queries. This ontology can be used as a source of information and reference in learning and utilizing Balinese folklore.

Keywords: *Ontology, Balinese folklore, Methontology, Query SPARQL.*

1. Pendahuluan

Cerita rakyat Bali merupakan bagian penting dari warisan budaya pulau Bali. Kisah-kisah rakyat Bali diturunkan dari generasi ke generasi dan sering digunakan untuk mengajarkan pelajaran moral atau menjelaskan fenomena alam, serta menjadi sumber inspirasi seni, musik, dan pertunjukan tari [1]. Namun, ada kebutuhan untuk melestarikan dan mempromosikan kisah-kisah tersebut di era digital, di mana informasi dapat diakses dengan mudah melalui berbagai media. Ontologi adalah cabang ilmu komputer yang berhubungan dengan representasi formal pengetahuan yang menyediakan cara untuk mengatur dan mengkategorikan informasi secara terstruktur, membuatnya lebih mudah untuk mencari, mengambil, dan menganalisis [2]. Ontologi telah diterapkan dalam berbagai bidang, antara lain bahasa, budaya, dan seni. Dalam konteks cerita rakyat Bali, ontologi dapat digunakan untuk membuat model yang merepresentasikan struktur dan isi cerita tersebut.

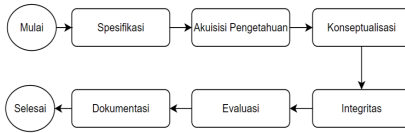
Kajian sebelumnya telah mengembangkan model ontologi untuk berbagai aspek budaya Bali, seperti model ontologi Bahasa Bali (Mahendra, dkk., 2023) dan seni pertunjukan (Virginia, dkk., 2022). Namun penelitian mengenai model ontologi cerita rakyat Bali masih sangat minim. Mengembangkan model ontologi cerita rakyat Bali dapat memberikan kerangka untuk mengatur dan mengkategorikan cerita-cerita ini, membuatnya lebih mudah diakses dan dipahami oleh khalayak yang lebih luas. Selain itu juga dapat memfasilitasi pengembangan aplikasi digital yang menggunakan cerita rakyat Bali sebagai sumber inspirasi.

Penelitian ini akan mengusulkan dan mengembangkan model ontologi rakyat Bali yang komprehensif dan terstruktur. Model ontologi tersebut akan memetakan cerita-cerita rakyat Bali, tokoh-tokoh, hubungan antara cerita, dan elemen-elemen penting lainnya dalam sebuah kerangka yang terstandarisasi. Langkah-langkah yang akan diambil meliputi analisis dan identifikasi ontologi yang relevan, pengembangan skema ontologi, pengorganisasian dan klasifikasi informasi, serta pengujian dan validasi model ontologi yang dibangun. Diharapkan,

dengan adanya model ontologi ini, akan memudahkan akses, pemahaman, dan penggunaan cerita rakyat Bali dalam berbagai konteks seperti pendidikan, pariwisata, dan pelestarian budaya.

2. Metode Penelitian

Pada penelitian ini, metode terstruktur methontologi digunakan untuk membangun ontologi dari awal. Metode ini mencakup serangkaian aktivitas, teknik, dan hasil yang dihasilkan oleh setiap aktivitas yang dilakukan dengan tekniknya masing-masing. Methontologi sangat menganjurkan untuk memanfaatkan ontologi yang telah ada. Berikut tahapan Methontologi.



Gambar 1. Metode Methontologi adopsi Cokorda Prammartha, dkk., 2017

2.1 Spesifikasi

Tujuan dari tahap spesifikasi adalah untuk menghasilkan dokumen spesifikasi ontologi formal, semi-formal, dan informal yang ditulis dalam bahasa alami dengan menggunakan pertanyaan kompetensi atau representasi menengah. Tahap ini akan memberikan spesifikasi terkait ontologi yang telah dibangun berikut merupakan deskripsi dari ontologi "Cerita Rakyat Bali":

- Domain: Cerita Rakyat Bali
- Tanggal: 09 Juni 2023
- Dirancang Oleh: Ni Made Julia Budiantari
- Diimplementasikan Oleh: Ni Made Julia Budiantari
- Level Formalitas: Formal
- Ruang Lingkup: Cerita Rakyat Daerah Bali
- Sumber Pengetahuan: Internet (E-book dan Website cerita Rakyat Bali)

2.2 Akuisisi Pengetahuan

Tahap ini merupakan tahap terpisah dalam pembangunan ontologi. Sebagian besar, tahap ini telah selesai bersamaan dengan tahap spesifikasi dan berlanjut seiring dengan proses pengembangan ontologi. Tujuan dari tahap ini adalah untuk memperoleh informasi yang relevan tentang ontologi Cerita Rakyat Bali yang akan dibangun. Tahapan akuisisi pengetahuan dalam penelitian ini adalah sebagai berikut:

- Membaca literatur untuk mendapatkan informasi tentang domain Cerita Rakyat Bali.
- Analisis teks informal untuk mempelajari konsep-konsep utama yang ditemukan dalam buku dan studi pegangan; dan
- Analisis teks formal. Struktur yang akan dideteksi (seperti definisi dan penegasan) dan jenis pengetahuan yang diberikan masing-masing (seperti konsep, atribut, nilai, dan hubungan).

2.3 Konseptualisasi

Pada tahap ini, model konseptual pengetahuan domain akan dibangun. Model ini akan menggambarkan masalah dan solusinya dalam kosakata domain yang telah diidentifikasi pada tahap spesifikasi. Membangun *Glossary of Terms* (GT) yang lengkap meliputi konsep, contoh, kata kerja, dan properti. GT mencari dan mengumpulkan semua pengetahuan yang mungkin digunakan tentang domain dan artinya. Tujuan konseptualisasi adalah untuk menyusun domain

pengetahuan dalam bentuk konseptual dan mempertahankan dan mengelola pengetahuan yang diperoleh selama proses akuisisi pengetahuan. Setelah membangun model konseptual, metodologi akan berubah untuk mengubah model tersebut menjadi model formal, yang akan digunakan dalam bahasa ontologi. Konsep, instance, verba, dan properti termasuk dalam membangun domain pengetahuan. Oleh karena itu, glosarium menemukan dan mengumpulkan semua pengetahuan tentang domain yang dapat digunakan, dan kemudian digunakan untuk membuat kelas dan sub-kelas.

2.4 Integrasi dan Impementasi

Menggunakan definisi ontologi yang sudah ada untuk membangun ontologi baru memungkinkan integrasi, sehingga pembangunan ontologi baru tidak perlu dimulai dari awal. Hasil dari tahap ini melalui perangkat lunak Protégé akan dilakukan pendefinisian kembali dan penerapan perancangan ontologi. Penelitian ini memanfaatkan perangkat lunak bernama Protégé 4.3, yang dibuat oleh Pusat Penelitian Informatika Biomedis Stanford di Fakultas Kedokteran Universitas Stanford dan bertujuan untuk membantu dalam pembentukan ontologi yang berbasis pada pengetahuan dasar system [2]. Setiap bagian ontologi didefinisikan dengan hasil dari setiap langkah tugas dalam Metodologi. Aplikasi Protégé 4.3 Ontography digunakan untuk memformalkan rancangan konseptual yang telah dibuat. Ini memungkinkan pembuatan model ontologi di atas laporan ini. Tujuan dari tahap implementasi adalah untuk menerapkan perancangan ontologi yang dibangun. Tahap ini mencakup:

- a. Properti objek mendefinisikan konsep atau makna dari suatu domain.
- b. Properti data mendefinisikan atribut class; dan
- c. Properti individu mendefinisikan atribut individu.

2.5 Evaluasi

Pada tahap evaluasi dilakukan penilaian teknis dari ontologi, lingkungan perangkat lunak, dan dokumentasi yang berkaitan dengan kerangka referensi pada setiap tahap dan di antara tahap kehidupan cycle mereka. Evaluasi terdiri dari dua proses: verifikasi dan validasi. Verifikasi mengacu pada proses teknis yang menjamin kebenaran ontologi, lingkungan perangkat lunak, dan dokumentasi yang berkaitan dengan kerangka referensi pada setiap tahap dan di antara tahap kehidupan cycle mereka. Validasi menjamin lingkungan perangkat lunak dan dokumentasi telah sesuai sistem yang ingin diwakili. Pada titik ini, penulis melakukan proses evaluasi pada ontologi yang telah dibuat sebelumnya. Untuk melakukan evaluasi, pengujian dilakukan dengan query SPARQL pada aplikasi Protégé 4.3. Query yang diinginkan dapat diubah menjadi query SPARQL, yang memungkinkan penampilan hasil ontologi yang dihasilkan.

2.6 Dokumentasi

Pada tahap terakhir ini, proses dokumentasi dilakukan baik dalam kode ontologi, teks bahasa alami yang disertakan dengan definisi formal, maupun makalah yang diterbitkan dalam konferensi dan jurnal yang mengatur masalah penting dari ontologi yang sudah dibangun. Laporan ini mencakup hasil dokumentasi dari penelitian tentang pengembangan ontologi cerita rakyat Bali.

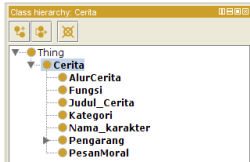
3. Hasil dan Diskusi

Dalam penelitian ini, sebuah ontologi berdomain Cerita Rakyat Bali dibangun. Dalam rangka melestarikan cerita rakyat sebagai bagian penting dari identitas budaya, peneliti merancang dan membangun sebuah model ontologi yang dapat menggambarkan elemen-elemen cerita rakyat secara sistematis. Dalam penjabaran hasil, akan membahas proses perancangan ontologi, implementasi data ke ontologi yang dibuat, serta evaluasi model ontologi.

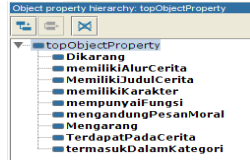
3.1. Pembangunan Ontologi

Bagian ini akan melakukan desain ontologi domain Cerita Rakyat Bali. Hal pertama yang harus dilakukan adalah membuat hirarki kelas dan subkelas dari ontologi, seperti yang ditunjukkan pada

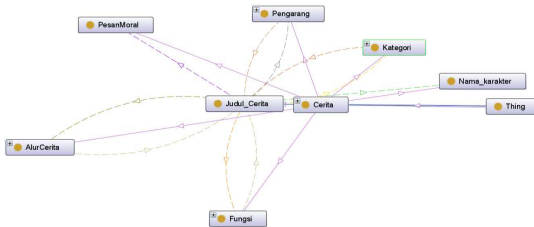
Gambar 2. Selanjutnya, rancang objek properti cerita dari hirarki domain yang menunjukkan implementasi properti objek yang digunakan untuk menghubungkan *instances* atau individu satu sama lain, ditunjukkan pada gambar 3. Kemudian dibuatkan relasi entitas kelas dan subkelas ontograf, yang ditunjukkan pada gambar 4.



Gambar 2. Hirarki Domain Cerita Rakyat Bali



Gambar 3. Hirarki Objek Properti

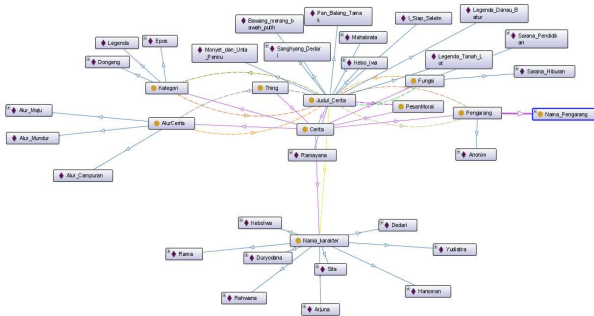


Gambar 4. Konseptualisasi Domain Cerita Rakyat Bali

Berdasarkan gambar 2, ditunjukkan bahwa terdapat class-class Cerita, Judul Cerita, Pengarang, Alur Cerita, Fungsi Cerita, Kategori Cerita, Nama karakter Cerita, Pesan Moral Cerita. Kemudian pada gambar 3, ditunjukkan properti penghubung antar individunya. Sedangkan Gambar 4, menunjukkan relasi antar *class*. Terlihat bahwa semua class terhubung pada judul cerita dan *superclass* cerita yang merupakan tolak ukur parameter pembangunan ontologinya.

3.2. Implementasi

Berdasarkan perancangan ontologi yang telah dibuat, selanjutnya peneliti mengimplementasikan rancangan ontologi pada data yang telah dikumpulkan melalui berbagai sumber internet seperti e-book dan website cerita rakyat nusantara. Tahap implementasi juga menggunakan bantuan tool protégé 4.3. Berikut hasilnya implementasinya, ditunjukkan pada gambar 5.



Gambar 5. Ontograf Hasil Implementasi Data Terhadap Pembangunan Ontologi.

3.3 Evaluasi

Pada tahap evaluasi ini dapat dilakukan dengan mengeksekusi query menggunakan query SPARQL pada aplikasi Protégé versi 4.3. Pertanyaan yang diinginkan dapat diubah menjadi formulir kueri SPARQL. Kemudian, hasil yang terkandung dalam model ontologi yang dibuat atau dibangun ditampilkan. Berikut hasil evaluasi ontology yang ditunjukkan pada gambar 6 dan gambar 7.

```

SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX cerita: <http://www.semanticweb.org/alltw/ontologies/2023/5/untitled-ontology-25#>

SELECT * WHERE {
?CeritaBali cerita:memilikiAlurCerita cerita:Alur_Maju.
?CeritaBali cerita:Dikarang cerita:Anonim
}

CeritaBali
Kebo_Iwa
SangHyang_Dedari
    
```

Gambar 6. Hasil 1 Pencarian Menggunakan Query SPARQL

Pada gambar 6, ditunjukkan bahwa pencarian cerita yang fokusnya untuk mencari judul cerita yang memiliki "alur maju" dan juga dikarang oleh pengarang "anonim" menggunakan Query SPARQL berhasil didapatkan. Hasilnya adalah "Kebo Iwa" dan "Sanghyang Dedari".

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX cerita: <http://www.semanticweb.org/altw/ontologies/2023/5/untitled-ontology-25#>

SELECT * WHERE { ?CeritaBali cerita:memilikiAlurCerita cerita:Alur_Maju }
```

```
CeritaBali
SangHyang_Dedari
Ramayana
Mahabarata
Kebo_Iwa
```

Gambar 7. Hasil 2 Pencarian Menggunakan Query SPARQL

Pada gambar 7, ditunjukkan bahwa pencarian cerita yang fokusnya untuk mencari judul cerita yang memiliki "alur maju" menggunakan Query SPARQL berhasil didapatkan. Hasilnya adalah "Kebo Iwa", "Sanghyang Dedari", "Mahabrata", "Ramayana".

4. Kesimpulan

Berdasarkan hasil perancangan yang telah dilakukan, Ontologi terkait ranah cerita rakyat Bali telah selesai berdasarkan hasil dan pembahasan yang dilakukan. Ontologi dikembangkan menggunakan aplikasi Protégé 4.3 dengan metode Methontology, menghasilkan 1 superclass, 7 class, 9 object properties, dan 100 instance untuk setiap class. Model yang diusulkan juga dievaluasi dengan mengajukan beberapa pertanyaan kepada pengguna untuk mengakses informasi tentang cerita rakyat Bali. Oleh karena itu diharapkan model ontologi dapat memberikan informasi yang sistematis terkait cerita rakyat Bali. Pengembangan ontologi cerita rakyat Bali diharapkan dapat dijadikan landasan untuk mengembangkan sistem yang berkaitan dengan cerita rakyat Indonesia.

Daftar Pustaka

- [1] I. Nyoman Suwija, I. Darmada, and I. Mulyawan, "Kumpulan Satua (Dongeng Rakyat Bali)," Pelawa Sari, 2019.
- [2] I. Made Yoga Mahendra, Cokorda Pramatha, I Putu Gede Hendra Suputra, I Made Widiartha, I Gede Arta Wibawa, and I Wayan Supriana, "Pengembangan Model Ontologi Pada Sistem Informasi Bahasa Bali," Jurnal Elektronik Ilmu Komputer Udayana, vol. 11, no. 3, pp. 489, Feb. 2023, p-ISSN: 2301-5373, e-ISSN: 2654-5101.
- [3] Gloria Virginia, Budi Susanto, Umi Probeyekti, dan Silvanus Satno Nugraha, "Semantic Web Seni Pertunjukan Indonesia," TRANSFORMATIKA, vol. 19, tidak. 2, hlm. 124-137, Jan. 2022, p-ISSN: 1693-3656, e-ISSN: 2460-6731.
- [4] Cokorda Pramatha, Joseph Davis, dan Kevin Kuan, "Digital Preservation of Cultural Heritage: An Ontology-Based Approach," dalam Konferensi Australasia tentang Sistem Informasi, Desember 2017, hlm. 1-12.
- [5] C. Pramatha, "Pengembangan Ontologi Tujuan Wisata Bali Dengan Pendekatan Kulkul Knowledge Framework," Jurnal SINTECH (Sains dan Teknologi Informasi), vol. 3, tidak. 2, hlm. 77-89, 2020.

Implementasi Aplikasi “SahabatAnabul” Berbasis User Centered Design Sebagai Website Pangkalan Informasi Anabul

Ni Luh Putu Ayu Siwastuti Cayadewi^{a1}, Made Agung Raharja^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia

¹ayunathan1609@gmail.com

²made.agung@unud.ac.id

Abstract

Currently, many people keep animals in their homes. Most of the animals they have are dogs and cats. Thus, they need a lot of information about the animals they keep. Therefore, to make it easier for people to find information about Anabul, an information website about Anabul was created, starting from the types of Anabul, its characteristics, what Anabul needs, and others. This journal produces a website based SahabatAnabul application. SahabatAnabul is a website to share all information about anabul, especially for dogs and cats. The resulting website can help people who have pets or want to know information about the types of dogs and cat breeds that exist in the world.

Keywords: Friends of Anabul, Information, Anabul, Applications, Websites.

1. Pendahuluan

Manusia selama hidupnya tidak hanya berinteraksi dengan manusia lainnya saja. Terkadang manusia juga berinteraksi dengan hewan baik itu hewan peliharaan atau hewan liar yang mereka temui saat berpergian. Interaksi ini juga dapat membuat manusia melepas penat, sebagai hobi, ataupun sebagai teman bermain agar tidak merasa sendiri dan kesepian. Orang – orang pastinya mencari semua informasi seputaran anabul mulai dari jenis anabulnya, apasaja yang dibutuhkan anabul dalam memeliharanya, dan informasi lainnya agar mereka siap untuk memelihara anabul ataupun hanya untuk menambah informasi terkait anabul. Untuk itu, dibuatlah website yang berisikan informasi seputaran anabul yang bermanfaat untuk orang – orang agar dapat mengetahui jenis – jenis anabul yang ada, kebutuhannya apa saja, dan apakah cocok untuk dipelihara atau tidak.

Berkembangnya teknologi saat ini, maka dibutuhkan sebuah sistem yang berisi segala informasi mengenai anabul yang mudah diakses dan bermanfaat bagi orang – orang. Dan untuk memudahkan orang – orang dalam mencari informasi tersebut, maka aplikasi SahabatAnabul ini dibuat berbasis website.

2. Metode Penelitian

2.1 Data Penelitian

Data yang digunakan dalam penulisan laporan ini adalah data sekunder. Data sekunder adalah data yang sudah ada sebelumnya yang dikumpulkan oleh peneliti lain. Keuntungan menggunakan data sekunder yaitu biaya dan waktu yang dikeluarkan lebih sedikit dari pada menggunakan data primer. Data yang didapat berupa berbagai macam informasi seputaran hewan yang difokuskan ke anjing dan kucing.

2.2 Teknik Pengumpulan Data

Untuk mengumpulkan data, penulis menggunakan metode studi pustaka. Metode Studi Pustaka dilakukan dengan membaca dan mencari tau segala informasi di jurnal atau buku – buku literatur yang ada hubungannya dengan informasi anjing dan kucing agar tidak ada informasi yang salah.

2.3 Metode yang Digunakan

Pada jurnal ini, penulis menggunakan metode prototype dan user centered design untuk membuat website. Adapun beberapa penjelasan dari metode ini yaitu sebagai berikut:

a. Metode Prototype

Prototype atau prototipe adalah sebuah metode dalam pengembangan produk dengan cara membuat rancangan, sampel, atau model dengan tujuan pengujian konsep atau proses kerja dari produk. Prototype sendiri bukanlah produk final yang nantinya akan diedarkan. Prototype dibuat untuk kebutuhan awal development software dan untuk mengetahui apakah fitur dan fungsi dalam program berjalan sesuai dengan kebutuhan yang telah direncanakan. Sehingga pengembang produk dapat mengetahui kekurangan dan kesalahan lebih awal sebelum mengimplementasikan fitur lain ke dalam produk dan menulis produk [1].

b. Tujuan Prototype

Tujuan dari dibuatnya prototype yaitu untuk mengidentifikasi masalah sejak dini agar dapat diperbaiki sebelum melakukan produksi secara penuh. Prototype juga dapat memunculkan ide – ide baru terkait design yang dibuat oleh pengguna sehingga produk yang dihasilkan sesuai dengan keinginan pembuat.

c. Manfaat Prototype

Metode prototype memiliki beberapa manfaat yang bisa dirasakan ketika mengembangkan produk, yaitu:

- Dapat menghemat biaya produksi
Banyak orang mungkin akan mengatakan membuat prototype membutuhkan banyak biaya, namun nyata nya membuat prototype sangat menghemat biaya produksi. Sehingga dapat menghemat biaya produksi dan menghemat waktu.
- Menjadi acuan atau patokan dalam pengembangan produk yang nantinya akan dihasilkan
Prototype yang dihasilkan merupakan hasil dari analisis dari kebutuhan pasar terhadap jenis produk yang dihasilkan nantinya. Prototype juga sering memunculkan ide – ide baru pada pembuatan ataupun update produk terbaru.
- Mengetahui kebutuhan konsumen lebih dulu
Dengan mengetahui kebutuhan konsumen lebih dulu, kita dapat mengetahui apa saja kebutuhan dari pengguna sehingga produk yang kita buat nantinya akan selesai lebih cepat.
- Menjadi wadah keinginan pengguna
Manfaat lainnya dari prototype yaitu menjadi wadah keinginan pengguna. Dengan ini, klien dapat memberikan masukan dan saran secara bebas agar produk yang dihasilkan lebih lengkap sehingga ada fitur – fitur baru yang akan berguna dimasa depan.

3. Hasil dan Diskusi

Pada penelitian ini menghasilkan website yang diberi nama SahabatAnabul. Website ini berfungsi sebagai wadah informasi dari anabul yaitu anjing dan kucing. Sehingga pengguna dapat mengetahui informasi mengenai anabul mulai dari ciri fisiknya, tingkah lakunya, dan kebutuhan anabul. Pada halaman ini, merupakan halaman awal pada website SahabatAnabul.



Gambar 1. Tampilan awal website

Dilanjutkan dengan tampilan berisikan informasi mengenai website SahabatAnabul. Jika memilih gambar kucing, maka user akan masuk ke halaman yang berisikan informasi mengenai kucing. Jika user memilih gambar anjing, maka user akan masuk ke halaman yang berisikan informasi seputaran anjing.



Gambar 2. Halaman informasi website

Berikutnya adalah halaman mengenai informasi seputaran kucing jika user memilih gambar kucing atau informasi mengenai anjing jika user memilih gambar anjing.



Gambar 3. Informasi Kucing



Gambar 4. Informasi Anjing

Setelah halaman mengenai informasi kucing atau informasi anjing, dilanjutkan dengan halaman berbagai ras kucing beserta foto dari kucing tersebut untuk halaman kucing dan untuk halaman anjing akan ada berbagai jenis anjing beserta fotonya. Jika salah satu dari foto tersebut di klik oleh user, maka user akan masuk ke halaman informasi dari ras kucing atau jenis anjing tersebut.



Gambar 5. Ras Kucing



Gambar 6. Jenis Anjing

Jika user memilih ras kucing ragdoll, maka informasi yang ditampilkan adalah informasi mengenai ras kucing ragdoll dan jika user memilih jenis anjing golden retriever maka informasi yang ditampilkan adalah informasi seputaran anjing golden retriever.



Gambar 7. Halaman Ras Kucing Ragdoll



Gambar 8. Halaman Jenis Anjing Golden Retriever

4. Kesimpulan

Dari penjelasan diatas, kesimpulan dari jurnal ini bahwa pembuatan website SahabatAnubul sangat bermanfaat. Website berisikan informasi mengenai jenis anjing dan ras kucing yang tentu saja dapat menambah informasi dan wawasan mengenai anubul. Terutama yang ingin memelihara anubul, pastinya dengan mengetahui informasi tersebut, mereka dapat menentukan ingin memelihara anubul jenis apa dan anubul mana yang cocok dipelihara oleh mereka. Adapun beberapa saran yang dapat penulis harapkan dalam upaya pengembangan dan perbaikan terhadap website.

Daftar Pustaka

- [1] S. Rony, 11 Agustus 2021, "Apa Itu Prototype? Kenapa Itu Penting?," diambil dari <https://www.dicoding.com/blog/apa-itu-prototype-kenapa-itu-penting/>.
- [2] N. Renaningtias, D. Apriliani, " Penerapan Metode Prototype Pada Pengembangan Sistem Informasi Tugas Akhir Mahasiswa," Jurnal Rekursif, vol. 9, no. 1, pp 92-98, 2021.
- [3] E. W. Fridayanthie, Haryanto, and T. Tsabitas, "Penerapan Metode Prototype Pada Perancangan Sistem Informasi Penggajian Karyawan (Persis Gawan) Berbasis Web," Paradigma, vol. 23, no. 2, pp 151-157, 2021.
- [4] A. A. Murwanti, K. N. Ramadhani, and P. E. Yunanto, "Pengenalan Ras Kucing Menggunakan Metode Ekstraksi Ciri Pyramid Histogram of Oriented Gradients (P-HOG)," e-Proceeding of Engineering, vol. 7, no. 3, pp 9877-9894, 2020.
- [5] M. Sari, 23 Oktober 2015, "43 Jenis Jenis Anjing beserta Gambarnya," diambil dari <https://dosenbiologi.com/hewan/jenis-jenis-anjing>.

Halaman ini sengaja dibiarkan kosong

Penyusunan Sistem Rekomendasi Produk Diecast Mobil Dengan Metode Content-Based Filtering (CBF)

Anak Agung Aditya Nugraha^{a1}, Ngurah Agus Sanjaya ER^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹gungwah1927@gmail.com
²agus_sanjaya@unud.ac.id

Abstract

The growing popularity of diecast car collections has created a demand for efficient recommendation systems to assist collectors in discovering new products. This study focuses on the development of a content-based filtering (CBF) recommendation system for diecast car products. The system employs the TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity techniques to calculate the relevance between products and user preferences. By analyzing the textual features of diecast car products, such as brand, model, and specifications, the CBF system generates personalized recommendations based on similarity scores. The evaluation of the system's performance demonstrates its effectiveness in providing accurate and relevant recommendations, which enhance the user experience and facilitate the exploration of the diecast car market.

Keywords: CBF, Diecast cars, Recommendation System, TF-IDF, Cosine Similarity

1. Pendahuluan

Dengan munculnya berbagai platform e-commerce serta pesatnya pertumbuhan industri diecast mobil, masyarakat semakin bergantung kepada teknologi informasi dengan tujuan memperoleh rekomendasi produk yang sesuai dengan preferensi mereka. Diecast mobil telah menjadi tren yang cukup populer di berbagai kalangan, khususnya kolektor serta penggemar otomotif dengan variasi model dan brand yang terus berkembang seiring zaman. Akan tetapi dengan semakin banyaknya pilihan yang tersedia, konsumen sering menghadapi kesulitan dalam mengelola berbagai opsi serta menemukan produk yang tepat untuk mereka. Kebutuhan akan sistem rekomendasi yang efektif dan efisien menjadi semakin meningkat. Sistem rekomendasi dapat membantu konsumen dalam menemukan produk diecast mobil yang sesuai dengan keinginan mereka serta mempermudah proses pemilihan dan pembelian. Dengan terciptanya sistem rekomendasi yang andal, konsumen dapat menghemat waktu serta usaha dalam mencari produk yang sesuai serta meningkatkan kepuasan mereka dengan hasil pembelian. Agar dapat meningkatkan akurasi sistem rekomendasi, pendekatan yang efektif dalam menganalisis konten produk yaitu menggunakan metode Content-Based Filtering (CBF). Metode CBF berfokus pada karakteristik atau atribut dari produk tersebut, seperti deskripsi, fitur, ataupun kategori. Maka dari itu, pada penelitian ini akan dilakukan penyusunan sistem rekomendasi produk diecast mobil menggunakan metode CBF yang akan dibantu oleh penggunaan Term Frequency-Inverse Document Frequency (TF-IDF) dan juga Cosine Similarity berdasarkan dataset deskripsi produk diecast mobil yang tersedia. Penggabungan metode TF-IDF dan Cosine Similarity ini bertujuan mengembangkan sistem rekomendasi yang dapat memberikan rekomendasi produk diecast mobil yang akurat dan juga relevan. Dengan adanya sistem rekomendasi ini, diharapkan konsumen dapat menemukan produk diecast mobil dengan mudah dan sesuai dengan preferensi, serta meningkatkan efisiensi dan kepuasan dalam proses pembelian.

2. Metode Penelitian

2.1 Sistem Rekomendasi

Sistem rekomendasi didefinisikan sebagai suatu sistem yang di mana bertugas memberikan saran/rekomendasi kepada pengguna sistem yang bersifat personal (tiap pengguna memiliki saran yang berbeda) [1]. Sistem ini dapat ditemukan dalam berbagai konteks, seperti e-commerce, layanan streaming musik, platform media sosial, dan lain sebagainya. Sistem rekomendasi memiliki tujuan dalam memahami preferensi pengguna serta menyajikan konten dan produk yang relevan. Pemahaman preferensi tersebut dilakukan dapat dengan pengumpulan data secara eksplisit (preferensi yang diberikan secara langsung) atau secara implisit (melacak perilaku saat menggunakan sistem). Sistem rekomendasi dibagi dalam beberapa klasifikasi berdasarkan metode rekomendasi yang sering digunakan, yaitu content-based recommendation, collaborative filtering recommendation, knowledge-based recommendation, serta hybrid recommendation [2].

2.2 Content-Based Filtering

Content Based Filtering (CBF) didefinisikan sebagai salah satu metode dalam sistem rekomendasi yang di mana menggunakan informasi mengenai atribut ataupun karakteristik item untuk memberikan rekomendasi kepada pengguna. Metode ini berfokus pada kesesuaian konten item dengan preferensi pengguna [3]. Content-Based Filtering memiliki keuntungan dalam kemampuannya untuk memberikan rekomendasi yang personal dan relevan berdasarkan preferensi pengguna. Sistem tidak bergantung pada data pengguna dan dapat bekerja dengan baik untuk pengguna baru yang belum memberikan banyak umpan balik.

2.3 Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) didefinisikan sebagai algoritma/metode yang digunakan dalam pemrosesan bahasa alami (NLP) dan juga sistem informasi dengan tujuan mengukur seberapa penting sebuah kata dalam sebuah dokumen atau korpus teks. TF-IDF berasal dari bidang information retrieval, namun pada saat ini banyak digunakan dalam perbandingan dokumen [4]. Metode ini menggabungkan dua buah komponen utama, yaitu frekuensi kata dalam dokumen (TF) dan invers dari frekuensi kata dalam korpus (IDF). Metode TF-IDF sering digunakan dalam berbagai aplikasi pemrosesan bahasa alami seperti klasifikasi teks, analisis sentimen, dan juga sistem rekomendasi berbasis teks. Rumus dari TF-IDF yaitu :

$$TF - IDF = TF * IDF \quad (1)$$

- a. Term Frequency (TF), didefinisikan sebagai pengukuran seberapa sering sebuah kata muncul pada dokumen. Salah satu rumus dari TF yaitu:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah total kata dalam dokumen}} \quad (2)$$

- b. Inverse Document Frequency (IDF), didefinisikan sebagai pengukuran seberapa penting sebuah kata dengan memperhatikan seberapa umum atau jarang kata tersebut dalam keseluruhan korpus teks. Rumus yang digunakan dalam IDF yaitu:

$$IDF = \log \frac{\text{jumlah dokumen pada korpus}}{\text{jumlah dokumen yang mengandung kata}} \quad (3)$$

dalam rumus tersebut dihitung logaritma dari rasio jumlah total dokumen dalam korpus dengan jumlah dokumen yang mengandung kata tersebut yang dilakukan untuk memberikan bobot yang lebih besar pada kata yang jarang muncul dalam korpus, sehingga kata tersebut dianggap lebih penting.

2.4 Cosine Similarity

Cosine similarity didefinisikan sebagai metode yang digunakan dalam mengukur sejauh mana dua vektor berada dalam arah yang sama atau memiliki kesamaan dalam ruang vektor. Metode ini sering digunakan dalam pemrosesan bahasa alami serta sistem rekomendasi untuk membandingkan kemiripan antar dua dokumen atau dua vektor teks. Rumus cosine similarity mengukur sejauh mana dua vektor berada dalam arah yang sama atau memiliki kesamaan dalam ruang vektor. Nilai cosine similarity berkisar antara -1 hingga 1, dengan 1 menunjukkan arti kesamaan sempurna antar dua vektor, sedangkan 0 menunjukkan ketidaksamaan atau ketidakberhubungan antar dua vektor, dan -1 menunjukkan kebalikan atau keberlawanan antar dua vektor. Cosine similarity memiliki rumus sebagai berikut :

$$\cos \theta = \frac{(A \cdot B)}{\|A\| \|B\|} \quad (4)$$

- A dan B, merupakan vektor yang mewakili dua dokumen atau dua vektor teks yang ingin dibandingkan. Setiap komponen vektor mewakili nilai frekuensi atau bobot kata dalam dokumen.
- $(A \cdot B)$, merupakan hasil perkalian dot (dot product) antara vektor A dan vektor B (jumlah dari perkalian tiap komponen vektor yang sesuai).
- $\|A\| \|B\|$, merupakan panjang Euclidean (magnitude) dari masing-masing vektor A dan B yang dihitung dengan mengambil akar kuadrat dari jumlah kuadrat tiap komponen vektor.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data dikumpulkan dengan tujuan untuk memperoleh dataset yang mencakup informasi produk diecast mobil yang diperlukan dalam menyusun sistem rekomendasi. Dalam penelitian ini, data dikumpulkan menggunakan metode web scrapping dari website HobbyDB (<https://www.hobbydb.com/>) yang menyediakan berbagai list diecast dari berbagai model dan brand. Data yang diambil berupa link gambar, model, brand, skala, dan tahun rilis. Data tersebut disimpan dalam format CSV bernama hobbyDB.

3.2. Preprocessing Data

Setelah data dikumpulkan, dataset dipreprocessing dengan melalui beberapa proses, yaitu [5]:

- Case folding, yaitu mengubah data yang digunakan menjadi huruf kecil (lowercase) serta menghilangkan atau menghapus seluruh tanda baca yang terdapat pada data.
- Tokenizing, yaitu pemotongan string input dari data berdasarkan tiap kata yang menyusun data.
- Filtering (stopwords removal), yaitu menghilangkan atau menghapus kata yang kurang penting (tidak memiliki arti) pada data yang diproses.
- Stemming, yaitu mengubah kata menjadi kata dasar dengan menghilangkan imbuhan awalan, akhiran, ataupun kombinasinya.

3.3. Menghitung TF- IDF

Setelah dipreprocessing, dilakukan perhitungan TF-IDF. Perhitungan dilakukan dengan bobot TF-IDF untuk setiap kata dalam deskripsi produk berdasarkan frekuensi kemunculan kata dalam dataset serta seluruh dataset. Representasi tiap produk dapat dilakukan dengan berbagai bentuk, yaitu representasi matriks, representasi graf, representasi vektor, representasi model statistik, dan lain sebagainya.

3.4. Menghitung Cosine Similarity

Cosine similarity digunakan dalam tujuan mengukur tingkat kesamaan antara vektor representasi produk diecast mobil. Dengan membandingkan vektor produk diecast mobil tersebut dengan produk lainnya, maka akan dapat dihitung skor cosine similarity yang mencerminkan sejauh mana kedua produk memiliki kesamaan.

3.5. Pembentukan Sistem Rekomendasi

Dari hasil cosine similarity, pengguna dapat mempertimbangkan produk yang paling mirip dengan identifikasi produk diecast mobil yang memiliki skor tertinggi/ Dengan demikian, disusunlah sistem rekomendasi produk diecast mobil yang paling relevan dan sesuai dengan preferensi pengguna.

3.6. Evaluasi Sistem Rekomendasi

Dilakukan evaluasi kinerja sistem rekomendasi yang telah disusun. Evaluasi ini dilakukan dengan matriks evaluasi yaitu precision, recall, atau F1-score dengan tujuan mengukur sejauh mana sistem rekomendasi yang telah disusun memberikan rekomendasi yang akurat dan relevan.

4. Kesimpulan

Dalam penelitian ini telah disusun sebuah sistem rekomendasi produk diecast mobil yang menggunakan metode Content-Based Filtering (CBF) dengan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity. Sistem rekomendasi yang dikembangkan ini bertujuan untuk membantu konsumen dalam menemukan produk diecast mobil yang sesuai dengan preferensi mereka. Sistem rekomendasi ini diharapkan dapat memberikan kontribusi positif dalam memenuhi kebutuhan konsumen dan mendorong pertumbuhan industri diecast mobil.

Daftar Pustaka

- [1] N. Ula, C. Setianingsih, and R. A. Nugrahaeni, "SISTEM REKOMENDASI LAGU DENGAN METODE CONTENT-BASED FILTERING BERBASIS WEBSITE" e-Proceeding of Engineering, vol. 8, no. 6, pp. 12193–12199, 2021.
- [2] I. Diah, I. F. Rahmad, and A. Saleh, "IMPLEMENTASI RECOMMENDER SYSTEM PADA PEMILIHAN KAMERA MENGGUNAKAN CONTENT BASED DAN COLLABORATIVE FILTERING" IT (INFORMATIC TECHNIQUE) JOURNAL, vol. 7, no. 1, pp. 1-12, 2019.
- [3] Daniel, B. Muiyawan, and T. Sutrisno, "PEMBUATAN APLIKASI E-COMMERCE BERBASIS WEB DENGAN FITUR REKOMENDASI MENGGUNAKAN METODE CONTENT-BASED FILTERING" JURNAL ILMU KOMPUTER DAN SISTEM INFORMASI, vol. 10, no. 1, pp. 1-6, 2022.
- [4] A. R. Harischandra, M. F. A. Pratama, Felix, and A. P. Laia, "APLIKASI PENDUKUNG DESAIN INTERIOR DENGAN SISTEM REKOMENDASI BERDASARKAN NAMA BRAND PERABOT MENGGUNAKAN ALGORITMA CONTENT-BASED FILTERING BERBASIS WEB" Jurnal SIFO Mikroskil 23, vol. 23, no. 1, pp. 1-16, 2022.
- [5] M. Alkaff, H. Khatimi, and A. Eriady, "Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering" Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 1, pp. 193-202, 2020.

Sistem Penjualan Merchandise Berbasis Aplikasi Mobile

I Dewa Gede Partha Wijaya^{a1}, I Made Widiartha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹parthawijaya04@gmail.com
²madewidiartha@unud.ac.id

Abstract

The rapid growth of mobile technology has revolutionized various industries, including merchandise sales. This paper presents a study on the development of a mobile application-based merchandise sales system aimed at replacing manual recording reports by merchandisers and expanding the range of reportable data. By harnessing the capabilities of mobile devices, the research focuses on analyzing merchandisers' requirements and designing user-friendly interfaces. Through thorough testing and evaluation, the system showcases its reliability and functionality. This study contributes to the existing body of knowledge by offering valuable insights into the advantages and challenges of adopting mobile technology in merchandise sales.

Keywords: Mobile Application, Sales Report, SDLC, Flowchart, UML.

1. Pendahuluan

Perusahaan distribusi barang mengandalkan peran penting seorang merchandiser dalam menjalankan kegiatan penjualan dan distribusi produk ke berbagai outlet. Merchandiser bertugas untuk memastikan ketersediaan barang di setiap outlet, mencatat laporan penjualan, serta menjaga hubungan baik dengan pemilik toko. Namun, pada umumnya, metode pencatatan laporan yang masih menggunakan cara tertulis menjadi tantangan tersendiri bagi para merchandiser. Merchandiser harus mencatat secara manual setiap transaksi penjualan dan mengirimkan laporan secara fisik ke kantor pusat. Selain itu, proses ini juga rentan terhadap kesalahan manusia dan kehilangan data yang penting. Dalam era teknologi yang semakin maju, penggunaan metode yang lebih efisien dan efektif menjadi kebutuhan mendesak. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem penjualan merchandise berbasis aplikasi mobile yang akan menggantikan metode pencatatan laporan yang masih tertulis.

2. Metode Penelitian

Dalam kajian ilmiah ini, metode penelitian yang digunakan pada penelitian ini adalah metode penelitian kualitatif dan studi kasus. Metode penelitian kualitatif meliputi beberapa teknik pengumpulan data, seperti observasi dan wawancara. Observasi dilakukan untuk mengamati perilaku atau fenomena secara langsung, sedangkan wawancara dilakukan untuk mendapatkan informasi dari narasumber yang terkait dengan fenomena yang diteliti [1]. Penulis juga akan menggunakan flowchart, untuk menggambarkan langkah-langkah yang digambarkan dengan bentuk, garis, atau panah untuk melakukan sebuah proses dalam suatu program untuk mendokumentasikan, merencanakan, dan mengomunikasikan proses yang cukup rumit menjadi sebuah diagram [5]. Tujuannya agar proses yang rumit ini lebih mudah dibaca dan dipahami oleh orang awam sekalipun.

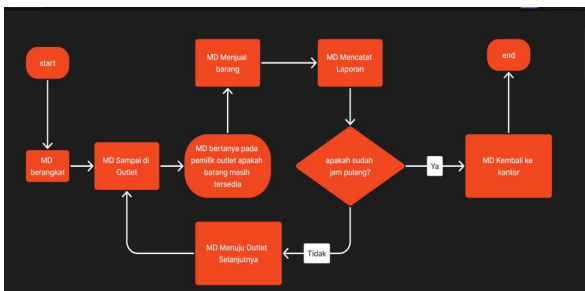
2.1. Wawancara dengan Pegawai Perusahaan

Selama wawancara dengan salah seorang pegawai di perusahaan distributor, saya menggali tentang harapan mereka terkait data laporan yang dihasilkan oleh para merchandiser. Dalam

percakapan tersebut, pegawai tersebut menyampaikan bahwa mereka mengharapkan data laporan yang lebih akurat, terperinci, dan data yang relevan untuk mengatasi permasalahan-permasalahan merchandiser adalah informasi terkait konsumsi bensin pada setiap perjalanan merchandiser ke outlet, foto merchandiser di outlet, serta waktu yang dihabiskan oleh merchandiser pada masing-masing outlet. Data ini dianggap penting karena akan membantu perusahaan untuk melakukan analisis biaya operasional, mengalokasikan bahan bakar dengan lebih efisien, serta memperbaiki rencana rutinitas kerja merchandiser sehingga mereka dapat mengoptimalkan waktu dan produktivitas mereka di lapangan. Penulis juga melihat bahwa, pegawai pada perusahaan distribusi ini bekerja menggunakan excel, jadi saat data laporan tertulis merchandiser sampai, data tersebut akan mereka ubah menjadi data digital dengan mengetiknyanya secara manual.

2.2. Observasi

Penulis juga turun ke lapangan untuk melakukan observasi kepada merchandiser dan penulis menyimpulkan alur kerja merchandiser adalah seperti gambar berikut



Gambar 1. Alur Kerja Merchandiser

2.3. Wawancara dengan Merchandiser

Selama wawancara dengan seorang merchandiser di perusahaan tersebut, penulis membahas tentang peralihan dari metode pencatatan laporan tertulis menjadi laporan menggunakan aplikasi mobile. Merchandiser tersebut menyampaikan bahwa mereka sangat mendukung peralihan ini karena melihat potensi besar dalam meningkatkan efisiensi dan kualitas pekerjaan mereka. Dalam percakapan tersebut, merchandiser menekankan pentingnya memiliki aplikasi yang mudah digunakan. Mereka menginginkan antarmuka yang intuitif dan navigasi yang sederhana agar dapat dengan cepat mengakses dan memasukkan data yang dibutuhkan.

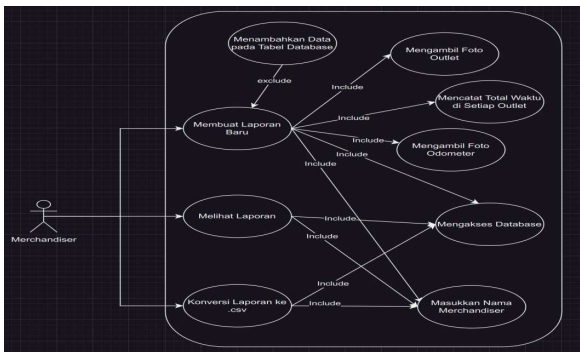
2.4. Strategi Penyelesaian Masalah

Penulis mengatasi permasalahan terkait bensin dan foto merchandiser di outlet dengan memanfaatkan fitur kamera pada aplikasi. Merchandiser dapat memotret odometer kendaraan untuk menghitung jumlah bensin yang digunakan secara akurat dan memotret foto outlet sebagai bukti visual dalam laporan mereka. Untuk mengoptimalkan waktu merchandiser, kami akan mengimplementasikan fitur pencatatan waktu otomatis dalam aplikasi, yang mencatat waktu kedatangan dan waktu selesai secara otomatis. Selain itu, untuk memudahkan penggunaan data, kami akan menyediakan opsi untuk menghasilkan file CSV yang dapat dibaca dan diolah dengan mudah menggunakan aplikasi spreadsheet seperti Excel.

2.5. SDLC Agile dengan Aplikasi Android

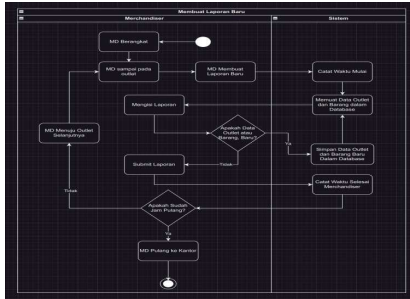
SDLC Agile adalah salah satu jenis metode dalam pengembangan perangkat lunak yang memungkinkan sebuah tim untuk mengambil keputusan dengan cepat, menghasilkan kualitas yang baik, dan memiliki potensi yang baik. Metode SDLC Agile didasarkan pada 12 prinsip utama dalam pengembangan perangkat lunak, di mana pengembangan dilakukan dalam jangka pendek dan membutuhkan adaptasi yang cepat dari para pengembang terhadap perubahan dalam berbagai bentuk dengan tujuan memaksimalkan nilai bisnis dan kepuasan pelanggan dengan cara merilis aplikasi secara cepat dan bertahap [2] [3] [4]. Dalam konteks pembuatan aplikasi pelaporan ini, aplikasi Android memiliki fitur-fitur yang diperlukan, seperti pelacakan waktu kedatangan, waktu selesai, dan pengambilan foto di lokasi outlet. Fitur-fitur ini akan memberikan bukti konkret tentang kehadiran dan kegiatan yang dilakukan oleh merchandiser.

2.6. Rancangan Alur Aplikasi



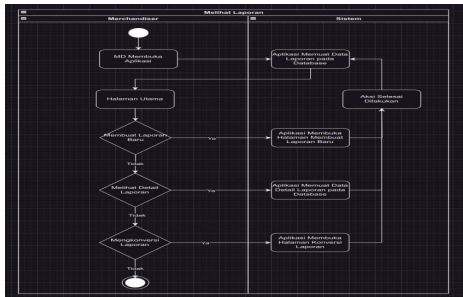
Gambar 2. Alur Aplikasi

UML Use Case yang telah dibuat menggambarkan interaksi antara aktor (Merchandiser) dengan sistem aplikasi dalam konteks "Sistem Penjualan Merchandise Berbasis Aplikasi Mobile". Pada Use Case ini, Merchandiser berperan sebagai aktor utama yang menggunakan beberapa Use Case untuk menjalankan tugas-tugasnya. Use Case "Membuat Laporan Baru" digunakan oleh Merchandiser untuk mencatat penjualan dan informasi lainnya. Dalam proses ini, Merchandiser juga menggunakan Use Case "Memotret Odometer", Use Case "Memotret Foto Outlet", dan Use Case "Mencatat Total Waktu di Setiap Outlet". Selain membuat laporan, Merchandiser juga bisa menggunakan Use Case "Menambahkan Data pada Tabel Database" jika data outlet maupun barang belum ada pada database.



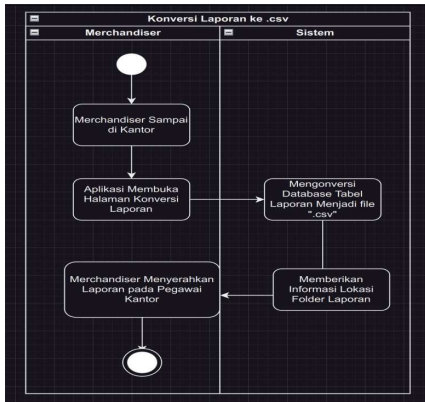
Gambar 3. Activity Diagram Membuat Laporan Baru

Gambar diatas menunjukkan proses detail tentang bagaimana tahapan-tahapan dalam Membuat Laporan baru.



Gambar 4. Activity Diagram Melihat Laporan

Gambar diatas menunjukkan proses detail tentang bagaimana tahapan-tahapan dalam melihat laporan.

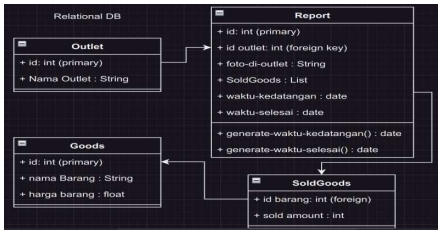


Gambar 5. Activity Diagram Konversi Laporan

Gambar diatas menunjukkan proses detail tentang bagaimana tahapan-tahapan dalam Mengonversi laporan.

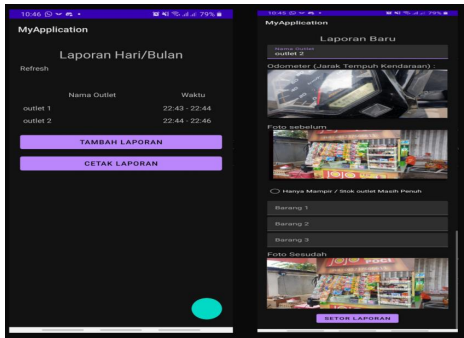
2.7. Database dan Relasi Data

Pada penelitian ini, diperlukan perencanaan yang matang terhadap hubungan data pada sistem. Sebagai bagian dari pengembangan aplikasi, perencanaan ini mencakup desain struktur database yang efisien dan terorganisir dengan baik. Hal ini memastikan bahwa data yang dihasilkan oleh aplikasi, seperti laporan penjualan, stok barang, dan informasi merchandiser, dapat saling terhubung dan diakses dengan mudah. Dengan perencanaan yang baik, hubungan data yang konsisten dan akurat akan memungkinkan analisis data yang lebih efektif serta pengambilan keputusan yang lebih baik dalam operasional perusahaan distribusi merchandise. berikut tabel yang penulis buat



Gambar 6. Relasi Data

2.8. Tampilan Aplikasi



Gambar 7. Tampilan Main Menu Aplikasi dan Tampilan Membuat Laporan Baru

Gambar diatas adalah tampilan mainmenu dan tampilan membuat laporan baru pada aplikasi yang dikembangkan. tampilan mainmenu memberikan informasi dan aksi cepat dengan memperlihatkan daftar laporan dan 2 tombol yaitu tambah laporan dan cetak laporan. Sedangkan, tampilan membuat laporan baru menampilkan antarmuka yang user-friendly untuk merchandiser dalam mengisi informasi terkait penjualan dan kegiatan di outlet.

3. Hasil dan Diskusi

Black-Box Testing adalah metode pengujian yang berfokus pada spesifikasi fungsional perangkat lunak. Tester dapat mendefinisikan kondisi input dan menguji spesifikasi fungsional program. Black-Box Testing tidak menggantikan White-Box Testing, tetapi merupakan pelengkap untuk menguji hal-hal yang tidak dicakup oleh White-Box Testing. Black-Box Testing cenderung menemukan fungsi yang tidak benar atau tidak ada, kesalahan antarmuka, kesalahan pada struktur data dan akses basis data, kesalahan performansi, dan kesalahan inialisasi dan terminasi [6].

3.1. Pengujian Perangkat Lunak Mobile dengan Blackbox

Metode pengujian blackbox dalam jurnal ini mencakup langkah-langkah sebagai berikut pengujian sistem atau komponen yang bersifat blackbox, penggunaan test case sebagai instruksi untuk menguji fitur atau fungsi tertentu, harapan akan hasil yang diharapkan dari sistem berdasarkan test case, hasil pengujian yang sebenarnya setelah menjalankan test case yang memberikan analisis dan penilaian terhadap hasil pengujian untuk mengevaluasi kualitas sistem atau komponen.

Tabel 1. Pengujian Aplikasi dengan Metode Blackbox

Pengujian	Test case	Hasil yang Diharapkan	Hasil Pengujian
Nama Merchandiser	Merchandiser Memasukkan nama pada menu utama	Nama merchandiser akan diingat dan dijadikan referensi nama	Pengujian Berhasil

Pengujian	Test case	Hasil yang Diharapkan	Hasil Pengujian
		untuk folder laporan, file foto maupun .csv	
Menu Daftar Laporan	Membuka Menu Daftar Laporan	Menampilkan histori laporan, tombol "Membuat Laporan Baru", dan tombol "Konversi Laporan".	Pengujian Berhasil
Membuat Laporan Baru	Mengisi semua data laporan baik berupa data atau foto.	Sistem mencatat waktu kedatangan, waktu selesai dan laporan baru tersimpan dengan benar.	Pengujian Berhasil
Mengambil Gambar	Mengambil foto odometer dan foto outlet langsung dengan aplikasi	URI Gambar tersimpan dalam database dan gambar folder eksternal.	Pengujian Berhasil
Konversi Laporan	Menekan tombol "Cetak Laporan"	Database Laporan berhasil dikonversi menjadi format .csv.	Pengujian Berhasil
Hasil Laporan Akhir	Melihat Folder laporan yang dibuat otomatis oleh aplikasi pada smartphone	Data pada file .csv beserta semua gambar yang ada disana cocok dengan data pada database laporan.	Pengujian Berhasil

3.2. Perbandingan Laporan Pencatatan Manual dengan Laporan Aplikasi

Tabel 2. Perbandingan Laporan Pencatatan Manual dengan Laporan Aplikasi

Jenis Data Laporan	Laporan Pencatatan Manual	Laporan Aplikasi
Laporan data waktu kedatangan, waktu selesai, dan total waktu	tidak tersedia	tersedia
Laporan data penjualan barang tiap outlet	tersedia	tersedia
Laporan data foto di lokasi outlet	tidak tersedia	tersedia
Laporan data odometer (angka jarak tempuh dalam kilometer pada motor)	tidak tersedia	tersedia

Penggunaan aplikasi Android untuk pelaporan telah meningkatkan kualitas dan keakuratan data yang dihasilkan. Dalam perbandingan antara laporan pencatatan manual dan laporan aplikasi, terdapat beberapa jenis data laporan yang mengalami peningkatan. Pada laporan pencatatan manual, data seperti waktu kedatangan, waktu selesai, dan total waktu tidak tersedia, sementara pada laporan aplikasi data tersebut tersedia secara otomatis. Selain itu, laporan aplikasi juga menyediakan data penjualan barang tiap outlet, yang sebelumnya tidak tersedia dalam laporan manual. Selain itu, laporan aplikasi juga mencakup data foto di lokasi outlet, yang tidak dapat dilakukan dalam laporan manual. Selain itu, laporan aplikasi juga menyediakan data odometer, yang merupakan angka jarak tempuh dalam kilometer pada motor, yang sebelumnya tidak tersedia dalam laporan manual. Dengan adanya aplikasi Android, pengguna dapat mengakses informasi yang lebih lengkap dan akurat untuk analisis dan pengambilan keputusan yang lebih baik.

4. Kesimpulan

Aplikasi Android sebagai aplikasi laporan pencatatan dalam sistem penjualan merchandise memberikan improvisasi yang signifikan. Selain itu, aplikasi ini juga memperhatikan kebutuhan merchandiser dengan menyediakan antarmuka yang mudah digunakan, fitur-fitur seperti pemotretan odometer, pemotretan foto outlet, dan pencatatan waktu otomatis. Aplikasi ini berhasil menyediakan data penting dengan mengurangi beban pengerjaan laporan merchandiser. Dengan demikian, aplikasi ini memberikan solusi yang efektif dalam meningkatkan produktivitas dan pengelolaan data pada perusahaan distributor merchandise.

Secara keseluruhan, penelitian ini memberikan wawasan yang berharga mengenai penerapan teknologi mobile dalam sistem penjualan merchandise. Diharapkan hasil dari penelitian ini dapat memberikan kontribusi positif dalam pengembangan sistem penjualan merchandise berbasis aplikasi mobile di perusahaan-perusahaan distributor lainnya.

Daftar Pustaka

- [1] I. Isnawati, N. Jalinus, and R. Risfendra, "Analisis Kemampuan pedagogi guru SMK yang Sedang Mengambil Pendidikan profesi Guru Dengan metode Deskriptif Kuantitatif Dan Metode Kualitatif," *INVOTEK: Jurnal Inovasi Vokasional dan Teknologi*, vol. 20, no. 1, pp. 37–44, 2020. doi:10.24036/invotek.v20i1.652
- [2] M. R. Adani, "Metode Agile: Pengertian, Jenis, Manfaat, Dan 12 prinsipnya," Software House & System Integrator di Malang, Indonesia, <https://www.sekawanmedia.co.id/blog/metode-agile-development/> (accessed Jun. 10, 2023).
- [3] R. Setiawan, "Metode SDLC Dalam Pengembangan software," Dicoding Blog, <https://www.dicoding.com/blog/metode-sdlc/> (accessed Jun. 10, 2023).
- [4] R. Setiawan, "Konsep Agile Pada software development," Dicoding Blog, <https://www.dicoding.com/blog/konsep-agile-pada-software-development/> (accessed Jun. 10, 2023).
- [5] R. Setiawan, "Flowchart Adalah: Fungsi, Jenis, Simbol, Dan Contohnya," Dicoding Blog, <https://www.dicoding.com/blog/flowchart-adalah/> (accessed Jun. 11, 2023).
- [6] J. Supardi, M.S. Slide, and R.A. Sukamto, "Black-Box Testing," Hal, pp. 2-16, 2017.

Integrasi Chatbot Berbasis GPT Terhadap Website E-Commerce (Studi Kasus Toko JSA)

^{1a}Made Putra Teguh Pramana, ^{2a}I Gede Surya Rahayuda

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana

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

^{1a}baruputra111@gmail.com

^{2a}gedesuryarahayuda@unud.ac.id

Abstract

A website is a series of sites with related hyperlinks that individual, organization, company, or government made. E-commerce websites are websites that have features for transaction through online services. Toko jsa emas perak is one of the companies that want to develop its company's presence on the internet using e-commerce websites. uniqueness is important to have an edge between online competition. By implementing chatbot system in the e-commerce websites it can increase the effect on user in each interaction. the propose of this research is to implement a chatbot service in the toko jsa websites.

Keywords: balckbox, e-commerce websites, chatbot, GPT.

1. Pendahuluan

Perkembang teknologi di dalam zaman sekarang semakin cepat dan banyak. Mulai dari perkembangan internet yang makin cepat menggunakan sinyal 5G hingga perkembangan dari artificial intelligence yang semakin cerdas dalam menangani pekerjaan-perjaan yang makin sulit. Terlebih lagi menurut laporan survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia pengguna internet di Indonesia mencapai dengan jumlah 143,26 juta pada tahun 2017 atau sama dengan 54,68% jumlah penduduk Indonesia [1]. Oleh karena itu perusahaan – perusahaan perlu mengikuti perkembangan zaman dengan mengikuti perkembang teknologi tersebut. Toko Jsa merupakan toko yang menyediakan jasa dalam menjual dan pemesanan perhiasaan – perhiasaan. Demi meningkatkan brand online usaha sekaligus membuat pemesanan berbasis online, Toko jsa dapat membuat website e-commerce. Website tersebut dapat menyediakan pilihan perhiasaan-perhiasaan yang ada pada toko serta bisa dijadikan sebagai tempat pemesanan barang dengan pilihan perhiasaan dan opsi – opsi dari kustomisasi yang disediakan. Terlalu banyak opsi – opsi dan Langkah-langkah pemesanan dapat membingungkan pelanggan yang kurang mahir di dalam bidang teknologi. Oleh sebab itu website ini patut memiliki cara untuk menjelaskan alur dari Langkah – Langkah pemesanan dan opsi – opsi kustomisasi perhiasaan. Salah satu caranya adalah dengan menggunakan customer service menggunakan chatbot. Chatbot dapat membantu interaksi antara manusia dan computer serta memiliki kemampuan untuk memeriksa dan mempengaruhi perilaku pengguna dengan memberikan pertanyaan dan menanggapi terhadap pertanyaan pengguna [2].

1.1. Chatbot

Pengertian dari chatbot sudah ada sejak tahun 19-a yang didefinisikan sebagai suatu system software yang berguna untuk mensimulasi dan membuat pembicaraan antara AI dan user [3]. Zaman sekarang chatbot sudah semakin canggih dikarenakan adanya AI yang mampu membaca dan memberikan dialog yang semakin mirip dengan manusia. Sehingga chatbot sering digunakan di dalam bidang industri agar mendapatkan kelebihan dari competitor. Chatbot biasanya paling berguna di dalam customer service, dikarenakan chatbot mampu mengelola jumlah data yang sangat banyak [3].

1.2. Database

Database adalah suatu kumpulan data yang normal berukuran besar dengan racangan yang dibuat untuk pencarian cepat dan pengembalian [4]. Database memiliki berbagai macam struktur yang dapat dimiliki untuk menyimpan data. Struktur database diatur dengan bahasa yang populer yaitu SQL dan NOSQL. Database SQL diatur dengan table atau entitas yang terdiri dari kolom – kolom yang merepresentasikan suatu attribute dan element-element dari data tersebut. SQL berguna untuk melakukan query untuk memasukan data, membuat schema dan table, menghapus data-data tersebut. Query digunakan agar mendapatkan komputasi yang efektif dan efisien karena data – data memiliki relasi dari attribute – attributenya.

1.3. Website

Web site adalah kumpulan dari halaman World Wide Web (WWW) yang memiliki hyperlink terhadap masing – masing halaman dibuat oleh Individu, Perusahaan, Lembaga Pendidikan, Pemerintah, atau Organisasi [5].

1.4. Django

Django merupakan sebuah framework web python yang menggunakan rapid development dan clean, pragmatic design [6]. Django didesign dengan algoritma yang sering digunakan oleh developer agar developer yang menggunakan Django akan lebih cepat mendevelop software yang akan dibuatnya. Django memiliki architecture model-view-controller (MVC) dengan beberapa perbedaan dimana Django menggunakan model-view-template (MVT) atau model yang menggunakan structure dari template model. Django menyediakan access pada banyak sekali feature seperti (ORM) dan Engine yang dapat digunakan untuk membuat html template yang dinamis.

1.5. Postgresql

Postgresql adalah salah satu jenis open-source object-relational database system yang menggunakan dan menambahkan Bahasa pemrograman SQL dengan kombinasi berbagai macam fitur yang mampu dengan aman menyimpan dan mengskala pekerjaan data [7]. Postgresql merupakan database yang cocok untuk menyimpan data dengan jumlah besar atau kecil karena Postgresql memiliki kecepatan yang lebih cepat dalam melakukan query pada data besar.

1.6. GPT

Generative pre-trained transformers (GPT) adalah suatu large language model (LLM) dan framework untuk melakukan generative artificial intelligence [8]. GPT dibuat oleh OpenAI yaitu suatu artificial intelligence research laboratory yang bisa digunakan untuk menjawab pertanyaan, translate pertanyaan, dan membuat text menjawab [8].

2. Metode Penelitian

Dalam metodologinya penelitian ini penulis menggunakan metodologi agile. Agile merupakan metodologi yang fleksibel dan dapat dilakukan dalam jangka waktu yang pendek. Namun apabila terjadi perubahan developer harus dapat bertindak dengan cepat untuk menangani perubahan tersebut. Berikut adalah langkah – langkah dari pengembangan aplikasi menggunakan metodologi agile:

2.1. Teknik Pengumpulan Data

a. Observasi

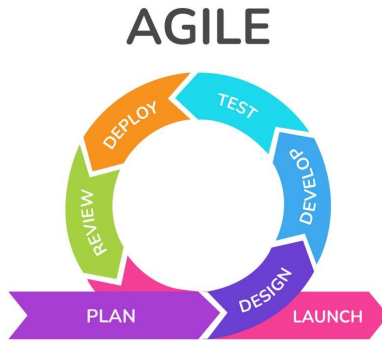
Observasi adalah teknik memperoleh data primer dengan cara mengamati objek data secara langsung [9]. Observasi dilakukan di dalam tempat penelitian yaitu toko jsa emas perak dengan cara melihat proses dari bisnis berlanjut antara pelanggan, penjual, dan tukang di toko.

b. Wawancara

Wawancara adalah bentuk komunikasi dua arah untuk mendapatkan data dari narasumber [9]. Wawancara merupakan bagian penting untuk dapat memvalidasi hasil dari proses observasi. Wawancara akan dilaksanakan kepada penjual, pembeli, dan tukang di toko.

2.2. Tahapan Penelitian

Adapun tahapan – tahapan dari agile development untuk pengembangan website e-commerce yang dapat dilihat pada gambar 1.



Gambar 1 . Agile Development Source: Krasamo

a. Plan

Planing atau analisis sistem adalah tahapan dilakukan asesmen terhadap user untuk mengetahui secara detail kebutuhan dan kemampuan perangkat lunak user [9]. Analisis yang dilakukan berkaitan dengan dua hal yaitu analisis dari kemampuan perangkat keras dan kemampuan user lalu analisis tentang kebutuhan dan kemampuan perangkat keras server. Analisis system ini akan menghasilkan data tentang keperluan spesifikasi dari perangkat keras dan lunak yang dibutuhkan.

b. Design

Pada Tahapan perancangan atau Designing terdapat perancangan design arsitektur, perancangan proses bisnis dan perancangan desain database [9]. Perancangan ini akan dibuat menggunakan pemodelan Entity Relationship Diagram (ERD). Pemodelan

pendekatan secara OOP juga dilakukan dengan menggunakan pemodelan use case diagram.

c. Development Aplikasi

Development aplikasi adalah proses pengimplementasian dari perancangan aplikasi yaitu pengkodean dan database [9]. Pengimplemtasikan perancangan aplikasi ini akan menggunakan Django sebagai framework fullstack. Database yang digunakan dalam perancangan aplikasi ini akan menggunakan database Postgresql. Library OpenAI api merupakan salah satu cara openai memberikan endpoint untuk mengakses AI dari Chatgpt. Kita dapat menggunakan library langchain untuk memasukan prompt agar chatbot mendapatkan data yang harus dilatihnya.

d. Testing

Dalam testing dapat dilakukan beberapa hal untuk mengecek kelayakan dari suatu aplikasi atau websites. Salah satu yang dapat dilakukan adalah dengan menggunakan blackbox testing. Blackbox testing merupakan teknik pengujian perangkat lunak yang berfokus pada spesifikasi fungsional dari perangkat lunak [10]. Hasil pengujian nantinya akan ditampilkan di dalam table yang memuat keberhasilan dari fungsionalitas suatu feature.

e. Deploy Aplikasi

Deploy aplikasi merupakan tahapan dimana aplikasi yang dibuat akan dapat diakses oleh user. Pembukaan akses tersebut dapat dilakukan dengan berbagai macam cara . salah satu cara tersebut adalah menggunakan web hosting yang membuat website dapat diakses dengan internet.

f. Review

Review berguna sebagai penjaga dari kualitas yang paling efektif di dalam bidang software engineering [11]. Ada berbagai macam tipe review yang memiliki kegunaannya masing – masing. Untuk membuat feedback dari pengguna diperlukan review yang dihasilkan oleh pengguna. Review yang didapatkan maka akan dievaluasi dan dikembangkan.

g. Maintance

Pada fase maintance terjadi loop dimana aplikasi akan secara terus menerus mengalami perbaikan kesalahan atau kelemahan yang dimiliki. Selain itu aplikasi mungkin mengalami perubahan apabila mendapatkan perubahan update dari salah satu dependency nya.

3. Hasil dan Pembahasan

Adapun hasil yang telah didapatkan oleh penelitian ini. Berikut adalah hasil dari penelitian tersebut:

a. Plan

Pada perancangan aplikasi suatu sistem akan dijabarkan secara utuh ke dalam seluruh komponen sistem dengan tujuan mengidentifikasi masalah atau hambatan yang dapat terjadi serta kebutuhan yang dilakukan untuk memperbaiki hal-hal tersebut. Telah dilakukan observasi yang telah diverifikasi dengan wawancara untuk menghasilkan data kemampuan perangkat keras yang dimiliki oleh pelanggan. Berdasarkan observasi pelanggan kebanyakan tidak memiliki perangkat computer dan lebih dominan menggunakan handphone. Sehingga dapat perkiraan standar minimal yang dimiliki oleh end-user.

Tabel 1. Spesifikasi Minimal Perangkat User

Spesifikasi	Keterangan
Processor	MT6769H Helio G85 (12nm)
Ram	3 GB
Ukuran Layar	6.71 inch
OS	Android OS 10
ROM	32GB

Selanjutnya adalah analisis hosting yang akan digunakan untuk mengetahui infrastruktur yang digunakan oleh backend dari server website. Hosting yang digunakan sudah sesuai dengan aplikasi yaitu server yang dibuat dengan backend Django sekaligus viewnya, SSH, FTP, Postgresql. adapun spesifikasi dari hosting yang digunakan dapat dilihat pada table 2.

Tabel 2. Spesifikasi dari Hosting

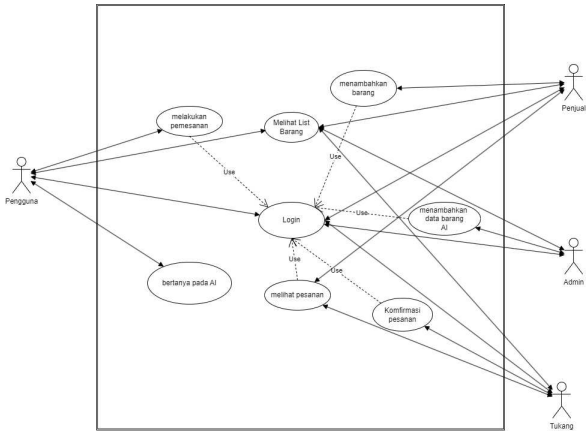
Spesifikasi	Keterangan
Domain	Include
Disk Space	100GB
Ram	2 GB
OS	Linux
Fitur Tambahan	SSH Akses, FTP Akses, Docker

b. Design

Design pertama dilakukan dengan membuat arsitektur model dari UML yaitu usecase diagram dan schema diagram, lalu akan dibuatkan perancangan database menggunakan entity relational diagram (ERD).

1. Usecase Diagram

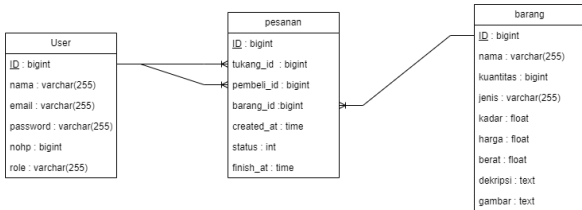
Usecase Diagram digunakan untuk menggambarkan penggunaan pada sistem yang terkait dengan beberapa pengguna. Adapun usecase diagram yang digunakan pada website e-commerce Toko JSA emas perak yang dapat dilihat pada gambar 2. Di dalam usecase diagram terdapat empat buah actor yaitu Pengguna, Admin, Penjual, Tukang. Terdapat activity juga yang berada pada sistem di usecase diagram yaitu melihat list barang dan login yang dapat diakses oleh semua actor. Pengguna memiliki activity yang dapat digunakan oleh pengguna yaitu bertanya pada ai dan melakukan pemesanan, dimana melakukan pemesanan harus melakukan login terlebih dahulu. Activity melihat pesanan dapat dilakukan oleh actor penjual dan tukang jika harus mengetahui pesanan yang ada.



Gambar 2. Usecase Diagram

2. Schema Diagram

Schema diagram merupakan notasi untuk menggambarkan model logis basis data. Model ini digunakan untuk mendeskripsikan basis data yang terdapat pada aplikasi website e-commerce. Schema Diagram dijelaskan pada gambar 3.



Gambar 3. Schema Diagram Server

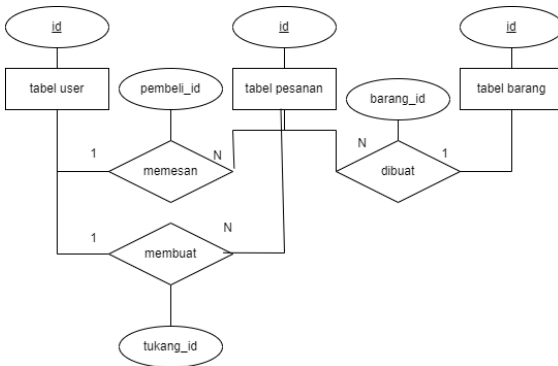
Schema diagram terdiri dari 3 entiti atau table yang akan di terapkan pada database di dalam aplikasi. Nama dari table ketiga database tersebut adalah user, pesanan, barang. Kardinatilitas dari diagram ada pada tabel 3.

Tabel 3. Kardinalitas Pada Entity

Tabel 1	Tabel 2	Kardinalitas (foreign Key)
user	pesanan	One to many (tukang_id)
user	pesanan	One to many (pembeli_id)
barang	pesanan	One to many(barang_id)

3. Entity Relational Diagram

ERD merupakan himpunan entitas dan himpunan relasi yang digambarkan lebih jauh melalui sejumlah atribut-atribut yang menggambarkan seluruh fakta dari sistem yang ada. Adapun ERD dari aplikasi website toko jsa emas perak yang ada pada gambar 4.



Gambar 4. Entity Relationship Diagram

c. Development Aplikasi

Proses development aplikasi terbagi menjadi dibagi menjadi tiga bagian yaitu pembuatan kode di atas framework Django berdasarkan usecase yang telah dibuat , pembuatan database menggunakan schema dari ERD dan Schema diagram , pelatihan chatbot yang dilatih menggunakan data dari barang dan alur – alur pemesanan. Django memiliki fitur migration yang lasung terkoneksi kedalam database tanpa harus membuka atau terkoneksi dengan database secara langsung.

Didalam implementasi frontend websites dibuat berdasarkan alur yang ada pada usecase diagram. Pertama, fitur login harus memiliki fitur registrasi untuk menyimpan data email, password, nama, password, no hp. untuk attribute role di dalam tabel user akan disi element pelanggan dan hanya bisa dirubah oleh admin. Nantinya setelah data tersimpan ada pemberian authorization token yang beruna untuk mengingat status login pada server.

Pada alur melakukan pesanan terjadi pembuatan MVT untuk pengumpulan dari form pemesanan dari barang yang ingin dipesan. Itu akan meminta data yang seperti barang ,jumlah, besar barang. Nantinya pemesanan akan masuk ke dalam database server dan dapat dilihat oleh actor tukang dan penjual di dalam activity melihat pesanan. Activity

melihat pesanan akan memiliki MVTnya sendiri yang akan digunakan ketika ingin melihat pemesanan.

Konfirmasi pemesanan adalah activity yang hanya bisa dilakukan oleh tukang. Pesanan yang diminta oleh pelanggan nantinya akan diberikan kepada salah satu tukang. Tukang dapat melihat pesanan mana saja yang merupakan tugasnya dia dan melakukan konfirmasi apabila pekerjaan tersebut telah selesai melalui MVT.

Perkerjaan admin adalah mengelola data dan menjaga keberlangsungan dari server. Salah satu perkerjaan yang dilakukan oleh admin pada usecase diagram adalah melatih data barang pada AI. Setiap ada perubahan baru atau penambahan barang agar ai dapat mengikuti perubahan tersebut admin harus memasukan data dari barang yang dapat membuat AI mengeluarkan data yang diinginkan.

Process pelatihan AI dilakukan dengan awalnya mencari data-data barang yang akan ditaruh ke dalam database websites. Setelah itu menggunakan API yang telah disediakan oleh OpenAI dan library langchain AI dapat mempelajari barang-barang yang kita berikan. AI tersebut sekarang dapat menjawab tentang hal-hal sekitar data yang telah kita berikan. Setelah AI dibilang cukup mahir menjawab pertanyaan maka AI akan di launch ke dalam suatu container dari docker yang dapat ditanyakan oleh user melalui API.

d. Testing

Pada tahap testing website aplikasi akan digunakan metode black box dalam menguji fungsionalitas dari website yang telah dibuat. Skenario pengujian black box menggunakan inputan data masuk dan memiliki tujuan hasil dari data yang harusnya dikeluarkan.

Tabel 4. Skenario Pengujian Black Box.

Skenario (input)	Pengamatan
Data benar	Output Benar
Data Salah	Output Salah
Data kosong	Output kosong

Setelah semua inputan data blackbox dan hasil yang dikeluarkan oleh program sama dengan harapan keluaran hasil dari blackbox sama pada saat itu baru dapat dibilang bahwa program sudah lolos dari uji blackbox.

e. Deploy Aplikasi

Website dan struktur database yang sudah deprogram akan nantinya di deploy dalam sebuah hostingan. Langkah – langkah dari mendeploy website yang akan dibuat adalah menkonfigurasi hostingan atau vps yang telah disewa agar dapat terhubung dengan computer kita, melakukan kongufurasi file-file yang ada, melakukan pembuatan dan setup dari schema di server.

Konfigurasi dari hostingan biasanya cloud server akan memintak untuk menentukan service apa saja yang diperlukan dalam suatu virtual environment. Setelah itu akan dilakukan pembuatan ssh kay agar dapat dilakuakn pengaksesan virtual environment tersebut.

Setelah melakukan hal tersebut kita selanjutnya mengclone semua file yang telah dikoding kedalam server yang kita miliki. File ini nantinya akan di run untuk menjadi server dari backend dan frontend kita. Akan tetapi untuk melakukan hal tersebut perlunya mengaktifasi database server terlebih dahulu.

Database di deploy kedalam server yang kita miliki apabila menggunakan VPS atau harus di aktifkan apabila menggunakan hostingan. Database yang di deploy masih belum isi table atau schema apapun sehingga harus dilakukan migration yang ada pada file Django. Untuk melakukan migration cukup dengan menjalankan command migrasi di Django. Setelah semua ini maka server harusnya sudah dapat dijalankan.

f. Review

Pada tahapan ini akan terjadi evaluasi dan perbaikan apabila terjadi kesalahan atau kelemahan didalam program. Perubahan kodingan dapat dilakukan melalui apliakasi seperti git untuk dapat melakukan Version Control.

g. Maintance.

Maintance (pemeliharaan) adalah serangkaian tindakan yang dilakukan untuk memelihara atau meningkatkan kondisi optimal suatu sistem atau objek agar tetap berfungsi dengan baik. Tujuan pemeliharaan adalah untuk mencegah kerusakan, memperpanjang umur dan menjaga peralatan atau barang dalam kondisi terbaiknya.

4. Kesimpulan

Kesimpulan yang didapatkan dari hasil penelitian yang dilakukan pada penjelasan sebelumnya yaitu, Integrasi dari chatgpt dalam pembuatan websites dapat dilakukan dengan process pembaruan AI yang cukup efisien. Website juga dapat dibidang cukup efisien dalam implementasinya.

Daftar Pustaka

- [1] Sakin Rakhma Diah Setiawan, "Pengguna Internet di Indonesia Mencapai 143,26 Juta Orang," <https://ekonomi.kompas.com>, 19 Februari 2018.
- [2] S. A. dan Dr. John, "Survey on Chatbot Design Techniques in Speech Conversation Systems," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 7, 2015, doi: 10.14569/IJACSA.2015.060712.
- [3] S. Di Gaetano dan P. Diliberto, "Chatbots and conversational interfaces: Three domains of use," 2018. Diakses: 10 Juni 2023. [Daring]. Tersedia pada: <https://ceur-ws.org/Vol-2101/paper8.pdf>
- [4] Merriam-webster, "Database," Merriam-Webster Online Dictionary. <https://www.merriam-webster.com/dictionary/database> (diakses 10 Juni 2023).
- [5] Merriam-Webster, "Website," Merriam-Webster Online Dictionary. <https://www.merriam-webster.com/dictionary/websites> (diakses 10 Juni 2023).
- [6] Django Software Foundation, "Django." 2019. [Daring]. Tersedia pada: <https://djangoproject.com>
- [7] PostgreSQL Global Development Group (or the organization responsible for PostgreSQL), "PostgreSQL." PostgreSQL Global Development Group, 10 Juni 2023.
- [8] cambridge, "GPT," Cambridge Dictionary. <https://dictionary.cambridge.org/us/dictionary/english/gpt> (diakses 10 Juni 2023).
- [9] I. Mahendra dan D. T. Eby Yanto, "Sistem Informasi Pengajuan Kredit Berbasis Web Menggunakan Agile Development Methods Pada Bank Bri Unit Kolonel Sugiono," *Jurnal Teknologi Dan Open Source*, vol. 1, no. 2, hlm. 13–24, Des 2018, doi: 10.36378/jtos.v1i2.20.
- [10] S. R. Jan, S. T. U. Shah, Z. U. Johar, Y. Shah, dan F. Khan, "An innovative approach to investigate various software testing techniques and strategies," *International Journal of Scientific Research in Science, Engineering and Technology (JSRSET)*, Print ISSN, vol. 23951990, 2016.
- [11] M. Ciolkowski, O. Laitenberger, S. Vegas, dan S. Biffi, *Practical experiences in the design and conduct of surveys in empirical software engineering*. Springer, 2003.

Halaman ini sengaja dibiarkan kosong

Implementasi Metode Certainty Faktor Untuk Rekomendasi Pembelian Hp

^{a1} Komang Arsa Wiguna, ^{a2} I Gede Arta Wibawa

"Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
komangarsa23@email.com
gede.arta@unud.com

Abstract

In this study, the implementation of the Certainty Factor (CF) method was carried out to provide consumers with more accurate cell phone purchase recommendations. This method combines expert knowledge and user data to produce more reliable recommendations. The implementation process involves collecting relevant HP feature data, establishing rules based on expert knowledge, calculating confidence factors (CF) based on the level of expert confidence and suitability of user data, and merging the CFs of all rules to produce a final recommendation. Through testing and evaluation, the Certainty Factor method is proven to be able to provide accurate HP purchase recommendations and assist consumers in choosing a HP that suits their preferences and needs. However, this research still requires further development to improve the accuracy of recommendations and to conduct tests with larger and more diverse datasets to ensure the effectiveness and validity of this method in the context of H purchase recommendations.

Keywords: rekomendasi, HP, certainty factor

1. Pendahuluan

Teknologi modern telah banyak mengubah kebiasaan dan merevolusi cara kita bekerja, hidup, dan bermain [1]. Salah satu perkembangan yang sangat memengaruhi hal tersebut adalah terciptanya komputer. Teknologi tersebut memungkinkan sejumlah besar informasi disimpan dalam ruang kecil [2]. Komputer juga memungkinkan seseorang untuk menghitung masalah matematika dengan mudah. Terakhir, komputer memungkinkan orang untuk berkomunikasi satu sama lain melalui situs internet [3]. Hal ini mengakibatkan banyak orang tergantung dengan penggunaan komputer. Komputer dan HP kini tersedia dalam fitur dan spesifikasi yang beragam. Sebagian besar dari customer tidak tahu untuk memilih HP yang sesuai dengan kebutuhan [4]. Dalam pembelian HP konsumen hanya membeli HP berdasarkan merek dan harga. Jika disesuaikan dengan kebutuhan penggunaan komputer pemilihan berdasarkan hal tersebut sering tidak sesuai [5]. Yang terjadi ada kemungkinan HP yang dipilih memiliki spesifikasi yang terlalu tinggi untuk pekerjaan pengguna atau bahkan performanya kurang untuk pekerjaan yang lebih besar [6]. Berdasarkan permasalahan tersebut akan diimplementasikan sebuah metode yang mampu merekomendasikan HP yang sesuai dengan kebutuhan konsumen. Sistem pakar merupakan program komputer yang menggunakan metode kecerdasan buatan untuk memecahkan masalah dalam permasalahan khusus yang biasanya membutuhkan keahlian manusia [7]. Salah satu metode Expert System adalah metode Certainty Factor. Certainty Factor (CF) merupakan perhitungan numerik yang digunakan untuk mencari seberapa besar kemungkinan suatu peristiwa atau pernyataan seharusnya benar berdasarkan fakta dan aturan serta menggambarkan penilaian suatu ahli [8]. Penelitian ini berdasarkan penelitian sebelumnya yang menggunakan metode Certainty Factor namun memiliki objek yang berbeda. Peneliti [8] yang mendeteksi penyakit yang menyerang tanaman jagung menggunakan metode Certainty Factor mendapatkan hasil akurasi 91 % berdasarkan hasil dari validitas masing-masing penyakit. Penelitian bertujuan untuk mendiagnosa penyakit jerawat pada wajah dan mendapatkan hasil uji akurasi sebesar 85 % [9]. Penelitian selanjutnya yaitu penelitian tentang menentukan keyakinan

penyakit tulang pada seseorang dan berdasarkan hasil uji akurasi mendapatkan nilai sebesar 90 % [10]. Dari beberapa penelitian tersebut bahwa metode Certainty Factor layak dijadikan metode untuk rekomendasi HP. Berdasarkan dari permasalahan dan beberapa penelitian yang sudah dibuat maka penelitian ini memiliki tujuan yaitu menerapkan metode Certainty Factor untuk menentukan rekomendasi yang cocok untuk konsumen berdasarkan ruang lingkup fitur dan pekerjaan yang mereka gunakan.

2. Metode Penelitian

Penentuan rekomendasi HP dengan menggunakan metode Certainty Factor melewati beberapa tahap dalam mengimplementasikan metode [11]. Adapun tahapan yang dilakukan adalah pengumpulan data, penentuan aturan CF, penerapan metode CF, penarikan kesimpulan, dan pengujian yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Metode Penelitian

2.1. Pengumpulan Data

Data yang dikumpulkan pada penelitian ini didapatkan dari hasil wawancara dengan ahli yaitu seorang ahli yang memiliki pendidikan IT dan bekerja dalam penjualan HP di Bali. Data yang didapatkan dari hasil wawancara dengan ahli berupa data fitur yang disediakan oleh HP, pekerjaan yang sering dilakukan oleh konsumen terhadap HP dan HP jenis terbaru per desember 2022. Berdasarkan pengumpulan data dari hasil wawancara, maka pada penelitian ini akan digunakan 12 jenis fitur atau pekerjaan yang dapat digunakan pada HP dan 4 data HP tipe terbaru per desember 2022. Adapun data fitur atau pekerjaan yang sering diperlukan konsumen pada HP ditunjukkan pada Tabel 1, sedangkan data HP terbaru ditunjukkan pada Tabel 2.

Tabel 1. Pekerjaan atau Fitur

F1	Performa Tinggi
F2	Kamera Berkualitas
F3	Kapasitas Baterai Tahan Lama
F4	Memori Internal Besar
F5	Layar Resolusi Tinggi
F6	Desain Elegan
F7	Koneksi 5G
F8	Fitur Keamanan Tinggi
F9	Sistem Operasi Terbaru
F10	Harga Terjangkau

F11 Fitur AI (Kecerdasan Buatan)
F12 Daya Tahan Baterai

Tabel 1 merupakan data fitur yang dibutuhkan atau pekerjaan yang sering dilakukan konsumen dalam pemakaian HP. Fitur tersebut berdasarkan pengalaman ahli dalam melayani konsumen saat pembelian HP.

Tabel 2. HP

Kode HP	Nama HP
L1	Samsung Galaxy S22
L2	iPhone 14 Pro
L3	Google Pixel 7
L4	Xiaomi Mi 12

Tabel 2 merupakan data sampel HP terbaru yang diperoleh dari pakar dan beberapa situs resmi penjualan HP periode bulan desember 2022.

2.2. Penentuan Aturan CF

Aturan CF merupakan metode yang digunakan untuk menjelaskan hubungan antara fitur atau pekerjaan dengan HP dan nilai keyakinan pada masing-masing kondisi. Penerapan metode Certainty Factor dalam sistem pakar membutuhkan aturan berupa variabel (fitur dengan simbol F) dan nilai bobot yang diberikan oleh ahli [12]. Nilai bobot dibutuhkan untuk setiap fitur pada setiap HP. Ahli komputer memberikan skala nilai bobot untuk tiap fitur antara 0,2 – 1,0. Aturan CF HP dan hubungannya dengan fitur ditampilkan pada Tabel 3.

Tabel 3. Aturan CF

No	Nama HP
1	if F1 AND F6 AND F7 AND F8 AND F10 then L1
2	if F2 AND F3 AND F4 then L2
3	if F3 AND F4 AND F5 AND F11 AND F12 Then L3
4	if F3 AND F4 AND F5 AND F11 AND F12 Then L4

Langkah selanjutnya setelah menentukan hubungan antara fitur dan HP, adalah penentuan nilai CF fitur terhadap HP. Adapun nilai yang sudah ditentukan oleh ahli ditunjukkan pada Tabel 4.

Tabel 4. Nilai CF Fitur terhadap HP

Pekerjaan	Nilai CF			
	L1	L2	L3	L4
F1	0.8	0	0	0
F2	0	1	0	0
F3	0	0.8	0.9	1
F4	0	0.6	0.8	1
F5	0	0.7	1	0.7

Pekerjaan	Nilai CF			
	L1	L2	L3	L4
F6	0.8	0	0	0.6
F7	0.8	0	0	0
F8	0.8	0	0	0
F9	0.4	0	0	0
F10	0.7	0	0	0
F11	0	0	0.8	0.8
F12	0.2	0	0.6	0.5

2.3. Penerapan Metode CF

Pasti atau tidaknya sebuah fakta dapat ditentukan menggunakan metode CF. Metode CF merupakan metode yang digunakan pada sistem pakar untuk menentukan suatu ketidakpastian [7]. Rumus CF berdasarkan teori David Allister adalah sebagai berikut

$$CF[H, E] = MB[H, E] - MD[H, E] \quad (1)$$

Dengan

- CF [H, E] = Certainty Factor / Faktor Kepastian
- MB [H, E] = Tingkat kepercayaan berdasarkan pekerjaan / fitur terhadap HP
- MD [H, E] = Tingkat ketidakpercayaan berdasarkan pekerjaan / fitur terhadap HP

Nilai MD atau nilai CF fitur dalam penelitian ini telah didapatkan dari ahli secara langsung, dan selanjutnya disebut dengan CF (Pakar) sehingga untuk pencarian nilai CF HP berdasarkan fitur atau pekerjaan, dapat dimulai sebagai berikut:

- a. CF dengan keinginan konsumen tunggal

$$CF(H, E) = CF_{ahli} \times CF_{konsumen} \quad (2)$$

- b. Jika terdapat aturan memiliki lebih dari satu fitur/pekerjaan, nilai CF dihitung menggunakan persamaan berikut

$$CF_{combine} = CF_{lama} + CF_{fitur} \times (1 - CF_{lama}) \quad (3)$$

- c. Selanjutnya menghitung persentase rekomendasi HP

$$CF_{persentase} = CF_{combine} \times 100\% \quad (4)$$

MB pada penelitian ini diperoleh dari penilaian ahli terhadap premis yang terdapat pada setiap aturan dan diubah menjadi angka dengan skala. Hasil tersebut ditunjukkan pada Tabel 5.

Tabel 5. Nilai CF User

Pernyataan	Nilai CF User
Pasti	1
Hampir Pasti	0,8
Kemungkinan besar ya	0,6

Pernyataan	Nilai CF User
Kemungkinan	0,4
Kemungkinan Tidak	0,2
Tidak	0

2.4. Penarikan Kesimpulan

Langkah selanjutnya setelah mendapatkan nilai CF dari masing-masing HP adalah menarik kesimpulan dari hasil tersebut. HP yang akan dijadikan rekomendasi ke konsumen merupakan HP yang mendapatkan nilai CF terbesar

2.5. Pengujian

Tahap selanjutnya adalah tahap pengujian untuk menilai tingkat keberhasilan metode Certainty Factor dalam menentukan rekomendasi HP untuk konsumen. Untuk mengetahui keberhasilan akan dilakukan pengujian terhadap hasil dengan menghitung akurasi menggunakan rumus akurasi. Pengujian akurasi dilakukan dengan membandingkan jumlah data uji yang benar dengan seluruh data uji dikali 100 %. Data yang benar ditentukan oleh ahli untuk setiap fitur atau pekerjaan yang dipilih dari hasil rekomendasi menggunakan metode Certainty Factor. Adapun rumus dari pengujian akurasi adalah sebagai berikut

$$\text{Akurasi \%} = \frac{\sum \text{Data uji benar}}{\sum \text{data uji}} \times 100\% \tag{5}$$

3. Hasil dan Pembahasan

3.1. Penerapan Metode CF

Bagian ini memuat hasil proses perhitungan sistem pakar rekomendasi HP dengan menggunakan metode Certainty Factor. Perhitungan ini diperoleh berdasarkan dari hasil wawancara dengan pelanggan dan mendapatkan data kebutuhan pelanggan yang akan digunakan pada HP. Ilustrasi berikut merupakan tampilan hasil perhitungan sistem pakar kecerdasan rekomendasi pembelian HP dengan menggunakan metode Certainty Factor.

Tabel 6. Hasil Wawancara Dengan Pelanggan

Kode Fitur	Hasil Wawancara
F1	Pasti
F2	Tidak
F3	tidak
F4	Tidak
F5	Tidak
F6	Kemungkinan besar ya
F7	Hampir Pasti
F8	Kemungkinan besar ya
F9	Kemungkinan besar ya
F10	Kemungkinan
F11	Kemungkinan
F12	Kemungkinan besar ya

Setelah mendapatkan nilai keyakinan fitur pelanggan terhadap masing-masing fitur, akan dicari nilai CF Fitur yang didapatkan dari pengalian CF user pada setiap fitur dan nilai CF pakar pada masing-masing HP. Berikut merupakan perhitungan manual untuk menentukan nilai CF Fitur untuk Samsung Galaxy S22 dengan Kode L1

$$\begin{aligned} \text{CF Fitur 1} &= \text{CF User} * \text{CF Pakar} \\ &= 0,8 * 1 \\ &= 0,8 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 6} &= \text{CF User} * \text{CF Pakar} \\ &= 0,8 * 0,6 \\ &= 0,48 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 7} &= \text{CF User} * \text{CF Pakar} \\ &= 1 * 0,8 \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 8} &= \text{CF User} * \text{CF Pakar} \\ &= 0,8 * 0,6 \\ &= 0,48 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 9} &= \text{CF User} * \text{CF Pakar} \\ &= 0,4 * 0,5 \\ &= 0,2 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 10} &= \text{CF User} * \text{CF Pakar} \\ &= 0,7 * 0,4 \\ &= 0,28 \end{aligned}$$

$$\begin{aligned} \text{CF Fitur 12} &= \text{CF User} * \text{CF Pakar} \\ &= 0,2 * 0,6 \\ &= 0,12 \end{aligned}$$

Studi kasus pada penelitian ini HP memiliki banyak fitur dan pekerjaan yang mendukung, sehingga untuk mendapatkan nilai CF menggunakan perhitungan persamaan 3 adalah berikut ini:

$$\begin{aligned} \text{CFComb1} &= \text{CFFitur1} + \text{CFFitur6} * (1 - \text{CFFitur1}) \\ &= 0,8 + 0,48 * (1 - 0,8) \\ &= 0,8 + 0,48 * 0,2 \\ &= 0,8 + 0,96 \\ &= 0,896 \end{aligned}$$

$$\begin{aligned} \text{CFCom2} &= \text{CFCom1} + \text{CFFitur7} * (1 - \text{CFComb1}) \\ &= 0,896 + 0,8 * (1 - 0,896) \\ &= 0,896 + 0,8 * 0,106 \\ &= 0,896 + 0,0832 \\ &= 0,9792 \end{aligned}$$

$$\begin{aligned} \text{CFCom3} &= \text{CFCom2} + \text{CFFitur8} * (1 - \text{CFComb2}) \\ &= 0,9792 + 0,48 * (1 - 0,9792) \\ &= 0,9792 + 0,48 * (0,0208) \\ &= 0,9792 + 0,01 \\ &= 0,9892 \end{aligned}$$

$$\begin{aligned} \text{CFCom4} &= \text{CFCom3} + \text{CFFitur9} * (1 - \text{CFComb3}) \\ &= 0,9892 + 0,24 * (1 - 0,9892) \end{aligned}$$

$$\begin{aligned}
 &= 0,9892 + 0,24 * (0,010816) \\
 &= 0,9892 + 0,0026 \\
 &= 0,99178
 \end{aligned}$$

$$\begin{aligned}
 \text{CFCom5} &= \text{CFCom4} + \text{CFFitur10} * (1 - \text{CFComb4}) \\
 &= 0,99178 + 0,28 * (1 - 0,99178) \\
 &= 0,99178 + 0,28 * (0,00822) \\
 &= 0,99178 + 0,002301 \\
 &= 0,994081
 \end{aligned}$$

$$\begin{aligned}
 \text{CFCom6} &= \text{CFCom5} + \text{CFFitur12} * (1 - \text{CFComb5}) \\
 &= 0,994081 + 0,12 * (1 - 0,994081) \\
 &= 0,994081 + 0,12 * (0,005918) \\
 &= 0,994081 + 0,000710 \\
 &= 0,994791
 \end{aligned}$$

Berdasarkan dari hasil inputan user dan nilai kepercayaan dari ahli didapatkan nilai CF konsumen terhadap Samsung Galaxy S22 adalah sebesar 0,994791 dan dibulatkan menjadi 0,9948. Untuk dapat menentukan persentase keyakinan konsumen terhadap HP. Nilai CF Combine dikalikan dengan 100 %, dengan rumus sebagai berikut.

$$\begin{aligned}
 \text{Persentase Keyakinan} &= \text{CF Combine} * 100 \% \\
 &= 0,9948 * 100 \% \\
 &= 99,48\%
 \end{aligned}$$

Semua HP akan dilakukan perhitungan Certainty Factor untuk mendapatkan nilai keyakinan konsumen terhadap masing-masing HP tersebut. Adapun hasil nilai CF pada masing-masing HP ditunjukkan pada Tabel 6

Tabel 7. Nilai CF Masing – Masing HP

Kode HP	Persentase Keyakinan
L1	99,48
L2	0
L3	56,48
L4	69,54

Berdasarkan dari hasil Tabel 6 bahwa HP dengan kode L1 mendapatkan nilai kepercayaan tertinggi, sehingga tipe HP tersebut akan dijadikan sebagai rekomendasi untuk konsumen untuk membeli HP.

3.2. Pengujian Akurasi

Pengujian akurasi merupakan pengujian yang didapatkan dengan cara menghitung jumlah hasil yang benar dibagi dengan seluruh data uji. Pada penelitian ini data benar merupakan data hasil rekomendasi berdasarkan fitur yang dipilih konsumen dan dibenarkan hasilnya oleh seorang ahli. Penelitian ini menggunakan 20 kali uji hasil rekomendasi dan mendapatkan data benar sejumlah 16 sehingga nilai pengujian akurasi berdasarkan persamaan 5 adalah sebagai berikut

$$\begin{aligned}
 \text{Akurasi} &= \frac{16}{20} \times 100\% \\
 &= 0,8 \times 100\% \\
 &= 80\%
 \end{aligned}$$

Nilai akurasi dari perhitungan pada penelitian ini adalah 80% sehingga metode ini layak

digunakan untuk memberikan rekomendasi pada pembelian HP dengan menggunakan variabel fitur atau pekerjaan yang dapat dilakukan pada HP.

4. Kesimpulan

Perkembangan zaman membuat beberapa pekerjaan dituntut untuk menggunakan HP. Tetapi ketika ingin membeli HP konsumen sering bingung harus memilih HP karena banyak HP dengan berbagai macam jenis, tipe serta harga yang beragam. Untuk konsumen yang tidak memiliki latar belakang akan merasa kebingungan dalam memilih. Berdasarkan hal tersebut digunakan metode Certainty Factor untuk dapat memberikan rekomendasi pembelian HP. Berdasarkan dari hasil uji akurasi yang mendapatkan nilai 80 % membuktikan bahwa metode Certainty Factor dapat digunakan sebagai acuan untuk menentukan rekomendasi HP berdasarkan pekerjaan atau fitur yang diharapkan oleh konsumen.

Daftar Pustaka

- [1] A. R. Al-Mursyid, "Pengaruh Pembangunan Teknologi Informasi Dan Komunikasi Terhadap Pembangunan Kawasan Timur Indonesia," *J. Ekon. Pembang. STIE Muhammadiyah Palopo*, vol. 5, no. 2, 2020, doi: 10.35906/jep01.v5i2.372.
- [2] Mpssoft, "Sejarah Komputer Dari Awal Perkembangannya Hingga Sekarang," *Www.Mpssoft.Co.Id*. 2019.
- [3] S. Z. Dewi and I. Hilman, "Penggunaan TIK sebagai Sumber dan Media Pembelajaran Inovatif di Sekolah Dasar," *Indones. J. Prim. Educ.*, vol. 2, no. 2, 2019, doi: 10.17509/ijpe.v2i2.15100.
- [4] A. S. Perdhana, W. L. Y. Saptomo, and S. Siswanti, "Sistem Pendukung Keputusan Pemilihan Jenis HP Dengan Menggunakan Metode Analytical Hierarchy Process," *J. Teknol. Inf. dan Komun.*, vol. 1, no. 1, 2013.
- [5] G. P. Sanyoto, R. I. Handayani, and E. Widanengsih, "Sistem Pendukung Keputusan Pemilihan HP Untuk Kebutuhan Operasional Dengan Metode AHP (Studi Kasus: Direktorat Pembinaan Kursus Dan Pelatihan Kemdikbud)," *J. Pilar Nusa Mandiri*, vol. 13, no. 2, 2017.
- [6] F. N. Khasanah, "Metode Simple Additive Weighting Untuk Mendukung Pemilihan HP," *PIKSEL. Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 7, no. 1, 2019, doi: 10.33558/piksel.v7i1.1657.
- [7] Sumiati, H. Saragih, T. K. A. Rahman, and A. Triayudi, "Expert system for heart disease based on electrocardiogram data using certainty factor with multiple rule," *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, 2021, doi: 10.11591/ijai.v10.i1.pp43-50.
- [8] A. Sucipto, S. Ahdan, and A. Abyasa, "Usulan Sistem untuk Peningkatan Produksi Jagung menggunakan Metode Certainty Factor," *Prosiding-Seminar Nas. Tek. Elektro UIN Sunan Gunung Djati Bandung*, no. November 2019, 2020.
- [9] K. A. Aryani, D. G. H. Divayana, and I. M. A. Wirawan, "Sistem Pakar Diagnosis Penyakit Jerawat di Wajah dengan Metode Certainty Factor," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 2, 2017, doi: 10.23887/janapati.v6i2.11496.
- [10] A. Sucipto, Y. Fernando, R. I. Borman, and N. Mahmuda, "Penerapan Metode Certainty Factor Pada Diagnosa Penyakit Saraf Tulang Belakang," *J. Ilm. FIFO*, vol. 10, no. 2, 2019, doi: 10.22441/ffifo.2018.v10i2.002.
- [11] I. H. Santi and B. Andari, "Sistem Pakar Untuk Mengidentifikasi Jenis Kulit Wajah dengan Metode Certainty Factor," *INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*, vol. 3, no. 2, 2019, doi: 10.29407/intensif.v3i2.12792.
- [12] K. E. Setyaputri, A. Fadlil, and S. Sunardi, "Analisis Metode Certainty Factor pada Sistem Pakar Diagnosa Penyakit THT," *J. Tek. Elektro*, vol. 10, no. 1, 2018, doi: 10.15294/jte.v10i1.14031.

Klasifikasi Lagu Daerah di Indonesia dengan Metode *Machine Learning*

Gst. Ayu Vida Mastrika Giri^{a1}, Made Leo Radhitya^{a2}

^aProgram Studi Informatika, Universitas Udayana
Badung, Bali, Indonesia
¹vida@unud.ac.id

^bProgram Studi Teknik Informatika, Institut Bisnis dan Teknologi Indonesia
Denpasar, Bali, Indonesia
²leo.radhitya@instiki.ac.id

Abstract

Keunikan dari lagu daerah yang mencerminkan daerah asal adalah diiringi dengan alat musik daerah dan dinyanyikan dengan bahasa daerah masing-masing. Ciri khas lagu daerah dapat dilihat dari fitur-fitur musik seperti *spectral centroid* dan *Mel Frequency Cepstral Coefficients (MFCC)* karena dimainkan dengan alat musik berbeda dan memiliki timbre yang berbeda pula. Dengan menggunakan fitur-fitur musik tersebut dan algoritma *machine learning*, lagu-lagu daerah dapat diklasifikasi berdasarkan daerah asalnya. Pada penelitian ini digunakan sebuah dataset lagu daerah Indonesia yang bernama IRSD: Indonesian Regional Song Dataset yang terdiri dari 67 fitur musik yang diantaranya adalah MFCC, energy, dan *spectral centroid* dari 500 lagu daerah dari 10 provinsi di Indonesia. Metode *machine learning* yang akan digunakan untuk klasifikasi adalah SVM dan K-NN untuk menghasilkan nilai klasifikasi yang baik dengan waktu eksekusi yang cepat. Dengan menggunakan nilai $K=3$ dan 5-fold cross validation, metode K-NN menghasilkan nilai akurasi 0,69. Klasifikasi dengan metode SVM menggunakan kernel RBF dan 5-fold cross validation menghasilkan nilai akurasi 0,73. Pada penelitian kali ini, metode SVM dapat mengklasifikasi lagu daerah lebih baik daripada metode K-NN.

Keywords: K-Nearest Neighbor, klasifikasi, lagu daerah Indonesia, machine learning, Support Vector Machine

1. Pendahuluan

Wilayah negara Indonesia berbentuk kepulauan dari Sabang sampai Merauke yang memiliki keragaman suku dan kebudayaan. Setiap daerah di Indonesia memiliki kebudayaannya masing-masing yang unik jika dibandingkan dengan daerah lainnya. Salah satu jenis kebudayaan yang dimiliki setiap daerah adalah lagu daerah. Lagu daerah diciptakan di hampir setiap daerah di Indonesia dan memiliki sifat serta keunikan tersendiri yang mencerminkan ciri khas dari daerah asalnya [1]. Keunikan dari lagu daerah yang mencerminkan daerah asal adalah diiringi dengan alat musik daerah dan dinyanyikan dengan bahasa daerah masing-masing. Ciri khas lagu daerah dapat dilihat dari fitur-fitur musik seperti *spectral centroid* dan *Mel Frequency Cepstral Coefficients (MFCC)* karena dimainkan dengan alat musik berbeda dan memiliki timbre yang berbeda pula. Saat ini, ada ratusan lagu daerah Indonesia yang bisa didengarkan dan dipelajari. Dengan menggunakan fitur-fitur musik tersebut dan algoritma *machine learning*, lagu-lagu daerah dapat diklasifikasi berdasarkan daerah asalnya.

Penelitian tentang klasifikasi musik sebelumnya telah dilakukan oleh beberapa peneliti menggunakan fitur musik dan metode yang berbeda. Fitur MFCC telah digunakan untuk mengklasifikasi lagu daerah Indonesia berdasarkan *genre* menggunakan metode *machine learning Support Vector Machine (SVM)* dan menghasilkan akurasi sebesar 83% [2]. Pada penelitian yang dilakukan oleh Cettiar & Selvakumar tentang klasifikasi musik berdasarkan *genre* yang juga menggunakan fitur MFCC dan berbagai metode *machine learning* dan *deep learning*, metode SVM menghasilkan nilai akurasi tertinggi, yaitu 70,66% [3]. Metode *machine learning*

juga digunakan untuk klasifikasi bunyi alat musik dengan fitur MFCC pada penelitian yang telah dilakukan oleh Anuz, dkk. Metode SVM dan *K-Nearest Neighbor* (K-NN) yang digunakan pada penelitian tersebut menghasilkan nilai akurasi 97% dengan menggunakan metode SVM dan 96% dengan menggunakan K-NN [4].

Pada penelitian ini digunakan sebuah dataset lagu daerah Indonesia telah disusun oleh Mahardhika, dkk yang bernama IRSD: *Indonesian Regional Song Dataset* yang terdiri dari 67 fitur musik yang diantaranya adalah MFCC, *energy*, dan *spectral centroid* dari 500 lagu daerah dari 10 provinsi di Indonesia [5]. Metode *machine learning* yang akan digunakan untuk klasifikasi adalah SVM dan K-NN untuk menghasilkan nilai klasifikasi yang baik dengan waktu eksekusi yang cepat.

2. Metode Penelitian

Bagian ini berisi penjelasan tentang dataset yang digunakan serta alur yang digunakan dalam penelitian ini secara umum.

2.1. Dataset

Dataset yang digunakan pada penelitian ini adalah IRSD: Indonesian Regional Song Dataset [5] yang terdiri dari 500 lagu daerah dari 10 provinsi di Indonesia dengan pembagian jumlah 50 lagu daerah pada setiap provinsi. Provinsi yang terdaftar dalam dataset ini adalah DI Aceh, Sumatera Barat, Riau, Jawa Barat, DKI Jakarta, Jawa Tengah, Kalimantan Barat, Sulawesi Utara, Maluku, dan Papua. Sebanyak 67 fitur telah diekstrak dari setiap lagu berdasarkan domain waktu dan frekuensi. Tabel 1 menunjukkan daftar fitur yang diekstrak, domain, dan dimensi dari fitur tersebut. Dalam setiap fitur terdapat nilai *mean* dan *standard deviation* sehingga semua fitur berjumlah 67.

Tabel 1. Daftar Fitur [5]

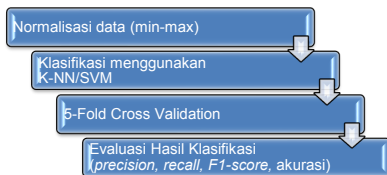
Nama Fitur	Domain	Dimensi
<i>Energy</i>	Waktu	2
<i>Energy Entrophy</i>		2
<i>Zero Crossing Rate (ZCR)</i>		2
<i>Spectral Centroid</i>		2
<i>Spectral Spread</i>	Frekuensi	2
<i>Spectral Entropy</i>		2
<i>Spectral Flux</i>		2
<i>Spectral Rolloff</i>		2
<i>MFCC 1 to MFCC 13</i>		26
<i>Chroma 1 to Chroma 12</i>		24
<i>Chroma Deviation</i>		1

Fitur yang termasuk pada domain waktu pada dataset IRSD adalah *energy*, *energy entrophy*, dan ZCR. Energi terkait dengan intensitas suara yang dirasakan, fitur ini digunakan untuk memperkirakan kenyaringan dan sebagai indikator kejadian baru dalam segmentasi audio. Entropi energi dapat diartikan sebagai perubahan yang terjadi secara spontan atau tiba-tiba [5]. ZCR adalah frekuensi saat sinyal waktu diskrit berubah tanda dari positif ke negatif atau sebaliknya [6].

Fitur domain frekuensi diekstrak dengan mengimplementasikan *Fourier Transform* pada sinyal audio musik. Pada dataset IRSD terdapat 8 fitur domain frekuensi seperti yang tertera pada Tabel 1. *Spectral centroid* adalah frekuensi dimana energi terpusat pada sebuah *frame* [7], hal ini dapat digunakan untuk mengukur "kecerahan" suara dan berhubungan dengan *timbre* musik. *Spectral spread* merupakan turunan dari *spectral centroid*, yang dapat diartikan sebagai varian dari frekuensi rata-rata dalam sinyal. *Spectral entropy* digunakan untuk mengukur besar kecilnya distribusi daya atau daya spektral. *Spectral flux* digunakan untuk menggambarkan perubahan daya atau spektrum daya secara berturut-turut antara setiap *frame* [5].

2.2. Gambaran Umum Penelitian

Penelitian ini melakukan klasifikasi lagu daerah berdasarkan daerah asalnya menggunakan fitur yang tersedia pada dataset IRSD. Alur penelitian ini, ditunjukkan pada Gambar 1, diawali dengan normalisasi data, klasifikasi lagu dengan K-NN/SVM yang dilakukan secara terpisah, *5-Fold Cross Validation* untuk memastikan bahwa dataset terbagi dengan baik, dan diakhiri dengan evaluasi hasil klasifikasi dengan menghitung nilai *precision*, *recall*, *F1-Score*, dan nilai akurasi.



Gambar 1. Alur Umum Penelitian

Data fitur yang terdapat pada dataset IRSD belum berada pada *range* yang sama, sehingga perlu dilakukan normalisasi. Persamaan 1 digunakan untuk menormalisasi data dengan metode min-max, agar semua fitur berada dalam *range* 0-1 [8].

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

Klasifikasi dengan metode *machine learning* K-NN pada penelitian ini menggunakan *python library* Scikit Learn, *Grid Search* untuk mencari nilai K optimal, serta menggunakan *5-Fold Cross Validation* untuk membagi sampel data secara acak. Klasifikasi dengan metode *machine learning* SVM pada penelitian ini menggunakan *kernel* RBF. Sama dengan klasifikasi dengan metode K-NN, *5-Fold Cross Validation* juga digunakan pada klasifikasi dengan SVM.

Evaluasi hasil klasifikasi dilakukan dengan menghitung nilai *precision*, *recall*, dan *F1 score*, dan nilai akurasi. *Precision* adalah jumlah prediksi kasus positif yang benar/sesuai dengan kelas sebenarnya (*true positive*) yang dihitung dengan persamaan (2), dimana TP adalah *true positive*, dan FP adalah *false positive* [9].

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

Nilai *recall* adalah jumlah prediksi kasus positif yang diklasifikasi benar/sesuai kelas sebenarnya oleh metode klasifikasi dibandingkan dengan semua data kasus positif yang ada pada dataset. Nilai *recall* dihitung dengan persamaan (3), dimana TP adalah *true positive*, dan FN adalah *false negative* [9].

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

F1-Score didapat dengan menggabungkan nilai *precision* dan *recall* yang dikalkulasi menggunakan persamaan (4) [9].

$$F1\ Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

Nilai akurasi, berbeda dengan nilai *precision*, *recall*, dan *F1-Score*, menghitung perbandingan secara keseluruhan jumlah kasus positif yang diprediksi benar dengan jumlah total prediksi. Persamaan (5) digunakan untuk menghitung nilai akurasi [9].

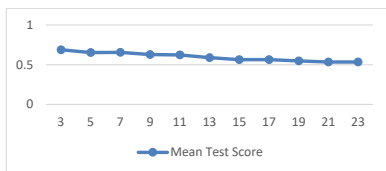
$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ prediction} \tag{5}$$

3. Hasil dan Pembahasan

Bagian ini memuat hasil dan pembahasan dari penelitian yang telah dilakukan. Penjelasan akan dibagi menjadi dua, yaitu hasil pembahasan klasifikasi menggunakan K-NN dan SVM.

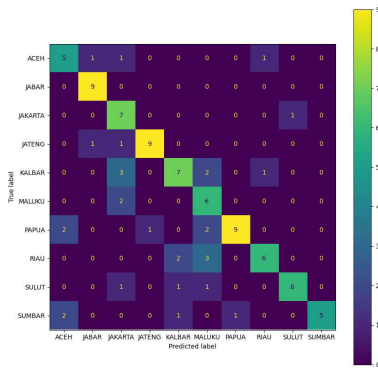
3.1. Klasifikasi Menggunakan K-Nearest Neighbor (K-NN)

Klasifikasi dengan menggunakan metode K-NN dimulai dengan menggunakan *grid search* untuk menentukan nilai K terbaik dengan mencoba keseluruhan kemungkinan yang ada (pencarian menyeluruh/*exhaustive search*). Nilai K yang akan digunakan dipilih berdasarkan hasil *cross validation* terbaik dari setiap percobaan. *Grid search* dilakukan dalam *range* nilai K=3 hingga K=50. Pada Gambar 2 diperlihatkan hasil *mean test score*, hasil rata-rata dari tiap *5-fold cross validation* dari 10 nilai K pertama yaitu K=3 sampai K=23. Nilai *means test score* tertinggi ada pada nilai K=3, yaitu 0,688. Nilai K setelahnya menghasilkan nilai yang tidak lebih tinggi dari 0,55.



Gambar 2. Hasil *Grid Search* untuk 10 Nilai K Pertama.

Selanjutnya, klasifikasi dengan K-NN menggunakan nilai K=3. Hasil klasifikasinya ditunjukkan pada Gambar 3. Provinsi yang paling banyak lagu daerahnya benar diklasifikasi adalah Jawa Barat, Jawa Tengah, dan Papua. Jumlah prediksi tertinggi tersebut adalah 9 lagu daerah. Klasifikasi terendah ada pada Provinsi Aceh dan Sumatera Barat, dengan 5 lagu daerah. Lebih jelas tentang hasil klasifikasi menggunakan K-NN dapat dilihat pada Tabel 2.



Gambar 3. Confusion Matrix Hasil Klasifikasi dengan Metode K-NN

Hasil *precision*, *recall*, dan *F1-Score* untuk setiap provinsi ditunjukkan pada Tabel 2. Nilai *F1-Score* tertinggi ada pada Provinsi Jawa Barat, dengan nilai 0,90, disusul dengan Provinsi Jawa Tengah dengan Nilai 0,86, dan Papua dengan nilai 0,75. Nilai *F1-Score* terendah ada pada Provinsi Aceh, dengan nilai 0,59. Rata-rata nilai *F1-Score* yang didapat adalah 0,70. Nilai akurasi total dari semua provinsi adalah 0,69, yang menyatakan bahwa 69% lagu daerah terklasifikasi dengan benar berdasarkan provinsi/daerah asal lagu tersebut dengan menggunakan metode *machine learning* K-NN.

Tabel 2. Nilai *Precision*, *Recall*, dan *F1-Score* Hasil Klasifikasi Lagu Daerah dengan K-NN

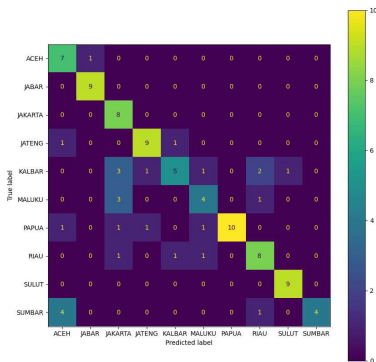
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Aceh	0,56	0,62	0,59
Jabar	0,82	1,00	0,90
Jakarta	0,47	0,88	0,61
Jateng	0,90	0,82	0,86
Kalbar	0,64	0,54	0,58
Maluku	0,43	0,75	0,55
Papua	0,90	0,64	0,75
Riau	0,75	0,55	0,63
Sulut	0,86	0,67	0,75
Sumbar	1,00	0,56	0,71
Macro Average	0,73	0,70	0,70

3.2. Klasifikasi Menggunakan *Support Vector Machine* (SVM)

Klasifikasi dengan metode *machine learning* SVM pada penelitian ini menggunakan *kernel* RBF dan sama dengan klasifikasi dengan K-NN, klasifikasi dengan SVM juga menggunakan 5-fold

cross validation. Hasil *means* dari seluruh *cross validation* adalah 0,686, dengan masing-masing hasil *cross validation* 0,65, 0,8, 0,65, 0,71, dan 0,62.

Confusion matrix dari hasil klasifikasi dengan SVM dapat dilihat pada Gambar 4. Lagu daerah yang paling banyak benar diklasifikasi berdasarkan daerah asalnya adalah lagu daerah Papua, dengan jumlah 10 lagu, selanjutnya adalah Jawa Barat, Jawa Tengah, dan Sulawesi Utara dengan 9 lagu. Lagu daerah yang paling banyak salah diklasifikasi adalah lagu daerah Sumatera Barat, sebagian besar lagunya salah diklasifikasi berasal dari Aceh. Lebih jelas tentang hasil klasifikasi dapat dilihat pada Tabel 3.



Gambar 4. *Confusion Matrix* Hasil Klasifikasi dengan Metode SVM

Nilai *F1-score* tertinggi ada pada klasifikasi lagu daerah Jawa Barat dan Sulawesi Utara dengan nilai 0,95, disusul dengan Jawa Tengah dengan nilai 0,82. Nilai rata-rata *F1-score* untuk keseluruhan hasil klasifikasi adalah 0,72. Perhitungan akurasi juga telah dilakukan dan menghasilkan 73% lagu daerah dapat diklasifikasi dengan benar dengan menggunakan metode *machine learning* SVM.

Tabel 3. Nilai *Precision*, *Recall*, dan *F1-Score* Hasil Klasifikasi Lagu Daerah dengan SVM

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Aceh	0,54	0,88	0,67
Jabar	0,90	1,00	0,95
Jakarta	0,50	1,00	0,67
Jateng	0,82	0,82	0,82
Kalbar	0,71	0,38	0,50
Maluku	0,57	0,50	0,53
Papua	1,00	0,71	0,83
Riau	0,67	0,73	0,70
Sulut	0,90	1,00	0,95

	Precision	Recall	F1-Score
Sumbar	1,00	0,44	0,62
Macro Average	0,76	0,75	0,72

4. Kesimpulan

Klasifikasi lagu daerah di Indonesia telah berhasil dilakukan dengan menggunakan dataset IRSD dua metode *machine learning*, yaitu K-NN dan SVM. Dataset IRSD yang digunakan dinormalisasi terlebih dahulu sebelum masuk ke metode klasifikasi. Klasifikasi dengan metode K-NN menggunakan *grid search* untuk menemukan nilai K terbaik. Dengan menggunakan nilai K=3 dan *5-fold cross validation*, metode K-NN menghasilkan nilai akurasi 0,69. Klasifikasi dengan metode SVM menggunakan *kernel RBF* dan *5-fold cross validation* menghasilkan nilai akurasi 0,73. Pada penelitian kali ini, metode SVM dapat mengklasifikasi lagu daerah lebih baik daripada metode K-NN.

Daftar Pustaka

- [1] G. Santoso, R. Sakinah, A. S. Hidayat, A. Ramadhania, S. N. Tiara, D. Safitri and G. Geifira, "Mengenal Lagu Daerah dan Lagu Nasional Republik Indonesia Sebagai Pendidikan Multikultural bagi Mahasiswa," *Jurnal Pendidikan Transformatif (Jupetra)*, vol. 2, no. 2, pp. 325-335, 2021.
- [2] S. Y. Yehezkiel and Y. Suyanto, "Music Genre Identification Using SVM and MFCC Feature Extraction," *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, vol. 12, no. 2, pp. 115-122, 2022.
- [3] G. Chettiar and K. Selvakumar, "Music Genre Classification Techniques," in *International Conference on Innovation Challenges and Advances in Engineering & Technology: A road to self-reliant India (ICAET-2021) - Virtual Conference*, 2021.
- [4] H. Anuz, A. K. M. Masum, S. Abujar and S. A. Hossain, "Musical Instrument Classification Based on Machine Learning Algorithm," in *Emerging Technologies in Data Mining and Information Security*, Springer, 2021, pp. 57-67.
- [5] F. Mahardhika, H. L. H. S. Warnars, A. S. Nugroho and W. Budiharto, "IRSD: Indonesian Regional Song Dataset," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 1107-1117, 2023.
- [6] P. Sahil, V. Chetan and M. Nikita, "10.1007/978-981-19-2177-3_83," *10.1007/978-981-19-2177-3_83*, vol. 13, no. 1, pp. 15-21, 2022.
- [7] D. Lau and R. Ajooda, "Music Genre Classification: A Comparative Study Between Deep-Learning And Traditional Machine," in *Proceedings of Sixth International Congress on Information and Communication Technology*, Singapore, 2021.
- [8] I. N. Y. T. Giri, L. A. A. R. Putri, G. A. V. M. Giri, G. N. A. C. Putra, I. M. Widiartha and I. W. Supriyana, "Music Genre Classification Using Modified K-Nearest Neighbor (MK-NN)," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 10, no. 3, pp. 261-270, 2022.
- [9] G. A. V. M. Giri and M. L. Radhitya, "Klasifikasi Genre Musik Menggunakan Teknik Pembelajaran Mesin," *Jurnal Teknologi Informasi dan Komputer (JuTIK)*, vol. 9, no. 1, pp. 1-9, 2023.
- [10] R. Luis and N. Rokhman, *Klasifikasi Daerah Asal Musik Tradisional menggunakan Convolutional Neural Network (CNN) dan Mel Frequency Cepstrum Coefficients (MFCC)*, Yogyakarta: Universitas Gadjah Mada, 2022.
- [11] P.-N. Tan, M. Steinbach, A. Karpatne and V. Kumar, *Introduction to Data Mining*, Harlow: Pearson Education Limited, 2019.

Halaman ini sengaja dibiarkan kosong

