

Bit

Fakultas Teknologi Informasi
Universitas Budi Luhur



**Fakultas Teknologi Informasi
Universitas Budi Luhur**

Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260
<https://journal.budiluhur.ac.id/index.php/bit/index>

Penanggung Jawab

Achmad Solichin

Ketua Redaksi

Achmad Solichin

Wakil Ketua Redaksi

Atik Ariesta

Redaksi Pelaksana

Kukuh Harsanto

Painem

Alamat Redaksi

Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)

Fakultas Teknologi Informasi, Universitas Budi Luhur

Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260

Telp: 021-585 3753 Fax: 021-585 3752

MITRA BESTARI

1. Albar Rubhasy, Universitas Nasional, Indonesia
2. Andhika Octa Indarso, UPN Veteran Jakarta, Indonesia
3. Anita Ratnasari, Universitas Mercu Buana, Indonesia
4. Arief Wibowo, Universitas Budi Luhur, Indonesia
5. Dwi Pebrianti, Universitas Malaysia Pahang, Malaysia
6. Falahah, Universitas Telkom, Indonesia
7. Gandung Triyono, Universitas Budi Luhur, Indonesia
8. Grace Gata, Universitas Budi Luhur, Indonesia
9. Hari Soetanto, Universitas Budi Luhur, Indonesia
10. Hendra Cipta, Universitas Islam Negeri Sumatera Utara, Indonesia
11. Imelda, Universitas Budi Luhur, Indonesia
12. Indra, Universitas Budi Luhur, Indonesia
13. Iwan Setiawan, Universitas Nusa Putra, Indonesia
14. Jan Everhard Riwurohi, Universitas Budi Luhur, Indonesia
15. Kelik Sussolaikah, Universitas PGRI Madiun, Indonesia
16. Mardi Hardjianto, Universitas Budi Luhur, Indonesia
17. Mayanda Mega Santoni, UPN Veteran Jakarta, Indonesia
18. Mohammad Syafrullah, Universitas Budi Luhur, Indonesia
19. Painem, Universitas Budi Luhur, Indonesia
20. Rohmat Indra Borman, Universitas Teknokrat, Indonesia
21. Rusdah, Universitas Budi Luhur, Indonesia
22. Safitri Juanita, Universitas Budi Luhur, Indonesia
23. Setyawan Widyarto, Universiti Selangor, Malaysia
24. Siswanto, Universitas Budi Luhur, Indonesia
25. Windu Gata, Universitas Nusa Mandiri, Indonesia

Implementasi Smartphone Sebagai Smart Security Berbasis Cloud Storage <i>Gufron Gufron</i>	105-111
Penerapan E-Commerce Berbasis CMS dengan Metode Business Model Canvas Pada Noru Label <i>Bianca Vania Adeline, Agnes Aryasanti, Anita Diana, Mohammad Anif</i>	112-119
Implementasi Website E-Commerce Berbasis Content Management System Pada Toko Vape.Bill Indonesia <i>Muhammad Rafi Pratama, Lauw Li Hin</i>	120-130
Klasterisasi Daerah Kriminalitas di Indonesia dengan Metode K-Means Clustering <i>T.Sofia Chairani, Hijka Listia, Sabina Wardaniah, Siti Wulandari, Putri Tasya Agustina Tampubolon, Arnita Piliang</i>	131-138
Implmentasi Algoritma Lattern Dirichlet Allocation (LDA) Dengan Gibss Sampling Untuk Topic Modelling Menggunakan Data X Terkait IKN <i>Raihan Labib Hanif, Mohammad Syafrullah</i>	139-144
Deteksi Website Phishing Dari Analisis Url Menggunakan Algoritma Random Forest <i>Damar Bambang Suwarno, Mardi Hardjianto</i>	145-152
Penerapan Algoritma Agglomerative Clustering Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Pendidikan <i>Hafiih Taufiqul Hakim, Wendi Usino</i>	153-160
Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Gojek Dan Indrive Pada Google Playstore Dengan Metode Algoritma K-Nearest Neighbor <i>Teguh Febriyanto, Achmad Solichin</i>	56-64
Sistem Monitoring dan Controlling Kualitas Air pada Aquarium Ikan Hias Berbasis Internet Of Things <i>Wahyu Dwi Ramadhan, Subandi Subandi</i>	161-168
Klasterisasi Tingkat Kemiskinan Menggunakan Metode K-Means di DKI Jakarta <i>Muhamad Khaerul Rafli, Utomo Budiyanto</i>	169-177

IMPLEMENTASI SMARTPHONE SEBAGAI SMART SECURITY BERBASIS CLOUD STORAGE

Gufron^{1*}

¹Pendidikan Teknik Informatika & Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bung Hatta
Email: ¹gufron@gufron.com

(Naskah masuk: 2 September 2024, diterima untuk diterbitkan: 30 September 2024)

Abstrak

Keamanan merupakan salah satu aspek penting dalam kehidupan sehari-hari, baik dalam konteks pribadi maupun profesional, terutama dalam melindungi properti dari ancaman pencurian, perusakan, dan aktivitas kriminal lainnya. Untuk itu, banyak pihak memasang sistem pengawasan seperti CCTV. Akan tetapi, CCTV tradisional dengan penyimpanan lokal seperti harddisk atau dvr memiliki kelemahan, terutama karena rentannya terhadap perusakan fisik oleh pelaku kejahatan. Seiring perkembangan teknologi informasi dan komunikasi, akses ke peralatan CCTV juga berkembang dari kabel ke Wi-Fi, dan media penyimpanan digantikan dengan SD-card yang lebih kecil. Meski demikian, kelemahan terkait perusakan fisik pada media penyimpanan masih tetap ada. Untuk mengatasi hal tersebut, penggunaan smartphone sebagai alat keamanan kini semakin populer. Dilengkapi dengan kemampuan komputasi tinggi, berbagai sensor, kamera berkualitas, dan konektivitas internet yang baik, smartphone dapat diubah menjadi perangkat keamanan yang dapat memonitor situasi secara *real-time* dan mengunggah rekaman langsung ke penyimpanan awan, sehingga aman dari upaya perusakan atau pencurian perangkat fisik. Pengguna juga dapat mengakses rekaman kapan saja melalui internet, sehingga meningkatkan fleksibilitas dan kenyamanan dalam menjaga keamanan properti mereka. Penelitian ini menggunakan pendekatan deskriptif dengan metode studi kasus untuk mengkaji implementasi smartphone sebagai sistem keamanan cerdas berbasis penyimpanan awan. Berdasarkan hasil implementasi dan pengujian yang dilakukan, penggunaan smartphone terbukti efektif dalam mendukung pemantauan dan pengelolaan keamanan secara aktif. Penggunaan sensor bawaan pada smartphone memungkinkan deteksi ancaman yang cepat dan akurat, sementara integrasi dengan penyimpanan awan memberikan penyimpanan yang aman, fleksibel, dan mudah diakses. Dengan demikian, kombinasi smartphone dan cloud storage dapat menjadi solusi untuk kebutuhan keamanan modern.

Kata kunci: *cctv, keamanan cerdas, penyimpanan awan, smartphone*

IMPLEMENTATION OF SMARTPHONE AS SMART SECURITY BASED ON CLOUD STORAGE

Abstract

Security is one of the important aspects in everyday life, both in personal and professional contexts, especially in protecting property from the threat of theft, vandalism, and other criminal activities. For this reason, many parties install surveillance systems such as CCTV. However, traditional CCTV with local storage such as hard disks or DVRs has weaknesses, especially because it is vulnerable to physical damage by criminals. Along with the development of information and communication technology, access to CCTV equipment has also developed from cables to Wi-Fi, and storage media has been replaced with smaller SD cards. However, weaknesses related to physical damage to storage media still exist. To overcome this, the use of smartphones as security tools is now increasingly popular. Equipped with high computing capabilities, various sensors, quality cameras, and good internet connectivity, smartphones can be turned into security devices that can monitor situations in real-time and upload recordings directly to cloud storage, so they are safe from attempts to damage or theft of physical devices. Users can also access recordings at any time via the internet, increasing flexibility and convenience in maintaining the security of their property. This study uses a descriptive approach with a case study method to examine the implementation of smartphones as intelligent security systems based on cloud storage. Based on the results of the implementation and testing conducted, the use of smartphones has proven effective in supporting active security monitoring and management. The use of built-in sensors on smartphones allows for fast and accurate threat detection, while integration with cloud storage provides secure, flexible, and easily accessible storage. Thus, the combination of smartphones and cloud storage can be a solution for modern security needs.

Keywords: *cctv, cloud storage, smart security, smartphone*

1. PENDAHULUAN

Keamanan rumah dan tempat usaha merupakan aspek yang semakin penting dalam kehidupan modern. Untuk memantau dan melindungi properti mereka dari potensi ancaman seperti pencurian, perusakan, dan aktivitas kriminal lainnya, banyak pihak memasang sistem pengawasan seperti *closed-circuit television* (cctv) [1]. Akan tetapi, sistem cctv tradisional yang menggunakan penyimpanan lokal seperti harddisk atau *digital video recorder* (dvr) memiliki beberapa kelemahan, terutama terkait dengan kerentanannya terhadap perusakan fisik. Pelaku kejahatan yang menyadari keberadaan cctv sering kali merusak atau mencuri perangkat penyimpanan untuk menghilangkan bukti yang dapat digunakan untuk mengidentifikasi pelaku kejahatan dan membawa mereka ke ranah hukum [2].

Seiring perkembangan teknologi informasi dan komunikasi yang semakin pesat, sistem pengawasan dan keamanan juga mengalami perkembangan. Cctv yang sebelumnya memerlukan kabel dalam pemasangan dan komunikasinya, saat ini mulai menggunakan wifi, dan penggunaan hard disk mulai diganti dengan media penyimpanan yang lebih kecil dalam bentuk sd-card. Walau begitu, kelemahan terkait dengan perusakan fisik media penyimpanannya tetap ada. Cctv dengan penyimpanan *cloud* juga sudah mulai tersedia [3]. Cctv ini dikenal dengan nama *IP Camera*. Namun pada umumnya penyimpanan awan (*storage cloud*) yang dapat digunakan merupakan *storage cloud* berbayar dari vendor cctv atau rekanan vendor cctv itu sendiri, sehingga memberikan biaya operasional tambahan yang tidak sedikit.

Sebagai tanggapan terhadap kelemahan ini dan untuk meningkatkan sistem keamanan dengan cara yang lebih efektif dan efisien, penggunaan smartphone sebagai alat untuk keamanan semakin populer [4],[5]. Dilengkapi dengan kemampuan komputasi yang cukup tinggi mencakup pemrosesan data yang cepat, terintegrasi dengan berbagai sensor, kamera berkualitas tinggi, konektivitas internet yang baik dan aplikasi yang tepat, smartphone dapat diubah menjadi perangkat keamanan yang dapat mengunggah rekaman langsung ke *cloud storage*, sehingga aman dari upaya perusakan atau pencurian perangkat fisik. Dengan memanfaatkan sensor yang ada pada smartphone, seperti kamera, mikrofon, dan GPS, sistem keamanan dengan smartphone dapat memonitor situasi dan kondisi lingkungan secara *real-time*. Pengguna juga dapat mengakses rekaman kapan saja melalui internet, yang meningkatkan fleksibilitas dan kenyamanan.

Beberapa studi telah menunjukkan bahwa penggunaan teknologi *cloud* dalam sistem keamanan dapat meningkatkan efektivitas dan keandalan sistem tersebut. Penelitian yang dilakukan [6] membahas tentang implementasi *cloud computing* dalam layanan publik, terutama terkait keamanan data dan sistem. Penelitian ini berfokus pada bagaimana

layanan publik yang berbasis *cloud* menghadapi tantangan dan risiko keamanan. Penelitian yang dilakukan [7] yang membahas tentang implementasi cctv berbasis *internet of things* sebagai *smart security* untuk menanggulangi angka kejahatan dalam lingkungan pendidikan. Sementara penelitian ini fokus pada penggunaan smartphone yang terhubung ke *cloud storage* sebagai sistem keamanan pintar, di mana data dapat diakses dan dipantau dari jarak jauh, serta disimpan dengan aman di *cloud*. Dengan demikian, penelitian ini menawarkan keamanan yang lebih terjangkau dan mudah diakses oleh individu, sehingga cocok untuk diterapkan dalam berbagai lingkungan tanpa memerlukan infrastruktur fisik yang kompleks seperti IoT CCTV.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk mengkaji implementasi smartphone sebagai cctv berbasis *cloud storage* yang lebih aman dan fleksibel. Penelitian ini akan mengeksplorasi keuntungan, tantangan, dan solusi potensial untuk meningkatkan keamanan dan efektivitas sistem ini. Penelitian ini juga akan membahas bagaimana teknologi ini dapat diakses oleh masyarakat luas, dengan mempertimbangkan faktor-faktor seperti biaya, kemudahan penggunaan, dan ketersediaan infrastruktur pendukung seperti layanan internet yang memadai. Selain itu, penelitian ini akan meneliti berbagai aplikasi dan platform *cloud storage* yang dapat digunakan dalam konteks ini, serta bagaimana mereka dapat diintegrasikan dengan smartphone untuk menciptakan sistem pengawasan yang handal dan aman.

Penelitian ini diharapkan dapat memberikan kontribusi praktis dengan menghasilkan panduan bagi pengguna dalam memilih dan mengimplementasikan solusi cctv berbasis smartphone dan *cloud storage*, serta memberikan rekomendasi tentang praktik terbaik dalam penggunaan teknologi ini. Serta menjadi alternatif terutama di kalangan individu dan usaha kecil yang mungkin tidak mampu atau tidak ingin berinvestasi dalam sistem cctv konvensional.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif dengan metode studi kasus untuk mengkaji implementasi smartphone sebagai *smart security* berbasis *cloud storage*. Penelitian deskriptif merupakan strategi penelitian dimana di dalamnya peneliti menyelidiki kejadian, fenomena kehidupan individu-individu dan meminta seorang atau sekelompok individu untuk menceritakan kehidupan mereka. Informasi ini kemudian diceritakan kembali oleh peneliti dalam kronologi deskriptif [8], [9]. Karakteristik dari deskriptif sendiri adalah data yang diperoleh berupa kata-kata, gambar, dan bukan angka-angka seperti penelitian kuantitatif. Pada dasarnya penelitian deskriptif adalah penelitian yang berusaha untuk mendeskripsikan dan menginterpretasikan sesuatu, misalnya situasi dan

kondisi dengan hubungan yang ada, pendapat-pendapat yang berkembang, akibat atau efek yang terjadi dan sebagainya [10].

Untuk tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Berikut penjelasan tahapan penelitian:

a. Pengumpulan Data

Penelitian dimulai dengan pengumpulan data sekunder yang mencakup literatur terkait, jurnal, dan studi sebelumnya tentang penggunaan smartphone dalam sistem keamanan dan teknologi *cloud*. Tujuan dari tahapan ini adalah untuk memperoleh pemahaman yang mendalam mengenai teknologi dan konsep yang ada, serta untuk mengidentifikasi tren dan tantangan dalam implementasi smartphone sebagai perangkat keamanan berbasis *cloud*.

b. Analisis Kebutuhan

Berdasarkan data yang telah dikumpulkan, dilakukan analisis kebutuhan yang bertujuan untuk mengidentifikasi aplikasi yang akan digunakan, fitur dan fungsi utama yang harus ada dalam sistem keamanan berbasis smartphone serta layanan *cloud storage* yang tersedia.

c. Implementasi Sistem

Hasil analisis digunakan sebagai dasar dalam implementasi sistem yang melibatkan pemilihan aplikasi yang tepat sesuai dengan kebutuhan serta integrasi smartphone dengan *cloud storage* yang digunakan.

d. Pengujian Sistem

Dalam tahapan ini, sistem akan diuji dalam lingkungan yang terkendali untuk memastikan bahwa semua fitur dan fungsi berjalan sesuai dengan yang direncanakan. Uji coba ini mencakup proses smartphone merekam video dan mengunggahnya secara otomatis ke *cloud storage*, manajemen video rekaman di *cloud storage* dan kemudahan penggunaan aplikasi. Pengujian dilakukan menggunakan metode *blackbox*, dimana pengujian hanya pada fungsional sistem yang telah dibangun [11]. Pengujian ini penting untuk memastikan bahwa sistem yang dikembangkan tidak hanya berfungsi dengan baik tetapi juga memberikan pengalaman pengguna yang optimal.

3. HASIL DAN PEMBAHASAN

Implementasi penelitian ini menggunakan aplikasi CamOn *live streaming* yang tersedia di

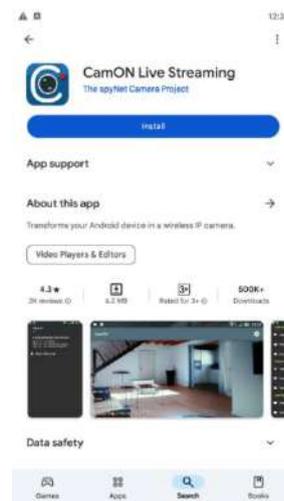
Google playstore dan dapat diunduh gratis oleh semua orang. Dari penelusuran yang dilakukan, banyak aplikasi di playstore yang dapat digunakan untuk menjadikan smartphone sebagai perangkat cctv. Pemilihan CamOn *live streaming* didasarkan pada ukurannya yang kecil (hanya 6,22 MB) sehingga tidak membutuhkan ruang penyimpanan yang besar di smartphone. Selain sebagai cctv dan webcam, sesuai dengan namanya, CamOn juga dapat digunakan dalam kegiatan *live streaming*.

Kelebihan lain yang ditawarkan CamOn adalah pilihan lokasi penyimpanan yang akan digunakan, terutama tersedianya pilihan untuk menggunakan Google drive sebagai *cloud storage*. Dengan akun Google gratis yang kita gunakan, kita sudah mendapatkan kapasitas penyimpanan sampai 15 GB. Kapasitas penyimpanan sebesar itu memadai karena kita bisa menentukan proses penghapusan rekaman secara otomatis dalam waktu mulai dari per jam sampai maksimal per minggu.

Selain fitur yang tersedia lengkap, CamOn dapat digunakan secara gratis tanpa ada iklan yang mengganggu. Tidak seperti aplikasi cctv lainnya, walaupun bisa digunakan secara gratis, ketika aplikasi dijalankan, maka secara berkala akan menampilkan iklan yang tentunya dapat mengganggu kinerja smartphone dan menghabiskan paket data Internet yang digunakan.

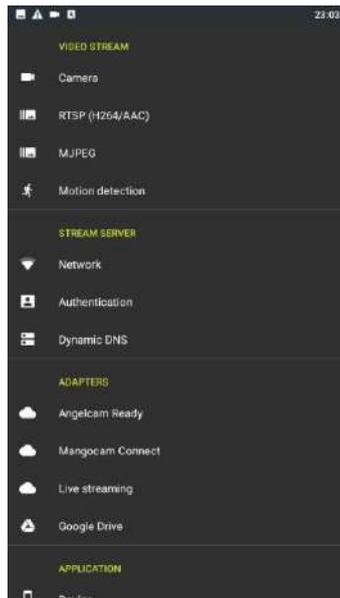
3.1 Instalasi dan Pengaturan

Agar dapat menjadikan smartphone sebagai cctv, langkah pertama adalah mengunduh CamOn *live streaming* langsung di playstore seperti terlihat pada gambar 2 dan menginstallnya pada perangkat smartphone.



Gambar 2. Instalasi CamON Live Streaming

Langkah selanjutnya melakukan pengaturan aplikasi. Berikut beberapa pengaturan yang perlu dilakukan agar dapat memaksimalkan fungsi CamOn *live streaming* seperti terlihat pada gambar 3:



Gambar 3. Pengaturan CamOn live streaming

a. Camera

Di bagian kamera, kita dapat menentukan apakah menggunakan kamera belakang atau kamera depan dari smartphone sebagai input. Kita juga dapat menentukan resolusi video yang akan diambil, dimana opsi yang tersedia secara otomatis menyesuaikan dengan kemampuan kamera smartphone. Kita juga menambahkan watermark catatan waktu rekaman.

b. RTSP (H264/AAC)

CamOn live streaming mendukung *Real Time Streaming Protocol* (RTSP) sehingga kita dapat memantau kondisi area secara *real time* menggunakan aplikasi media player biasa. Untuk mengatur kualitas video dapat dilakukan pada menu RTSP (H264/AAC)

c. MJPEG

Untuk kualitas tangkapan gambar di atur pada bagian MJPEG (*Motion Joint Expert Group*).

d. *Motion detection*

Aplikasi CamOn juga memiliki kemampuan untuk melakukan perekaman ketika menangkap adanya gerakan. Di bagian ini juga dilengkapi username dan password agar hanya orang yang berhak yang dapat mengakses kamera cctv kita.

e. Network

Pengaturan bagian network berguna untuk menentukan port (saluran) akses ke kamera. Dengan adanya fitur ini, kita dapat menggunakan satu alamat IP publik untuk mengakses berbagai perangkat lain. Di bagian ini kita juga dapat memastikan agar aplikasi CamOn tidak menggunakan paket data, dan hanya jaringan lokal.

f. *Authetication*

Agar akses kamera hanya dapat diakses oleh orang yang berwenang, maka kita dapat menambahkan user dan password

g. Siaran Langsung (*Live streaming*)

Sesuai namanya, kita dapat melakukan *live streaming* dengan CamOn. Agar dapat melakukan *live streaming* dengan CamOn, kita harus memiliki akun di Youtube dan akun tersebut sudah diizinkan oleh Youtube untuk melakukan *live streaming*.

h. Google Drive

Salah satu kelebihan CamOn *live streaming* adalah tersedianya opsi untuk menggunakan Google drive sebagai lokasi penyimpanannya. Dengan demikian, kita tidak perlu menyewa *cloud storage* khusus untuk menyimpan hasil perekaman. Pengiriman hasil rekaman ke Google drive dilakukan secara berkala sesuai dengan waktu yang kita tentukan. Agar Google drive yang kita gunakan tidak penuh, kita dapat menentukan berapa lama hasil rekaman tersimpan di Google drive sebelum dihapus secara otomatis.

3.2. Implementasi Smartphone Sebagai Pusat Kendali Keamanan

Pengujian smartphone sebagai pusat kendali sistem keamanan menunjukkan bahwa perangkat ini mampu menjalankan fungsi utama seperti deteksi, pemantauan, dan pengelolaan keamanan dengan efektif. Berikut adalah temuan utama dari hasil implementasi tersebut:

a. Kemampuan deteksi pergerakan (*motion detection*)

Smartphone yang digunakan dalam sistem keamanan ini mampu memanfaatkan sensor-sensor bawaan seperti kamera, GPS, dan accelerometer untuk mendeteksi aktivitas yang mencurigakan. Pada beberapa skenario uji, ketika opsi pengaturan *Motion Detection* diaktifkan, saat ada gerakan tidak terduga di area yang dipantau oleh kamera smartphone, perangkat secara otomatis akan menghidupkan alarm dan penerangan (fungsi senter pada smartphone) dan aplikasi akan langsung mengambil *screenshot* dan video rekaman *motion detection* yang segera dikirimkan ke cloud. Ini membuktikan bahwa smartphone tidak hanya berfungsi sebagai alat monitoring pasif, tetapi juga berperan sebagai perangkat aktif yang memproses informasi di tingkat awal (*local processing*).

b. Kecepatan respon dan pemantauan *real-time*

Penggunaan smartphone memungkinkan sistem untuk memberikan respon *real-time*. Dari hasil pengujian, ancaman atau aktivitas mencurigakan akan memicu alarm dan menghidupkan lampu senter pada smartphone dalam waktu rata-rata 1 detik. Sistem juga mendukung pemantauan secara langsung (*live feed*) melalui browser atau aplikasi media player yang tersedia, seperti vlc media player, sehingga memungkinkan pengguna untuk melihat kondisi terkini dari lokasi yang dipantau. Hal ini memberikan fleksibilitas yang tidak dimiliki oleh sistem keamanan konvensional yang sering kali memerlukan alat khusus atau server lokal untuk pemantauan atau ip camera yang memerlukan aplikasi khusus yang harus

diinstall di smartphone untuk melakukan pemantauan langsung.

c. Mobilitas dan aksesibilitas

Dengan menggunakan jaringan internet, pengguna dapat memantau keamanan rumah atau properti mereka dari mana saja menggunakan smartphone. Mobilitas ini menawarkan kenyamanan dan fleksibilitas yang lebih baik dibandingkan dengan sistem keamanan tradisional atau ip camera. Pengujian juga menunjukkan bahwa browser bawaan smartphone atau aplikasi media player yang tersedia dapat digunakan tanpa kendala yang berarti, menunjukkan interoperabilitas dan ketersediaan lintas platform yang penting dalam sistem keamanan modern.

d. Kewenangan akses

Agar kamera hanya dapat diakses oleh orang yang berwenang, maka aplikasi memiliki fitur untuk melakukan pengaturan username dan password. Username dan password akan diminta ketika seseorang ingin mengakses kamera cctv melalui browser atau video player. Dengan demikian, sistem keamanan yang dibangun hanya dapat diakses oleh seseorang yang memiliki kewenangan.

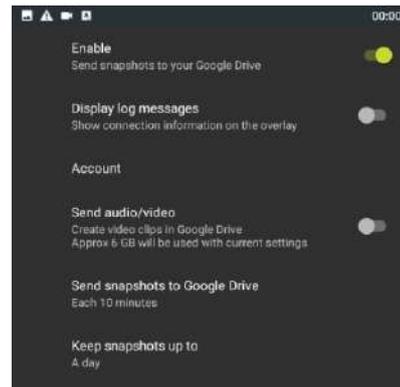


Gambar 3. Akses Kamera Melalui Browser

3.2. Penyimpanan Berbasis Cloud untuk Data Keamanan

CamOn live streaming mendukung beberapa penyedia layanan penyimpanan cloud, mulai Angelcam, Mangocam dan Google drive. Dari 3 lokasi cloud storage yang didukung, pilihan terbaik adalah menggunakan Google drive, karena dengan akun gratis yang kita buat, kita sudah mendapatkan kapasitas penyimpanan sebesar 15 GB. Sementara 2 lokasi cloud storage lainnya bersifat gratis hanya untuk beberapa hari, setelah habis masa ujicoba (trial), maka kita diharuskan untuk membayar jika ingin menggunakan layanan cloud tersebut.

Untuk menggunakan Google drive sebagai lokasi penyimpanan, di bagian pengaturan aplikasi (Setting), pilih Google drive. Selanjutnya, pada opsi Enable diaktifkan. Setelah itu klik bagian Account untuk menambahkan account Google yang akan digunakan seperti terlihat pada gambar 4.



Gambar 4. Pengaturan Google drive sebagai cloud storage

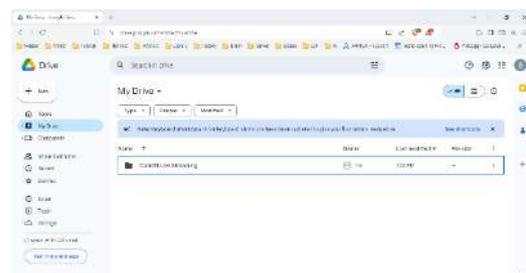
Analisis hasil penerapan penyimpanan cloud menunjukkan beberapa keuntungan signifikan:

a. Keamanan dan pemulihan data

Sistem penyimpanan cloud menawarkan keamanan yang lebih baik dibandingkan penyimpanan lokal. Pada beberapa skenario uji, perangkat yang berfungsi sebagai sensor (smartphone) dapat dimatikan atau dicabut tanpa mempengaruhi integritas data karena semua data penting sudah tersimpan di cloud. Hal ini memberikan keuntungan besar dalam perlindungan data, di mana rekaman penting tetap aman meskipun terjadi gangguan fisik pada perangkat. Selain itu, kemampuan cloud dalam mendukung pemulihan data yang cepat juga terbukti penting. Data dapat dipulihkan dengan mudah ke perangkat baru setelah terjadi kegagalan pada perangkat lama.

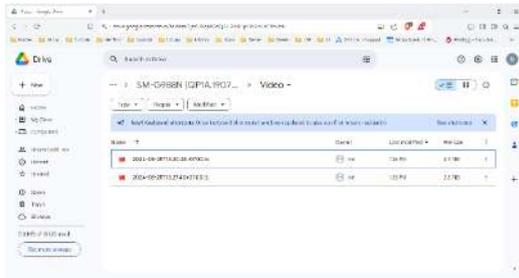
b. Kapabilitas penyimpanan yang fleksibel dan terukur

Pengujian menunjukkan bahwa aplikasi ini mampu menyimpan rekaman video dari beberapa perangkat smartphone secara simultan selama periode yang lama seperti terlihat pada gambar 5.



Gambar 5. Folder rekaman CamOn Live Streaming berdasarkan merek dan spesifikasi smartphone

Selain itu, fleksibilitas aplikasi memungkinkan pengguna untuk mengatur agar proses penyimpanan dapat dihapus secara otomatis dalam rentang waktu yang ditentukan. Dengan demikian, pengguna dapat menggunakan akun google drive versi standar yang tanpa biaya, sehingga tidak menambah cost terhadap sistem keamanan yang dibangun.



Gambar 6. Hasil tangkapan video disimpan pada cloud

Berdasarkan pengujian yang dilakukan, seperti terlihat pada gambar 6, dengan pengaturan jeda waktu 5 menit untuk setiap pengiriman rekaman, diperoleh ukuran file video berkisar antara 2,8-3 MB, sehingga dalam waktu 1 jam jumlah video yang dikirim sebanyak 12 video dengan total ukuran 36 MB. Dalam 1 hari, maka kapasitas penyimpanan cloud yang digunakan sebesar 864 MB untuk satu smartphone.

Tabel 1. Kapasitas penggunaan cloud storage

Unit	Storage (MB)		Durasi Simpan
	per hari	per minggu	
1	864	6.048	1 minggu
2	1.728	12.096	1 minggu
3	2.592	18.144	2 hari

Berdasarkan data pada tabel 1, jika menggunakan dua smartphone sebagai kamera keamanan, data rekaman pada akun Google drive gratis dengan kapasitas 15 GB dapat diatur untuk dihapus dalam waktu 1 minggu, sesuai dengan waktu maksimal pengaturan yang ada pada aplikasi. Sementara ketika menggunakan smartphone antara 3-7 unit, maka durasi waktu simpan rekaman sebelum dihapus otomatis dapat diatur pada waktu per dua hari.

Perhitungan ini dilakukan ketika smartphone menggunakan akun Google drive yang sama. Sementara, masing-masing smartphone bisa menggunakan akun Google drive yang berbeda, sehingga terlihat bahwa pengaturan *cloud storage* CamOn *live streaming* tidak menimbulkan biaya (*cost*) tambahan dalam penggunaannya.

c. Akses data global dan *real-time*

Hasil implementasi menunjukkan bahwa dengan penyimpanan *cloud*, pengguna dapat mengakses data keamanan kapan saja dan dari mana saja, selama terhubung dengan internet. Ini memberikan keuntungan dalam hal aksesibilitas global, yang tidak dapat dicapai oleh sistem keamanan berbasis penyimpanan lokal.

Berdasarkan hasil implementasi dan pengujian, smartphone sebagai pusat kendali keamanan terbukti efektif dalam mendukung pemantauan dan pengelolaan keamanan secara aktif. Penggunaan sensor bawaan pada smartphone memungkinkan deteksi ancaman yang cepat dan akurat, sementara integrasi dengan *cloud storage* memberikan

penyimpanan yang aman, fleksibel, dan mudah diakses secara global.

Penggunaan Google drive sebagai *cloud storage* juga memperlihatkan keuntungan dalam hal efisiensi ruang penyimpanan dan pemulihan data, serta ketersediaan sistem yang lebih baik dan semuanya tanpa biaya. Kombinasi antara smartphone dan *cloud storage* ini menciptakan sistem keamanan yang lebih responsif, terdistribusi, dan fleksibel dibandingkan dengan sistem keamanan konvensional yang berbasis perangkat keras lokal atau ip camera yang membatasi penyimpanan cloud yang berbayar.

Dengan demikian, sistem ini dapat menjadi solusi yang menjanjikan untuk kebutuhan keamanan modern, di mana mobilitas, *real-time monitoring*, dan keamanan data menjadi prioritas utama.

4. KESIMPULAN

Berdasarkan implementasi dan pengujian yang dilakukan, penggunaan smartphone sebagai cctv berbasis *cloud storage* tidak hanya mengatasi masalah kerentanan penyimpanan lokal tetapi juga menawarkan solusi yang lebih praktis dan ekonomis untuk menjaga keamanan rumah dan tempat usaha. Selain keamanan yang lebih tinggi, penggunaan smartphone sebagai perangkat cctv juga menawarkan fleksibilitas dan efisiensi biaya, karena orang dapat dengan mudah merubah smartphonenya menjadi perangkat pengawasan tanpa harus membeli perangkat cctv khusus. Dengan bantuan aplikasi yang tepat, smartphone dapat digunakan untuk merekam video secara *real-time* dan mengunggahnya langsung ke cloud. Ini tidak hanya mengurangi kebutuhan akan perangkat CCTV yang mahal, tetapi juga memungkinkan pengguna untuk mengawasi properti mereka dari jarak jauh melalui akses internet. Dengan teknologi ini, pemilik properti dapat merasa lebih tenang, mengetahui bahwa rekaman keamanan mereka aman dan dapat diakses kapan saja dan dari mana saja, memberikan perlindungan yang lebih efektif terhadap potensi ancaman. Dengan demikian, penelitian ini akan dapat mendorong inovasi lebih lanjut dalam bidang keamanan digital dan memberikan manfaat yang nyata bagi masyarakat dan dunia usaha dalam menjaga keamanan properti mereka.

DAFTAR PUSTAKA

- [1] F. R. Doni, "Akses Kamera CCTV dari Jarak Jauh Untuk Monitoring Keamanan dengan Penerapan PSS". *Evolusi: Jurnal Sains dan Manajemen*, vol. 8, no. 1, p. 1-9, 2020, <https://doi.org/10.31294/evolusi.v8i1.7142>.
- [2] H. Mubaraq, "Rekaman CCTV Sebagai Alat Bukti Tindak Pidana". *Sekolah Tinggi Hukum Militer AHM-PTHM*. 2021. [Online]. Available: <https://sthmahmptm.ac.id/detailpost/rekaman-cctv-sebagai-alat-bukti-tindak-pidana>. [Accessed: 25 Agustus 2024].
- [3] S. Indriyanto and B. Rahardjo, "Taksonomi Tinjauan Keamanan Pada Jaringan IP Camera", *SENTER* 2018,

- pp. 164–172, Jan. 2019.
- [4] A. Pratama and S. Wahyuni, “Implementasi Sistem Keamanan Berbasis Cloud pada Kamera CCTV untuk Monitoring Rumah”. *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 3, p. 200-210, 2021.
- [5] A. Rahman and D. Kusumaningtyas, “Sistem Keamanan Berbasis Internet of Things (IoT) Menggunakan Smartphone dan Cloud Storage”. *Jurnal Rekayasa Sistem & Teknologi Informasi*, vol. 4, no. 2, p. 75-85, 2020.
- [6] Suhendar, E. “Tinjauan Sistematis : Implementasi Cloud Computing Terhadap Keamanan Layanan Publik”, *Smart Comp : Jurnalnya Orang Pintar Komputer*, 11(4), 599-606, April 2021, <https://doi.org/10.30591/smartcomp.v11i4.4245>.
- [7] Pindarwati, A., et.Al. “Implementasi Penggunaan Cctv Berbasis Internet Of Things (Iot) Sebagai Smart Security Untuk Menanggulangi Angka Kejahatan Studi Kasus: SMK Insan Cita”. *Jurnal Multidisiplin Indonesia*, 1(2), 431-439, Oktober 2022, <https://doi.org/10.58344/jmi.v1i2.45>.
- [8] A. Kusumastuti and A. M. Khoiro, “Metode Penelitian Kualitatif”. Semarang: Lembaga Pendidikan Sukarno Pressindo, 2019.
- [9] M. Sari, H. Rachman, N.J. Astuti, M.W. Afgani & R.A. Siroj, “Explanatory survey dalam metode penelitian deskriptif kuantitatif”. *Jurnal Pendidikan Sains Dan Komputer*, vol. 3, no. 01, p. 10-16, 2023, <https://doi.org/10.47709/jpsk.v3i01.1953>.
- [10] Rusandi and M. Rusli, “Merancang Penelitian Kualitatif Dasar/Deskriptif dan Studi Kasus. Al-Ubudiyah”, *Jurnal Pendidikan dan Studi Islam*, vol. 2, no. 1, 2021, <https://doi.org/10.55623/au.v2i1.18>.
- [11] Gufron and P. S. Syahriarti, “Perancangan Private Cloud Storage dan Cloud Office Menggunakan OnlyOffice dan Nextcloud Pada Pusat Pelatihan Djamboe Training Center”. *Jurnal Intelek Insan Cendikia*, vol. 1, no. 6, p. 2120-2128, 2024.

PENERAPAN E-COMMERCE BERBASIS CMS DENGAN METODE BUSINESS MODEL CANVAS PADA NORU LABEL

Bianca Vania Adeline^{1*}, Agnes Aryasanti², Anita Diana³, Mohammad Anif⁴

^{1,2,3}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur

⁴Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur

Email: ¹2012500902@student.budiluhur.ac.id, ²agnesaryasanti@budiluhur.ac.id, ³anitadiana@budiluhur.ac.id,
⁴muhammad.anif@budiluhur.ac.id

(Naskah masuk: 10 September 2024, diterima untuk diterbitkan: 30 September 2024)

Abstrak

Pemasaran produk dalam bisnis saat ini dapat dilakukan secara mudah dengan adanya teknologi informasi yang berkembang sangat pesat. Banyak toko yang melakukan kegiatan jual beli secara *online*. Noru Label adalah toko tas yang menjual totebag dan tas laptop yang mengandalkan media sosial dan *marketplace* sebagai media penjualan. Beberapa masalah yang dialami oleh toko Noru Label yaitu lamanya proses transaksi karena tidak adanya informasi produk seperti stok, harga dan deskripsi produk pada sosial media, calon pelanggan pun ada yang membatalkan pesanan karena admin tidak cepat merespon pesan. Masalah lainnya yaitu tingginya tingkat pembatalan pada penjualan di *marketplace* sebesar 34% karena pelanggan salah membaca diskripsi produk, sedangkan produk sudah dikemas dan siap untuk dikirim sehingga merugikan toko Noru Label. Selain itu, adanya permintaan pengembalian barang karena barang yang diterima pelanggan rusak saat proses pengiriman dan kesalahan pelanggan dalam melakukan pesanan. Penelitian ini membuat website *E-commerce* menggunakan *Content Management System* (CMS) pada toko Noru Label untuk mempermudah pemilik toko dan pelanggan dalam proses transaksi dan tidak merugikan toko. *Search Engine Optimization* (SEO) diimplementasikan pada website *E-Commerce* untuk meningkatkan promosi agar mencakup segmen pasar yang lebih luas. Hasil dari penelitian ini adalah dengan adanya website *E-commerce* pelanggan dapat menjelajahi produk dengan informasi yang lebih komprehensif dan deskripsi yang lebih jelas, proses transaksi lebih cepat dan tidak adanya retur produk. Penggunaan SEO sebagai tambahan strategi promosi toko Noru Label dapat melakukan pemasaran dengan jangkauan pasar yang lebih luas. Hasilnya untuk penggunaan SEO, pada tanggal 17 juli 2024 di mesin pencari google dengan kata kunci produk toko Noru Label yaitu luna gingham sleeves , Maka toko Noru label berada pada urutan ke enam di halaman pertama. Hal ini dapat meningkatkan visibilitas toko dan menarik lebih banyak pelanggan.

Kata kunci: *E-Commerce*, CMS, BMC.

IMPLEMENTATION OF CMS-BASED E-COMMERCE USING THE BUSINESS MODEL CANVAS METHOD AT NORU LABEL

Abstract

Currently, product marketing in business can be easily conducted thanks to the rapid advancement of information technology. Many stores engage in online buying and selling activities, commonly referred to as *E-Commerce*. Noru Label is a bag store that sells tote bags and laptop bags, primarily relying on social media and marketplaces as sales platforms. However, in conducting business through social media and marketplaces, Noru Label faces several issues. Customers often struggle to obtain product information such as stock availability, prices, and product descriptions on social media, requiring them to inquire first and wait for the admin's response, which can sometimes take a long time. Another issue is the high cancellation rate on Shopee, which stands at 34%, causing losses for Noru Label as products are already packaged and ready for shipment. Additionally, there are instances of product returns due to customers ordering the wrong items or receiving damaged goods during shipping, despite Noru Label always ensuring that products are thoroughly checked and free of defects before dispatch. Moreover, when searched on Google, Noru Label's products have yet to appear on the first page of results. The aim of this study is to design an *E-Commerce* website using a *Content Management System* (CMS) to facilitate customers in ordering and transaction processes, implement *E-Commerce* at Noru Label to boost sales and marketing without causing losses to the seller, and utilize *Search Engine Optimization* (SEO) to enhance promotion and reach a wider market segment. The results of this study indicate that the *E-Commerce* website allows customers to explore

products with more comprehensive information and clearer descriptions. Additionally, with SEO as a supplementary promotional strategy, Noru Label can market its products to a broader audience, and the high cancellation rate can be reduced due to the different features available on the E-Commerce website.

Keywords: E-Commerce, CMS, BMC.

1. PENDAHULUAN

Seiring berkembangnya teknologi di zaman modern ini, banyak inovasi baru yang memudahkan para pelaku usaha dalam memasarkan produknya. Hal tersebut tentu menambah efisiensi dan efektivitas suatu toko dalam proses penjualan, sehingga banyak pelaku usaha yang memanfaatkan teknologi tersebut dalam memasarkan produknya. Dengan adanya internet dan penjualan secara *online*, toko berpotensi menjual produk ke seluruh Indonesia bahkan luar negeri karena calon pelanggan mendapatkan kemudahan dalam mengakses produk yang diinginkan kapanpun dan dimanapun. Penggunaan teknologi informasi untuk mendukung bisnis sangat beragam, salah satunya adalah perdagangan elektronik, atau yang biasa dikenal dengan *e-commerce* [1]. *E-commerce* adalah penggunaan teknologi digital berbasis internet yang diterapkan dalam proses bisnis, melibatkan transaksi jual beli antara konsumen dan produsen [2]. Dengan *e-commerce*, para pelaku usaha dapat mengelola produk mereka dengan lebih optimal. Banyak pelaku usaha yang mulai mengintegrasikan *e-commerce* sebagai upaya untuk memanfaatkan teknologi guna meningkatkan kecepatan dan kelincihan dalam proses operasional bisnis serta meningkatkan kepuasan pelanggan [3].

Noru label adalah sebuah toko tas yang berdiri sejak 2022, Noru label menjual tas seperti *totebag* dan tas laptop dengan desain yang trendi. Awalnya, Noru label menjual dan memasarkan produknya melalui sosial media lalu merambah ke *marketplace*. Dengan hanya memasarkan produk melalui sosial media tentu banyak calon pembeli kesulitan untuk melihat ketersediaan produk, karena hanya menampilkan gambar dari produk tanpa menampilkan deskripsi produk, harga, dan ketersediaan stok produk. Dengan begitu, proses pemesanan pun menjadi lebih memakan banyak waktu karena calon pembeli harus menghubungi penjual untuk bertanya tentang produk melalui pesan *chat* dan harus menunggu respon penjual untuk menjawab. Dengan metode seperti itu tentu akan berdampak calon pelanggan bisa saja mencari toko lain dengan produk yang serupa. Sedangkan penjualan melalui *marketplace* ditemukan masalah yaitu tingginya tingkat pembatalan pemesanan oleh pelanggan sebesar 34% hal ini merugikan penjual karena produk sudah diproses untuk dikemas bahkan beberapa sudah siap untuk dikirim. Masalah lainnya adanya permintaan pengembalian barang karena pelanggan salah memesan ukuran dan barang rusak saat proses pengiriman, sedangkan sebelum dilakukan

pengiriman toko Noru Label memastikan kondisi barang baik dan tidak ada cacat.

Pada penelitian sebelumnya dilakukan oleh [4], yaitu membantu pemasaran produk agar mencapai potensi yang maksimal menggunakan metode *Business Model Canvas* untuk merancang model bisnis dan *Search Engine Optimization* untuk mempromosikan *website e-commerce*. Hasil dari penelitian toko memperoleh keuntungan dalam menjual produk dan dapat memfasilitasi promosi produk dengan pasar yang lebih luas. Penelitian [5] mengembangkan sistem *e-commerce* sebagai alternatif penjualan bagi toko. Hasil dari penelitian ini adalah *e-commerce* dapat memudahkan pelanggan dan penjual dalam melakukan transaksi. Penelitian selanjutnya [6] membuat *website e-commerce* berbasis *Business to Business (B2B)* menggunakan metode *fishbone diagram*, *Business Model Canvas (BMC)*, *Unified Modeling Language* dan *Interaction Flow Modeling*. Hasilnya, *website E-commerce* terbukti menghemat biaya SDM dan dapat memluas jangkauan pasar. Santosa [7] melakukan penelitian membuat *website e-commerce* agar memudahkan proses penjualan dan meningkatkan promosi. Hasilnya, *website e-commerce* efektif berperan sebagai strategi pemasaran bagi toko dan mengurangi biaya operasional. Penelitian [8] menjelaskan bahwa penerapan *website e-commerce* menggunakan metode *fishbone diagram*, wawancara dan observasi menghasilkan *website e-commerce* yang dapat menjadi solusi bagi pemasaran produk yang lebih luas dan efisien. Penelitian selanjutnya oleh Fuad [9] yaitu meningkatkan penjualan dan dapat mencetak laporan yang akurat, hasilnya *website e-commerce* dapat memasarkan produk lebih luas dan dapat mencetak laporan dengan efisien dan akurat. Penelitian selanjutnya [10] membuat *website e-commerce* dengan metode *Business Model Canvas* dan *Content Management System (CMS)*. Hasil dari penelitian ini menunjukkan bahwa *website e-commerce* berhasil membantu toko meningkatkan penjualan dan memperluas target pasar.

Berdasarkan penelitian sebelumnya, terdapat beberapa masalah yang berbeda dengan penelitian ini yaitu, pelanggan mengalami kesulitan memperoleh informasi di media sosial karena deskripsi produk yang kurang jelas, promosi yang hanya dilakukan melalui media sosial sehingga jangkauan produk terbatas, serta tingginya tingkat pembatalan di *marketplace* yang sering merugikan penjual karena produk sudah siap untuk diproses.

Tujuan dari penelitian ini adalah merancang sebuah *website e-commerce* yang memudahkan

pelanggan dalam bertransaksi, menerapkan *e-commerce* pada toko Noru Label guna meningkatkan penjualan dan pemasaran, serta memanfaatkan *Search Engine Optimization* (SEO) untuk memperluas cakupan promosi ke segmen pasar yang lebih luas.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian dapat dilihat pada Gambar 1. Tahap awal penelitian dimulai dengan identifikasi masalah melalui serangkaian wawancara untuk memahami berbagai masalah yang dihadapi oleh toko Noru Label. Setelah masalah telah teridentifikasi, dilakukan tinjauan untuk menentukan fitur-fitur apa saja yang diperlukan dan akan dikembangkan. Setelah itu, dilakukan studi literatur yang digunakan sebagai referensi penting dalam penelitian ini. Pada tahap analisa masalah, digunakan *fishbone diagram* untuk mengidentifikasi akar dari penyebab masalah. Setelah itu dilakukan *Business Model Canvas* (BMC) untuk memvisualisasikan konsep bisnis dan mengidentifikasi sistem yang diperlukan. Setelah itu dilakukan Implementasi CMS, Tahap ini menggambarkan rancangan dalam membuat *e-commerce* dengan menggunakan diagram-diagram *Unified Modeling Language* (UML), seperti *Activity Diagram* dan *Use Case Diagram* untuk menggambarkan kebutuhan Toko Noru Label serta interaksi antara pelanggan dan admin dengan fitur-fitur dalam sistem. Kemudian, dilakukan perancangan layar atau UI/UX menggunakan Figma, sebagai langkah persiapan sebelum diimplementasikan sebagai situs *website* berbasis *e-commerce*. Pada tahap Implementasi, sistem diimplementasikan menggunakan *WordPress* sebagai *Content Management System* (CMS) dengan menggunakan *plugin WooCommerce* untuk membuat *e-commerce*. Selain itu, *website* akan dioptimalkan dengan teknik *Search Engine Optimization* (SEO) menggunakan *plugin Yoast SEO* guna meningkatkan jumlah pengunjung dan *website* toko muncul di halaman utama mesin pencari dan *plugin advance order export* untuk mencetak laporan. Setelah dilakukan implementasi *E-Commerce* langkah selanjutnya adalah melakukan strategi marketing dengan promosi di beberapa *platform* sosial media seperti Instagram dan TikTok sebagai media pemasaran. Dengan memanfaatkan fitur-fitur yang ada di Instagram dan TikTok untuk menarik perhatian dan mendorong interaksi dengan calon pembeli, diharapkan dapat meningkatkan lalu lintas ke situs *E-Commerce* dan pada akhirnya akan meningkatkan penjualan.

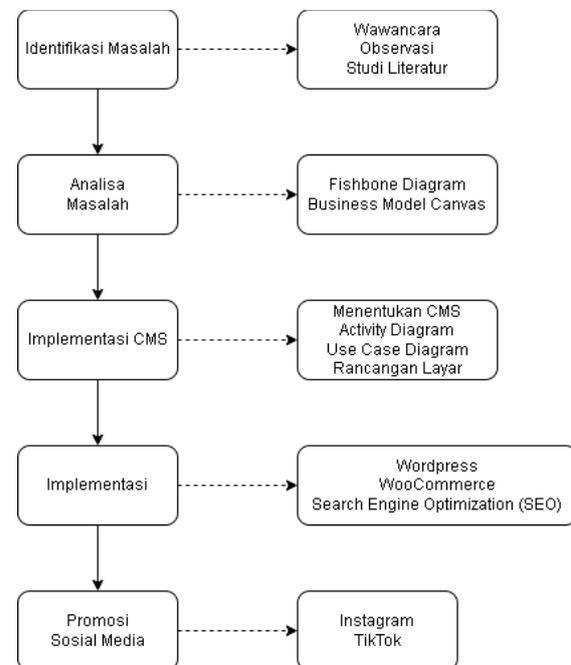
3. HASIL DAN PEMBAHASAN

3.1 Identifikasi Masalah

Identifikasi masalah dilakukan untuk mendapatkan informasi yang dapat dianalisis guna

mencapai tujuan penelitian. Berikut adalah langkah-langkah identifikasi masalah yang dilakukan:

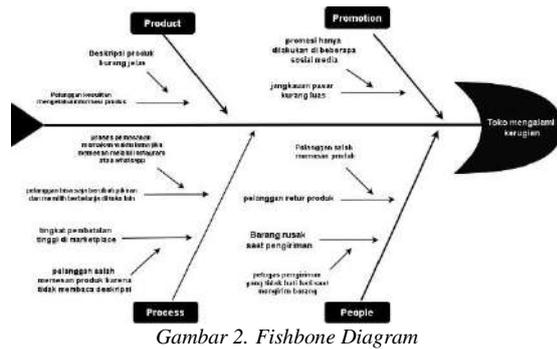
- Wawancara dilakukan agar mendapatkan informasi melalui tanya jawab dengan pemilik toko Noru Label. Pada tahap ini, Penulis mengajukan beberapa pertanyaan kepada pemilik toko Noru Label yang berkaitan dengan penelitian ini. Hasil yang diperoleh dari wawancara ini adalah informasi detail mengenai masalah yang ada di toko Noru Label, seperti tantangan mengenai masalah yang ada, tantangan dalam operasional sehari-hari dan hambatan yang dihadapi dalam sistem yang berjalan
- Observasi dilakukan untuk mendapatkan pemahaman terhadap obyek yang diteliti. Pada tahap ini, dilakukan pengamatan sistem yang berjalan pada toko Noru Label. Hasil dari observasi ini adalah pemahaman yang lebih mendalam tentang bagaimana sistem saat ini beroperasi, identifikasi sumber-sumber masalah, serta fitur-fitur yang perlu dioptimalkan.
- Studi Literatur dilakukan untuk mendapatkan pemahaman dari penelitian sebelumnya. Tahap ini bertujuan untuk membangun landasan teori yang kuat untuk mendukung penelitian yang sedang dilakukan. Hasil dari studi literatur yang dilakukan adalah diperolehnya referensi dan pemahaman yang mendalam dari penelitian-penelitian sebelumnya, yang menjadi dasar teoritis yang kuat serta memberikan arah yang jelas bagi penelitian ini.



Gambar 1. Database Mirroring Architecture

3.2 Fishbone Diagram

Fishbone diagram digunakan untuk mengidentifikasi permasalahan yang dihadapi dalam penelitian ini. *Fishbone* diagram dapat dilihat pada Gambar 2.



Gambar 2. Fishbone Diagram

produk yang menarik dan realistis, produk dengan banyak variasi dan multifungsi, serta kemudahan pemesanan kapanpun dan di manapun.



Gambar 3. Business Model Canvas

Kategori yang digunakan meliputi *product*, *process*, *promotion* dan *people*, yang membantu menjelaskan penyebab dan akibat dari masalah yang dihadapi oleh toko Noru Label. Dalam kategori *product*, pemasaran melalui media sosial sering kali membuat calon pembeli kesulitan untuk mendapatkan informasi rinci tentang ketersediaan produk. Ini karena media sosial umumnya hanya menampilkan gambar produk tanpa disertai dengan deskripsi lengkap, harga, dan ketersediaan stok, sehingga proses pemesanan menjadi lebih lambat. Calon pelanggan harus menghubungi penjual melalui pesan untuk menanyakan detail produk dan menunggu respon, yang dapat menyebabkan calon pelanggan mencari toko lain dengan proses yang lebih efisien. Pada kategori *process*, ditemukan tingginya tingkat pembatalan pesanan oleh pelanggan yang mencapai 34%, yang merugikan penjual. Hal ini disebabkan oleh kesalahan pelanggan dalam memesan produk karena tidak membaca deskripsi dengan teliti, sehingga tingkat pembatalan tinggi di *marketplace*, dan proses pemesanan yang memakan waktu melalui media sosial dapat membuat pelanggan berubah pikiran dan berbelanja di tempat lain. Pada kategori *promotion*, masalah muncul karena promosi hanya dilakukan di beberapa media sosial, sehingga jangkauan pasar menjadi terbatas. Sedangkan dalam kategori *people*, sering terjadi permintaan pengembalian barang karena pelanggan salah memesan ukuran atau produk rusak saat pengiriman, meskipun toko Noru Label selalu memastikan kondisi barang dalam keadaan baik sebelum dikirim, yang menyebabkan pelanggan melakukan retur produk.

3.3 Business Model Canvas

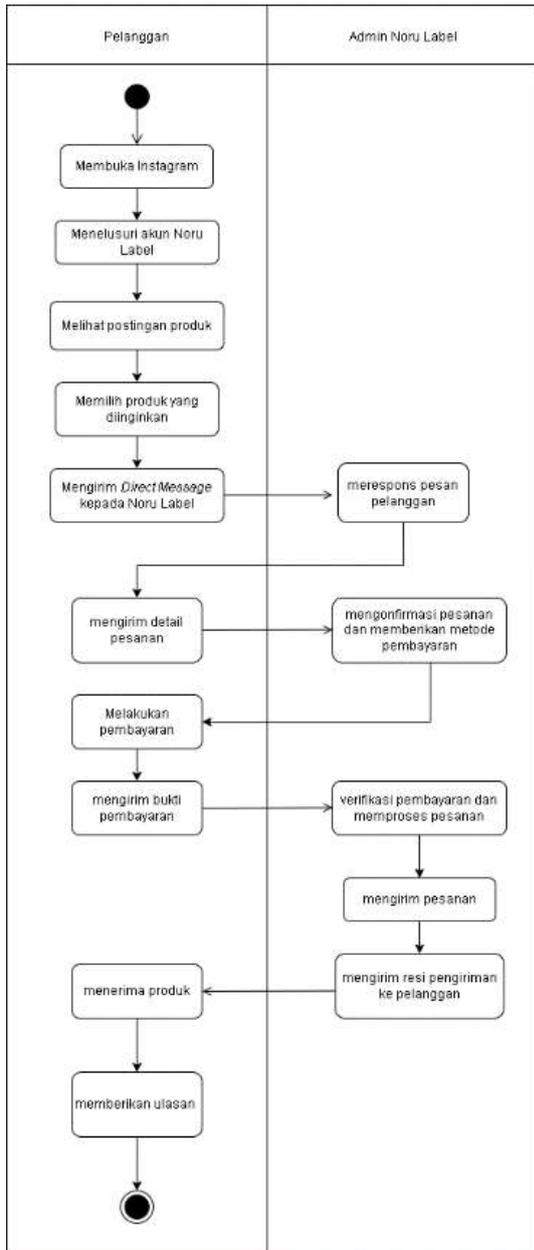
Business Model Canvas yang menggambarkan ide bisnis pada toko Noru Label yang dibagi menjadi sembilan bagian. *Business Model Canvas* pada toko Noru Label ditampilkan pada gambar 3.

- Customer Segments* dari toko Noru Label adalah wanita dan laki-laki berusia 16-45 tahun, pengguna media sosial dan *marketplace*, serta para pengguna laptop.
- Value Propositions* yang ditawarkan oleh toko Noru Label meliputi desain produk yang trendy dan kekinian, kualitas bahan premium, foto

- Channels* yang digunakan toko Noru Label dalam menyampaikan produknya adalah melalui *website E-Commerce*, Instagram, dan Shopee.
- Customer Relationship* Toko Noru Label dengan pelanggan dengan memberikan diskon harga dengan minimal pembelian dan memberikan *voucher* diskon untuk pengguna baru.
- Revenue Stream* yang didapat toko Noru Label berasal dari penjualan produk.
- Key Resources* dalam proses bisnis Toko Noru Label adalah produk toko, WooCommerce, Instagram, dan Shopee.
- Key Activities* yang dilakukan oleh toko Noru Label meliputi penjualan produk, promosi produk, dan mengelola pesanan.
- Key Partners* Toko Noru Label adalah beberapa mitra, seperti produsen bahan baku, penyedia layanan (WordPress & WooCommerce), dan ekspedisi pengiriman.
- Cost Structure* yang dikeluarkan toko Noru Label untuk menunjang bisnis meliputi biaya produksi produk, biaya *hosting* dan *maintenance website*, biaya iklan sosial media, serta biaya pengiriman dan pengemasan.

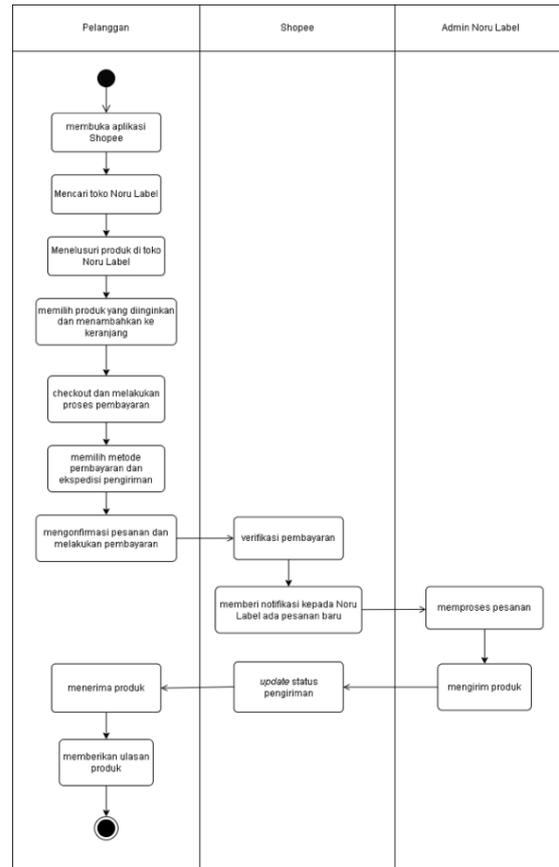
3.4 Activity Diagram

Proses di toko Noru Label di media sosial: pelanggan menelusuri akun Noru Label, memilih produk, dan mengirim pesan melalui *Direct Message* untuk *detail* dan stok. Admin merespons dengan informasi yang diminta. Pelanggan mengirim informasi produk yang dipilih, alamat, dan metode pembayaran. Admin mengonfirmasi total harga dan info pembayaran. Pelanggan membayar dan mengirim bukti pembayaran. Setelah verifikasi, pesanan diproses dan dikirim. Admin mengirim nomor resi. Pelanggan menerima produk dan dapat memberikan ulasan di *Instastory*. *Activity diagram* proses pemesanan melalui media sosial ditampilkan pada Gambar 4.



Gambar 4. ActivityDiagram order process on social media

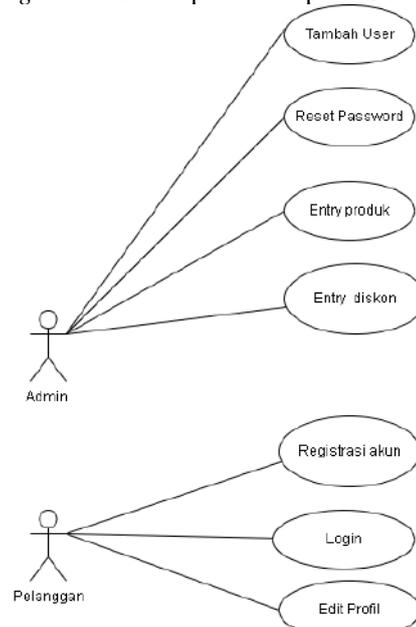
Proses sistem berjalan pada toko Noru Label di Shopee adalah pelanggan membuka aplikasi marketplace dan mencari toko Noru Label, setelah itu pelanggan menelusuri produk di toko Noru Label. Setelah mendapatkan produk yang diinginkan, pengguna menambahkan produk ke keranjang belanja dan checkout produk ke proses pembayaran. Lalu pelanggan memilih metode pembayaran dan memilih ekspedisi pengiriman yang diinginkan. Setelah pelanggan mengonfirmasi pesanan dan menyelesaikan pembayaran, pesanan akan segera diproses dan dikirim ke alamat pelanggan. Lalu pelanggan akan menerima produk dan dapat memberikan ulasan produk. Activity Diagram proses pemesanan melalui Shopee ditampilkan pada Gambar 5.



Gambar 5. Activity Diagram Order Process on Shopee

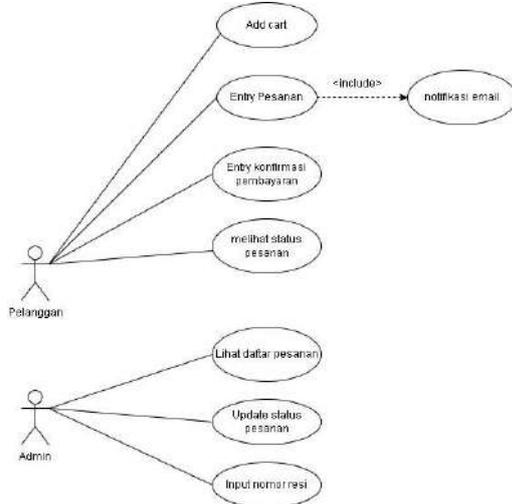
3.5 Use Case Diagram

Use Case Diagram pada website e-commerce toko Noru Label meliputi Use Case Diagram master, Use Case Diagram Transaksi dan Use Case Diagram laporan. Use case diagram master terdiri dari sisi admin mencakup tambah user, reset password, entry produk dan diskon. Lalu di bagian pelanggan mencakup registrasi akun, login dan edit profil. Use Case diagram Master dapat dilihat pada Gambar 6.



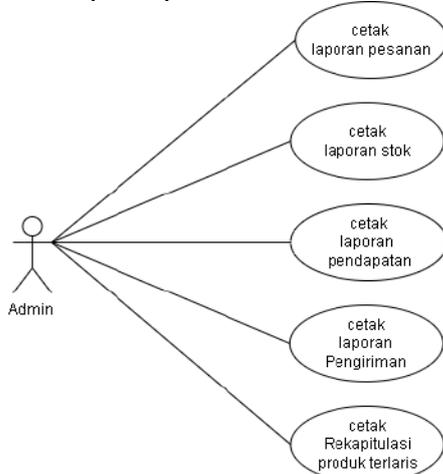
Gambar 6. Use Case Diagram Master

Use case diagram transaksi bagian admin mencakup melihat daftar pesanan, *update* status pesanan dan *input* nomor resi. Lalu di bagian pelanggan mencakup *add cart*, *entry* pesanan, *entry* konfirmasi pembayaran dan melihat status pesanan. *Use Case Diagram* transaksi dapat dilihat pada Gambar 7.



Gambar 7. Use Case Diagram Transaksi

Use case diagram laporan mencakup cetak laporan pesanan, cetak laporan stok, cetak laporan pendapatan, cetak laporan pengiriman, dan cetak laporan rekapitulasi produk terlaris. *Use case diagram* laporan ditampilkan pada Gambar 8.

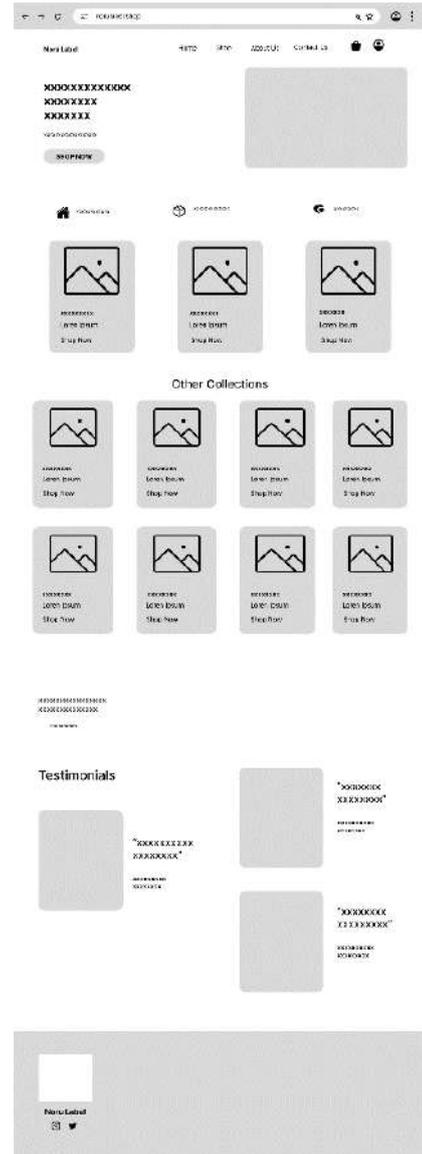


Gambar 8. Use Case Diagram Laporan

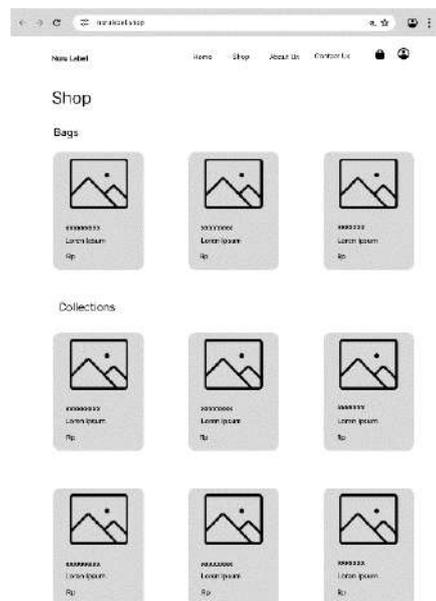
3.6 Rancangan Layar

Rancangan layar *website e-commerce* toko Noru Label. Rancangan layar bagian home ditampilkan pada gambar 9.

Rancangan layar pada halaman *home* yang berfungsi sebagai navigasi utama pelanggan saat berkunjung ke *website e-commerce* toko Noru Label. Rancangan layar ini berisi *navigation bar*, logo toko noru label, dan beberapa produk. Rancangan layar bagian *shop* ditampilkan pada gambar 10.



Gambar 9. Rancangan Layar Bagian Home



Gambar 10. Rancangan layar bagian shop

User Interface bagian Shop website E-Commerce toko Noru Label yang menampilkan harga dan produk yang di jual di toko Noru Label. Halaman ini berfungsi agar pengguna bisa menjelajahi berbagai produk yang dijual di toko Noru Label.

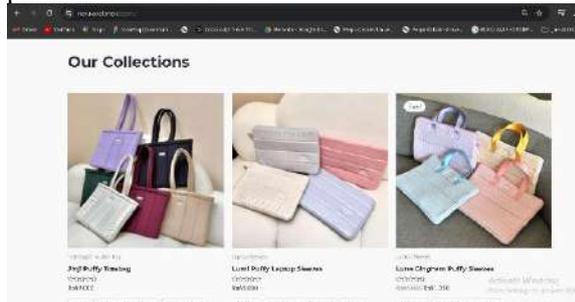
3.7 Implementasi

Setelah dibuat rancangan layar, website e-commerce diimplementasikan menggunakan Content Management System (CMS) Wordpress. Tampilan layar website e-commerce ditampilkan pada gambar 11.



Gambar 11. Tampilan Website E-Commerce Toko Noru Label

Pada gambar 12 ditampilkan tampilan website e-commerce bagian shop yang menampilkan produk-produk toko Noru Label.



Gambar 12. Tampilan Website E-Commerce Bagian Shop

3.8 Search Engine Optimization

SEO diimplementasikan sebagai salah satu strategi untuk memperluas cakupan promosi dan dapat meningkatkan visibilitas website noru label, sehingga mampu menjangkau berbagai segmen audiens dengan lebih efektif serta memastikan website ditemukan oleh target audiens yang tepat dan relevan di mesin pencari.

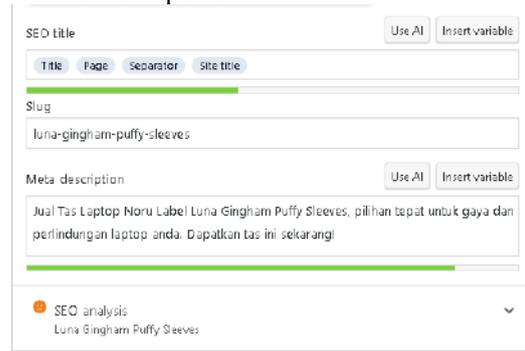
Beberapa langkah dan strategi SEO yang diaplikasikan pada website e-commerce toko Noru Label antara lain mengoptimasi kata kunci yang relevan dengan produk yang dijual. Kata kunci diidentifikasi menggunakan alat riset seperti Google Keyword Planner dan plugin YoastSEO untuk menemukan kata kunci SEO yang tepat. Kata kunci yang digunakan meliputi nama produk yang dijual serta kata kunci relevan lainnya, seperti "tas laptop Tangerang", "tas laptop wanita", dan "tas laptop murah Tangerang". Selain itu, optimasi gambar yang ditampilkan pada gambar 13 dilakukan dengan menggunakan nama file yang mengandung kata kunci

untuk gambar yang diunggah ke website e-commerce toko Noru Label.

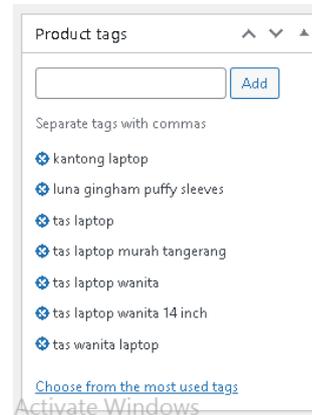


Gambar 13. Optimasi Gambar

Gambar 14 dan 15 menampilkan kata kunci dan tags yang digunakan pada produk agar produk dapat masuk ke mesin pencari.

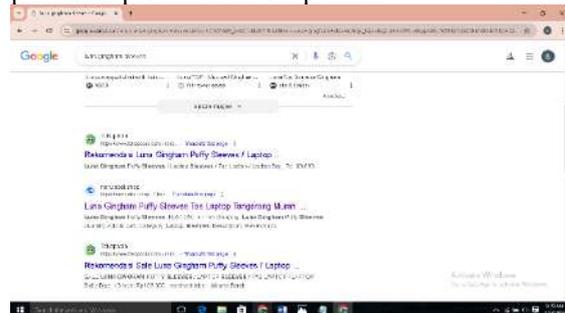


Gambar 14. SEO



Gambar 15. SEO Product Tags

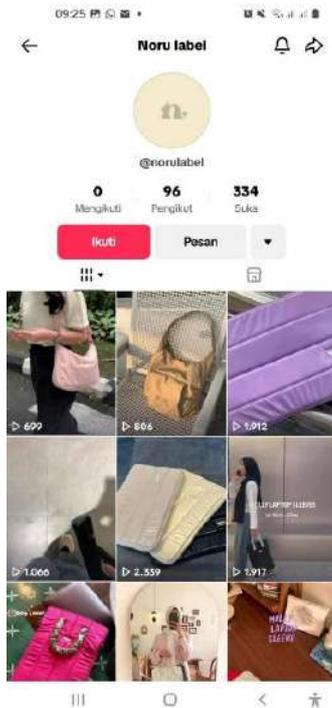
Kata kunci yang digunakan di mesin pencari adalah salah satu produk dari toko Noru Label yaitu "luna gingham sleeves" di akses pada tanggal 17 juli 2024 berada di urutan ke enam di halaman pertama mesin pencari. Pada gambar 16 ditampilkan hasil pencarian produk di mesin pencari



Gambar 16. Hasil di mesin pencari

3.9 Strategi Marketing

Strategi *marketing* yang akan diterapkan pada toko Noru Label untuk meningkatkan penjualan meliputi beberapa langkah yaitu, optimalisasi SEO pada halaman *website* dilakukan agar *website e-commerce* mudah ditemukan di mesin pencari. Lalu melakukan promosi di media sosial TikTok untuk menarik *audiens* yang lebih luas. Setelah itu dilakukan pembuatan video TikTok yang diharapkan dapat menarik perhatian pelanggan. Strategi *marketing* melalui tiktok ditampilkan pada gambar 17.



Gambar 17. Strategi marketing Tiktok

4. KESIMPULAN

Agar dapat memberikan saran untuk penelitian selanjutnya, dapat disimpulkan dengan adanya *website e-commerce*, pelanggan dapat menjelajahi produk dengan informasi yang lebih komprehensif dan deskripsi yang lebih jelas, proses transaksi lebih cepat dan tidak adanya retur produk. Penerapan SEO sebagai tambahan strategi promosi memungkinkan Noru Label untuk memperluas jangkauan pasar.

DAFTAR PUSTAKA

- [1] A. Tirtana, A. Zulkarnain, B. K. Kristanto, S. Suhendra, and M. A. Hamzah, "Rancang Bangun Aplikasi E-Commerce Untuk Meningkatkan Pendapatan UMKM," *J. Ilm. Teknol. Inf. Asia*, vol. 14, no. 2, p. 101, 2020, doi: 10.32815/jitika.v14i2.473.
- [2] M. L. Ramdhani, Nurleli, and A. Anandya, "Pengaruh Literasi Keuangan, Inklusi Keuangan, Dan Inovasi Terhadap Kinerja Ukm," *Curr. J. Kaji. Akunt. dan Bisnis Terkini*, vol. 3, no. 1, pp. 1–10, 2022, doi: 10.31258/current.3.1.1-10.
- [3] P. L. Rini, "Penggunaan e-Marketplace sebagai Mediasi antara Keunggulan e-Commerce dan

- Kinerja Manajemen Rantai Pasokan," *Jurnal Manajemen Bisnis dan Kewirausahaan*, vol. 5, no. 6, p. 673, 2021, doi: 10.24912/jmbk.v5i6.11345.
- [4] A. Rosyana, G. Gata, Y. Santoso, and ..., "Penerapan E-Commerce Untuk Memperluas Pemasaran Pada Toko Shafira," *Pros. Semin. ...*, vol. 2, no. April, pp. 899–907, 2023, [Online]. Available: <http://senafti.budiluhur.ac.id/index.php/senafti/article/view/579%0Ahttp://senafti.budiluhur.ac.id/index.php/senafti/article/download/579/345>
- [5] R. H. Nurhuda and H. Irawan, "Penerapan E-Commerce Business To Consumer (B2C) Menggunakan Content Management System Wordpress Studi Kasus Jocysprei," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 1, pp. 17–26, 2023, doi: 10.36080/idealis.v6i1.2987.
- [6] N. Sriwahyuni, A. Ariesta, and D. R. Utari, "Penerapan E-commerce Bisnis Pada PT Agiva Indonesia Guna Mencapai Target Pendapatan Penjualan," *G-Tech J. Teknol. Terap.*, vol. 7, no. 2, pp. 627–636, 2023, doi: 10.33379/gtech.v7i2.2325.
- [7] S. Santosa and H. Ismaya, "Perancangan Sistem Informasi Penjualan Menggunakan Cms Wordpress Pada Toko Importir Laptop Bandung," *J. Bisnis dan Pemasar.*, vol. 11, no. 1, pp. 1–8, 2021.
- [8] Susana Ayu Handayani and Enty Nur Hayati, "Penerapan E-Commerce Berbasis Cms Pada Umkm Herbal Bu Pini Sebagai Upaya Peningkatan Jangkauan Pemasaran Dalam Menghadapi Kebijakan Ppkm," *Elkom J. Elektron. dan Komput.*, vol. 15, no. 1, pp. 16–24, 2022, doi: 10.51903/elkom.v15i1.633.
- [9] A. Fuad and S. Juanita, "Penerapan E-Commerce Menggunakan Cms Wordpress Untuk Meningkatkan Penjualan Pada Online Shop Flanelis," *IDEALIS Indones. J. Inf. Syst.*, vol. 3, no. 1, pp. 511–518, 2020, doi: 10.36080/idealis.v3i1.2309.
- [10] I. R. Arief and I. Novita, "Penerapan E-Commerce Untuk Mengoptimalkan Penjualan Pada Toko Gita Giza," *IDEALIS Indones. J. Inf. Syst.*, vol. 4, no. 1, pp. 127–136, 2021, doi: 10.36080/idealis.v4i1.2823.

IMPLEMENTASI WEBSITE E-COMMERCE BERBASIS CONTENT MANAGEMENT SYSTEM PADA TOKO VAPE.BILL INDONESIA

Muhammad Rafi Pratama^{1*}, Lauw Li Hin²

^{1*}²Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta
Email: ^{1*}rafipratama@gmail.com, ²lihin@budiluhur.ac.id

(Naskah masuk: 21 September 2024, diterima untuk diterbitkan: 30 September 2024)

Abstrak

Dalam era digital yang semakin maju, penerapan teknologi informasi menjadi elemen kunci dalam meningkatkan daya saing bisnis, termasuk di sektor industri vape. Penelitian ini berfokus pada pengembangan dan implementasi website e-commerce berbasis Content Management System (CMS) pada Toko Vape.Bill dengan tujuan meningkatkan efisiensi operasional, manajemen stok, serta aksesibilitas konsumen. Masalah utama yang dihadapi oleh Toko Vape.Bill adalah keterbatasan promosi, pengelolaan stok yang tidak efisien, dan kesulitan menjangkau konsumen di luar daerah. Metode yang digunakan dalam penelitian ini melibatkan analisis sistem menggunakan Fishbone Diagram dan Business Model Canvas (BMC) untuk mengidentifikasi penyebab masalah, serta pengembangan dan perancangan sistem berbasis CMS dengan pendekatan Unified Modeling Language (UML) untuk pemodelan fungsional. Hasil penelitian menunjukkan bahwa implementasi CMS memungkinkan pengelolaan produk, promosi, dan transaksi secara lebih mudah, serta memperluas jangkauan pasar dan meningkatkan penjualan. CMS terbukti sebagai solusi yang tepat bagi usaha kecil dan menengah untuk membangun platform e-commerce yang adaptif terhadap perkembangan teknologi dan kebutuhan bisnis.

Kata kunci: *e-commerce, content management system, website, digitalisasi bisnis, pengembangan toko online, toko vape.*

IMPLEMENTATION OF AN E-COMMERCE WEBSITE BASED ON A CONTENT MANAGEMENT SYSTEM AT VAPE.BILL STORE INDONESIA

Abstract

In the rapidly advancing digital era, the implementation of information technology has become a key element in enhancing business competitiveness, including in the vape industry. This research focuses on the development and implementation of an e-commerce website based on a Content Management System (CMS) for Vape.Bill Store, with the aim of improving operational efficiency, stock management, and consumer accessibility. The main issues faced by Vape.Bill Store include limited promotion, inefficient stock management, and difficulty in reaching customers outside its local area. The methodology used in this study involves system analysis using Fishbone Diagram and Business Model Canvas (BMC) to identify the root causes of the problems, as well as the development and design of a CMS-based system using the Unified Modeling Language (UML) approach for functional modeling. The results of the research show that the implementation of CMS allows for easier management of products, promotions, and transactions, while also expanding market reach and increasing sales. CMS has proven to be an effective solution for small and medium-sized businesses to build e-commerce platforms that are adaptive to technological developments and business needs.

Keywords: *e-commerce, content management system, website, business digitalization, online store development, vape store*

1. PENDAHULUAN

Informasi teknologi saat ini memiliki peran yang informasi berfungsi untuk meningkatkan efektivitas sangat signifikan di berbagai sektor organisasi, termasuk Pendidikan, Kesehatan, pemerintah, bisnis dan lainnya. Dalam dunia bisnis, teknologi informasi

berfungsi untuk meningkatkan efektivitas promosi serta mempermudah proses transaksi[1].

Dengan kemajuan teknologi dan kehadiran internet di Indonesia, internet telah menjadi kebutuhan penting dalam kehidupan sehari-hari masyarakat. Melalui teknologi internet, masyarakat

Indonesia dapat dengan mudah mencari informasi dan mengakses konten yang mereka inginkan. Akibatnya, sebagian besar masyarakat kini memiliki telepon pintar (smartphone) yang selalu tersedia di gengaman mereka[2].

Dalam dunia bisnis modern, persaingan semakin intensif karena banyak perusahaan telah mengadopsi sistem informasi. Sistem informasi memainkan peran yang sangat penting, terutama dengan pesatnya pertumbuhan perusahaan atau organisasi, menjadikannya semakin krusial[3]. memahami pentingnya teknologi informasi dan dampaknya terhadap bisnis mereka. Selain itu, mereka perlu mengetahui cara memanfaatkan sistem informasi secara efektif untuk mendukung operasional dan strategi bisnis mereka[4].

Para pelaku usaha terus berinovasi untuk tetap kompetitif di era modern ini, dengan memperkenalkan berbagai inovasi produk dan strategi promosi untuk meningkatkan penjualan. Kehadiran teknologi informasi membuka peluang lebih besar bagi pelaku usaha untuk menjangkau konsumen secara efektif. Salah satu upaya yang dapat dilakukan adalah dengan mengimplementasikan sistem penjualan elektronik atau e-commerce. Oleh karena itu, e-commerce menjadi sistem yang penting untuk dikembangkan oleh perusahaan, terutama bagi perusahaan yang menawarkan layanan atau pemasaran kepada individu yang menjalankan bisnis mereka sendiri[5].

Namun, dengan meningkatnya persaingan di dunia e-commerce, manajemen konten yang efisien dan dinamis menjadi faktor kunci untuk mempertahankan daya saing. Di sinilah peran Content Management System (CMS) menjadi sangat penting[6]. CMS adalah sistem yang memungkinkan pengelolaan konten website secara mudah dan fleksibel, tanpa memerlukan pengetahuan teknis yang mendalam[7]. Dalam konteks e-commerce, CMS membantu pengusaha dalam mengelola katalog produk, melakukan pembaruan informasi, dan mengoptimalkan pengalaman pengguna secara efisien[8].

Industri vape di Indonesia telah mengalami perkembangan yang signifikan dengan semakin banyaknya konsumen yang beralih dari rokok konvensional ke produk vape. Pasar vape tumbuh dengan cepat dan menjadi alternatif yang lebih diterima, terutama di kalangan generasi muda[9]. Masyarakat semakin menyadari berbagai pilihan rasa dan jenis produk vape, serta klaim bahwa vape dapat menjadi pilihan yang lebih aman dibandingkan dengan merokok[10].

Namun, pertumbuhan industri vape juga menghadapi tantangan, termasuk regulasi yang ketat dan stigma negatif yang masih melekat pada produk ini. Untuk bersaing secara efektif, pelaku usaha di sektor vape perlu memanfaatkan teknologi digital guna meningkatkan daya saing dan efisiensi operasional. Toko Vape.Bill, sebagai salah satu

pemain di industri ini, mengalami kesulitan dalam hal promosi, pengelolaan stok, dan aksesibilitas terhadap konsumen. Sebelum implementasi sistem e-commerce, Toko Vape.Bill hanya mengandalkan metode pemasaran konvensional yang membatasi jangkauan pasar dan menyebabkan kesulitan dalam manajemen produk.

Berdasarkan pemahaman tersebut, penelitian ini bertujuan untuk membangun sebuah website e-commerce yang akan digunakan oleh Toko Vape.Bill dalam menjalankan operasional bisnisnya. Dengan adanya platform ini, diharapkan berbagai permasalahan yang dihadapi oleh toko dapat diselesaikan secara efektif melalui pemanfaatan teknologi digital yang tepat.

Vape.Bill merupakan salah satu pelaku usaha di segmen rokok elektrik, atau yang biasa disebut vape. Vape.Bill adalah toko vape yang berlokasi di Kota Tangerang, tepatnya di daerah Cirendeui. Toko ini menyediakan berbagai kebutuhan alat vape, seperti liquid, kapas, perangkat, cartridge, dan kawat. Seiring dengan meningkatnya minat konsumen terhadap produk vape, terutama di kalangan anak muda, Vape.Bill menghadapi tantangan, yaitu pelanggan yang enggan datang langsung ke toko hanya untuk membeli liquid atau cartridge. Dengan berkembangnya teknologi, adaptasi diperlukan untuk meningkatkan daya saing mereka di pasar.

Selama ini, Toko Vape.Bill hanya melakukan upaya promosi melalui metode jaringan perkenalan, yang menyebabkan keterbatasan dalam menjangkau pelanggan di luar daerah Cirendeui. Hal ini mengakibatkan potensi pasar yang lebih luas belum sepenuhnya dapat dimanfaatkan. Selain itu, Toko Vape.Bill juga menghadapi kendala dalam pengelolaan stok produk, di mana informasi terkait ketersediaan barang sering kali tidak tersampaikan dengan baik kepada pelanggan. Keterbatasan ini berdampak pada kepuasan pelanggan dan efisiensi operasional toko secara keseluruhan.

Penelitian ini bertujuan untuk merancang dan mengimplementasikan website e-commerce berbasis CMS yang akan digunakan oleh Toko Vape.Bill. Dengan memanfaatkan CMS, diharapkan toko ini dapat mengatasi permasalahan terkait promosi, manajemen stok, dan aksesibilitas pelanggan. Implementasi ini diharapkan tidak hanya meningkatkan efisiensi operasional tetapi juga memperbaiki pengalaman pelanggan dan meningkatkan daya saing di pasar yang semakin kompetitif.

Beberapa penelitian sebelumnya telah membahas implementasi e-commerce berbasis Content Management System (CMS) di masyarakat. Salah satunya adalah penerapan e-commerce dalam bentuk website untuk penjualan produk di Toko F2T Sport By Yovis, yang bertujuan untuk mempermudah proses penjualan dan meningkatkan efisiensi bisnis di toko tersebut[5].

Selain itu, terdapat pula penelitian mengenai pengembangan model e-commerce berbasis CMS untuk meningkatkan penjualan di Toko Parfum Herbal XYZ. Penelitian ini menggunakan metode Business Model Canvas untuk menganalisis model bisnis yang akan diterapkan dalam CMS, dengan menerapkan strategi Search Engine Optimization (SEO) dan Marketing Mix 4P untuk memaksimalkan hasilnya[11]. Terdapat juga pengembangan model e-commerce berbasis CMS untuk meningkatkan penjualan di Toko Pesona Tanaman. Penelitian ini menggunakan metode Business Model Canvas (BMC) untuk merancang dan menganalisis model bisnis yang akan diimplementasikan dalam CMS[8].

Penelitian selanjutnya adalah penelitian mengenai Pengaruh Citra Merek, Atmosfer Toko, dan Kualitas Produk Terhadap Keputusan Pembelian di Panda Vapestore yang membahas mengenai pengaruh citra merek, suasana toko, dan kualitas produk terhadap keputusan pembelian di Panda Vapestore baik secara parsial maupun simultan[12]. Perbedaan penelitian ini dengan penelitian sebelumnya terdapat pada objek yang diteliti serta goals dari penelitian. Seperti perbedaan dengan penelitian [8], yaitu objek yang diteliti adalah toko yang memiliki fokus bisnis dibidang tanaman hias, sedangkan pada penelitian ini objek yang diteliti adalah toko vape yang termasuk dalam Usaha Mikro Kecil Masyarakat (UMKM). Serta tujuan yang ada pada penelitian ini untuk mengatasi permasalahan terkait promosi, manajemen stok, dan aksesibilitas pelanggan, sedangkan penelitian[8] fokus kepada peningkatan penjualan produk. Penelitian ini berkontribusi dalam meningkatkan efisiensi operasional, memperluas jangkauan pasar, dan daya saing Toko Vape.Bill melalui implementasi website e-commerce berbasis Content Management System (CMS) yang mudah dikelola, serta memberikan bukti bahwa CMS merupakan solusi yang efektif untuk usaha kecil dan menengah dalam membangun platform e-commerce yang fleksibel dan mudah dikelola.

2. METODE PENELITIAN

Penelitian yang berkualitas harus dilakukan secara sistematis, mengikuti serangkaian langkah penelitian yang terstruktur dengan baik. Ini termasuk penetapan metode dan tahapan yang jelas untuk menyelesaikan masalah yang sedang diteliti[13]. Gambar 1 menunjukkan langkah-langkah dalam menyelesaikan penelitian ini, yang mencakup beberapa tahapan. Langkah pertama adalah menganalisis masalah yang ada di lokasi penelitian, diikuti dengan pengumpulan data yang diperlukan. Selanjutnya, dilakukan analisis sistem, perancangan sistem, dan implementasi e-commerce. Proses ini berlanjut hingga mencapai hasil yang diinginkan.



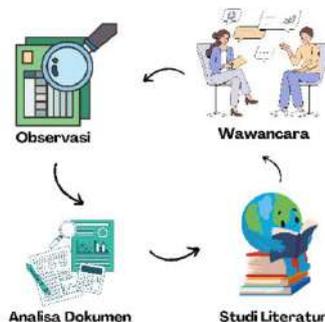
Gambar 1. Metode Penelitian

2.1 Identifikasi Masalah

Tahap identifikasi masalah adalah langkah awal penting dalam penelitian, di mana peneliti fokus pada memahami dan menentukan masalah yang ada. Pada tahap ini, peneliti melakukan observasi mendalam, mengumpulkan informasi dari berbagai sumber, dan menganalisis kondisi untuk mengidentifikasi isu utama yang memerlukan solusi[14]. Selain itu, pada tahap ini juga dilakukan analisa studi literatur guna menemukan metode dan tools yang tepat untuk digunakan dalam penelitian.

2.2 Pengumpulan Data

Gambar 2 menunjukkan alur tahap pengumpulan data dalam penelitian ini, yang melibatkan beberapa metode. Pertama, wawancara dilakukan untuk berinteraksi langsung dengan pemilik dan konsumen Toko Vape.Bill guna bertukar informasi. Kedua, observasi dilakukan dengan mengamati langsung operasional Toko Vape.Bill yang berlokasi di Jl. Raya Cirendeui, Kec. Ciputat Timur, Kota Tangerang Selatan, Banten 15419, untuk memahami situasi dan kegiatan sehari-hari toko. Ketiga, analisis dokumen dilakukan dengan menilai dokumen-dokumen yang relevan dari lokasi penelitian yang berkaitan dengan topik penelitian. Hanya dokumen yang berhubungan dengan topik yang dipilih untuk analisis data. Terakhir, studi literatur dilakukan untuk meninjau topik penelitian mengenai e-commerce dengan menggunakan Content Management System. Studi ini berfungsi memberikan pemahaman mendalam tentang konsep dan teori yang relevan, serta memastikan penelitian didasarkan pada pengetahuan yang solid dan sesuai dengan perkembangan terbaru di bidang e-commerce[15].



Gambar 2. Tahapan Pengumpulan Data

2.3 Analisa Sistem

Analisa sistem dalam penelitian ini dilakukan dengan menggunakan Fishbone Diagram dan Business Model Canvas (BMC) untuk mengidentifikasi dan mengevaluasi berbagai faktor yang mempengaruhi kinerja Toko Vape.Bill. Fishbone Diagram membantu menguraikan penyebab utama masalah dalam sistem yang meliputi proses pengelolaan stok yang tidak efisien, keterbatasan teknologi informasi, dan kekurangan dalam pelatihan staf[16]. Sementara itu, BMC memberikan perspektif tentang bagaimana elemen- elemen model bisnis seperti saluran distribusi, proposisi nilai, dan hubungan pelanggan berkontribusi pada masalah sistem yang ada. Kombinasi kedua alat ini memungkinkan pemahaman yang mendalam tentang bagaimana ketidakmampuan sistem yang ada mempengaruhi keseluruhan operasional, efisiensi promosi, dan kepuasan pelanggan, serta menawarkan panduan untuk perbaikan yang diperlukan[17].

2.4 Perancangan Sistem

Pada tahap ini, Use Case Diagram digunakan dengan bantuan alat Draw.io untuk menjelaskan persyaratan fungsional sistem dari perspektif pengguna[18]. Unified Modeling Language (UML) adalah seperangkat teknik diagram standar yang menyediakan representasi grafis yang kaya untuk memodelkan setiap tahap proyek pengembangan sistem, mulai dari analisis hingga implementasi. Selain itu, model Use Case Diagram digunakan untuk membuat Deskripsi Use Case yang merinci aktivitas atau perilaku sistem dari sudut pandang pengguna[19]. Peneliti juga memanfaatkan alat Balsamiq untuk merancang antarmuka situs web e-commerce Toko Vape.Bill, memastikan bahwa desain antarmuka memenuhi kebutuhan pengguna dan fungsionalitas sistem[20].

2.5 Implementasi Sistem

Tahapan implementasi sistem dimulai dengan perencanaan dan analisis kebutuhan, di mana data dikumpulkan melalui wawancara dan analisis dokumen untuk menyusun spesifikasi sistem e-commerce berbasis Content Management System (CMS)[21]. Selanjutnya, sistem dirancang menggunakan Use Case Diagram dengan Draw.io untuk mendefinisikan persyaratan fungsional dan Deskripsi Use Case untuk menjelaskan aktivitas sistem, serta antarmuka pengguna dirancang dengan Balsamiq

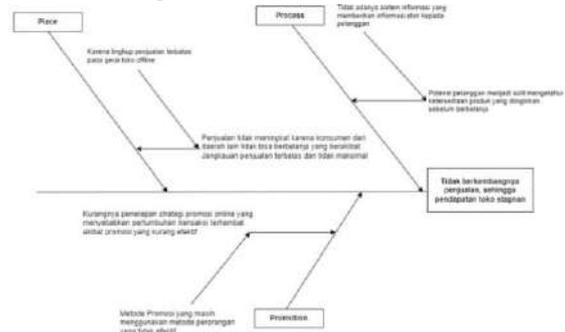
3. HASIL PENELITIAN

3.1 Analisa Masalah

Pada analisis masalah dilakukan dengan pendekatan Fishbone Diagram dan Business Model Canvas (BMC). Fishbone Diagram digunakan untuk mengidentifikasi faktor-faktor penyebab utama masalah yang dihadapi[22], sementara BMC digunakan untuk mengevaluasi dampak masalah

terhadap elemen-elemen penting dalam model bisnis Toko Vape.Bill[23].

a. *Fishbone Diagram*



Gambar 3. Tahapan Pengumpulan Data

Pada Gambar 3 merupakan Fishbone Diagram yang menjelaskan masalah yang dialami oleh toko Vape.Bill, yakni fokus penjualan hanya pada toko fisik, membatasi jangkauan konsumen dari daerah yang lebih jauh. Dalam hal proses, tidak adanya sistem informasi yang memberikan data stok kepada pelanggan membuat pembeli sulit mengetahui ketersediaan produk sebelum datang. Dari sisi promosi, kurangnya strategi pemasaran online serta masih digunakannya metode promosi tradisional menyebabkan pertumbuhan transaksi lambat dan tidak efektif[24]. Masalah-masalah ini berkontribusi pada stagnasi pendapatan toko dan menunjukkan kebutuhan mendesak untuk implementasi strategi digital dan e-commerce yang lebih baik.

b. *Business Model Canvas (BMC)*



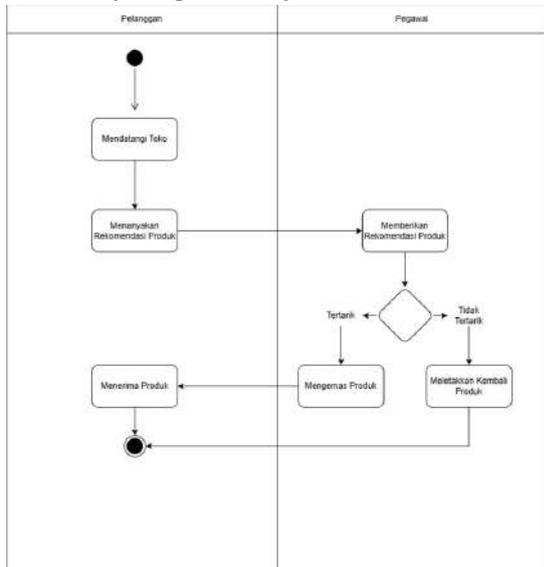
Gambar 4. Analisa Masalah Menggunakan Business Model Canvas

Pada Gambar 4 menunjukkan Analisa *Business Model Canvas* (BMC) Toko Vape.Bill menunjukkan bahwa bisnis bergantung pada pemasok produk vape dan platform e-commerce sebagai Key Partners, dengan Key Activities mencakup manajemen stok dan penjualan online. Key Resources meliputi produk vape dan platform online. Value proposition berfokus pada kemudahan akses informasi produk dan layanan konsultasi. Pelanggan dilayani melalui toko fisik dan online, dengan Customer Segments terbagi antara pelanggan lokal dan online. Cost Structure mencakup pembelian produk dan operasional, sedangkan

Revenue Streams berasal dari penjualan produk dan layanan konsultasi.

3.2 Analisa Proses Berjalan

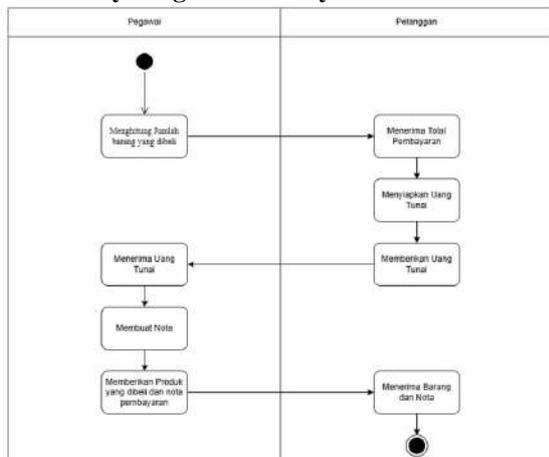
a. Activity Diagram Penjualan



Gambar 5. Activity Diagram Penjualan

Pada Gambar 5 menunjukkan Activity diagram yang menggambarkan alur proses interaksi antara pelanggan dan pegawai di toko Vape.Bill dalam hal pemilihan dan penerimaan produk. Proses dimulai dengan pelanggan mendatangi toko, lalu menanyakan rekomendasi produk kepada pegawai. Pegawai kemudian memberikan rekomendasi produk kepada pelanggan. Jika pelanggan tertarik dengan produk yang ditawarkan, pegawai akan mengemas produk tersebut, dan pelanggan akan menerima produk yang telah dipilih. Namun, jika pelanggan tidak tertarik, pegawai akan mengembalikan produk tersebut ke tempat semula. Diagram ini menunjukkan bagaimana alur interaksi dan pengambilan keputusan dalam proses penjualan di toko, dengan fokus pada layanan rekomendasi produk dan penerimaan barang

b. Activity Diagram Pembayaran

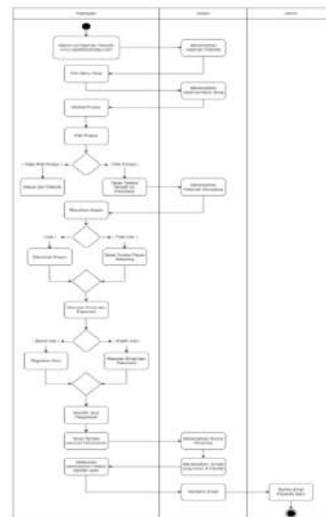


Gambar 6. Activity Diagram Pembayaran

Pada Gambar 6 menunjukkan Activity diagram yang menggambarkan proses pembayaran dan penyerahan barang di toko Vape.Bill antara pegawai dan pelanggan. Proses dimulai ketika pegawai menghitung jumlah barang yang dibeli oleh pelanggan. Setelah itu, pelanggan menerima total pembayaran dan menyiapkan uang tunai. Pelanggan kemudian memberikan uang tunai kepada pegawai, yang menerima uang tersebut dan membuat nota pembayaran. Selanjutnya, pegawai menyerahkan produk yang telah dibeli beserta nota pembayaran kepada pelanggan. Akhirnya, pelanggan menerima barang dan nota sebagai bukti pembelian. Diagram ini menunjukkan alur proses transaksi pembayaran tunai yang dilakukan secara langsung di toko.

3.3 Proses Bisnis Usulan

a. Proses pemesanan produk

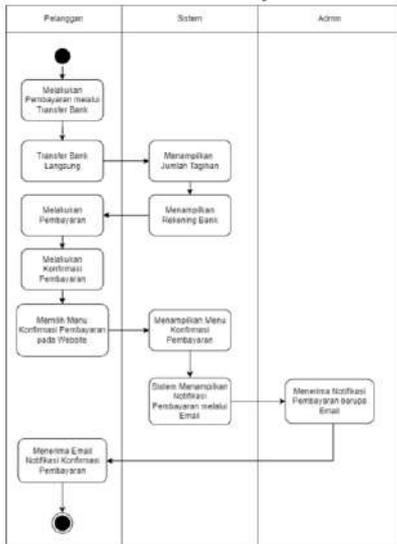


Gambar 7. Proses Usulan Pemesanan Produk

Pada Gambar 7 menunjukkan proses pemesanan produk di website e-commerce Vape.Bill. Pelanggan memulai dengan memilih produk dari menu "Shop", menambahkannya ke keranjang, dan memasukkan kupon jika tersedia. Setelah itu, pelanggan login atau registrasi akun, memilih jasa pengiriman, dan melakukan pemesanan. Sistem menampilkan informasi pembayaran, dan pelanggan melakukan transfer. Terakhir, sistem mengirimkan notifikasi email ke admin untuk memproses pesanan baru

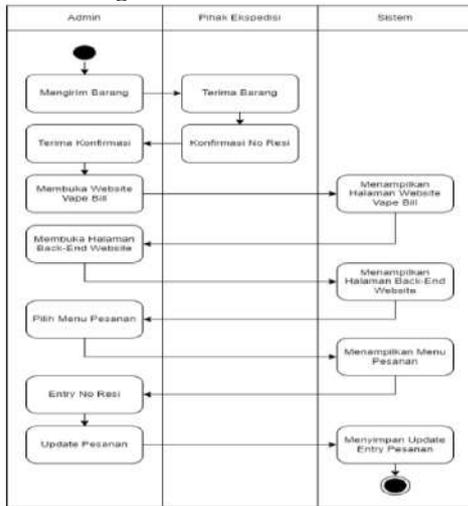
b. Proses Konfirmasi Pembayaran

Pada Gambar 8 menjelaskan alur konfirmasi pembayaran pada sistem e-commerce Vape.Bill. Pelanggan melakukan pembayaran melalui transfer bank, setelah itu sistem menampilkan jumlah tagihan dan informasi rekening bank. Setelah pembayaran selesai, pelanggan memilih menu konfirmasi pembayaran di website, dan sistem mengirimkan notifikasi konfirmasi pembayaran melalui email kepada admin. Admin kemudian menerima email notifikasi yang mengonfirmasi bahwa pembayaran telah diterima.



Gambar 8. Proses Usulan Konfirmasi Pembayaran

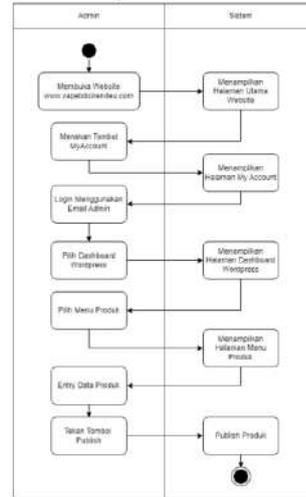
c. Proses Pengiriman



Gambar 9. Proses Pengiriman

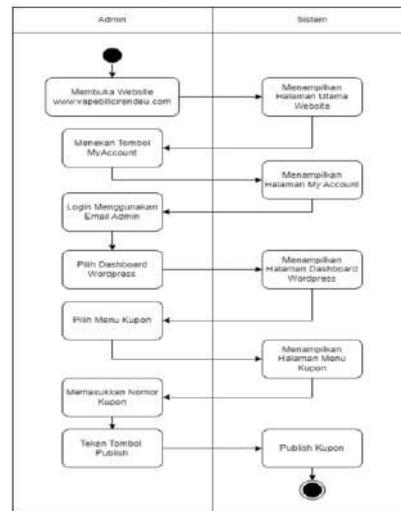
d. Proses Entry Produk

Pada Gambar 10 menunjukkan proses Entry produk hanya dapat dilakukan oleh admin, dengan melakukan login terlebih dahulu melalui website, menggunakan email admin. Kemudian masuk ke halaman dashboard wordpress, lalu pilih menu produk, selanjutnya data produk dan sistem akan menyimpan ke menu admin.



Gambar 10. Proses Usulan Entry Produk

e. Proses Entry Kupon

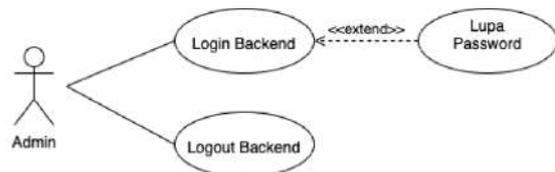


Gambar 11. Proses Entry Kupon

Pada Gambar 11 menunjukkan proses Entry kupon hanya dilakukan oleh admin, dengan melakukan login terlebih dahulu melalui website menggunakan email admin. Selanjutnya admin akan masuk kedalam halaman dashboard dan memilih menu produk. Selanjutnya entry data kupon dan sistem akan menyimpannya

3.4 Proses Bisnis Usulan

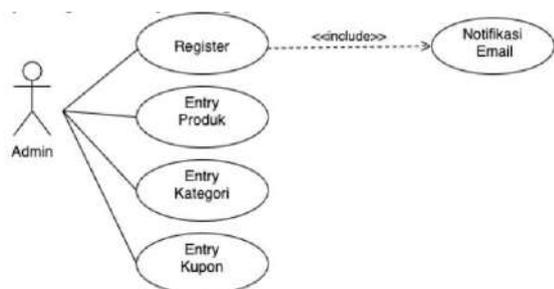
Hasil dari tahap ini adalah rancangan sistem berupa Use Case Diagram dimulai dari Use Case Login hingga Transaksi pelanggan maupun admin.



Gambar 12. Use Case Diagram Login Admin

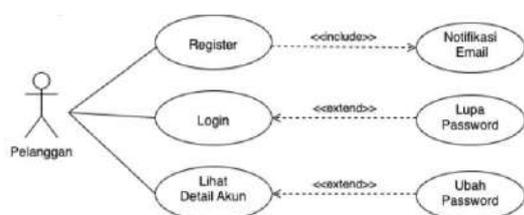
Gambar 12 menggambarkan interaksi antara seorang Admin dengan sistem untuk melakukan login

dan logout pada backend. Admin dapat menggunakan fitur Login Backend untuk masuk ke sistem dan Logout Backend untuk keluar. Jika Admin lupa password saat login, ada opsi tambahan "Lupa Password" yang dapat diakses melalui hubungan extend dari use case "Login Backend", yang memungkinkan Admin untuk mengatur ulang kata sandi. Hubungan extend ini bersifat opsional dan hanya digunakan jika diperlukan.



Gambar 13. Use Case Diagram Master Admin

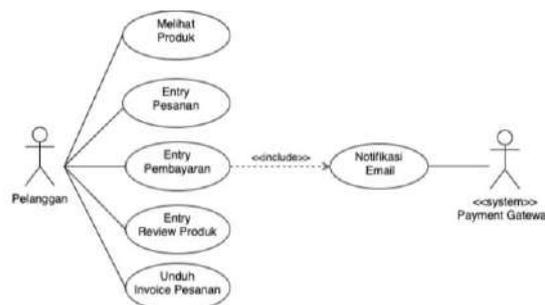
Pada Gambar 13 menunjukkan interaksi seorang Admin dengan sistem untuk melakukan berbagai tugas, termasuk register dan pengelolaan data. Admin dapat melakukan register, yang memiliki hubungan include dengan notifikasi email, artinya setiap kali proses register dilakukan, sistem juga akan mengirimkan email notifikasi secara otomatis. Selain itu, Admin juga dapat memasukkan (entry) data terkait produk, kategori, dan kupon melalui masing-masing use case (Entry Produk, Entry Kategori, dan Entry Kupon). Setiap aktivitas ini tidak bergantung satu sama lain, tetapi semuanya dikelola oleh Admin sebagai aktor utama.



Gambar 14. Use Case Diagram Transaksi Admin

Gambar 14 menunjukkan use case diagram yang menunjukkan pelanggan dapat register dengan notifikasi email otomatis (include), login dengan opsi Lupa Password (extend), serta melihat detail akun dengan pilihan untuk Ubah Password (extend). Diagram ini menggambarkan fitur dasar manajemen akun pelanggan.

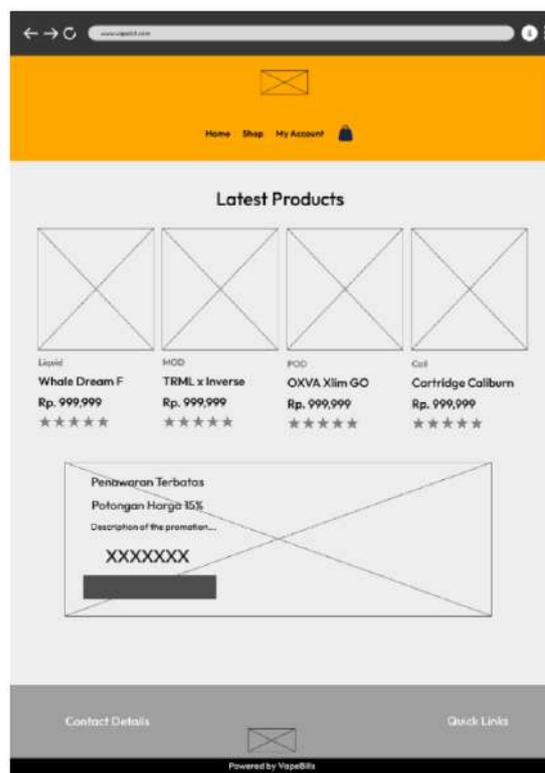
Diagram use case seperti pada Gambar 15 menunjukkan bagaimana Pelanggan berinteraksi dengan sistem e-commerce untuk melihat produk, membuat pesanan, melakukan pembayaran, memberikan ulasan produk, dan mengunduh invoice. Sistem juga mengirimkan notifikasi email setelah pembayaran diproses melalui Payment Gateway.



Gambar 15. Use Case Diagram Transaksi Pelanggan

3.5 Rancangan Layar

3.5.1 Rancangan Layar Home

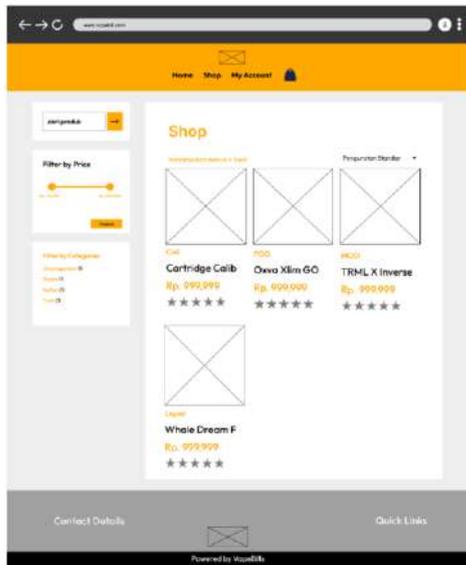


Gambar 16. Rancangan Layar Home

Halaman Beranda berfungsi sebagai tampilan utama bagi pengguna untuk menjelajahi produk dan penawaran yang tersedia. Beberapa elemen penting pada halaman ini meliputi: Banner Promosi, yang menampilkan berbagai promosi dari toko; Menu Navigasi, yang memudahkan pengguna untuk berpindah antar halaman; serta tombol (Button), yang memungkinkan pengguna berinteraksi dengan pesan dan menavigasi ke halaman produk.

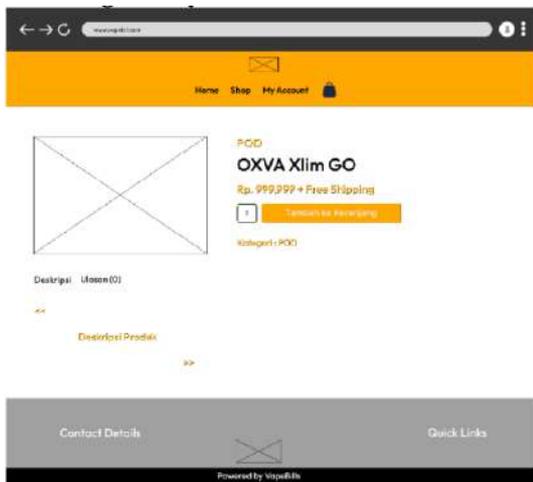
3.5.2 Rancangan Halaman Produk

Gambar 17 menampilkan daftar produk dalam kategori tertentu, dengan beberapa elemen utama yang disajikan, seperti Thumbnail Produk yang menampilkan gambar produk, Nama Produk, dan Harga Produk



Gambar 17. Rancangan Layar Halaman Produk

3.5.3 Rancangan Layar Halaman Detail Produk



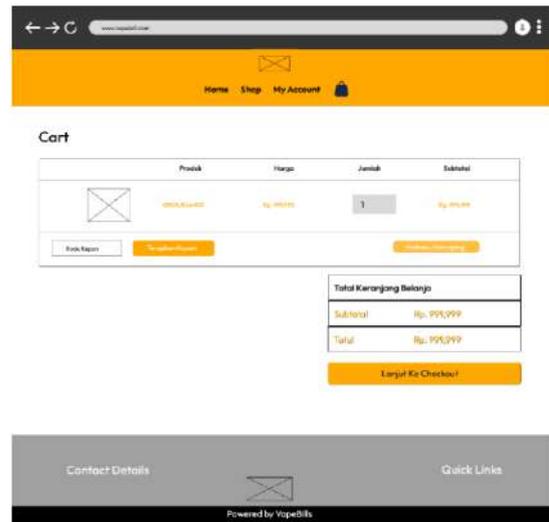
Gambar 18. Rancangan Layar Detail Produk

Halaman Detail Produk menampilkan informasi terkait produk yang dipilih, dengan beberapa elemen penting yang terdapat di dalamnya. Elemen-elemen tersebut meliputi thumbnail produk yang menampilkan gambar produk, nama produk, harga produk, dan deskripsi produk yang memberikan penjelasan lebih rinci. Selain itu, terdapat tombol untuk menambahkan produk ke keranjang serta fitur untuk menambah atau mengurangi jumlah produk yang dipilih.

3.5.4 Rancangan Layar Keranjang

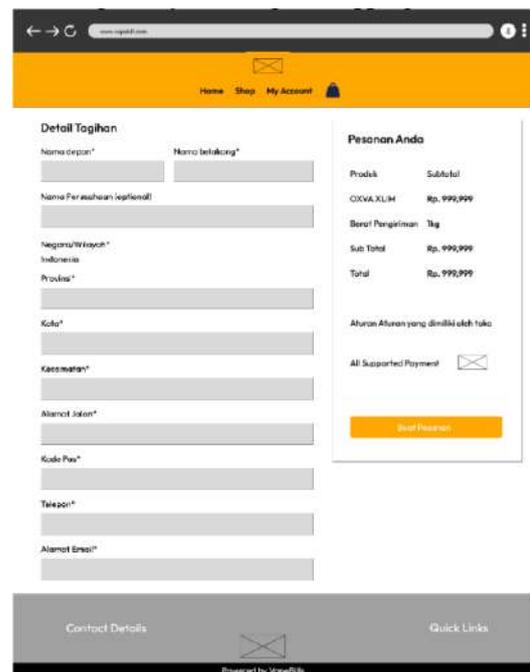
Halaman Keranjang dirancang untuk menampilkan produk-produk yang dipilih pelanggan untuk dibeli, dengan beberapa elemen utama yang terdapat di dalamnya. Elemen tersebut meliputi daftar produk yang ada di keranjang belanja, total harga yang merupakan jumlah biaya dari semua produk yang dipilih, tombol checkout untuk melanjutkan ke proses pembayaran, serta fitur kupon yang

memungkinkan pengguna menerapkan diskon pada pembelian mereka



Gambar 19. Rancangan Layar Keranjang

3.5.5 Rancangan Layar Billing & Shopping



Gambar 20. Rancangan Layar Billing & Shopping

Halaman ini menyediakan informasi terkait proses pembayaran kepada pengguna, dengan beberapa elemen penting di dalamnya. Elemen-elemen tersebut mencakup formulir informasi pengiriman, di mana pengguna mengisi nama, alamat, dan kontak penerima, serta detail pembayaran yang menampilkan metode pembayaran yang tersedia. Selain itu, terdapat ringkasan pesanan yang mencantumkan total biaya dan rincian produk yang dipesan, serta tombol untuk melanjutkan proses pemesanan

3.5.6 Rancangan Layar Detail Pesanan

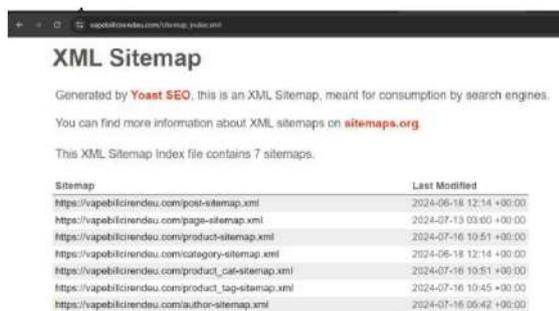


Gambar 21. Rancangan Layar Detail Pesanan

Halaman Detail Pesanan menampilkan ringkasan dan status pesanan kepada pengguna, dengan beberapa elemen penting di dalamnya. Elemen-elemen tersebut meliputi nomor pesanan yang memungkinkan pengguna melihat pesanan yang telah tervalidasi, daftar produk pesanan yang berisi rincian produk yang telah dipesan, status pesanan yang menunjukkan tahap pemrosesan pesanan, serta informasi pengiriman yang mencantumkan nomor pelacakan untuk pesanan yang telah dikirim

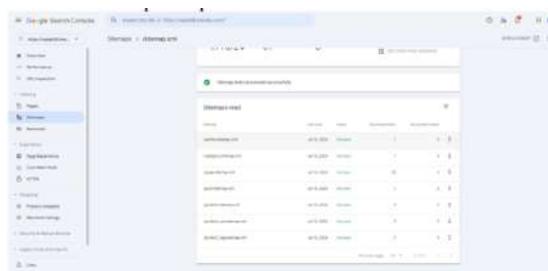
3.6 Strategi SEO

Setelah selesai membangun website Toko Vape.Bill, tahapan selanjutnya adalah melakukan optimalisasi SEO (Search Engine Optimization). Penggunaan plugin Yoast SEO menjadi langkah yang efektif karena memberikan saran dan panduan langkah-langkah yang diperlukan untuk mempercepat proses indeksasi oleh mesin pencari Google. Tahap berikutnya adalah mengonfigurasi sitemap pada website www.vapebillcirendeu.com ke dalam Google Search Console. Plugin Yoast SEO digunakan sebagai alat bantu dalam menghasilkan sitemap. Gambar 22 menunjukkan hasil dari XML Sitemap.



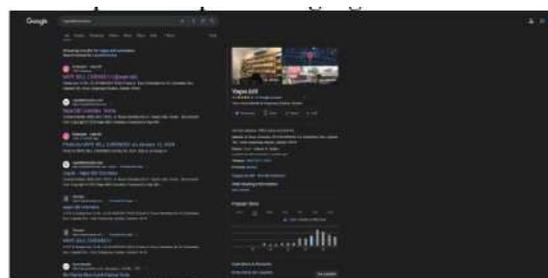
Gambar 22. XML Sitemap

Setelah mendapatkan sitemap, diperlukan konfigurasi pada google search console dengan memasukkan sitemap dari XML Sitemap ke add a new sitemap pada halaman google search console seperti pada Gambar 23



Gambar 23. XML Sitemap Per Kategori

Pada Gambar 24 merupakan hasil dari keyword yang bisa digunakan untuk mencari website Vape Bill di pencarian google



Gambar 24. Hasil Keyword

3.7 Strategi Marketing

Strategi pemasaran yang diterapkan pada Toko Vape.Bill adalah penggunaan platform e-commerce seperti Instagram. Media sosial Instagram dipilih sebagai alat pemasaran karena dapat mengarahkan pengguna ke halaman website e-commerce untuk melakukan transaksi pembelian. Selain berfungsi sebagai media promosi, Instagram juga dimanfaatkan karena adanya beberapa komunitas yang bisa dijadikan sarana beriklan. Selain itu, media sosial juga dapat menjadi sarana interaksi antara pelanggan dan penjual, yang pada akhirnya meningkatkan kepercayaan pelanggan. Strategi pemasaran yang diterapkan pada website Vape.Bill juga mencakup penggunaan kupon. Strategi ini bertujuan untuk menarik minat pelanggan serta memperluas jangkauan pemasaran. Penggunaan kupon diharapkan dapat meningkatkan jumlah transaksi, kepercayaan, dan loyalitas pelanggan.

4. KESIMPULAN

Berdasarkan hasil penelitian, yang mencakup perancangan antarmuka dan penerapan strategi Search Engine Optimization (SEO) serta pemasaran, implementasi sistem e-commerce berbasis CMS di Toko Vape.Bill telah berhasil meningkatkan efisiensi dalam pengelolaan produk dan proses transaksi. Sistem ini memungkinkan interaksi yang lebih optimal dengan pelanggan melalui fitur pencarian produk yang lebih sederhana dan proses pembayaran

yang lebih efisien. Selain itu, CMS memberikan fleksibilitas dalam pengaturan produk dan memudahkan penyesuaian serta pengembangan agar tetap selaras dengan perkembangan pasar. Secara keseluruhan, implementasi ini telah berkontribusi pada peningkatan kinerja Toko Vape.Bill, baik dalam hal peningkatan penjualan maupun perluasan jangkauan pasar melalui platform online

DAFTAR PUSTAKA

- [1] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," PRISMA, *Prosiding Seminar Nasional Matematika*, vol. 3, no. 1, pp. 64–71, Feb. 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [2] F. Ramadhan and L. Fajarita, "Implementasi Rancangan E-Commerce Pada Surya Timur Collection Dengan Metode Business Model Canvas (Bmc) Berbasis Content Management System (CMS)," IDEALIS : InDonEsiA journal Information System, vol. 3, no. 1, pp. 251–259, Feb. 2020, doi: 10.36080/idealis.v3i1.1508.
- [3] Y. Wahyudin and D. N. Rahayu, "Analisis Metode Pengembangan Sistem Informasi Berbasis Website: A Literatur Review," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 15, no. 3, pp. 26–40, Oct. 2020, doi: 10.35969/interkom.v15i3.74.
- [4] Farhatun Nisaul Ahadiyah, "Perkembangan Teknologi Infomasi Terhadap Peningkatan Bisnis Online," INTERDISIPLIN: Journal of Qualitative and Quantitative Research, vol. 1, no. 1, pp. 41–49, Dec. 2023, doi: 10.61166/interdisiplin.v1i1.5.
- [5] A. N. Kulyubi and H. Patrie, "Implementasi Website E-Commerce Berbasis Content Management System (CMS) Pada Toko F2T Sport By Yovis," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, vol. 1, no. 1, pp. 1191–1200, Sep. 2022.
- [6] A. A. A. Ushud, I. Novita, and N. Juliasari, "Pelatihan Pemanfaatan CMS Untuk Pembuatan Website Bagi OrangTua Siswa Sekolah Alam Tangerang," JAM-TEKNO (Jurnal Pengabdian Kepada Masyarakat TEKNO), vol. 2, no. 1, pp. 20–25, Jun. 2021.
- [7] Basorudin, Gunarso, Erni Rouza, Luth Fimawahib, and Asep Supriyanto, "Perancangan dan Implementasi Sistem Operasi Linux Debian untuk Konfigurasi Content Management System (CMS) Wordpress Dengan Winscp," *Bulletin of Computer Science Research*, vol. 3, no. 1, pp. 21–29, Dec. 2022, doi: 10.47065/bulletincsr.v3i1.188.
- [8] Faisal, A. Diana, and D. R. Utari, "Implementasi Website E-Commerce Berbasis Content Management System Wordpress Pada Toko Pesona Tanaman," *Ikraith-Informatika*, vol. 5, no. 3, pp. 121–131, Oct. 2021.
- [9] A. S. Agustin, P. N. L. Maghfiroh, M. Firdaus, and S. Alfariisy, "Peran Inovasi Teknologi dalam Transformasi Industri Rokok: Dampaknya Terhadap Produksi, Distribusi, dan Penggunaan Produk Rokok," *El-Mujtama: Jurnal Pengabdian Masyarakat*, vol. 4, no. 5, Sep. 2024, doi: 10.47467/elmutjama.v4i5.3553.
- [10] H. N. Putri, E. Budianita, F. Syafria, and F. Insani, "Penerapan Algoritma Apriori Dalam Menentukan Pola Perilaku Dan Gaya Hidup Terhadap Penderita Hipertensi," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 3, pp. 450–458, 2022.
- [11] S. A. H. Munabari and A. U. Hamdani, "Model E-Commerce Berbasis Content Management System Untuk Meningkatkan Penjualan Pada Toko Parfum Herbal XYZ," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, no. 1, pp. 1777–1786, Sep. 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php>
- [12] I. A. Farah, A. Mansur, and E. Lovita, "The Effect of Brand Image, Packaging, Advertising and Product Quality on Purchasing Decisions," *Research of Business and Management*, vol. 2, no. 2, pp. 77–87, Aug. 2024, doi: 10.58777/rbm.v2i2.245.
- [13] R. Ramadhan and I. Novita, "Perancangan E-Commerce Untuk Meningkatkan Penjualan Pada Space Vapor Store," IDEALIS : InDonEsiA journal Information System, vol. 3, no. 1, pp. 152–156, Feb. 2020, doi: 10.36080/idealis.v3i1.2111.
- [14] A. Nuraini and A. A. Rachmawati, "Perancangan User Interface Aplikasi M Parkir Universitas Widyatama Menggunakan Metode Design Thinking," JUSTINFO | Jurnal Sistem Informasi dan Teknologi Informasi, vol. 1, no. 1, pp. 21–33, Dec. 2023, doi: 10.33197/justinfo.vol1.iss1.2023.1249.
- [15] R. A. Santoso et al., "Analisis Literature Review Tentang Efektivitas Perencanaan Dan Pengendalian Anggaran Biaya Produksi Di Perusahaan," *Nusantara Journal of Multidisciplinary Science*, vol. 1, no. 6, pp. 333–341, Jan. 2024, [Online]. Available: <https://jurnal.intekom.id/index.php/njms>
- [16] Ahadya Silka Fajaranie and A. N. Khairi, "Pengamatan Cacat Kemasan Pada Produk Mie Kering Menggunakan Peta Kendali Dan Diagram Fishbone Di Perusahaan Produsen Mie Kering Semarang, Jawa Tengah," *Jurnal Pengolahan Pangan*, vol. 7, no. 1, pp. 7–13, Jun. 2022, doi: 10.31970/pangan.v7i1.69.
- [17] A. Suarni, F. Adzim, M. N. Abdi, and A. A. Adiningrat, "Training Bisnis Model Canvas (BMC) Bagi Petani Buah Naga Masa Pandemi Covid-19 Desa Caramming Bulukumba," *Jurnal Abmas Negeri (JAGRI)*, vol. 1, no. 1, pp. 11–19, Dec. 2020, doi: 10.36590/jagri.v1i1.88.
- [18] S. Pranoto, S. Sutiono, Sarifudin, and D. D. Nasution, "Penerapan UML Dalam Perancangan Sistem Informasi Pelaporan Dan Evaluasi Pembangunan Pada Bagian Administrasi Pembangunan Sekretariat Daerah Kota Tebing Tinggi," SURPLUS : Jurnal Ekonomi dan Bisnis, vol. 2, no. 2, pp. 384–401, Jun. 2024.
- [19] M. Munawir Arpan and A. D. M. Arif, "Rancang Bangun Sistem Informasi Pengarsipan Dokumen Berbasis Web Di Kantor Kecamatan Siak Hulu," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 1, no. 2, pp. 115–123, Sep. 2021, doi: 10.57152/ijirse.v1i2.53.

- [20] P. Saputra, Saputra Zulhipni Reno, and Karnadi, “Perancangan User Interface Dan User Experience Administrasi Desa Di Kayu Agung Menggunakan Balsamiq,” *KOhesi : Jurnal Multidisiplin Saintek*, vol. 2, no. 12, pp. 86–96, Apr. 2024, doi: 10.3785/koehesi.v2i12.2822.
- [21] A. Imam Solihin and Hadiq, “Pengembangan Website Promosi Pada CV Starmedia Technology Menggunakan Waterfall,” *Jurnal Teknologi Informasi Mura* Ahmad Imam Solihin, Hadiq, vol. 16, no. 1, Jun. 2024.
- [22] A. A. Dewi and F. Yuamita, “Pengendalian Kualitas Pada Produksi Air Minum Dalam Kemasan Botol 330 ML Menggunakan Metode Failure Mode Effect Analysis (FMEA) Di PDAM Tirta Sembada,” 2022.
- [23] M. H. Wardana and F. D. Sitania, “Strategi Pengembangan Bisnis Melalui Pendekatan Business Model Canvas pada Kedai Kabur Bontang,” *Jurnal Teknik Industri (JATRI)*, vol. 1, no. 1, pp. 10–17, Feb. 2023, doi: 10.30872/jatri.v1i1.492.
- [24] T. Erika, M. Tinneke, M. Tumbel, J. V. M. Program, S. Ilmu, and A. Bisnis, “Pengaruh Promosi Terhadap Keputusan Pembelian pada Konsumen Dapur Sedap Om Dol di Manado,” *Productivity*, vol. 2, no. 3, pp. 203–207, Jul. 2021.

KLASTERISASI DAERAH KRIMINALITAS DI INDONESIA DENGAN METODE K-MEANS CLUSTERING

T.Sofia Chairani^{1*}, Hijka Listia², Sabina Wardaniah³, Siti Wulandari⁴, Putri Tasya Agustina Tampubolon⁵, Arnita Piliang⁶

^{1,2,3,4,5,6}Universitas Negeri Medan

Email: ^{1*}tsofiachairani@gmail.com, ²hsbtia62@mhs.unimed.ac.id, ³sabinawardaniah@mhs.unimed.ac.id, ⁴sitiwulandari271023@gmail.com, ⁵ciyatampu@gmail.com, ⁶arnita@unimed.ac.id

(Naskah masuk: 3 September 2024, diterima untuk diterbitkan: 30 September 2024)

Abstrak

Kriminalitas di Indonesia menjadi masalah signifikan yang memengaruhi berbagai aspek seperti keamanan, sosial, dan ekonomi. Namun, upaya mitigasi sering terkendala oleh kurangnya informasi yang terstruktur tentang daerah rawan kriminalitas. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan mengelompokkan provinsi di Indonesia berdasarkan pola kriminalitas menggunakan algoritma K-Means Clustering. Data statistik kriminal tahun 2014-2023 dari Badan Pusat Statistik (BPS) dianalisis dengan menentukan jumlah cluster optimal yang menghasilkan lima cluster dengan evaluasi menggunakan Python dan library Scikit-learn. Hasil penelitian menunjukkan Silhouette Score sebesar 0.593, yang mencerminkan pembentukan cluster cukup baik. Klasterisasi ini memberikan panduan berbasis data bagi pemerintah dalam menyusun kebijakan keamanan yang lebih terarah untuk mengurangi tingkat kriminalitas di Indonesia.

Kata kunci: kriminalitas, klasterisasi, k-means clustering, analisis data, keamanan.

CLUSTERING CRIME-PRONE AREAS IN INDONESIA USING THE K-MEANS METHOD

Abstract

Crime in Indonesia is a significant problem that affects various aspects such as security, social, and economic. However, mitigation efforts are often hampered by a lack of structured information about crime-prone areas. This study aims to overcome this problem by grouping provinces in Indonesia based on crime patterns using the K-Means Clustering algorithm. Crime statistics data for 2014-2023 from the Central Statistics Agency (BPS) were analyzed by determining the optimal number of clusters that resulted in five clusters with evaluation using Python and the Scikit-learn library. The results showed a Silhouette Score of 0.593, which reflects the formation of a fairly good cluster. This clustering provides data-driven guidance for the government in developing more targeted security policies to reduce crime rates in Indonesia.

Keywords: criminality, clustering, k-means clustering, data analysis, security

1. PENDAHULUAN

Kriminalitas berasal dari istilah "kriminal" yang merujuk pada serangkaian tindakan atau perilaku yang terkait dengan pelanggaran hukum atau kejahatan yang dapat dikenai sanksi pidana menurut peraturan undang-undang. Istilah kriminalitas merujuk pada tindakan yang melanggar hukum pidana atau dianggap sebagai kejahatan. Dalam hal ini, kriminalitas mencakup berbagai tindakan yang dianggap sebagai pelanggaran hukum atau perbuatan kriminal [1],[2]. Kejahatan seperti pembunuhan, penyerangan, pemerkosaan, dan lainnya semakin meningkat saat ini. Semakin meningkat angka kejahatan, semakin terancam rasa aman manusia. Indonesia menempati peringkat 70 sebagai negara dengan tingkat kriminalitas tinggi di dunia menurut

Numbeo 2022, sementara di Asia, Indonesia berada di peringkat 15. Berada di peringkat 5 di antara negara-negara di Asia Tenggara [3]. Di Indonesia, tingkat kejahatan berbeda-beda di setiap daerah, dipengaruhi oleh hal-hal seperti jumlah penduduk, kondisi ekonomi, tingkat pendidikan, dan situasi sosial [4]. Tingginya tingkat kejahatan seperti pembunuhan dan penipuan berdampak negatif yang signifikan. Kejahatan dapat menyebabkan kemiskinan dan mengganggu pertumbuhan ekonomi serta mengurangi rasa aman masyarakat [5].

Maka, penting untuk menganalisis dan mengelompokkan daerah-daerah yang rentan terhadap kejahatan guna mendukung penetapan kebijakan pencegahan dan penanganan yang lebih efektif. Penelitian ini berkontribusi dengan

mengidentifikasi pola kriminalitas di Indonesia secara sistematis, memberikan peta kerawanan yang dapat mendukung pengambilan kebijakan berbasis data. Berbeda dengan penelitian sebelumnya, studi ini memanfaatkan algoritma K-Means Clustering untuk memberikan analisis yang lebih komprehensif termasuk evaluasi menggunakan Silhouette Score, yang memberikan gambaran akurasi pengelompokan. Hasil penelitian diharapkan membantu pemerintah dan lembaga terkait dalam menyusun strategi keamanan yang lebih efektif sesuai dengan kebutuhan setiap wilayah.

Data Mining mengacu pada proses mengungkap informasi tersembunyi dalam basis data untuk menemukan wawasan baru dan pola yang bermakna dalam data. Disiplin ini menggunakan berbagai metodologi yang disesuaikan dengan tujuan khusus yang terkait dengan pemanfaatan kumpulan data, yang meliputi estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi. [6].

Clustering adalah teknik yang digunakan untuk mengkategorikan atau mengelompokkan sekumpulan objek berdasarkan atribut atau karakteristik yang sama dalam kaitannya dengan data lainnya. Ini adalah metode mendasar dalam bidang penambangan data, yang dicirikan oleh sifatnya yang tidak diawasi, yang menunjukkan bahwa pendekatan ini tidak memerlukan pelatihan sebelumnya. [7].

Algoritma K-Means Clustering adalah metode yang digunakan untuk mempartisi data ke dalam kluster-kluster berbeda berdasarkan kesamaan titik-titik data. Titik-titik data yang menunjukkan karakteristik serupa ditetapkan ke dalam kluster yang sama, sehingga memudahkan pengorganisasian dan analisis kumpulan data yang kompleks [8]. Metode ini sering digunakan untuk data besar yang proses pengelompokannya lebih cepat dibandingkan dengan metode hierarki [9]. Dengan algoritma ini, dilakukan pengelompokan berbagai wilayah di Indonesia berdasarkan tingkat kriminalitas. Ini tidak hanya membantu masyarakat untuk tetap waspada terhadap kriminalitas atau kejahatan, tetapi juga dapat mengevaluasi pemerintah dan memengaruhi kebijakan yang akan diterapkan oleh pemerintah [10].

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengkategorikan wilayah-wilayah di Indonesia yang menunjukkan kerentanan terhadap aktivitas kriminal. Pengelompokan ini akan membantu pemerintah dan pihak keamanan dalam merancang strategi penanganan kriminalitas berdasarkan tingkat kerawanan di setiap daerah [11]. Analisis pola kejahatan melalui pengelompokan sistematis menawarkan wawasan berharga bagi lembaga pemerintah dan pasukan keamanan, yang memungkinkan mereka merancang tanggapan strategis yang disesuaikan dengan tingkat kerentanan kejahatan tertentu yang ada di wilayah geografis yang berbeda [12].

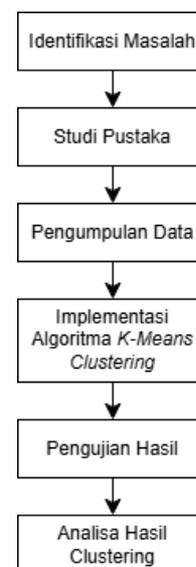
Penelitian terkait dilakukan oleh Mardhani Dwi Novianto dan Sri Andayani dalam penelitiannya yang mengelompokkan provinsi berdasarkan indikator kriminalitas nasional dengan metode SOM. Pengukuran kualitas kluster dilakukan menggunakan Davies-Bouldin Index untuk menilai seberapa baik kluster yang dihasilkan [13].

Sedangkan Jumadil, dkk mencoba mengidentifikasi daerah-daerah rawan kriminalitas di Sulawesi Tenggara menggunakan metode K-Means Clustering. Metode ini merupakan teknik data mining tanpa pengawasan (*unsupervised*) yang mengelompokkan data berdasarkan kesamaan karakteristiknya. Data yang digunakan diperoleh dari Badan Pusat Statistik (BPS) Sulawesi Tenggara, mencakup jumlah kejahatan dan persentase korban kejahatan di berbagai wilayah [1].

Penelitian ini memanfaatkan data kriminalitas terbaru dan mempertimbangkan berbagai faktor yang memengaruhi kriminalitas di setiap wilayah [12]. Diharapkan hasil pengelompokan ini bisa memberikan rekomendasi kepada pemerintah agar lebih mudah menentukan prioritas dan kebijakan dalam penanganan kriminalitas yang efektif dan efisien.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi kuantitatif dengan pendekatan data mining untuk meneliti provinsi-provinsi di Indonesia yang rawan terhadap tindak pidana, dengan menggunakan algoritma K-Means Clustering untuk analisis. Tahapan penelitian digambarkan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

2.1 Identifikasi Masalah

Pada tahap awal penelitian, fokusnya adalah mengidentifikasi isu-isu yang relevan dan melakukan tinjauan menyeluruh terhadap literatur yang ada,

termasuk jurnal ilmiah, artikel, sumber daring, buku, dan publikasi ilmiah. Tahap ini penting untuk mengumpulkan informasi relevan yang berkaitan dengan Data Mining, Klasifikasi, Algoritma K-Means, dan topik lainnya. Tujuan merumuskan masalah penelitian adalah untuk memastikan bahwa penelitian tetap fokus dan tidak menyimpang dari parameter penyelidikan yang ditentukan [14].

2.2 Pengumpulan Data

Data pada penelitian ini diperoleh dari Publikasi Statistik Kriminal 10 tahun terakhir yaitu pada tahun 2014-2023 yang diterbitkan oleh Bada Pusat Statistik (BPS) guna mengumpulkan data data jumlah tindak kriminal berdasarkan jenis kejahatan di setiap Kepolisian Daerah (Polda).

Penggunaan data yang mencakup hampir satu dekade ini memberikan gambaran yang komprehensif tentang tren kriminalitas di berbagai wilayah Indonesia. Pengambilan data dari BPS memastikan keakuratan dan validitas informasi, mengingat BPS adalah lembaga resmi pemerintah yang bertugas mengelola dan menyajikan data statistik nasional. Dataset ini sangat relevan untuk analisis tren jangka panjang serta perencanaan strategis dalam upaya pencegahan dan penanggulangan kriminalitas.

2.3 Implementasi Metode K-Means Clustering

Pengelompokan K-Means merupakan algoritma pengelompokan yang membagi himpunan data ke dalam kelompok-kelompok berbeda, sehingga titik-titik data dalam setiap kelompok memperlihatkan karakteristik yang serupa, sementara secara bersamaan memperlihatkan perbedaan dari titik-titik data di kelompok lain. 1. Pada implementasinya nanti akan menggunakan Python. Python adalah bahasa skrip berorientasi objek yang banyak digunakan untuk pengembangan perangkat lunak dan kompatibel dengan berbagai sistem operasi. Dalam konteks kontemporer, Python telah muncul sebagai bahasa yang menonjol di bidang ilmu data dan analisis data. Tren ini dapat dikaitkan dengan dukungannya yang kuat terhadap berbagai pustaka yang memfasilitasi analisis data, pembelajaran mesin, praproses data, dan visualisasi data [16]. Scikit-learn adalah pustaka sumber terbuka untuk analisis data yang mencakup berbagai algoritme penambangan data. Pustaka ini mencakup metodologi untuk klasifikasi, regresi, dan pengelompokan, serta berbagai teknik lainnya [7]. Langkah-langkah algoritma K-Means Clustering [15]:

1. Pemilihan acak sejumlah k titik data digunakan sebagai pusat kluster.
2. Jarak antara data dan pusat kluster dihitung menggunakan jarak Euclidean. Untuk menghitung jarak semua data ke setiap pusat kluster, kita dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

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

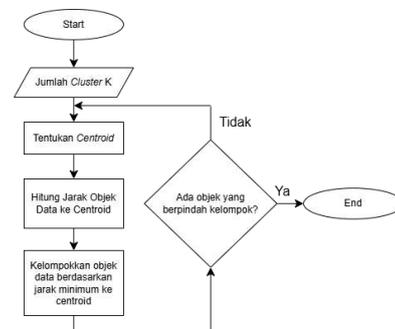
dimana, misalkan x_i adalah objek ke-i dalam himpunan objek x , y_i melambangkan objek ke-i di dalam himpunan objek y dan n melambangkan jumlah total objek yang sedang dipertimbangkan.

3. Titik-titik data ditetapkan ke kluster terdekat berdasarkan kedekatannya dengan titik pusat kluster tersebut.
4. Bila semua data telah ditetapkan ke kluster terdekat, maka akan ditentukan titik pusat kluster yang baru. Rumus perhitungan untuk titik pusat kluster yang baru adalah:

$$v = \sum_{i=1}^n x_i ; i = 1, 2, 3, \dots, n \dots\dots\dots(3)$$

Dimana, misalkan v melambangkan titik pusat kluster, X_i melambangkan objek ke- i , dan n melambangkan jumlah total objek yang termasuk dalam *cluster*.

5. Proses berulang dalam mengidentifikasi sentroid kluster dan menetapkan titik data ke klasternya masing-masing berlanjut hingga tidak ada lagi perubahan pada nilai sentroid. Diagram alir dari langkah langkah proses *K-Means Clustering* dapat dilihat pada gambar 2.



Gambar 2. Flowchart K-Means Clustering

2.4 Pengujian Hasil

Selama fase pengujian, dua metodologi digunakan: evaluasi manual terhadap hasil kluster dan pemanfaatan pemrograman Python. Selain itu, proses evaluasi menggabungkan Skor Silhouette, metrik statistik yang dirancang untuk menilai jumlah kluster optimal dalam kumpulan data tertentu [16]. *Skor Silhouette mengukur kohesi dan pemisahan kluster dengan menilai jarak setiap titik data dalam kluster relatif terhadap kluster lainnya. Metrik ini diperoleh dengan merata-ratakan jarak total, sehingga memberikan ukuran keseluruhan efektivitas pengelompokan* [17].

$$Silhouette\ Score = SI_k = \frac{1}{n} \sum_{i=1}^n \frac{(b_i - a_i)}{\max(a_i, b_i)} \dots\dots\dots(4)$$

2.5 Analisa Hasil Clustering

Analisis hasil dilakukan untuk menginterpretasikan temuan yang diperoleh dari data aktual. Hasil penilaian ini memberikan dasar bagi para pengambil keputusan, memfasilitasi pilihan yang tepat terkait kategorisasi area berdasarkan prevalensi aktivitas kriminal.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Dataset ini mencakup statistik kejahatan yang dikategorikan menurut provinsi untuk tahun 2014-2023, yang mencakup total 34 provinsi. Kriteria klasifikasi meliputi kejahatan terhadap nyawa, kejahatan terhadap integritas fisik, kejahatan terhadap moralitas, kejahatan terhadap harta benda, tindak pidana narkoba, serta penipuan dan korupsi. Berikut ini data jumlah tindakan kriminalitas dari tahun 2014-2023 untuk provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau dan Jambi yang disajikan pada tabel 1-10.

Tabel 1. Tindakan Kriminalitas 2014

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	37	48	89	22	18	74	82	91
	6	0			2	8	0	5
sumatera utara	34	60	19	13	84	70	21	22
	05		0	0	3	14	06	39
sumatera barat	74	38	70	27	41	21	31	97
	3	7			5	39	6	3
riau	56	26	39	31	38	12	10	51
	7	5			4	59	37	4
jambi	29	84	41	2	20	79	23	27
	9				3	6	8	6

Tabel 2. Tindakan Kriminalitas 2015

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	21	18	18	62	16	33	68	11
		53	8		1	23	8	07
sumatera utara	94	63	24	17	12	16	27	45
		37	0	2	07	11	32	31
						9		
sumatera barat	12	28	35	69	43	82	30	22
		34	0		7	53	5	07
riau	30	15	27	15	45	40	94	14
		02	7		3	62	3	69
jambi	22	92	80	13	23	34	27	14
		4			4	00	0	29

Tabel 3. Tindakan Kriminalitas 2016

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	40	20	23	64	49	37	94	91
		75	5			28	8	1
sumatera utara	11	55	15	10	15	13	47	50
	8	82	4	5	91	78	11	23
						0		
sumatera barat	18	33	36	12	52	83	97	27
		46	9	7	0	53	4	42
riau	49	13	28	4	50	37	12	12
		86	4		3	47	94	67
jambi	23	11	95	23	38	45	37	20
		67			7	72	1	83

Tabel 4. Tindakan Kriminalitas 2017

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	43	2.0	29	60	19	4.0	14	1.3
		03	1		1	13	02	78
sumatera utara	16	5.7	18	21	1.1	13.	5.5	5.1
	1	02	6	2	60	52	91	33
						6		
sumatera barat	21	2.7	34	14	53	8.2	72	2.0
		33	0	8	2	11	4	87
riau	20	1.1	17	3	39	3.1	1.4	1.1
		99	3		1	29	51	17
jambi	27	1.1	11	82	37	3.8	46	1.7
		43	1		5	39	2	90

Tabel 5. Tindakan Kriminalitas 2018

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	27	1.7	31	38	21	3.2	1.5	1.3
		64	1		9	38	44	38
sumatera utara	92	5.6	38	19	90	12.	5.9	5.2
		33	4		8	36	07	97
						1		
sumatera barat	14	2.3	34	15	46	7.5	59	1.8
		71	3	8	3	11	2	81
riau	28	79	97	11	26	2.3	1.1	96
		1			3	89	73	6
jambi	24	1.0	11	22	27	3.9	71	1.8
		15	3	5	4	13	2	65

Tabel 6. Tindakan Kriminalitas 2019

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	22	1.4	26	41	15	2.9	1.6	1.4
		75	2		6	59	25	71
sumatera utara	10	5.2	32	88	70	10.	6.2	4.6
	7	40	1		2	81	62	8
						8		
sumatera barat	14	2.1	39	13	38	7.2	80	1.8
		53	2	2	7	30	6	47
riau	30	85	98	16	30	2.3	1.7	1.0
		2			1	33	17	60
jambi	23	57	90	53	11	2.2	43	95
		1			9	83	9	9

Tabel 7. Tindakan Kriminalitas 2020

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	26	12	26	28	14	25	15	11
		56	7		0	04	82	20
sumatera utara	10	48	21	10	68	86	62	49
	4	17	6	5	5	93	01	04
sumatera barat	16	19	33	10	27	62	80	13
		88	3	4	7	78	6	52
riau	19	91	11	8	25	19	16	98
		8	8		4	68	71	7
jambi	15	73	69	31	16	28	40	12
		9			0	70	7	06

Tabel 8. Tindakan Kriminalitas 2021

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	15	12	23	29	14	24	15	11
		06	2		5	08	96	30
sumatera utara	99	62	77	55	78	10	59	55
		07	4		0	91	32	62
						6		
sumatera barat	20	17	29	97	18	37	91	98
		24	4		5	32	3	9
riau	21	10	14	11	41	31	16	10
		13	3		3	52	46	95
jambi	11	53	69	9	87	17	61	76
		1				70	9	7

Tabel 9. Tindakan Kriminalitas 2022

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	29	10	20	23	10	23	12	10
		74	4		7	80	08	23
sumatera utara	96	42	90	86	54	12	59	53
		87	4		0	43	49	96
						3		
sumatera barat	10	12	34	48	98	19	99	73
		99	0			71	4	9
riau	24	72	41	5	25	26	16	94
		8			5	22	70	6
jambi	15	39	46	3	73	15	36	59
		0				45	1	7

Tabel 10. Tindakan Kriminalitas 2023

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	24	11	21	51	57	26	12	10
		78	7			59	07	12
sumatera utara	84	59	45	12	63	16	41	53
		34	3	8	6	34	62	76
						7		
sumatera barat	8	12	15	30	52	15	92	55
		98	7			38	2	1
riau	36	10	17	46	32	49	14	12
		23	6		2	23	70	14
jambi	15	58	67	27	76	16	62	46
		2				42	6	9

Berikut adalah keterangan kode kriminalitas yang ditandai dengan A1-18 dan deskripsi kejahatan dari setiap klasifikasi kejahatan yang disajikan pada tabel 11.

Tabel 11. Keterangan kode Kriminalitas dan Klasifikasi Kejahatan

Kode	Klasifikasi Kejahatan	Jenis Kejahatan
A1	Kejahatan terhadap nyawa	Pembunuhan
A2	Kejahatan terhadap Fisik/Badan	Penganiayaan Berat Penganiayaan Ringan Kekerasan dalam Rumah Tangga
A3	Kejahatan terhadap Kesusilaan	Perkosaan Pencabulan
A4	Kejahatan terhadap Kemerdekaan Orang	Penculikan Mempekerjakan Anak di Bawah Umur
A5	Kejahatan terhadap Hak Milik/Barang dengan Penggunaan Kekerasan	Pencurian dengan Kekerasan Pencurian dengan Kekerasan Menggunakan Senjata Api (Senpi) Pencurian dengan Kekerasan Menggunakan Senjata Tajam (Sajam)
A6	Kejahatan Terhadap Hak Milik/Barang	Pencurian Pencurian dengan pemberatan Pencurian Kendaraan Bermotor Pengrusakan/Penghancuran Barang Pembakaran dengan Sengaja Penadahan
A7	Kejahatan Terkait Narkotika	Narkotika dan Psicotropika
A8	Kejahatan terkait Penipuan,	Penipuan/Perbuatan Curang Penggelapan

Penggelapan, dan Korupsi	Korupsi
--------------------------	---------

3.2 Jumlah Cluster (K)

Dalam penelitian ini, jumlah kluster yang ditetapkan adalah 2. Diasumsikan akan terbentuk dua kelompok yang memiliki karakteristik sangat berbeda satu sama lain dalam tingkatan kriminalitas dimana Cluster 1 adalah daerah dengan tingkat kriminalitas tinggi dan Cluster 2 adalah daerah dengan tingkat kriminalitas rendah.

Pemilihan ini didukung oleh metode evaluasi Elbow, yang menunjukkan bahwa k=2 memberikan hasil optimal dengan perbedaan yang jelas antar kelompok. Selain itu, dua kluster memudahkan interpretasi data dan penerapan kebijakan yang lebih sederhana, seperti fokus pengamanan intensif untuk wilayah berisiko tinggi dan pencegahan di wilayah dengan risiko rendah.

3.3 Centroid Data

Centroid data diambil secara acak untuk Pusat cluser 1 (C₁) diambil dari provinsi Jawa Barat dan cluser 2 (C₂) diambil dari provinsi Banten.

Tabel 12. Hasil Perhitungan Jarak Data

Provinsi	DC1	DC2
aceh	15438,28149	4779,340213
sumatera utara	22460,33429	35686,61043
sumatera barat	11873,94896	11290,93447
riau	14071,09971	6009,403573
jambi	15849,74987	3674,261577
sumatera selatan	10179,94287	11969,30166
bengkulu	19442,31504	2251,482579
lampung	16684,34085	5642,433533
kep. Bangka belitung	20991,93216	3844,34267
kepulauan riau	18948,49087	1751,664082
metro jaya	15152,04082	25715,64757
jawa barat	0	17908,39133
jawa tengah	11678,44509	8458,87229
di yogyakarta	17374,65733	2366,897152
jawa timur	14966,62629	18983,68885
banten	17908,39133	0
bali	19323,21042	2457,233252
nusa tenggara barat	18957,46665	2370,682873
nusa tenggara timur	18554,40764	4621,869704
kalimantan barat	17388,02107	1907,512921
kalimantan tengah	20334,60217	3279,559418
kalimantan selatan	19201,08198	3348,713811
kalimantan timur	16932,73889	3727,176407
kalimantan utara	22056,17569	4814,363543
sulawesi utara	18616,71316	6052,548487
sulawesi tengah	16098,79991	3876,378691
sulawesi selatan	11468,5924	14838,81828
sulawesi tenggara	19714,1334	3364,949911
gorontalo	20276,47015	3820,314888
sulawesi barat	21549,11514	4439,411352
maluku	21042,70026	4185,123145
maluku utara	21697,71942	4817,780053
papua	16416,39013	5233,956794
papua barat	20711,44107	3953,471599

3.4 Menghitung Jarak Terdekat Centroid

Rumus jarak Euclidean digunakan untuk menghitung jarak suatu objek ke pusat setiap centroid

guna menemukan jarak terdekat ke centroid setiap data. Setelah menyelesaikan perhitungan jarak antara titik data dan pusat kluster, seperti yang diuraikan dalam metodologi sebelumnya, nilai jarak yang dihasilkan disajikan dalam Tabel 12.

3.5 Mengelompokkan Data Kedalam Cluster

Setelah selesai melakukan perhitungan jarak dari setiap provinsi, maka selanjutnya dikelompokkan kedalam cluster yang disajikan pada tabel 13.

Tabel 13. Pengelompokan Data ke Centroid

Provinsi	Cluster
aceh	C2
sumatera utara	C1
sumatera barat	C2
riau	C2
jambi	C2
sumatera selatan	C1
bengkulu	C2
lampung	C2
kep. Bangka belitung	C2
kepulauan riau	C2
metro jaya	C1
jawa barat	C1
jawa tengah	C2
di yogyakarta	C2
jawa timur	C1
banten	C2
bali	C2
nusa tenggara barat	C2
nusa tenggara timur	C2
kalimantan barat	C2
kalimantan tengah	C2
kalimantan selatan	C2
kalimantan timur	C2
kalimantan utara	C2
sulawesi utara	C2
sulawesi tengah	C2
sulawesi selatan	C1
sulawesi tenggara	C2
gorontalo	C2
sulawesi barat	C2
maluku	C2
maluku utara	C2
papua	C2
papua barat	C2

Hasil klasterisasi dengan metode K-Means Clustering pada data daerah kriminalitas di Indonesia memberikan pengelompokan provinsi berdasarkan pola atau karakteristik tertentu terkait dengan tingkat kriminalitas.

Hasil klasterisasi menunjukkan pengelompokan provinsi di Indonesia berdasarkan karakteristik kriminalitas yang serupa. Cluster 1 mencakup provinsi seperti Sumatera Utara, Jawa Barat, dan Jawa Timur, yang kemungkinan memiliki tingkat kriminalitas lebih tinggi atau pola kejahatan khas urbanisasi dan ekonomi. Sementara itu, Cluster 2 mencakup sebagian besar provinsi lain, seperti Aceh, Bali, dan Papua, yang cenderung memiliki tingkat kriminalitas lebih rendah atau pola kejahatan terkait wilayah pedesaan dan tradisional. Informasi ini membantu memahami kesamaan antar provinsi

dalam satu kluster, seperti tingkat urbanisasi, kepadatan penduduk, atau jenis kejahatan dominan. Data ini bermanfaat untuk merancang kebijakan keamanan yang lebih spesifik dan efisien berdasarkan kebutuhan masing-masing wilayah.

3.6 Pengujian Hasil

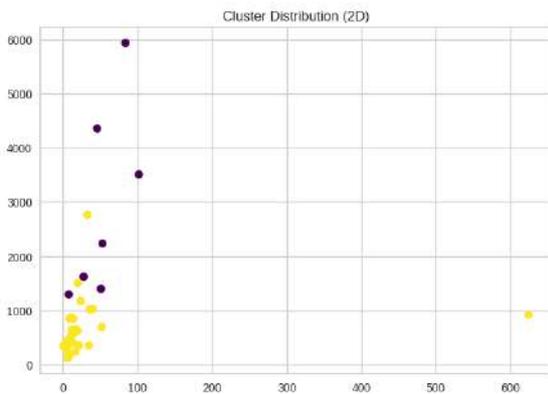
Berikut adalah hasil yang didapatkan dari analisis klasterisasi menggunakan Python yang disajikan pada tabel 14.

Tabel 14. Pengelompokan Cluster pada analisis menggunakan Python

Provinsi	Cluster
aceh	C2
sumatera utara	C1
sumatera barat	C1
riau	C2
jambi	C2
sumatera selatan	C1
bengkulu	C2
lampung	C2
kep. Bangka belitung	C2
kepulauan riau	C2
metro jaya	C1
jawa barat	C1
jawa tengah	C2
di yogyakarta	C2
jawa timur	C1
banten	C2
bali	C2
nusa tenggara barat	C2
nusa tenggara timur	C2
kalimantan barat	C2
kalimantan tengah	C2
kalimantan selatan	C2
kalimantan timur	C2
kalimantan utara	C2
sulawesi utara	C2
sulawesi tengah	C2
sulawesi selatan	C1
sulawesi tenggara	C2
gorontalo	C2
sulawesi barat	C2
maluku	C2
maluku utara	C2
papua	C2
papua barat	C2

Hasil dari pengelompokan ini menunjukkan bahwa mayoritas provinsi dikelompokkan dengan benar sesuai dengan pengelompokan manual, namun ada perbedaan dalam pengelompokan provinsi Sumatera Barat yang masuk dalam cluster C1 di Python, sedangkan dalam pengelompokan manual dikelompokkan dalam C2. Perbedaan ini bisa disebabkan oleh perbedaan dalam metode atau parameter yang digunakan dalam algoritma *clustering* Python, serta perbedaan cara centroid dihitung dan dipilih.

Visualisasi distribusi cluster dapat dilihat pada gambar 3. Titik berwarna kuning merupakan titik-titik data yang masuk ke dalam cluster C1 dan titik berwarna ungu merupakan titik-titik data yang masuk ke dalam cluster C2.



Gambar 3. Visualisasi Distribusi Cluster

Dalam pengujian hasil dilakukan pengujian Silhouette Score untuk mengevaluasi kualitas clustering. Hasilnya didapatkan Silhouette Score tertinggi untuk jumlah cluster terbaik adalah pada $k = 2$, dengan perolehan score 0,593. Hasil tersebut menunjukkan bahwa hasil clustering cukup baik. Sebagian besar objek dalam cluster berada cukup dekat dengan cluster mereka sendiri dan terpisah dari cluster lainnya.

3.7 Analisa Hasil Clustering

Dari data di atas didapatkan hasil clustering yaitu provinsi yang masuk ke dalam Cluster 1 yaitu Sumatera Utara, Sumatera Barat, Sumatera Selatan, Metro Jaya (Jakarta), Jawa Barat dan Sulawesi Selatan. Provinsi-provinsi pada cluster ini memiliki karakteristik yaitu dengan tingkat kejahatan cenderung jauh lebih tinggi dibandingkan Cluster C2. Kejahatan terhadap nyawa, fisik, dan penipuan memiliki angka rata-rata yang signifikan, dengan dominasi pada kejahatan ekonomi seperti penipuan dan barang tanpa kekerasan. Hal ini mencerminkan kompleksitas ekonomi di daerah urban yang memberikan lebih banyak peluang bagi tindak kejahatan. Selain itu, kejahatan narkoba juga menjadi salah satu isu besar, mencerminkan jaringan distribusi narkoba yang lebih kuat di kawasan ini. Tantangan sosial khas wilayah perkotaan, seperti ketimpangan sosial, kemiskinan, dan tekanan kehidupan kota, turut berkontribusi pada tingginya tingkat kriminalitas di cluster ini.

Sedangkan, sisanya merupakan bagian dari Cluster 2 di antaranya Aceh, Riau, Jambi, dan 24 provinsi lainnya. Provinsi-provinsi pada cluster ini memiliki karakteristik di antaranya dengan tingkat kejahatan lebih rendah, mencerminkan kondisi sosial yang lebih stabil dengan tekanan kriminalitas yang lebih ringan. Kejahatan barang tanpa kekerasan menjadi kategori yang paling dominan, tetapi jumlahnya masih jauh lebih rendah dibandingkan Cluster C1. Interaksi sosial yang lebih erat, pengawasan komunitas yang kuat, dan aktivitas ekonomi yang lebih sederhana menjadi faktor penting dalam menekan angka kejahatan. Selain itu, distribusi

narkoba di wilayah Cluster C2 juga lebih terbatas, sehingga kasus-kasus terkait narkoba jarang ditemukan dibandingkan dengan Cluster C1. Secara keseluruhan, Cluster C2 mencerminkan wilayah dengan dinamika sosial yang lebih harmonis dan tantangan kriminalitas yang lebih ringan.

4. KESIMPULAN

Metode *K-Means Clustering* berhasil diterapkan untuk mengelompokkan pola kriminalitas di Indonesia dengan mengelompokkan provinsi-provinsi ke dalam 2 cluster berdasarkan kesamaan karakteristik dan tingkat kejahatan. Dengan perolehan *Silhouette Score* yaitu 0.593 yang menunjukkan bahwa cluster sudah cukup baik.

Hasil penelitian ini dapat menjadi dasar bagi pihak kepolisian, pemerintah dan lainnya yang berwenang dalam pengambilan kebijakan agar lebih terfokus dan efisien yang disesuaikan dengan karakteristik setiap cluster.

DAFTAR PUSTAKA

- [1] Muh. A. Z. R. Ramadhan, R. A. Saputra, and J. Nangi, "Klasterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode K-Means Clustering," *JOINTER - Journal of Informatics Engineering*, vol. 05, no. 01, pp. 1–6, 2024.
- [2] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, Jan. 2020, doi: 10.30865/json.v1i2.1955.
- [3] J. Inayah, D. A. S. N. Maghfiroh, and D. C. R. Novitasari, "Clustering Daerah Rawan Kriminalitas Menggunakan Algoritma Fuzzy C-Means," *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 2, pp. 95–106, 2022, doi: 10.35760/ik.2022.v27i2.6019.
- [4] A. A. Dunda, A. Purwoto, A. Choirunnisa, R. A. Falah, and R. Y. Putri, "Faktor-Faktor yang Mempengaruhi Jumlah Tindak Kriminal di Sumatera Utara Tahun 2022," *Madani: Jurnal Ilmiah Multidisipliner*, vol. 2, no. 3, pp. 543–549, 2024, doi: 10.5281/zenodo.12171133.
- [5] T. Palokoto, E. Y. Purwanti, and Y. B. Mudakir, "Analisis Konsekuensi Ekonomi Akibat Kejahatan di Indonesia," *Jurnal Ekonomi dan Pembangunan Indonesia*, vol. 20, no. 2, pp. 146–159, Jul. 2020, doi: 10.21002/jepi.2020.09.
- [6] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap)," *Just IT: Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 12, no. 2, pp. 51–61, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [7] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering terhadap Data Pembayaran Transaksi

- Menggunakan Bahasa Pemrograman Python pada CV Digital Dimensi,” *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [8] L. Suriani, “Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, Jan. 2020, doi: 10.30865/json.v1i2.1955.
- [9] P. S. Rosiana, A. A. Mohsa, M. A. Fadila, and J. H. Jaman, “Visualisasi Data Kejahatan di Kabupaten Karawang Berdasarkan Jenis Kriminalitas dengan Algoritma Clustering K-Means,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 2830–7062, 2023, doi: 10.23960/jitet.v11i3%20s1.3347.
- [10] M. D. Novianto and S. Andayani, “Analisis Cluster dengan Metode SOM untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Kriminalitas,” *Jurnal Kajian dan Terapan Matematika*, vol. 9, no. 1, pp. 39–53, 2023, [Online]. Available: <https://www.numbeo.com/>
- [11] Statistik Kriminal oleh Badan Pusat Statistik 2023. 2023.
- [12] R. M. Salsabila and A. Syarief, “Analisis Spasial Tingkat Kerawanan Kriminalitas Menggunakan Metode Nearest Neighbor Analysis Dan Kernel Density Di Kota Pekanbaru,” *Jurnal Pendidikan Tambusai*, vol. 8, no. 2, pp. 28844–28850, 2024.
- [13] M. D. Novianto and S. Andayani, “Analisis Cluster dengan Metode SOM untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Kriminalitas,” *Jurnal Kajian dan Terapan Matematika*, vol. 9, no. 1, pp. 39–53, 2023, [Online]. Available: <https://www.numbeo.com/>
- [14] T. Hidayat, “Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering,” *Jurnal Sistim Informasi dan Teknologi*, vol. 4, no. 1, pp. 19–24, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [15] H. Kurniawan, S. Defit, and Sumijan, “Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal,” *Journal of Applied Computer Science and Technology (JACOST)*, vol. 1, no. 2, pp. 80–89, Dec. 2020, doi: 10.52158/jacost.v1i2.102.
- [16] D. A. S. Simamora, M. T. Furqon, and B. Priyambadha, “Clustering Data Kejadian Tsunami Yang Disebabkan Oleh Gempa Bumi Dengan Menggunakan Algoritma K-Medoids,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 8, pp. 635–640, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [17] A. H. Yusup and W. Maharani, “Model Prediksi Kepribadian Berdasarkan Tweet dan Kategori Big Five dengan Metode Agglomerative Hierarchical Clustering,” *TELKATIKA*, vol. 1, no. 1, p. 2021, 2021.

IMPLEMENTASI ALGORITMA *LATTERN DIRICHLET ALLOCATION (LDA)* DENGAN *GIBSS SAMPLING* UNTUK *TOPIC MODELLING* MENGGUNAKAN DATA X TERKAIT IKN

Raihan Labib Hanif¹, Mohammad Syafrullah^{2*}

^{1,2*} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta
Email: ¹2011501315@student.budiluhur.ac.id, ^{2*}mohammad.syafrullah@budiluhur.ac.id

(Naskah masuk: 7 Agustus 2024, diterima untuk diterbitkan: 7 September 2024)

Abstrak

Penelitian ini membahas tentang pemodelan topik terkait tweet masyarakat mengenai Ibu Kota Negara (IKN) di media sosial X. Masalah utama yang diangkat adalah kurangnya pemahaman mengenai tren topik yang sering dibahas masyarakat terkait IKN, yang dapat mempengaruhi persepsi dan opini publik. Untuk menyelesaikan masalah ini, metode Latent Dirichlet Allocation (LDA) digunakan dalam pemodelan topik, yang memungkinkan pengelompokan topik secara otomatis berdasarkan teks yang tersedia. Proses pemodelan dilakukan dengan menggunakan teknik Gibbs Sampling dan diukur menggunakan nilai coherence untuk mengevaluasi kualitas model yang dihasilkan. Tujuan dari penelitian ini adalah untuk mengidentifikasi dan mengelompokkan topik yang muncul dari tweet terkait IKN, sehingga dapat memberikan wawasan yang lebih mendalam mengenai pandangan masyarakat. Hasil penelitian menunjukkan bahwa model dengan 20 topik menghasilkan nilai coherence tertinggi sebesar 0.38389228712614887, yang menunjukkan bahwa model ini efektif dalam mengidentifikasi tren topik yang dibahas oleh masyarakat.

Kata kunci: pemodelan topik, gibss sampling, coherence score, IKN, latent dirichlet allocation (LDA)

IMPLEMENTATION OF *LATTERN DIRICHLET ALLOCATION (LDA)* ALGORITHM WITH *GIBSS SAMPLING* FOR *TOPIC MODELLING* USING X DATA RELATED TO IKN

Abstract

This research discusses topic modeling related to public tweets regarding the National Capital City (IKN) on X social media. The main problem raised is the lack of understanding of topic trends that are often discussed by the public related to the National Capital, which can affect public perception and opinion. To solve this problem, the Latent Dirichlet Allocation (LDA) method is used in topic modeling, which allows automatic clustering of topics based on available text. The modeling process is performed using the Gibbs Sampling technique and measured using the coherence value to evaluate the quality of the resulting model. The purpose of this research is to identify and categorize topics that arise from tweets related to IKN, so as to provide a deeper insight into the views of the public. The results show that the model with 20 topics produces the highest coherence value of 0.38389228712614887, which indicates that the model is effective in identifying trends in topics discussed by the public.

Keywords: topic modelling, gibss sampling, coherence score, IKN, latent dirichlet allocation (LDA)

1. PENDAHULUAN

Teknologi informasi mencakup apapun yang membantu manusia membuat, mengubah, menyimpan, mengomunikasikan, atau menyebarkan informasi. Tanpa teknologi ini, hubungan sosial tidak dapat terbentuk. Saat ini, perantara media cetak tidak lagi digunakan untuk menyampaikan informasi, tetapi diganti dengan media digital elektronik [1]. Fokus utama penelitian ini adalah teknologi teks yang dapat digunakan untuk mendapatkan topik dari kumpulan kalimat [2]. Tidak terlepas dari peran manusia yang terus menerus menciptakan dan

menyempurnakan teknologi, kemajuan teknologi terus berkembang dengan pesat [3].

Media sosial adalah *platform* yang memungkinkan individu untuk berbagi gagasan, pandangan, dan informasi melalui jaringan dan komunitas digital. X merupakan salah satu media sosial yang paling populer dan banyak digunakan oleh publik, yang menawarkan layanan jaringan sosial di mana pengguna dapat berbagi dan berinteraksi dengan pesan singkat yang dikenal sebagai "*tweet*" [4]. X Inc. mengoperasikan dan mengelola *platform* jejaring sosial berbasis microblogging yang dikenal sebagai X, yang

memungkinkan pengguna untuk mengirim dan membaca pesan singkat ke sesama pengguna [5].

IKN adalah singkatan dari Ibu Kota Negara. Dalam konteks Indonesia, IKN mengacu pada rencana pemindahan ibu kota pemerintahan dari Jakarta ke lokasi baru di Kalimantan Timur. Inisiatif ini diajukan oleh Presiden Jokowi dan disetujui oleh DPR RI melalui rapat paripurna pada 18 Januari 2022. Pemindahan ibu kota ini ke wilayah Sepaku, Penajam Paser Utara, diprediksi akan membawa dampak yang signifikan pada aspek ekonomi, sosial, dan budaya [6].

Pembahasan orang terkait IKN pada media sosial X cukup beragam, sehingga dalam penelitian ini merasa diperlukan adanya pemodelan topik untuk mengetahui topik apa yang sering dibahas oleh masyarakat.

Adapun penelitian terdahulu [7], Metode yang digunakan dalam penelitian ini adalah *Latent Dirichlet Allocation (LDA)*, yang berfungsi sebagai pemodelan topik untuk data *tweet* tentang tokoh publik politik. Penelitian ini memanfaatkan metode *Agglomerative Hierarchical Clustering* untuk mengelompokkan data *tweet* berdasarkan kemiripan, sehingga menghasilkan beberapa kluster topik. Setelah proses klustering, metode LDA diterapkan untuk mengidentifikasi distribusi topik dalam dokumen, dengan parameter yang telah diinisialisasi seperti jumlah topik, nilai α (*Dirichlet*), nilai β , dan jumlah iterasi. Hasil dari LDA kemudian diolah lebih lanjut menggunakan *TF-IDF* untuk mengetahui *tweet* mana saja yang memiliki kata-kata terbanyak dalam topik yang dihasilkan, dan menggunakan metode *Silhouette Coefficient* dalam evaluasi hasil klustering

2. METODE PENELITIAN

Dalam penelitian ini dilakukan pengambilan data dalam bentuk teks, berupa data *tweet* dari tanggal 1 Januari 2024 hingga 11 Juli 2024 yang diambil melalui situs resmi yaitu X, dengan total data sebanyak 216 *tweet* yang membahas tentang IKN.

2.1 Data Penelitian

Pada tahap pengumpulan data ini menggunakan *tweet harvest*. Pada tahapan *crawling*, kata kunci atau *keyword* yang digunakan berkaitan dengan IKN. Data yang diambil dari proses *crawling* akan disimpan dalam bentuk file CSV. Kemudian, data yang sudah menjadi format CSV akan disimpan pada *database* lokal untuk dijadikan dataset.



Gambar 1. Data Kotor

2.2 Preprocessing

Preprocessing adalah tahap awal dalam menyiapkan dataset untuk mempermudah pemrosesan data. Tujuannya adalah untuk meningkatkan kualitas data sehingga menghasilkan performa yang tinggi [8]. Tahap-tahap preprocessing yang digunakan antara lain sebagai berikut:

a) Stopword Removal

Proses penghapusan *stopword* adalah langkah untuk menghilangkan kata-kata penghubung dan kata-kata yang tidak relevan dalam dataset tersebut [9]. Langkah ini akan menghilangkan semua kata yang dianggap tidak memiliki makna signifikan, seperti kata penghubung [10].

b) Stemming

Stemming merupakan metode yang digunakan untuk menemukan bentuk dasar dari kata yang memiliki imbuhan dengan cara menghilangkan imbuhan tersebut. Proses stemming terdiri dari tiga langkah utama: pertama, mengidentifikasi dan mengumpulkan aturan-aturan yang terdapat dalam bahasa Balantak; kedua, mengumpulkan kata dasar yang akan disimpan dalam kamus data; dan ketiga, menghapus imbuhan yang ada pada kata dasar [11].

c) Case folding

Case folding adalah tahap proses yang mengubah kata menjadi bentuk seragam menggunakan metode lower pada string di Python. Tujuan dari *Case folding* adalah mengubah semua kata menjadi huruf kecil agar teks yang diproses berada dalam format yang konsisten. *Cleaning* teks adalah proses membersihkan dan memvalidasi kata-kata yang tidak diinginkan untuk mengurangi gangguan dalam proses klasifikasi. Contohnya, menghilangkan karakter tertentu [12].

2.3 Topic Modelling

Topik modeling adalah metode dalam text mining yang digunakan untuk menemukan pola dalam dokumen serta mengelompokkan kata-kata ke dalam beberapa topik yang mencerminkan konten dari dokumen tersebut. Teknik ini menggunakan pendekatan pengelompokan di mana kata-kata dalam teks dikategorikan menjadi topik-topik tertentu. Latent Dirichlet Allocation (LDA) merupakan salah satu teknik yang sering digunakan dalam topik modeling untuk mengekstrak topik-topik penting dari dokumen. LDA dikenal efektif dalam menangani data yang berukuran besar [13].

2.4 LDA dengan Gibbs Sampling

Mengaplikasikan LDA dengan *Gibbs Sampling* melibatkan kalkulasi variabel yang dibutuhkan, inialisasi acak untuk topik, serta menjalankan sejumlah iterasi yang telah ditentukan. Pada setiap iterasi, topik akan disampel untuk setiap kata dalam *corpus*. Setelah semua iterasi selesai, hasilnya digunakan untuk menghitung distribusi laten θ_d dan ϕ_k . Adapun variabel-variabel yang dibutuhkan adalah sebagai berikut:

- $n_{d_i k}$ = jumlah dokumen d pada topik k .
 $n_{k, w}$ = jumlah kata w yang diberikan ke topik k .
 n_k = jumlah total dari setiap kata yang diberikan ke topik k .
 n_d = jumlah total seluruh dokumen d dalam semua topik.

Karena prosedur pengambilan sampel *Gibbs* melibatkan pengambilan sampel dari distribusi yang dikondisikan pada semua variabel, kita harus menghapus penugasan topik saat ini sebelum membangun distribusi dari persamaan:

$$p(z_i = k | w_i, d_i, \alpha, \beta) \propto \frac{n_{d_i k} + \alpha}{\sum_{k'=1}^K n_{d_i k'} + K\alpha} \cdot \frac{n_{k w_i} + \beta}{\sum_{w=1}^W n_{k w} + W\beta} \dots (1)$$

Keterangan:

- $z_i = k$ = pemberian topik k untuk kata ke- i
 w_i = kata ke- i
 d_i = dokumen ke- i
 α = *hyperparameter* untuk prior *Dirichlet* pada distribusi topik-dokumen
 β = *hyperparameter* untuk prior *Dirichlet* pada distribusi kata-topik
 $n_{d_i k}$ = jumlah dokumen d pada topik k
 $\sum_{k'=1}^K n_{d_i k'}$ = total seluruh dokumen d dalam semua topik k
 $n_{k w_i}$ = jumlah kata w yang muncul pada topik k
 $\sum_{w=1}^W n_{k w}$ = jumlah kata pada topik k
 K = jumlah topik
 W = panjang dari kamus kata

Penghapusan dilakukan dengan mengurangi jumlah yang terkait dengan penugasan saat ini, karena urutan penugasan topik tidak mempengaruhi hasil akhir. Selanjutnya, probabilitas dari setiap penugasan topik dihitung menggunakan persamaan yang relevan. Distribusi ini kemudian diambil sampelnya, dan topik yang dipilih diatur dalam array z , serta jumlah yang sesuai ditambahkan kembali.

2.5 Nilai Coherence

Topic Coherence adalah metrik yang digunakan untuk menilai suatu topik dengan mengukur tingkat kesamaan semantik antara kata-kata dalam topik tersebut. Nilai *topic coherence* yang lebih tinggi menunjukkan bahwa topik lebih mudah diinterpretasikan oleh manusia, sehingga menghasilkan kualitas topik yang lebih baik [13].

Coherence score adalah ukuran yang digunakan untuk mengevaluasi pemodelan topik. Model yang baik akan menghasilkan topik dengan skor koherensi yang tinggi, yang menunjukkan bahwa topik-topik tersebut saling terkait dan memiliki keterkaitan yang logis dalam kumpulan dokumen yang dianalisis.

$$C = S \times M \times P \times \Sigma$$

Dimana:

- C = *Coherence Score*
 S = *Segmentation*
 M = *Confirmation Measure*
 P = *Probability Estimation*
 Σ = *Aggregation*

Rumus untuk mencari nilai S atau *Segmentation* di bawah ini

$$S_{set}^{one} = \{(w^1, w^*) | w^1 = \{w_i\}; w_i \in w; w^* = w\} \dots (2)$$

Rumus untuk mencari nilai M atau *Confirmation Measure* di bawah ini.

$$NPMI(w', w^*) = \frac{\log \frac{P(w', w^*) + \epsilon}{P(w') P(w^*)}}{-\log(P(w', w^*) + \epsilon)} \dots (3)$$

Rumus untuk mencari nilai P atau *Probability Estimation* di bawah ini.

$$P_{sw}(w_i, w_j) = \frac{\text{Number of documents in which } w_i \text{ and } w_j \text{ co-occur}}{\text{Number of documents in which } w_i \text{ appears}} \dots (4)$$

Rumus untuk mencari nilai Σ atau *Aggregation* di bawah ini.

$$\Sigma = \frac{\text{jumlah dari confirmation measure}}{\text{banyak confirmation measure}}$$

2.6 Evaluasi Model

Evaluasi model adalah proses untuk menilai kinerja dan efektivitas model yang telah dibangun dalam sebuah sistem, terutama dalam konteks pembelajaran mesin dan analisis data. Evaluasi ini bertujuan untuk memahami sejauh mana model mampu membuat prediksi yang akurat atau menghasilkan hasil yang diharapkan berdasarkan data yang digunakan.

3. HASIL PEMBAHASAN

Proses pemodelan LDA dimulai dengan tahap preprocessing untuk mempersiapkan data teks agar siap diproses. Setelah data siap, LDA diterapkan untuk memodelkan topik dengan menentukan distribusi kata dalam setiap topik dan distribusi topik dalam setiap dokumen. Proses ini melibatkan beberapa iterasi untuk menghasilkan model yang stabil.

Evaluasi model LDA dilakukan untuk menentukan seberapa baik topik yang dihasilkan mewakili struktur tersembunyi dalam data. Salah satu metrik yang sering digunakan untuk mengevaluasi kualitas pemodelan topik adalah nilai *coherence*. *Coherence* mengukur tingkat kesamaan atau keterkaitan antara kata-kata dalam topik yang dihasilkan. Nilai *coherence* yang lebih tinggi menunjukkan bahwa kata-kata dalam topik tersebut lebih sering muncul bersama dalam dokumen-

dokumen dalam corpus, sehingga topik tersebut lebih koheren dan relevan.

3.1 Hasil Preprocessing

Persamaan matematika dinomori dengan Angka Pada proses preprocessing data didalamnya terdapat beberapa fungsi yang digunakan yaitu *stopword removal*, *stemming*, *case folding*. Sehingga data yang sudah di *preprocessing* dapat memudahkan sistem algoritma *lattern dirichlet allocation* ini berjalan. Berikut adalah contoh data yang sudah di *preprocessing*.

No	Kata Bersih
1	Anda akan bertemu dengan banyak orang yang berbeda-beda
2	Anda akan bertemu dengan banyak orang yang berbeda-beda
3	Anda akan bertemu dengan banyak orang yang berbeda-beda
4	Anda akan bertemu dengan banyak orang yang berbeda-beda
5	Anda akan bertemu dengan banyak orang yang berbeda-beda
6	Anda akan bertemu dengan banyak orang yang berbeda-beda
7	Anda akan bertemu dengan banyak orang yang berbeda-beda
8	Anda akan bertemu dengan banyak orang yang berbeda-beda
9	Anda akan bertemu dengan banyak orang yang berbeda-beda
10	Anda akan bertemu dengan banyak orang yang berbeda-beda
11	Anda akan bertemu dengan banyak orang yang berbeda-beda
12	Anda akan bertemu dengan banyak orang yang berbeda-beda
13	Anda akan bertemu dengan banyak orang yang berbeda-beda
14	Anda akan bertemu dengan banyak orang yang berbeda-beda
15	Anda akan bertemu dengan banyak orang yang berbeda-beda
16	Anda akan bertemu dengan banyak orang yang berbeda-beda
17	Anda akan bertemu dengan banyak orang yang berbeda-beda
18	Anda akan bertemu dengan banyak orang yang berbeda-beda
19	Anda akan bertemu dengan banyak orang yang berbeda-beda
20	Anda akan bertemu dengan banyak orang yang berbeda-beda

Gambar 2. Data Bersih

3.2 Hasil Evaluasi Nilai Coherence

Dalam proses evaluasi, nilai coherence dihitung untuk setiap iterasi topik yang dilakukan. Hasil evaluasi ini digunakan untuk menentukan jumlah topik yang optimal dengan memilih iterasi yang memiliki nilai coherence tertinggi. Pemilihan jumlah topik yang optimal ini sangat penting karena mempengaruhi kualitas akhir dari pemodelan topik. Model LDA yang dihasilkan dengan nilai coherence tertinggi diharapkan mampu memberikan representasi topik yang lebih akurat dan informatif.

Secara keseluruhan, metode LDA dengan evaluasi nilai coherence memberikan kerangka kerja yang efektif untuk mengidentifikasi dan memahami topik-topik utama dalam kumpulan data teks yang besar. Evaluasi menggunakan nilai coherence memastikan bahwa topik-topik yang dihasilkan tidak hanya matematis signifikan tetapi juga relevan dan koheren dalam konteks data yang dianalisis

Tabel 1. Nilai Coherence

Num Topic	Coherence Score
1	0,35
2	0,30
3	0,29
4	0,33
5	0,34
6	0,32
7	0,31
8	0,34
9	0,32
10	0,32
11	0,32
12	0,32
13	0,36
14	0,34
15	0,35
16	0,38
17	0,33
18	0,37
19	0,36
20	0,38

Dari Tabel 1, terlihat bahwa iterasi topik ke-20 menghasilkan nilai coherence tertinggi. Hal ini menunjukkan bahwa dua puluh topik merupakan jumlah yang paling optimal untuk memberikan interpretasi atau pemahaman yang baik terhadap topik yang dihasilkan. Oleh karena itu, iterasi dengan dua puluh topik akan digunakan kembali dalam proses pemodelan untuk mendapatkan distribusi topik dalam dokumen serta distribusi kata dalam topik berdasarkan jumlah topik tersebut.

Dapat di gambarkan pada gambaran umum, untuk memperjelas hasil dari pemodan topic dengan menggunakan metode algoritma *lattern dirichlet allocation* dan pengujian *coherence score*, dan dipatkan nilai coherence tertinggi pada iterasi topik ke-20 dengan nilai 0.38389228712614887. Seperti yang dapat dilihat pada gambar 3.



Gambar 3. Grafik Coherence Score

3.3 Hasil Distribusi dengan LDA Gibbs Sampling

Pemodelan topik dengan menggunakan LDA dan Gibbs Sampling menghasilkan 217 kelompok distribusi topik-dokumen serta 20 kelompok distribusi kata-topik. Setiap kelompok mengandung istilah-istilah dengan bobot yang dijelaskan pada Tabel 2 dan Tabel 3. Bobot ini merepresentasikan probabilitas kemunculan istilah tersebut dalam setiap kelompok. Pada distribusi topik-dokumen, terdapat sampel dari seluruh dokumen yang mencakup kolom 'Topik', yang memuat istilah-istilah dengan bobot dari masing-masing kelompok dokumen. Sementara itu, dalam distribusi kata-topik, terlihat seluruh jumlah topik hasil evaluasi model dengan kolom 'Kata', yang memuat istilah-istilah dengan bobot dari masing-masing kelompok topik. Dapat di lihat seperti contoh tabel di bawah ini.

Tabel 2. Distribusi Topik-Dokumen

Dokumen	Topik	Bobot
Dokumen 1	Topik_16	1
Dokumen 2	Topik_7	0,6
	Topik_5	0,2
	Topik_12	0,2
Dokumen 4	Topik_1	0,5
	Topik_14	0,5

Pada tabel di bawah ini menjelaskan data terkait ditribusi kata-topik yang dimana terdapat nilai bobot pada seluruh kata yang ada di setiap topik. Seperti yang dapat kita lihat contohnya pada tabel 3.

Tabel 3. Distribusi Kata-Topik

Topik	Kata	Bobot
Topik 1	Ikn	0,43
	Proyek	0,34
	Bangun	0,09
	Bayar	0,04
	Paksa	0,04
Topik 3	Rakyat	0,02
	Rakyat	0,33
	Ikn	0,27
	Yang	0,16
	Eggak	0,11
Topik 4	Sudah	0,08
	Air	0,02
	Jabat	0,41
	Mundur	0,29
	Ya	0,20
Topik 9	Banget	0,04
	Eggak	0,04
	Jokowi	0,02
	Saja	0,29
	Ya	0,04
Topik 20	Bangun	0,04
	Jabat	0,04
	Kalo	0,04
	Ikn	0,40
	Bangun	0,23
	Rakyat	0,14
	Nusantara	0,08
	Cnnindonesia	0,06
	Yang	0,06

3.4 Hasil Pelabelan

Dalam proses evaluasi, nilai coherence dihitung untuk setiap iterasi topik yang dilakukan. Hasil evaluasi ini digunakan untuk menentukan jumlah topik yang optimal dengan memilih iterasi yang memiliki nilai coherence tertinggi. Pemilihan jumlah topik yang optimal ini sangat penting karena mempengaruhi kualitas akhir dari pemodelan topik. Model LDA yang dihasilkan dengan nilai coherence tertinggi diharapkan mampu memberikan representasi topik yang lebih akurat dan informatif. Berikut adalah contoh beberapa topik yang membahas tentang IKN di program yang sudah dibuat.

4. KESIMPULAN

Berdasarkan hasil uji coba dan analisis yang telah dilakukan, penerapan pemodelan topik dengan metode Latent Dirichlet Allocation (LDA) pada 216 data tweet tentang IKN menunjukkan hasil yang positif. Penelitian ini berhasil memberikan gambaran umum mengenai pemodelan topik terkait IKN, di mana pengujian menggunakan nilai coherence menghasilkan nilai tertinggi sebesar 0.38389228712614887 pada topik ke-20 dari 20 topik. Untuk meningkatkan efektivitas aplikasi ini, beberapa saran yang dapat dipertimbangkan meliputi penggabungan algoritma LDA dengan algoritma lain untuk mencapai hasil yang lebih optimal, penerapan confusion matrix untuk menghitung nilai accuracy dalam pemodelan topik, serta pengujian terlebih dahulu dalam pembobotan kata untuk menentukan apakah bag-of-words atau tf-idf yang lebih baik

digunakan. Kesimpulan ini sejalan dengan tujuan dan rumusan masalah yang diangkat dalam penelitian.

Tabel 4. Hasil Pelabelan Distribusi Kata-Topik

Topik	Kata	Jumlah
Proyek Infrastruktur	Ikn	11
	Proyek	
	Bangun	
	Bayar	
	Paksa	
Kebijakan Pemerintah	Kepala	13
	Otorita	
	Mundur	
	Ikn	
	Bambang	
Pendanaan Anggaran	Dana	6
	Eggak	
	Bangun	
	Sudah	
	Perintah	
Ekonomi Investasi	Ikn	18
	Bangun	
	Rakyat	
	Nusantara	
	Cnnindonesia	
Pendanaan Anggaran dan Perumahan Pemukiman	Ikn	6
	Indonesia	
	Otorita	
	Bayar	
	Tapera	

DAFTAR PUSTAKA

- [1] A. M. Dawis *et al.*, "Digital Signage As Interactive Information Of Nurul Iman Mosque Based On Responsive Web," vol. 1, pp. 8–16, 2023.
- [2] F. Nisrina, "Implementasi Deteksi Topik Putusan Hakim Dengan Latent Dirichlet Allocation (Lda)," pp. 1–64, 2020, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/23847>
- [3] A. Anas Azhar, "Analisis Pemanfaatan Aplikasi Gojek Berbasis Sistem Informasi Geografis (Persepsi Pelayanan Gojek Pada Mahasiswa Uinsu)," *J. Geogr.*, vol. 12, no. 02, p. 96, 2020, doi: 10.24114/jg.v12i02.17214.
- [4] Y. Afandi, "Gereja dan Pengaruh Teknologi Informasi 'Digital Ecclesiology,'" *Fidei J. Teol. Sist. dan Prakt.*, vol. 1, no. 2, pp. 270–283, 2019, doi: 10.34081/270033.
- [5] D. Z. T. Kannitha, M. Mustafid, and P. Kartikasari, "Pemodelan Topik Pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation Dalam Media Sosial Twitter," *J. Gaussian*, vol. 11, no. 2, pp. 266–277, 2022, doi: 10.14710/j.gauss.v11i2.35474.
- [6] Ripaldi, E. M. Lantigimo, I. Sasirais, and G. T. Yohannes, "Kesiapan Dayak Deah Di Desa Kinarum Dalam Menyambut Ibu Kota Negara Baru," *J. Teol. Pabelum*, vol. 2, no. 2, pp. 34–48, 2022, doi: 10.59002/jtp.v2i2.27.
- [7] F. N. Hikmah, S. Basuki, and Y. Azhar, "Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation (LDA)," *J. Repos.*, vol. 2, no. 4, pp. 415–426, 2020, doi: 10.22219/repositor.v2i4.52.
- [8] E. Laksono, A. Basuki, and F. Bachtiar, "Optimization of K Value in KNN Algorithm for Spam and Ham Email Classification," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no.

- 2, pp. 377–383, 2020, doi: 10.29207/resti.v4i2.1845.
- [9] T. A. Sari, E. Sinduningrum, and F. Noor Hasan, “KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes,” *Media Online*, vol. 3, no. 6, pp. 773–779, 2023, doi: 10.30865/klik.v3i6.884.
- [10] N. Nurzaman, N. Suarna, and W. Prihartono, “Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.
- [11] N. H. Hrp, M. Fikry, and Y. Yusra, “Algoritma Stemming Teks Bahasa Batak Angkola Berbasis Aturan Tata Bahasa,” *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 642–648, 2023, doi: 10.47065/josyc.v4i3.3458.
- [12] M. U. Albab, Y. Karuniawati, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” *J. Transform.*, vol. 20, no. 2, pp. 1–10, 2023, [Online]. Available: <https://journals.usm.ac.id/index.php/transformatika/page1>
- [13] S. Roiqoh, B. Zaman, and K. Kartono, “Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1582–1592, 2023, doi: 10.30865/mib.v7i3.6194.

DETEKSI WEBSITE PHISHING DARI ANALISIS URL MENGUNAKAN ALGORITMA RANDOM FOREST

Damar Bambang Suwarno^{1*}, Mardi Hardjianto²

^{1*,2} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Kota Jakarta Selatan, Indonesia
Email: ^{1*}sdamar321@gmail.com, ²mardi.hardjianto@budiluhur.ac.id

(Naskah masuk: 7 Agustus 2024, diterima untuk diterbitkan: 7 September 2024)

Abstrak

Serangan digital, atau yang lebih dikenal sebagai *cybercrime*, semakin sering terjadi di era modern ini, seiring dengan pesatnya perkembangan teknologi. Metode serangan yang paling sering dipakai adalah *phishing*, yang pertama kali muncul pada tahun 1996. *Phishing* merupakan salah satu bentuk kejahatan siber di mana penyerang berusaha mengelabui pengguna agar secara tidak sadar memberikan informasi sensitif seperti *username*, *password*, atau data keuangan. Berbagai penelitian telah dilakukan untuk mencari solusi dalam menangani serangan *phishing*, mulai dari penggunaan *tools* keamanan seperti menggunakan OS *Kali Linux* hingga mengedukasi para pegawai agar lebih waspada dan menggunakan alat bantu keamanan seperti antivirus. Salah satu solusi yang semakin relevan adalah penggunaan kecerdasan buatan (*Artificial Intelligence*) untuk secara otomatis mengidentifikasi apakah sebuah URL yang diakses aman atau berpotensi berbahaya. Penelitian ini bertujuan untuk mengembangkan aplikasi berbasis web yang mampu mendeteksi serangan *phishing* dengan menggunakan teknik *machine learning*, serta memanfaatkan *dataset* yang cukup besar dan representatif untuk melatih model deteksi URL *phishing*. Dalam penelitian ini kontribusi yang dilakukan ialah menggunakan algoritma *random forest* dalam pendeteksian *website phishing* dan penambahan fitur deteksi yang diintegrasikan pada *website* yang membahas *phishing*, algoritma klasifikasi *Random Forest* digunakan karena kemampuannya yang tinggi dalam memproses sejumlah besar fitur deteksi. Dengan menggunakan 30 fitur deteksi, hasil pengujian menunjukkan bahwa sistem yang dibangun mampu mencapai kinerja yang optimal, dengan tingkat prediksi sebesar 96%, *Recall* 92%, Akurasi 94%, dan *F-Score* 93%. Hasil ini menunjukkan bahwa metode yang diusulkan efektif dalam mendeteksi serangan *phishing* dengan tingkat akurasi yang tinggi, menjadikannya alat yang sangat berguna dalam mencegah pengguna dari ancaman siber dan dinilai dapat menyelesaikan permasalahan yang ada karena dapat bekerja dengan optimal.

Kata kunci: *cybercrime*, *phishing*, *random forest*, *machine learning*

PHISHING WEBSITE DETECTION FROM URL ANALYSIS USING RANDOM FOREST ALGORITHM

Abstract

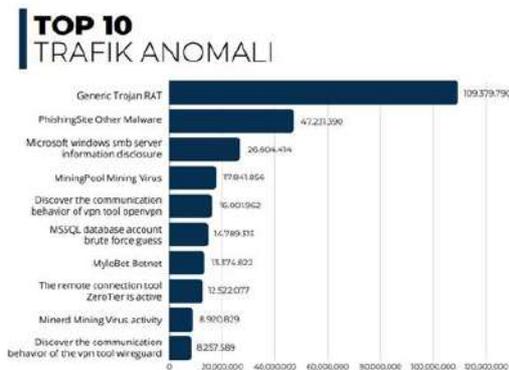
Digital attacks, or what is better known as *cybercrime*, are increasingly occurring in this modern era, along with the rapid development of technology. The most frequently used attack method is *phishing*, which first appeared in 1996. *Phishing* is a form of *cybercrime* in which attackers attempt to trick users into unknowingly providing sensitive information such as *usernames*, *passwords*, or financial data. Various studies have been carried out to find solutions for dealing with *phishing* attacks, starting from using security tools such as using the *Kali Linux* OS to educating employees to be more alert and use security tools such as anti-virus. One solution that is increasingly relevant is the use of artificial intelligence to automatically identify whether a URL to be accessed is safe or potentially dangerous. This research aims to develop a web-based application that is capable of detecting *phishing* attacks using machine learning techniques, as well as utilizing a fairly large and representative dataset to train a *phishing* URL detection model. In this research, the *Random Forest* classification algorithm is used because of its high ability to process a large number of detection features. By using 30 detection features, the test results show that the system built is able to achieve optimal performance, with a prediction level of 96%, *Recall* 92%, *Accuracy* 94%, and *F-Score* 93%. These results show that the proposed method is effective in detecting *phishing* attacks with a high degree of accuracy, making it a very useful tool in protecting users from cyber threats.

Keywords: *cybercrime*, *phishing*, *random forest*, *machine learning*

1. PENDAHULUAN

Internet adalah kepanjangan dari istilah "Interconnected Network", yang berarti hubungan komputer dengan berbagai jenis jaringan. Internet pada awalnya adalah proyek yang dirancang untuk kepentingan militer dari suatu negara [1]. Seiring cepatnya perkembangan internet terdapat pula kejahatan dunia maya yang terjadi atau biasa disebut *cybercrime*. *Cybercrime* adalah aktivitas kriminal yang melibatkan penggunaan komputer, perangkat digital, atau jaringan untuk melakukan tindakan ilegal [2]. Terdapat beberapa jenis serangan *cybercrime* yang umum meliputi. *Malware, ransomware, denial of service (DoS) attack, phishing*, dan lainnya.

Phishing adalah aktivitas kriminal yang menggunakan teknik rekayasa sosial [3]. *Phishing* pertama kali muncul pada tahun 1996. Jumlah anomali *phishing* yang tercatat sebanyak 47.231.390 pada tahun 2023 dari Badan Siber dan Sandi Negara Republik Indonesia Gambar 1.



Gambar 1. Top 10 trafik anomali [5]

Phishing telah menjadi masalah keamanan siber yang semakin mengkhawatirkan. Berbagai pihak telah melakukan upaya untuk mengatasi ancaman ini. Dijelaskan bahwa peneliti telah mengembangkan berbagai teknik deteksi *phishing*. Di antaranya URL di cek menggunakan API integrasi ke *website phistank* dan menggunakan algoritma *XGboost* [4] pada penelitian ini terdapat beberapa kekurangan *dataset* yang digunakan dan juga data yang tidak *real time* pada fitur deteksi dikarenakan tidak terintegrasi oleh *website* terpercaya yang membahas *phishing* juga, lalu lembaga keuangan dan penyedia layanan *online* juga terus meningkatkan keamanan sistem mereka dan mengedukasi pengguna tentang cara mengenali dan menghindari serangan *phishing*.

Kesimpulan yang didapat bahwa *phishing* adalah salah satu kasus *cyber crime* yang paling umum dan sering terjadi akhir-akhir ini. Serangan *phishing* menduduki peringkat dua berdasarkan data Badan Siber dan Sandi Negara Republik Indonesia pada tahun 2023 [5]. Peneliti mengusulkan bahwa detektor *phishing* dianggap dapat menyelesaikan masalah ini. Peneliti menggunakan data set API *public* yaitu *website similar web* dan *open page rank*,

penggunaan API digunakan agar data set selalu ter *update* dan juga penambahan *dataset* menjadi 11.055 data *phishing* dan *non-phishing*. Algoritma *random forest* juga dipakai di dalam penelitian kali ini bertujuan agar hasil lebih akurat dari sebelumnya.

2. METODE PENELITIAN

Pada metode penelitian ini peneliti menggunakan beberapa tahapan dalam penelitian di antaranya :

2.1 Understanding the problem

Pada tahap ini peneliti mencari berbagai jurnal terkait *phishing*, berbagai macam algoritma yang dipakai di antaranya algoritma KNN, *naive bayes*, *XGBoost* [6], dan lainnya. Dalam hal ini peneliti memilih algoritma *random forest* karena dinilai lebih cocok untuk menjadi algoritma dalam penelitian ini. Tahap ini juga tahap yang berfungsi untuk memahami masalah yang ada dan untuk menetapkan tujuan dari penelitian ini dengan masalah yang ada [7]. Keberhasilan dari penelitian ini diambil dari akurasi sistem mendeteksi suatu URL *phishing* dengan diukur menggunakan *confusion matrix*, semakin tinggi nilai akurasi semakin efektif pula model yang dibuat untuk mendeteksi dan mencegah URL *phishing*.

2.2 Understanding the data

Pada tahap ini pencarian data set dilakukan pada *website phistank.com, openphish.com, similarweb.com, archive.ics.uci.edu*, dan lainnya yang berhubungan dengan *phishing* [8]. Data dari *archive.ics.uci.edu* menjadi acuan program peneliti karena data set yang di hasilkan menjadi bahan belajar *machine learning* karena sudah di *extract* menjadi bilangan 1,0,-1 sesuai dengan *feature* deteksi yang digunakan mengeluarkan hasil yang serupa yaitu 1,0,-1, dan melihat isi data.

2.3 Pemodelan Algoritma Random Forest

Tahap ini merupakan tahap mengimplementasikan algoritma *random forest* ke dalam penelitian yang dibuat setelah memahami masalah yang ada dan mengumpulkan *dataset* yang ada selanjutnya proses algoritma *random forest* dilakukan yaitu membaca *dataset*, pembagian data, menggunakan klasifikasi *random forest*, dan penggunaan 30 fitur [9] pengtesan URL merupakan salah satu cara untuk mencari ciri dari suatu URL *phishing*, dan juga sebagai penentuan hasil klasifikasi. Pada Tabel 1 menampilkan 30 fitur deteksi dengan *value* dari fitur yang ada dan penjelasannya.

2.3.1 Having IP Address

Jika sebuah DNS diganti menjadi sebuah IP *address* ketika di akses seperti [HTTP://123.45.6.789/fake.html](http://123.45.6.789/fake.html) maka itu bisa dibidang *phishing* dan dapat mencuri informasi pribadi *user* yang mengakses URL tersebut. Selain itu, IP *address* terkadang diubah menjadi

hexadecimal, seperti
 HTTP://www.0x57.0xAA.0xyT.0x83.0x7t/3/paypal.
 ca/index.html jika URL berubah alamat IP ketika
 diakses, nilai -1. Sebaliknya, diberi nilai satu.

Tabel 1. Fitur Deteksi

Parameter	Fitur	Nilai Feature
	Have IP address	(-1,1)
	Panjang URL	(1,0,-1)
	URL pendek	(1,-1)
	Mempunyai simbol @	(-1,1)
	Menggunakan //	(-1,1)
	Menggunakan simbol -	(-1,1)
	Mempunyai <i>subdomain</i>	(-1,0,1)
	Mempunyai sertifikat SSL	(1,0,-1,-1)
	Panjang Domain	(-1,1,-1,-1-1)
	Favicon	(1,-1,-1)
	Port	(1,-1,1,1)
	Mempunyai HTTPS	(1,-1)
	Request URL	(1,-1)
	Mempunyai Anchor	(1,0,-1,-1)
1 = Valid	Link pada tags	(-1,1)
0 = Suspicious	SFH	(1,-1)
-1 = Phishing	Submit email	(-1,-1,-1)
	Link URL mencurigakan	(-1,1)
	Redirect	(1,-1)
	On mouse over	(1,-1)
	Klik kanan	(1,-1)
	Pop up window	(1,-1)
	Iframe	(1,-1)
	Umur domain	(1,-1,-1,-1)
	DNS record	(1,-1,-1,-1)
	Web traffic	(1,0,-1)
	Page rank	(1,-1,-1)
	Google indeks	(1,-1,-1)
	Link to page	(1,0,-1,-1)
	Statistic report	(-1,1,-1,-1)

2.3.2 URL Length

Hacker memakai URL yang panjang untuk menipu contohnya
 HTTP://www.paypal.ca/3f/aze/ab3ed5f46s64gasd65
 a8sd7/?cmd=homemp;dispatch=110278asd124379as
 d27134foe4ij123749asoieh1236419adfjsabeo19237
 @Phishing.html. Jika URL memiliki panjang kurang
 dari 54 maka diberi nilai 1, Jika URL memiliki
 panjang antara 54 dan 75 diberi nilai nol, dan
 selebihnya diberi nilai -1.

2.3.3 Shortening Service

URL dibuat menjadi pendek dan mengarahkan ke web yang dimaksud. Hal ini terjadi karena adanya “HTTP Redirect” pada domain URL dan mengarahkannya ke web yang sama yang memiliki URL panjang. Contoh, HTTP://www.sample.com dapat di *Shortening* menjadi HTTP://bit.ly/8as46e. Jika URL menggunakan *Shortening service* maka nilai -1. Jika tidak, nilai 1.

2.3.4 Having At Symbol

Mengecek URL apakah memiliki simbol ‘@’, Contoh, HTTP://con@toh.com, maka mengabaikan HTTP://con dan mengarahkannya ke “toh.com.” Jika URL mempunyai simbol ‘@’ diberi nilai -1. Kalau tidak diberi nilai satu.

2.3.5 Double Slash Redirecting

Jika ada ‘//’ pada URL setelah 6 karakter di dalam URL maka diarahkan ke *website* lain. Contohnya HTTP://contoh.com//HTTP://www.phishing.com maka diarahkan ke *phishing.com*, bukan “contoh.com.” Pada URL dengan HTTP, simbol ‘//’ ditemukan pada posisi keenam, sedangkan pada HTTPS, ditemukan pada posisi ketujuh. Jika posisi simbol ‘//’ pada URL berada pada posisi lebih dari tujuh maka nilainya -1 kalau tidak, maka diberi nilai satu.

2.3.6 Prefix Suffix

Penggunaan simbol (-) jarang dipakai pada URL sah *Hacker* menambahkan imbuhan atau akhiran dan memisahkannya dengan (-) untuk menipu, sehingga menganggap mereka mengakses URL yang asli. Contoh, HTTP://www.confirmed-paypal.com. Jika simbol ‘-’ ada pada URL maka nilai -1. Jika tidak, maka nilai satu.

2.3.7 Having Sub Domain

Contoh, HTTP://www.school.ac.uk/students URL tersebut mungkin punya *country-code-top-level-domain (ccTLD)*, yaitu “UK”. Bagian “AC” merupakan kependekan dari “academic”, apabila digabungkan “ac.uk” membentuk *second-level domain (SLD)* dan “school” merupakan nama domain tersebut. Agar tahu, maka bagian “WWW” harus diabaikan, kemudian menghilangkan *ccTLD* dan menghitung jumlah “.” tersisa. Jika jumlahnya > 1, maka URL dapat dikelompokkan “*Suspicious*” karena hanya mempunyai 1 *subdomain*. Jika jumlahnya > dua, maka dikelompokkan “*Phishing*” karena seharusnya memiliki *multiple* sub domain. Jika URL tidak punya sub domain, maka dapat dikelompokkan “*legitimate*”.

2.3.8 SSL Final State

Adanya HTTPS pada URL memberikan kesan kepada *users* bahwa *website* yang diakses *legitimate*, tapi itu belum cukup. Masih disarankan untuk mengecek *certificate* pada HTTPS, seperti otoritas pemberi *certificate* dan umur *certificate*. Pemberi *certificate* yang terkenal biasanya *GeoTrust*, *GoDaddy*, *Doster* dan *VeriSign*. Selain itu, umur *certificate* biasanya dua tahun. Jika URL memiliki HTTPS, *certificate* SSL terpercaya dan umur *certificate* memiliki umur satu tahun maka diberi nilai 1. Jika URL memiliki HTTPS tetapi *certificate* SSL tidak terpercaya, maka diberi nilai 0. Selebihnya, diberi nilai -1.

2.3.9 Domain Registration Length

Website phishing biasanya memiliki jangka umur yang pendek, sedangkan *website* yang asli biasanya dibayar secara *regular* untuk beberapa tahun ke depan. Umur *website phishing* terpanjang yang ditemukan adalah hanya satu tahun. Jika umur

website kurang dari satu tahun maka diberi nilai -1. Sebaliknya, maka diberi nilai 1, selebihnya -1.

2.3.10 Favicon

Favicon adalah sebuah icon yang terhubung dengan sebuah *webpage* yang spesifik. Kebanyakan *graphical browsers* dan *newsreaders* menampilkan *favicon* sebagai pengingat visual dari identitas *website* pada *address* bar. Jika *favicon* yang ditampilkan berbeda dengan yang muncul di *address* bar, maka kemungkinan *website* tersebut merupakan *phishing*. Jika *website* menggunakan *favicon external* diberi nilai -1 selebihnya 1.

2.3.11 Port

Fitur ini untuk validasi *service* tertentu seperti HTTP digunakan pada server. Disarankan hanya memakai *port* yang dibutuhkan dengan tujuan mengurangi adanya gangguan pada server. Maksudnya adalah, jika semua *port* dibuka, maka *Hacker* bisa menjalankan berbagai *service* yang diinginkan. Jika sebuah *website* hanya membuka *port* tertentu saja seperti *port* 21,22,23,80,443 maka diberi nilai satu. Sebaliknya, jika *port* tidak sesuai dengan *port* yang telah ditentukan maka diberi nilai -1.

2.3.12 HTTPS token

Hacker bisa saja menambahkan HTTPS di depan URL *phishing* dengan tujuan mengelabui pengguna. Contoh `HTTP://HTTPS-www-paypalwebapps.software.com`. Jika URL mempunyai HTTPS di depan DNS maka diberi nilai satu. Sebaliknya, maka diberi nilai -1.

2.3.13 Request URL

Fitur ini mengecek apakah ada objek eksternal yang dimasukkan ke dalam *website*, seperti *images*, *videos* dan *sounds* berasal dari domain lain. Dalam *website* yang *legitimate*. Jika sebuah *website* melakukan *Request* objek eksternal kurang dari 22% maka diberi nilai satu. Jika *website* melakukan *Request* objek eksternal $\geq 22\%$ dan $< 61\%$ diberi nilai nol. Selebihnya, maka diberi nilai -1.

2.3.14 URL of Anchor

Anchor adalah simbol `<a>` yang ada dalam *website*. Fungsi fitur ini yaitu *Request* URL. Tetapi fitur ini sebenarnya bertujuan untuk memeriksa: A. Jika *tags* `<a>` dan *website* memiliki domain yang berbeda. B. Jika *Anchor* tidak terhubung dengan *webpage* apa pun. Contohnya: - `<ahref="#">` - `<ahref="#content">` - `` - `` Jika *tags* `<a>` memiliki domain yang berbeda kurang dari 31% maka diberi nilai satu. Jika *tags* `<a>` memiliki domain yang berbeda $\geq 31\%$ dan $\leq 67\%$ maka diberi nilai nol. Selebihnya, maka diberi nilai -1.

2.3.15 URL in tags

Umumnya *website* sah mempunyai beberapa seperti *tag* `<meta>` yang menyediakan meta data dokumen HTML, *tag* `<script>` untuk membuat *script* dari sisi klien, dan *tag* `<URL>` untuk mengambil *resource* dari web lain. Ketiga, *tag* tersebut biasanya terhubung dari domain yang sama. Jika *tag* `<meta>`, *tag* `<script>`, dan *tag* `<URL>` terhubung dari domain yang berbeda kurang dari 17% maka diberi nilai nol. Jika *tag* `<meta>`, *tag* `<script>`, dan *tag* `<URL>` terhubung dari domain yang berbeda lebih dari sama dengan 17% dan kurang dari 81% maka diberi nilai nol. Selebihnya maka diberi nilai -1.

2.3.16 SFH (Server from Handler)

SFH mempunyai "*about:blank*" bisa dicurigai karena harus ada *action* yang dilakukan terhadap informasi yang di *submit*. Jika domain *name* dalam SFH berbeda dengan domain *name* dari *webpage*, patut dicurigai karena biasanya informasi yang di *submit* jarang dikendalikan *external domains*. Jika SFH pada URL berisi "*about:blank*" maka diberi nilai -1. Jika SFH pada URL berasal dari domain yang berbeda maka diberi nilai nol. Sebaliknya, maka diberi nilai satu.

2.3.17 Submitting to email

Web *form* biasanya diperbolehkan *submit* informasi pribadinya yang kemudian dikirim ke server. Tapi *Hacker* kemungkinan mengirim informasi pribadi *user* ke email pribadinya. Server - *side script language* seperti "*mail()*" *function* di PHP dan "*mailto:?*". Jika *website* mengandung *script* "*mail()*" atau "*mailto*" maka nilai -1. Selebihnya nilai satu.

2.3.18 Abnormal URL

Fitur ini mengambil *database* WHOIS. *Website* yang *legitimate* biasanya mempunyai identitas di bagian URL. Jika *hostname* tidak terdapat pada URL maka diberi nilai -1. Sebaliknya, maka diberi nilai satu.

2.3.19 Redirect

Perbedaan dari *website phishing* dan *website* asli adalah seberapa banyak jumlah *website* di *Redirect*. Dalam *datasets* yang digunakan, *website legitimate* biasanya di *Redirect* maksimal satu kali, sedangkan *website phishing* di *Redirect* paling sedikit empat kali. Jika, *website* hanya melakukan *Redirect* setidaknya sebanyak satu kali maka diberi nilai 1. Jika *website* melakukan *Redirect* sebanyak ≤ 2 dan > 4 diberi nilai nol. Selebihnya, maka diberi nilai -1.

2.3.20 On Mouseover

Peretas biasanya menggunakan *JavaScript* untuk menunjukkan URL palsu di status bar *users*. Untuk memakai fitur ini, maka harus mencari dari *source code* terutama di *event* "*onMouseOver*" untuk mengecek apakah terjadi perubahan di status bar. Jika

URL pada status bar berubah atau berbeda diberi nilai -1. Jika tidak, diberi nilai satu.

2.3.21 *Right Click*

Peretas biasanya memakai *JavaScript* untuk menghilangkan fungsi *right click* sehingga *users* tidak dapat melihat dan menyimpan *source code* dari *webpage* palsu tersebut. Fitur ini dapat digunakan dengan “*Using onMouseOver to hide the URL*”. Jika *website* melakukan *disable right click* diberi nilai -1. Sebaliknya nilai satu.

2.3.22 *PopUp Window*

Jarang sekali ditemukan *website legitimate* meminta *user* memasukkan informasi pribadinya melalui sebuah *pop-up window*. Hanya saja, fitur ini telah digunakan oleh beberapa *website legitimate*. Jika pada sebuah *website* terdapat *pop up*, diberi nilai -1. Sebaliknya nilai satu.

2.3.23 *Iframe*

Iframe ialah *tag HTML* dipakai agar bisa menampilkan halaman web tambahan jadi satu dengan web yang sedang dibuka. Peretas menggunakan *tag "iframe"* dan biasanya tidak terlihat atau tanpa menggunakan *border*. Dalam hal ini, pelaku *phishing* menggunakan atribut "*frame Border*" yang membuat browser menampilkan penggambaran visual. Jika *website* menggunakan *iframe* diberi nilai -1. Sebaliknya nilai satu.

2.3.24 *Age Of Domain*

Fitur ini dapat diekstrak dari *database WHOIS*. Kebanyakan situs web *phishing* biasanya memiliki umur yang lebih singkat, di mana *website legitimate* biasanya berumur enam bulan atau lebih. Jika umur domain setidaknya berumur enam bulan diberi nilai 1. Sebaliknya nilai -1

2.3.25 *DNSRecord*

Identitas *website* tidak ditemukan pada *database WHOIS* dan *hostname* juga tidak ditemukan maka bisa dikategorikan sebagai *phishing*. Jika domain tidak mempunyai *DNS Record*, maka diberi nilai -1. Sebaliknya, maka diberi nilai satu.

2.3.26 *Web Traffic*

Fitur ini mengukur popularitas *website* berdasarkan jumlah pengunjung dan jumlah halaman yang dikunjungi. Dari *datasets* yang digunakan, seburuk-buruknya *rank website legitimate*, masih masuk dalam Top 100.000. Jika sebuah domain tidak dikenali oleh *similar web*, maka domain tersebut patut dicurigai. Jika peringkat *website* kurang dari 100.000 maka diberi nilai satu. Jika peringkat *website* lebih dari sama dengan 100.000 maka diberi nilai nol. Sedangkan, jika *website* tidak terdaftar atau dikenali oleh *similar web* maka diberi nilai -1.

2.3.27 *Page Rank*

PageRank nilai yang berkisar dari "nol" hingga "satu". *PageRank* berguna mengukur seberapa penting halaman web di Internet. Semakin besar nilainya, semakin penting web tersebut. Peneliti menemukan bahwa sekitar 95% halaman web *phishing* tidak memiliki *PageRank*. Selain itu, peneliti menemukan bahwa sisa 5% dari laman web *phishing* dapat mencapai nilai *PageRank* hingga “2”. Jika *website* memiliki *PageRank* kurang dari 2, maka diberi nilai -1. Sebaliknya, maka diberi nilai satu.

2.3.28 *Google Index*

Fitur ini memeriksa apakah ada pada indeks Google atau tidak. Jika sebuah situs web terdapat pada indeks Google maka ditampilkan hasilnya. Biasanya, halaman web *phishing* hanya dapat diakses dalam waktu yang singkat dan akibatnya, banyak halaman web *phishing* mungkin tidak ditemukan di dalam indeks Google. Jika *website* terdapat di indeks Google diberi nilai 1. Sebaliknya nilai -1

2.3.29 *URL Pointing to Page*

Jumlah URL yang mengarah ke web menampilkan tingkat sahnya meskipun beberapa tautan berada di domain yang sama. Dalam kumpulan *datasets* yang digunakan, karena umur pakainya yang pendek, ditemukan bahwa 98% *website phishing* tidak memiliki tautan URL yang mengarah ke sana. Di sisi lain, *legitimate website* memiliki setidaknya dua tautan URL eksternal yang mengarah ke sana. Jika *website* tidak memiliki tautan URL eksternal maka diberi nilai -1. Jika *website* memiliki tautan URL kurang dari tiga maka diberi nilai nol. Sebaliknya, diberi nilai satu.

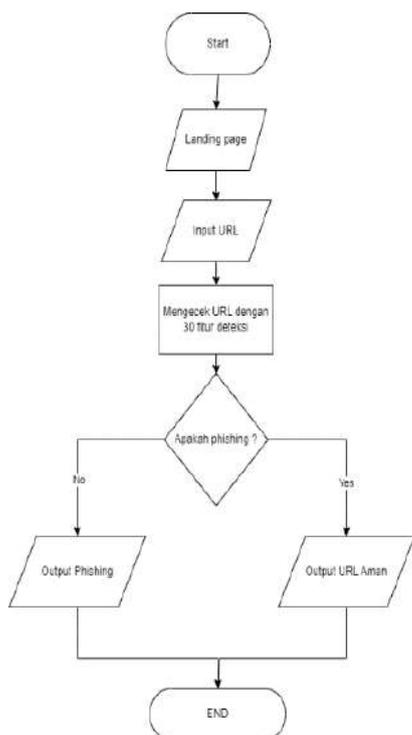
2.3.30 *Statistical Report*

Beberapa situs seperti *Phish Tank* dan *StopBadware* sering kali merilis berbagai laporan statistik mengenai *website phishing* dalam jangka waktu tertentu. Jika *website* termasuk dalam Top *phishing IP* atau Top *phishing Domain* maka diberi nilai -1. Sebaliknya nilai 1

3. HASIL DAN PEMBAHASAN

3.1 *Flowchart*

Flowchart di tunjukan pada Gambar 2 berisi langkah dari proses mendeteksi URL *phishing* menggunakan algoritma *random forest*. Untuk mendeteksi harus memasukkan URL saja setelah itu sistem mengecek ke 30 fitur deteksi dan menentukan hasil menggunakan algoritma *random forest*.



Gambar 2. Flowchart

3.2 Implementasi Metode

Pada tahap ini algoritma *random forest* digunakan untuk membuat pemodelan di antaranya yaitu :

- a. Membaca *dataset*. Data dibaca menggunakan *library pandas* dengan ketentuan. X membaca semua kolom kecuali kolom terakhir, dan Y label membaca semua kolom kecuali kolom terakhir seperti Gambar 3.

```

data = pd.read_csv('Dataset/dataset.csv')
X = data.iloc[:, :-1].values
y = data.iloc[:, -1].values
    
```

Gambar 2. Membaca data menggunakan library pandas

- b. Memisahkan data menggunakan *library sklearn* dengan kode *python* from *sklearn.model_selection import train_test_split* Gambar 4 menjelaskan *code*

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
    
```

Gambar 4. Memisahkan data uji dan latih

Pada Gambar 4 dijelaskan bahwa data dibagi menjadi 4 bagian yaitu X_{train} , X_{test} , Y_{train} , Y_{test} . Lalu data di *split* menjadi 20% data uji dan 80% data latih. Dengan fungsi *test_size=0.2*. Pembagian data model algoritma yang digunakan di sini umum karena memberikan cukup data untuk pelatihan model sambil menyisakan cukup data untuk evaluasi yang akurat [10].

- c. Penggunaan *random forest classifier* data yang sudah dibagi menjadi dua bagian x dan y disimpan pada folder *rf_final.pkl* untuk pembelajaran sistem berikut *code python* pada Gambar 5

```

from sklearn.ensemble import RandomForestClassifier
    
```

Gambar 3. Memanggil fungsi random forest classifier

Lalu setelah dipanggil *random forest classifier* maka ini dipanggil kembali di dalam *code* berikut pada Gambar 6.

```

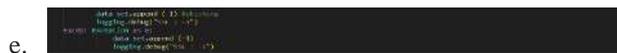
classifier = RandomForestClassifier()
classifier.fit(X_train, y_train)
joblib.dump(classifier, 'last_pkl/rf_final.pkl')
    
```

Gambar 6. Fungsi random forest

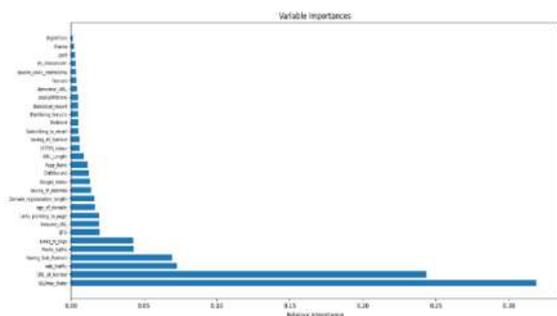
- d. Pengetesan *dataset* URL dengan fitur deteksi pada percobaan ini adalah salah satu contoh dari hasil deteksi pada satu URL [HTTPS://google.com](https://google.com) pada *link* ini *output* seperti Tabel 2

Tabel 1. Nilai pengecekan <https://google.com>

No.	Fitur deteksi	Tipe Data	URL aman	Phishing
1	Have IP address	Integer	1	
2	Panjang URL	Integer	1	
3	URL pendek	Integer	1	
4	Having Symbol	Integer	1	
5	Menggunakan //	Integer	1	
6	Prefix Suffix	Integer	1	
7	Having subdomain	Integer		-1
8	SSL final state	Integer	1	
9	Domain Length	Integer	1	
10	Favicon	Integer	1	
11	Port	Integer	1	
12	Mempunyai HTTPS	Integer	1	
13	Merequest URL	Integer		-1
14	Mempunyai Anchor	Integer	1	
15	Link in tags	Integer	1	
16	SFH	Integer		-1
17	Submit email	Integer	1	
18	Abnormal URL	Integer	1	
19	Redirect	Integer		-1
20	On mouse over	Integer		-1
21	Klik kanan	Integer		-1
22	Pop up window	Integer		-1
23	I frame	Integer		-1
24	Umur domain	Integer	1	
25	DNS catatan	Integer	1	
26	Web traffic	Integer	1	
27	Page rank	Integer	1	
28	Google indeks	Integer	1	
29	Link to page	Integer	1	
30	Statistic report	Integer		-1
	Hasil Deteksi			URL aman (1)



e. Penentuan hasil klasifikasi tahap ini setelah mengecek URL dan mengetahui hasil deteksi yang di cek pada 30 fitur deteksi yang ada maka sistem menentukan hasil prediksi, dengan cara menghitung jumlah dari banyaknya fitur deteksi jika hasil yang di tunjukan oleh sistem dominan - 1 maka sistem mengindikasikan URL yang di cek adalah *phishing*, begitu pun sebaliknya. Tetapi ada beberapa faktor juga yang menentukan hasil prediksi dari sistem yaitu nilai dari 30 fitur deteksi yang ada untuk mengetahui nilai dari 30 fitur peneliti membuat grafik nilai fitur *dataset* pada Gambar 7, dan cara membacanya.



Gambar 7. Nilai Fitur Deteksi

Grafik nilai pada Gambar 7 menunjukkan bahwa fitur *SSL_final_state* yang memiliki nilai deteksi paling besar di karena kan fitur ini sangat harus dipunyai oleh kebanyakan *website* yaitu fitur sertifikat *security* dari suatu *website*, karena *website* yang sah biasanya diawali dengan HTTPS dan situs *phishing* hanya diawali dengan HTTP, Jika terdapat situs yang mempunyai HTTPS tetapi tidak memiliki sertifikat SSL maka dilabeli *phishing*. Pada posisi kedua di duduki oleh fitur *URL_Anchor* umumnya situs *phishing* memiliki fungsi *tag* `<a href>` yang merujuk ke domain lainya dan *tag* ini tidak hanya terdapat 1 dalam penelitian kali ini peneliti membuat *tracehold* presentasi untuk *tag* `<a href>` sebanyak 31% karena jika melebihi itu maka di indikasikan situs *phishing*. Berikut salah satu program fitur SSL yang menduduki peringkat pertama pada Gambar 8.



Gambar 8. Fitur SSL Final State

Cara kerja dari *phishing* URL *detection* membedakan situs *phishing* dan sah seperti yang dijelaskan pada nilai fitur jika URL pada fitur SSL mendapatkan nilai 0,30. Diberikan kondisi juga jika

SSL terindikasi 1 maka menambahkan *point* sebanyak +0,30. Tetapi jika SSL mendapatkan nilai - 1 maka mendapatkan nilai -0,30. Lalu setelah mengetahui semua nilai fitur *dataset* maka dijumlahkan, jika hasil semua total berjumlah positif maka situs terindikasi URL aman.

f. Menghitung *score* akurasi dari satu URL yang dideteksi. Tahap ini setelah mengimplementasikan model algoritma *random forest* dan juga mencoba fitur deteksi maka data dievaluasi untuk mengetahui hasil *score* akurasi dari data yang diuji berikut *code* pada Gambar 9.

```
score = classifier.score(X_test, y_test) * 100
logging.info(f"Model accuracy: {score}%")
result_of_analysis = score
```

Gambar 9. Evaluasi score random forest

3.3 Pengujian

Pada Tabel 3 hasil pengecekan 50 URL menghasilkan hasil *true* positif URL *phishing* sebanyak 24 *true positive* URL tetapi *website* mendeteksi juga 1 *False positive* situs *phishing* yang di deteksi aman. Lalu mendeteksi 23 *true negative* URL aman yang dideteksi aman, dan mendeteksi 2 URL *false negative*.

Tabel 2. Tabel uji coba *confusion matrix*

Prediksi Keadaan Data	Keadaan Data Sebenarnya	
	Phishing	URL aman
Phishing	24	1
URL aman	2	23
Jumlah	26	25

Berikut perhitungan untuk menentukan prediksi, *recall*, *f-score*, akurasi.

$$\begin{aligned} \text{Prediksi} &= \frac{TP}{TP+FP} \times 100\% \\ &= \frac{24}{24+1} \times 100\% \\ &= 0.96 \text{ atau } 96\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN} \times 100\% \\ &= \frac{24}{24+2} \times 100\% \\ &= 0.92 \text{ atau } 92\% \end{aligned}$$

$$\begin{aligned} \text{Akurasi} &= \frac{(TP+TN)}{(TP + FP + FN + TN)} \times 100\% \\ &= \frac{(24+23)}{(24+1+2+23)} \times 100\% \\ &= 0.94 \text{ atau } 94\% \end{aligned}$$

$$\begin{aligned} \text{F - SCORE} &= \frac{2 \times P \times R}{(P + R)} \times 100\% \\ &= \frac{2 \times 0.96 \times 0.92}{(0.96 + 0.92)} \times 100\% \\ &= 93.95 \text{ atau } 93\% \end{aligned}$$

Hasil pengujian di sini dapat disimpulkan bahwa *website* deteksi *phishing* menggunakan algoritma *random forest* dan penambahan beberapa fitur deteksi

dapat menyelesaikan masalah dengan menunjukkan hasil akurasi yang cukup tinggi untuk memprediksi *website phishing*.

4. KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan pengujian, *website phishing URL detection* yang dibuat dapat berguna dengan baik walaupun terdapat beberapa URL yang terdeteksi *false negative* tetapi angka akurasinya masih menunjukkan cukup tinggi dan dinilai dapat menyelesaikan masalah. Penggunaan algoritma *random forest* sangat cocok untuk *website* ini karena mencapai tingkat akurasi 94%, prediksi 96%, *Recall* 92%, dan *f-score* 93%. Dengan penggunaan fitur deteksi sebanyak 30 fitur. Beberapa saran yang dapat dilakukan pada penelitian berikutnya ialah Menambahkan data set menggunakan URL terbaru baik URL *phishing* maupun *non-phishing*, Menambahkan fitur deteksi lainnya seperti melalui email DLL, Melakukan *update* fitur deteksi terbaru untuk mengecek URL *phishing*.

DAFTAR PUSTAKA

- [1] Saroji Ahmad, Harmini Triana, and Taqiyuddin Muhammad, 'Internet Evolution: A Historical View (SEJARAH EVOLUSI GENERASI INTERNET)', 2021, doi: 10.30598/Lanivol2iss2page65-75.
- [2] Acsany Philipp, 'Build a Scalable Flask Web Project From Scratch', <https://realpython.com/flask-project/>. Accessed: Jun. 07, 2024. [Online]. Available: <https://realpython.com/flask-project/>
- [3] Nurfitrianti Fifi, 'Apa Itu Phishing, Smishing, dan Vhishing?', <https://www.jenius.com/highlight/detail/apa-itu-phishing-smishing-dan-vhishing>. Accessed: Jun. 07, 2024. [Online]. Available: <https://www.jenius.com/highlight/detail/apa-itu-phishing-smishing-dan-vhishing>
- [4] A. Aljofey *et al.*, 'An effective detection approach for phishing websites using URL and HTML features', *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10841-5.
- [5] Badan Siber dan Sandi Negara, 'LANSKAP KEAMANAN SIBER INDONESIA', <https://www.bssn.go.id/wp-content/uploads/2024/03/Lanskap-Keamanan-Siber-Indonesia-2023.pdf>. Accessed: Jun. 07, 2024. [Online]. Available: <https://www.bssn.go.id/wp-content/uploads/2024/03/Lanskap-Keamanan-Siber-Indonesia-2023.pdf>
- [6] A. M. Veach and M. Abualkibash, 'Phishing Website Detection Using Several Machine Learning Algorithms: A Review Paper', *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, vol. 3, no. 2, pp. 219–230, Dec. 2022, doi: 10.34010/injiiscom.v3i2.8805.
- [7] I. Yurita, M. Kevin Ramadhan, and M. Candra, 'Pengaruh Kemajuan Teknologi Terhadap Perkembangan Tindak Pidana Cybercrime (Studi Kasus Phising Sebagai Ancaman Keamanan Digital)', 2023.
- [8] R. Zieni, L. Massari, and M. C. Calzarossa, 'Phishing or Not Phishing? A Survey on the Detection of Phishing Websites', *IEEE Access*, vol. 11, pp. 18499–18519, 2023, doi: 10.1109/ACCESS.2023.3247135.
- [9] W. Ali and S. Malebary, 'Particle Swarm Optimization-Based Feature Weighting for Improving Intelligent Phishing Website Detection', *IEEE Access*, vol. 8, pp. 116766–116780, 2020, doi: 10.1109/ACCESS.2020.3003569.
- [10] G. (Gareth M. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: with applications in R*. 2013.

PENERAPAN ALGORITMA AGGLOMERATIVE CLUSTERING UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR PENDIDIKAN

Hafizh Taufiqul Hakim^{1*}, Wendi Usino²

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ^{1*}hafizhtaufiqul10@gmail.com, ²wendi.usino@budiluhur.ac.id

(Naskah masuk: 7 Agustus 2024, diterima untuk diterbitkan: 7 September 2024)

Abstrak

Pendidikan memiliki peran yang penting dalam meningkatkan kualitas Sumber Daya Manusia (SDM) guna mendukung pembangunan suatu negara. Pentingnya pendidikan sebagai Indikator Pembangunan juga terbukti dengan adanya poin pendidikan di salah satu tujuan pada *Sustainable Development Goals* (SDGs). Namun, pendidikan di Indonesia saat ini masih belum merata, khususnya di jenjang Sekolah Menengah Atas (SMA) yang menunjukkan partisipasi lebih rendah dibandingkan dengan jenjang Sekolah Dasar (SD), dan Sekolah Menengah Pertama (SMP). Sehingga penelitian ini bertujuan untuk membangun model yang dapat pengelompokan provinsi di Indonesia berdasarkan Indikator Pendidikan pada jenjang Sekolah Menengah Atas (SMA). Penelitian ini menggunakan metode *Cross-Industry Standard Process for Data mining* (CRISP-DM) sebagai acuan dari proses *Data mining*. Penelitian ini menggunakan algoritma klastering untuk dapat melakukan pengelompokan provinsi dengan melakukan beberapa perbandingan algoritma seperti K-Means, *Agglomerative*, dan K-Medoids. Berdasarkan hasil perbandingan diketahui bahwa nilai algoritma *Agglomerative* lebih baik dibandingkan dengan algoritma lainnya. Parameter yang digunakan dalam model *Agglomerative* adalah metode *Average linkage* dan metrik *Euclidean distances*. Hasil penelitian menunjukkan bahwa penggunaan algoritma *Agglomerative* menghasilkan 2 *Cluster* dengan nilai *Davies-Bouldin Index* (DBI) terendah sebesar 0,497. *Cluster* 1 terdiri dari 33 provinsi dengan tingkat pendidikan yang lebih tinggi, sedangkan *Cluster* 2 terdiri dari 1 provinsi dengan tingkat pendidikan yang lebih rendah, yaitu Papua. Hasil penelitian ini mengindikasikan adanya ketimpangan yang signifikan dalam kualitas pendidikan di Indonesia.

Kata kunci: Indikator Pendidikan, Sekolah Menengah Atas (SMA), CRISP-DM, *Agglomerative Clustering*

APPLICATION OF AGGLOMERATIVE CLUSTERING ALGORITHM TO GROUP PROVINCES IN INDONESIA BASED ON EDUCATION INDICATORS

Abstract

Education has an important role in improving the quality of Human Resources (HR) to support the development of a country. The importance of education as a Development Indicator is also proven by the presence of education points in one of the goals in the Sustainable Development Goals (SDGs). However, education in Indonesia is currently still uneven, especially at the senior high school (SMA) level which shows lower participation compared to the elementary and junior high school levels. So this study aims to build a model that can group provinces in Indonesia based on Education Indicators at the Senior High School level. This research uses the Cross-Industry Standard Process for Data mining (CRISP-DM) method as a reference for the data mining process. This research uses Clustering algorithms to be able to group provinces by comparing several algorithms such as K-Means, *Agglomerative*, and K-Medoids. Based on the comparison results, it is known that the value of the *Agglomerative* algorithm is better than the other algorithms. The parameters used in the *Agglomerative* model are the *Average linkage* method and the *Euclidean distances* metric. The results showed that the use of the *Agglomerative* algorithm produced 2 *Clusters* with the lowest *Davies-Bouldin Index* (DBI) value of 0.497. *Cluster* 1 consists of 33 provinces with higher education levels, while *Cluster* 2 consists of 1 province with lower education levels, namely Papua. The results of this study indicate that there are significant inequalities in the quality of education in Indonesia.

Keywords: Education Indicators, High School, CRISP-DM, *Agglomerative Clustering*

1. PENDAHULUAN

Pendidikan memiliki peran yang penting dalam meningkatkan kualitas Sumber Daya Manusia (SDM) guna mendukung pembangunan suatu negara. Kualitas suatu negara dapat diukur dari peningkatan kualitas pendidikan yang diselenggarakan [1]. Pentingnya pendidikan sebagai indikator pembangunan tercermin dalam salah satu tujuan *Sustainable Development Goals* (SDGs), yaitu “Menjamin pendidikan yang inklusif dan berkualitas serta mendukung kesempatan belajar sepanjang hayat bagi semua.” Untuk mencapai tujuan ini, salah satu langkah yang dapat diambil adalah menerapkan program wajib belajar 12 tahun yang meliputi pendidikan dasar dan menengah. Selain itu, penting untuk melakukan tinjauan berkala terhadap kondisi pendidikan guna memastikan pencapaian dan perbaikan yang berkelanjutan [1]. Di Indonesia, sistem pendidikan terdiri dari beberapa jenjang, yaitu Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), dan Perguruan Tinggi.

Menurut Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, alokasi anggaran pendidikan dalam APBN harus minimal 20%. Alokasi dana ini mencerminkan komitmen pemerintah untuk mengoptimalkan kemajuan pendidikan. Dengan adanya undang-undang ini, diharapkan akan terjadi peningkatan dan pemerataan pendidikan di seluruh Indonesia. Anggaran pendidikan dalam APBN juga direncanakan untuk daerah terpencil atau tertinggal, sehingga masyarakat di kawasan tersebut dapat menikmati pendidikan yang layak seperti di daerah yang lebih maju. Upaya ini terus dilakukan untuk mengurangi kesenjangan pendidikan di Indonesia [2]. Bukti rendahnya mutu pendidikan di Indonesia dapat dilihat dari data UNESCO tahun 2000 mengenai Indeks Pengembangan Manusia (IPM). IPM adalah ukuran yang menggambarkan pencapaian suatu negara dalam berbagai bidang, termasuk pendidikan, kesehatan, dan pendapatan per kapita. Data tersebut menunjukkan bahwa IPM Indonesia mengalami penurunan yang konsisten dari tahun ke tahun. Menurut data *worldtop20.org* bahwa pendidikan di Indonesia menempati urutan ke-54 (2021) dari 203 negara, ke-67 (2022) dari 203 negara, dan ke-67 (2023) dari 209 negara yang ada di dunia

Berdasarkan Data dari Badan Pusat Statistik (BPS) [3] tahun 2023 menunjukkan adanya ketimpangan di setiap jenjang pendidikan. Jika dilihat berdasarkan angka partisipasi kasar (APK), tercatat bahwa jenjang SD memiliki persentase sebesar 105,62%, jenjang SMP sebesar 92,51%, dan jenjang SMA sebesar 86,34%. Sementara itu, angka partisipasi murni (APM) menunjukkan bahwa jenjang SD memiliki persentase sebesar 97,89%, jenjang SMP sebesar 81,35%, dan jenjang SMA sebesar 62,53%. Dari perbandingan angka-angka tersebut, dapat dilihat bahwa persentase partisipasi

untuk jenjang SMA lebih rendah dibandingkan dengan jenjang SD dan SMP.

Berdasarkan kondisi dari data di atas, Jenjang SMA masih tergolong rendah dibandingkan dengan jenjang SD dan SMP. Maka perlu dilakukan pengelompokan yang berfokus untuk jenjang SMA. Data yang akan digunakan berdasarkan indikator pendidikan untuk dapat mengukur kualitas pendidikan di setiap provinsi.

Dalam penelitian yang dilakukan [4] dengan judul “*Agglomerative Hierarchy Clustering* Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan” Menjelaskan bahwa metode *Agglomerative Hierarchical Clustering* dengan algoritma *average linkage* adalah yang paling optimal dengan nilai korelasi *cophenetic* sebesar 0,807. Validasi kluster, baik internal maupun stabilitas, serta karakteristik Kabupaten/Kota di Provinsi Jawa Timur menunjukkan bahwa jumlah kluster yang representatif adalah 2 kluster.

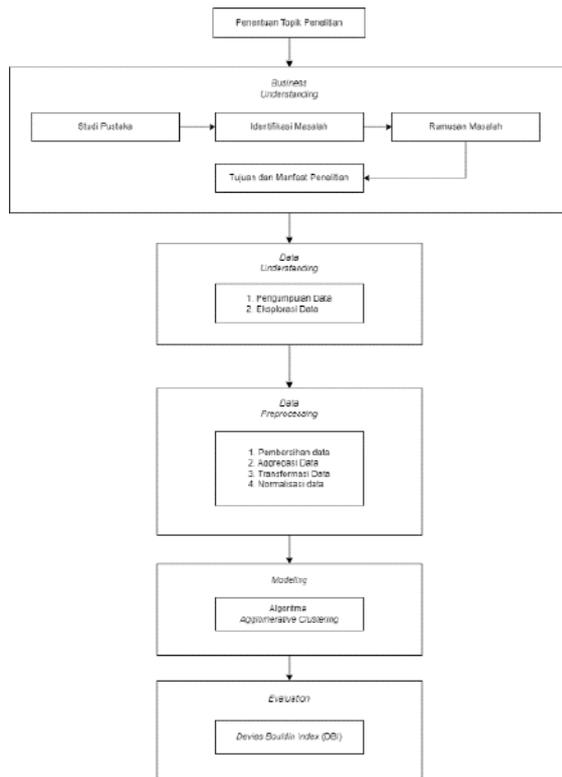
Selain itu, Pada penelitian yang dilakukan [5] dengan judul “Analisis *Cluster* Untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Pendidikan dengan Metode Ward”. Menjelaskan bahwa hasil *Cluster* yang optimal terdapat di 4 *Cluster* dengan nilai Indeks Dunn sebesar 0,499 yang menghasilkan nilai tertinggi dibandingkan *Cluster* lainnya. Sehingga penerapan algoritma *Agglomerative Clustering* dengan menggunakan metode ward dapat membagi kabupaten/kota menjadi 4 *Cluster*, yaitu *Cluster* 1 terdapat 4 anggota, *Cluster* 2 terdapat 14 anggota, *Cluster* 3 terdapat 5 anggota, dan *Cluster* 4 terdapat 1 anggota.

Menurut Penelitian yang dilakukan [6] dengan judul “Penerapan K-Means dan *Agglomerative Hierarchical Clustering* Untuk Pengelompokan Data Indikator Pendidikan (Studi Kasus Kabupaten/Kota di Wilayah Indonesia Timur)”. Menjelaskan bahwa hasil pengelompokan metode berhierarki *agglomeratif* memberikan hasil yang lebih baik dibandingkan pengelompokan dengan k-means sederhana ditinjau dari nilai DBI yang lebih kecil. Kemudian hasil pengelompokan kabupaten/kota di 13 provinsi yang ada di wilayah Indonesia timur berdasarkan Indikator Pendidikan terbentuk 3 kluster dan kluster ke 3 memiliki nilai Indikator Pendidikan yang rendah dibanding kluster 1 dan 2.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini menggunakan metodologi yaitu CRISP-DM sebagai metode untuk melakukan tahapan penelitian. Gambar 1 mengilustrasikan beberapa langkah dalam menentukan metode penelitian.



Gambar 1. Tahapan Penelitian

Tahapan ini dirancang untuk mengidentifikasi masalah yang akan diteliti. Dalam tahapan ini, penerapan metodologi CRISP-DM, yang terdiri dari beberapa langkah. Berikut adalah penjelasan singkat tentang langkah-langkah dalam metodologi CRISP-DM [7]:

a. *Business Understanding*

Pada tahap ini, diperlukan pemahaman mendalam mengenai latar belakang dan esensi dari aktivitas Data mining yang akan dijalankan. Ada dua masalah utama dalam penelitian ini, yaitu Angka kualitas pendidikan di jenjang Sekolah Menengah Atas (SMA) masih rendah dibandingkan dengan jenjang Sekolah Dasar (SD) dan Sekolah Menengah Pertama (SMP) dan Pendidikan di Indonesia masih belum merata. Sehingga penelitian ini bertujuan Membangun model yang dapat mengelompokkan provinsi di Indonesia berdasarkan Indikator Pendidikan di jenjang Sekolah Menengah Atas (SMA) menggunakan algoritma *Agglomerative Clustering*.

b. *Data Understanding*

Pada tahap ini, proses pengumpulan data dilakukan dengan mengambil sumber dari Badan Pusat Statistik (BPS) [3] dan Open Data Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud) [8] di tahun 2021-2023. Data yang digunakan sebanyak 102 baris dan 10 kolom berdasarkan Indikator Pendidikan dengan fokus pada jenjang Sekolah Menengah Atas (SMA) di seluruh provinsi di Indonesia. Atribut yang digunakan yaitu Provinsi, Rasio Murid-Guru, Rasio Murid-Kelas,

Mengulang, Putus Sekolah, Angka Partisipasi Kasar (APK), Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), Rata-Rata Lama Sekolah (RLS), dan Tingkat Penyelesaian Pendidikan.

c. *Data Preparation*

Pada tahapan ini, Data yang telah diperoleh dari proses pengumpulan akan dibersihkan melalui beberapa proses seperti Pembersihan Data, *Aggregasi Data*, *Transformasi Data*, dan *Normalisasi Data*. Tujuan dari tahap ini memastikan bahwa data yang akan digunakan dalam analisis atau pemodelan telah diproses dan dibersihkan dengan baik, sehingga hasil pemodelan yang diperoleh akan lebih akurat.

d. *Modeling*

Pada tahap ini, Implementasi model dilakukan menggunakan bahasa pemrograman *Python* di *Jupyter Notebook*. Penelitian melakukan pembuatan model melalui beberapa proses seperti Membandingkan 3 model, yaitu *K-Means*, *Agglomerative*, dan *K-Medoids* berdasarkan nilai *Davies Bouldin Index* (DBI). Hasil perbandingan menunjukkan bahwa model terbaik adalah *Agglomerative Hierarchical Clustering*. Selanjutnya, untuk memastikan kualitas model yang digunakan, perlu dilakukan penentuan korelasi yang baik. Pengukuran kualitas model dapat dilakukan dengan uji validitas menggunakan nilai *cophenetic correlation coefficient* dengan membandingkan berbagai metode *linkage*, seperti *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward Linkage*. Metode ini biasanya diterapkan pada model hierarki.

e. *Evaluation*

Pada tahap ini, Model terbaik akan dievaluasi dari hasil nilai *Davies Bouldin Index* (DBI). Tujuannya adalah untuk mendapatkan jumlah kluster yang optimal.

2.2 **Data Mining**

Data *mining* adalah disiplin ilmu yang mempelajari cara untuk mengekstrak pengetahuan atau pola dari sejumlah besar data. Ini bukan hanya pengumpulan data, tetapi juga pengolahan dan analisis data untuk menemukan informasi berharga. Oleh karena itu, Data *mining* adalah proses penting untuk mengubah data menjadi pengetahuan yang dapat digunakan untuk membuat keputusan yang lebih baik, menjadi lebih efisien, dan menemukan peluang baru dalam berbagai bidang [9].

Tujuan utama dari data *mining* adalah mengidentifikasi pola, hubungan, atau pengetahuan yang berharga dan tersembunyi dalam suatu set data besar atau kompleks. Proses Data *mining* bertujuan untuk menggali wawasan yang tidak dapat ditemukan secara langsung melalui pengamatan sederhana terhadap data. Data *mining* memiliki banyak manfaat dan dapat memberikan nilai tambah yang signifikan untuk berbagai industri dan organisasi [10].

Data *mining* atau kadang disebut juga *knowledge discovery in database* (KDD) merupakan proses mengumpulkan dan menganalisis data historis untuk mengungkap pengetahuan, informasi, pola, atau hubungan dalam kumpulan data besar. Hasil dari data *mining* dapat mendukung pengambilan keputusan di masa depan. Data *mining* tidak berdiri sebagai disiplin ilmu yang terpisah, melainkan sangat terkait dengan bidang lain seperti basis data, statistik, pencarian informasi, dan kecerdasan buatan [11].

2.3 Agglomerative Hierarchical Clustering

Algoritma Hierarki dengan metode *Agglomerative* merupakan metode statistika yang digunakan untuk pengelompokan data dengan banyak variabel. Tujuan utamanya adalah untuk mengelompokkan objek yang memiliki kemiripan [12].

Ada beberapa metode yang dapat digunakan untuk menentukan jarak antar *Cluster* [4]. Metode ini biasanya digunakan untuk jenis model *Clustering* yang hierarki untuk dapat menentukan jarak antar *Cluster* yang terbaik, Seperti:

a. *Single Linkage*

Metode *Single Linkage* merupakan metode analisis *Cluster* hierarki yang mengelompokkan data berdasarkan jarak terdekat antar satu sama lain. Pencarian jarak dilakukan dengan menggunakan jarak minimal.

$$d_{(uv)w} = \min\{d_{uw}, d_{vw}\} \dots\dots\dots(1)$$

Keterangan :

- $d_{(uv)w}$ = Jarak antara U dan V ke W
- d_{uw} = Jarak terpendek antara U dan W
- d_{vw} = Jarak terpendek antara V dan W

b. *Complete Linkage*

Metode *Complete Linkage* merupakan metode analisis *Cluster* hierarki yang mengelompokkan data berdasarkan jarak terjauh atau yang memiliki kemiripan terkecil. Pencarian jarak dilakukan dengan menggunakan jarak maksimal.

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\} \dots\dots\dots(2)$$

Keterangan :

- $d_{(uv)w}$ = Jarak antara U dan V ke W
- d_{uw} = Jarak terjauh antara U dan W
- d_{vw} = Jarak terjauh antara V dan W

c. *Average linkage*

Metode *Average linkage* merupakan metode analisis *Cluster* hierarki yang mengelompokkan data berdasarkan rata-rata jarak antar semua anggota. Pencarian jarak dilakukan dengan menggunakan jarak rata-rata

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{uv}N_w} \dots\dots\dots(3)$$

Keterangan :

- $d_{(uv)w}$ = Jarak antara U dan V ke W
- d_{ik} = Jarak antar objek i pada *Cluster* UV dan objek k pada *Cluster* W

- N_{uv} = Banyaknya item pada *Cluster* U dan V
- N_w = Banyaknya item pada *Cluster* W

d. *Ward Linkage*

Metode *Ward Linkage* merupakan metode analisis *Cluster* hierarki yang didasarkan pada prinsip bahwa informasi dapat hilang ketika objek digabungkan menjadi satu klaster. Metode ini mengutamakan perhitungan yang bertujuan untuk memaksimalkan homogenitas dalam satu kelompok dengan menggunakan *Error Sum of Squares* (ESS).

$$ESS = \sum_{j=1}^k \left(\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j} \left(\sum_{i=1}^{n_j} x_{ij} \right)^2 \right) \dots(4)$$

Keterangan :

- X_{ij} = Nilai objek ke-i
- $i = 1,2,3, \dots$ pada kelompok ke-j
- K = Jumlah kelompok setiap stage
- n_j = Jumlah kelompok ke-i pada kelompok ke-j

2.4 Davies Bouldin Index (DBI)

Davies Bouldin Indeks (DBI) adalah metrik evaluasi yang digunakan dalam analisis klaster untuk menilai seberapa baik klaster dapat terpisah satu sama lain. Tujuan utama DBI adalah mengukur perbedaan antara kluster dan memastikan setiap kluster memiliki pusat yang berbeda. Nilai DBI yang lebih rendah menunjukkan hasil yang lebih baik [13].

2.5 Cophenetic correlation coefficient

Pemilihan metode *Cluster* hierarki terbaik menggunakan Uji Validitas dengan koefisien korelasi *cophenetic*. *Cophenetic correlation coefficient* adalah sebuah koefisien korelasi yang mengukur kesesuaian antara matriks ketidakmiripan asli (matriks jarak *Euclidean*) dengan matriks *cophenetic* yang dihasilkan dari *dendrogram*. Rentang nilai *Cophenetic correlation coefficient* adalah dari -1 hingga 1. Semakin mendekati nilai 1 menunjukkan bahwa hasil *Clustering* memiliki kualitas yang baik [14].

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Penelitian ini mengidentifikasi masih adanya masalah dalam sistem pendidikan di Indonesia. Ada dua masalah utama dalam penelitian ini, yaitu Angka kualitas pendidikan di jenjang Sekolah Menengah Atas (SMA) masih rendah dibandingkan dengan jenjang Sekolah Dasar (SD) dan Sekolah Menengah Pertama (SMP) dan Pendidikan di Indonesia masih belum merata. Sehingga penelitian ini bertujuan Membangun model yang dapat mengelompokkan provinsi di Indonesia berdasarkan Indikator Pendidikan di jenjang Sekolah Menengah Atas (SMA) menggunakan algoritma *Agglomerative Clustering*.

3.2 Data Understanding

Penelitian ini menggunakan data publik dari Badan Pusat Statistik (BPS) dan Open Data Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud) di tahun 2021-2023. Data yang digunakan sebanyak 102 baris dan 10 kolom berdasarkan Indikator Pendidikan dengan fokus pada jenjang Sekolah Menengah Atas (SMA) di seluruh provinsi di Indonesia. Atribut yang digunakan yaitu

Provinsi, Rasio Murid-Guru, Rasio Murid-Kelas, Mengulang, Putus Sekolah, Angka Partisipasi Kasar (APK), Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), Rata-Rata Lama Sekolah (RLS), dan Tingkat Penyelesaian Pendidikan. Data yang ditampilkan pada tahap ini hanya 5 data sampel yang diambil dari sumber BPS dan Open Data Kemendikbud.

Tabel 1. Dataset Asli

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1.	Aceh	8	21	702	353	83	94	72	9	74
2.	Sumatera Utara	15	29	363	1721	79	98	69	10	74
3.	Sumatera Barat	12	28	648	204	84	92	69	9	69
4.	Riau	12	28	209	190	78	85	65	9	68
5.	Jambi	11	26	203	128	72	85	61	9	67
...
102.	Papua	14	25	0	586	64	75	44	7	33

Tabel 2. Penjelasan Atribut Data Penelitian

Variabel	Atribut	Deskripsi
X ₁	Rasio Murid-Guru	Hasil perbandingan dari data jumlah siswa dan jumlah kepala sekolah dan guru, dan jumlah tenaga kependidikan
X ₂	Rasio Murid-Kelas	Hasil perbandingan dari data jumlah siswa dan jumlah ruang kelas
X ₃	Mengulang	Jumlah siswa SMA yang mengulang di setiap provinsi.
X ₄	Putus Sekolah	Jumlah siswa SMA yang mengalami putus sekolah di setiap provinsi.
X ₅	Angka Partisipasi Kasar (APK)	Perbandingan antara jumlah penduduk yang sedang menempuh pendidikan di tingkat tertentu (tanpa memperhitungkan usia mereka) dengan jumlah penduduk yang memenuhi kriteria resmi sebagai usia sekolah pada tingkat pendidikan yang sama. Semakin tinggi APK, maka semakin banyak anak usia sekolah yang menempuh pendidikan di jenjang tertentu di suatu wilayah.
X ₆	Angka Partisipasi Sekolah (APS)	Persentase penduduk dalam kelompok usia sekolah tertentu yang sedang menempuh pendidikan (tanpa memperhitungkan jenjang pendidikan yang ditempuh) dibandingkan dengan jumlah total penduduk dalam kelompok usia sekolah tersebut. Semakin tinggi APS, maka semakin banyak anak usia sekolah yang bersekolah di suatu daerah.
X ₇	Angka Partisipasi Murni (APM)	Proporsi dari penduduk kelompok usia sekolah tertentu yang sedang bersekolah tepat di jenjang pendidikan yang seharusnya. APM selalu lebih rendah dibanding APK karena pembilangnya lebih kecil sementara penyebutnya sama.
X ₈	Rata-Rata Lama Sekolah (RLS)	Rata-rata jumlah tahun yang dihabiskan oleh penduduk dalam menjalani pendidikan formal (Maksimal 12 tahun).
X ₉	Tingkat Penyelesaian Pendidikan	Persentase jumlah siswa yang menyelesaikan suatu jenjang pendidikan tertentu dalam jangka waktu yang ditetapkan.

3.3 Preparation

Pada tahap ini akan dilakukan beberapa langkah seperti: Pembersihan data, dan Normalisasi data. Data yang ditampilkan pada tahap ini hanya 5 data sample yang diambil dari sumber BPS dan Open Data Kemendikbud:

a. Pembersihan Data

Data yang akan digunakan harus dipastikan bahwa tidak terdapat data yang hilang (*missing value*), duplikasi, data pencilon (*outliers*), dan tipe data yang tidak sesuai. Pada penelitian ini, keseluruhan data yang digunakan tidak memiliki data yang hilang atau duplikasi. Namun, terdapat beberapa atribut yang mengalami *outliers*, tetapi masalah *outliers* tidak ditangani di penelitian ini karena data

asli sangat penting untuk hasil model agar tidak ada informasi yang hilang.

b. *Aggregasi Data*

Pada tahap *aggregasi data* dilakukan untuk meringkas data tanpa mengurangi kualitas data tersebut. Atribut yang di *aggregasi* meliputi Rasio Murid-Guru, Rasio Murid-Kelas, Mengulang, Putus Sekolah, Angka Partisipasi Kasar (APK), Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), Rata-Rata Lama Sekolah (RLS), dan Tingkat Penyelesaian Pendidikan. Setelah di lakukan *agregasi*, jumlah data yang digunakan adalah 34 baris dari 102 baris data yang berasal dari data awal. Data yang dihasilkan mencakup 34 provinsi di Indonesia, dengan akumulasi dari tahun 2021 hingga 2023. Hasil

dari tahap ini yang akan di tampilkan hanya 5 sampel data yang dapat dilihat pada Tabel 3.

c. Transformasi Data

Pada tahap ini dilakukan perubahan data terhadap angka positif menjadi angka negatif yang terdapat

pada atribut Mengulang dan Putus Sekolah. Perubahan data ini perlu dilakukan karena setiap atribut harus memiliki karakter data yang sama yaitu angka semakin besar maka angka tersebut menjadi lebih bagus. Perubahan data dapat dilihat pada Tabel 4.

Tabel 3. Hasil Agregasi Data

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1.	Aceh	8	20	234	448	83	93	71	9	73
2.	Bali	13	31	1	27	84	91	75	9	76
3.	Banten	15	29	62	308	69	75	60	9	68
4.	Bengkulu	10	27	39	131	80	94	67	9	64
5.	DI Yogyakarta	11	26	26	9	90	90	74	10	89
...
34.	Sumatera Utara	15	29	121	1322	79	97	68	10	75

Tabel 4. Hasil Transformasi Data

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1.	Aceh	0	0	0,19	0,67	0,71	0,81	0,86	0,54	0,69
2.	Bali	0,54	1	1	0,99	0,77	0,72	1	0,54	0,76
3.	Banten	0,92	0,82	0,79	0,77	0,18	0	0,47	0,54	0,58
4.	Bengkulu	0,29	0,61	0,87	0,91	0,58	0,85	0,7	0,54	0,5
5.	DI Yogyakarta	0,33	0,52	0,91	1	1	0,67	0,94	0,77	1
...
34.	Sumatera Utara	0,88	0,76	0,58	0	0,56	0,99	0,76	0,77	0,72

Tabel 5. Hasil Normalisasi Data

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1.	Aceh	0	0	0,19	0,67	0,71	0,81	0,86	0,54	0,69
2.	Bali	0,54	1	1	0,99	0,77	0,72	1	0,54	0,76
3.	Banten	0,92	0,82	0,79	0,77	0,18	0	0,47	0,54	0,58
4.	Bengkulu	0,29	0,61	0,87	0,91	0,58	0,85	0,7	0,54	0,5
5.	DI Yogyakarta	0,33	0,52	0,91	1	1	0,67	0,94	0,77	1
...
34.	Sumatera Utara	0,88	0,76	0,58	0	0,56	0,99	0,76	0,77	0,72

d. Normalisasi Data

Pada tahap ini agar dapat mengatasi masalah rentang selisih yang besar pada data, perlu dilakukan Normalisasi Data. Normalisasi akan mengubah data ke dalam rentang angka yang lebih kecil, biasanya antara 0 hingga 1. Sehingga, bobot angka pada masing-masing data menjadi lebih terukur. Berikut adalah hasil normalisasi data dalam Tabel 5.

3.4 Modeling

Pada tahap ini dijelaskan hasil dari model terbaik yaitu pemodelan menggunakan algoritma *Agglomerative Clustering* dengan 2 Cluster. Hal ini sesuai dengan hasil komparasi model yang sudah dilakukan sebelumnya. Sehingga penelitian ini menggunakan algoritma *Agglomerative Clustering* karena hasil nilai *Davies Bouldin Index* (DBI) menunjukkan nilai yang terbaik dibandingkan dengan model lainnya. Hasil dari komparasi model dapat dilihat pada tabel 6 untuk melihat hasil DBI dari setiap Cluster, dan Tabel 7 untuk melihat hasil model terbaik dari setiap Cluster.

Tabel 6. Komparasi Model

Jumlah Cluster	Agglomerative	K-Means	K-Medoids
2	0,497	1,488	2,229
3	0,521	1,419	2,410
4	0,522	1,345	1,790
5	0,748	1,291	1,860
6	0,804	1,117	1,817
7	0,879	1,086	1,552
8	0,818	0,946	1,499
9	0,751	1,027	1,453

Tabel 7. Hasil Komparasi Model dengan DBI

Model	Jumlah Cluster	Davies Bouldin Index
Agglomerative	2	0,497
K-Means	8	0,946
K-Medoids	9	1,453

Setelah diketahui algoritma yang akan digunakan yaitu *Agglomerative Clustering*. langkah selanjutnya adalah melakukan pengelompokan menggunakan jarak *Euclidean* dengan pembentukan Cluster. Dalam hal ini, metode jarak yang akan dipilih adalah yang

menghasilkan nilai *Cophenetic correlation coefficient* tertinggi setelah membandingkan berbagai metode *linkage*, seperti *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward Linkage*. Penerapan metode ini umum digunakan dalam *Clustering* hierarki karena proses *Clustering* didasarkan pada pengelompokan nilai jarak terdekat antar anggota. Hasil dari perbandingan *Cophenetic correlation coefficient* dapat dilihat pada tabel 8.

Tabel 8. Hasil Komparasi Linkage

<i>Linkage Method</i>	<i>Cophenetic Correlation Coefficient</i>
<i>Single</i>	0,669
<i>Complete</i>	0,479
<i>Average</i>	0,736
<i>Ward</i>	0,501

Berdasarkan hasil perbandingan nilai *cophenetic correlation coefficient*, metode *Average linkage* menunjukkan nilai tertinggi yang mendekati 1. Hal ini menunjukkan bahwa metode *Average linkage* memiliki korelasi yang paling baik dalam menggambarkan jarak terdekat antar anggota dibandingkan dengan metode *linkage* lainnya. Sehingga dapat ditentukan parameter yang digunakan dalam model *Agglomerative Clustering* adalah metode *Average linkage* dan metrik *Euclidean distances*. Kemudian jumlah *Cluster* yang optimal yaitu 2 *Cluster* berdasarkan hasil dari *Davies Bouldin Index* (DBI).

Tabel 1. Pengelompokan Berdasarkan Hasil *Cluster*

<i>Cluster</i>	<i>Anggota</i>
1	Aceh, Maluku Utara, Maluku, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Utara, Kalimantan Timur, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Bali, Banten, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Papua Barat, Lampung, Kep. Riau, Dki Jakarta, Di Yogyakarta, Kep. Bangka Belitung, Jawa Barat, Jawa Timur, Sumatera Utara, Jawa Tengah
2	Papua

Berdasarkan hasil model yang telah dilakukan bahwa terdapat 2 klaster. Klaster 1 terdiri dari 33 provinsi yang tergolong dalam kelompok dengan tingkat pendidikan tinggi, sementara klaster 2 terdiri dari 1 provinsi yang tergolong dalam kelompok dengan tingkat pendidikan rendah.

3.5 Evaluation

Pada tahap ini, Model terbaik akan dievaluasi dengan membandingkan nilai *Davies Bouldin Index* (DBI). Tujuan dari evaluasi ini adalah untuk menentukan jumlah *Cluster* yang optimal. Jumlah *Cluster* yang dibandingkan antara 2 hingga 9 *Cluster*.

Tabel 2. Evaluasi *Agglomerative* dengan DBI

<i>Jumlah Cluster</i>	<i>Agglomerative</i>
2	0,497
3	0,521
4	0,522
5	0,748
6	0,804
7	0,879
8	0,818
9	0,751

Berdasarkan hasil nilai *Davies Bouldin Index* (DBI), *Cluster* yang optimal adalah 2 *Cluster* dengan nilai terbaik sebesar 0,497. Oleh karena itu, hasil dari model *Agglomerative Clustering* untuk dapat mengelompokkan provinsi di Indonesia berdasarkan Indikator Pendidikan untuk jenjang Sekolah Menengah Atas (SMA) dapat dibagi menjadi 2 *Cluster*. *Cluster* 1 terdiri dari 33 provinsi yang tergolong dalam kelompok dengan tingkat pendidikan tinggi. Sementara itu, *Cluster* 2 terdiri dari 1 provinsi yang tergolong dalam kelompok dengan tingkat pendidikan rendah.

4. KESIMPULAN

Berdasarkan hasil penelitian dengan menggunakan algoritma *Agglomerative Hierarchical Clustering* pada data Indikator Pendidikan untuk jenjang Sekolah Menengah Atas (SMA) di tahun 2021-2023. Didapatkan hasil nilai *Davies Bouldin Index* (DBI) paling rendah sebesar 0,497. Berdasarkan hasil dari perbandingan jarak bahwa metode *average linkage* memiliki nilai *Cophenetic correlation coefficient* terbesar yaitu 0,736. Sedangkan metode *Single linkage*, *Complete linkage*, dan *Ward linkage* memiliki nilai *Cophenetic correlation coefficient* lebih rendah yaitu 0,669; 0,479; dan 0,501. Sehingga hasil model menggunakan *Agglomerative Clustering* dengan jarak *Average linkage* dapat membagi provinsi di Indonesia menjadi 2 *Cluster*, yaitu: *Cluster* 1 terdiri dari 33 provinsi dengan tingkat pendidikan yang tinggi, sedangkan *Cluster* 2 terdiri dari 1 provinsi, yaitu Papua, dengan tingkat pendidikan yang rendah. Penyebab papua termasuk dalam kelompok dengan tingkat pendidikan rendah karena berdasarkan data bahwa tingkat partisipasi pendidikan dan kualitas siswa yang masih sangat rendah, yang berdampak pada rendahnya kualitas pendidikan di provinsi tersebut. Kesimpulan ini didasarkan pada aspek partisipasi pendidikan serta hasil dan pencapaian pendidikan.

Adapun saran dari penelitian yang dilakukan terkait klasterisasi provinsi di Indonesia berdasarkan Indikator Pendidikan di jenjang Sekolah Menengah Atas (SMA) sebagai berikut:

- a. Penelitian ini dapat dikembangkan lebih lanjut dengan menggunakan atribut, metode, dan studi kasus yang berbeda untuk memperoleh hasil yang lebih baik dan relevan.

- b. Hasil *Clustering* yang diperoleh dapat memberikan masukan bagi pemerintah dalam mengevaluasi provinsi yang masih memiliki kualitas pendidikan rendah.

DAFTAR PUSTAKA

- [1] A. P. Fialine, D. A. Alodia, D. Endriani, and E. Widodo, 'Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan', *Sepren*, vol. 2, no. 2, pp. 1–13, Nov. 2021, doi: 10.36655/sepren.v2i2.606.
- [2] M. R. Putri, G. Satya Nugraha, and R. Dwiyanaputra, 'Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering', *J-COSINE (Journal of Computer Science and Informatics Engineering)*, vol. 7, no. 1, Jun. 2023, doi: <https://doi.org/10.29303/jcosine.v8i1>.
- [3] 'Dataset APK, APM, APS, dan Penyelesaian Pendidikan', Badan Pusat Statistik. [Online]. Available: <https://www.bps.go.id/id>
- [4] A. F. Dewi and K. Ahadiyah, 'Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan', *Zeta - Math Journal*, vol. 7, no. 2, pp. 57–63, Nov. 2022, doi: 10.31102/zeta.2022.7.2.57-63.
- [5] Adiatma and N. Fitrah, 'Analisis Cluster Untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Pendidikan dengan Metode Ward', *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 12, no. 1, 2024.
- [6] M. Dearivany, 'Penerapan K-Means dan Agglomerative Hierarchical Clustering Untuk Pengelompokan Data Indikator Pendidikan (Studi Kasus Kabupaten/Kota di Wilayah Indonesia Timur)', 2020.
- [7] N. HOTZ, 'What is CRISP DM?', *datascience-pm.com*. Accessed: Jul. 06, 2024. [Online]. Available: <https://www.datascience-pm.com/crisp-dm-2/>
- [8] 'Dataset Siswa, Guru, Mengulang, dan Putus Sekolah', Portal Data Kemendikbudristek. [Online]. Available: <https://data.kemdikbud.go.id/>
- [9] R. S. Wahono, *Data Mining*. 2020.
- [10] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. Jambi: PT. Sonpedia Publishing Indonesia, 2024.
- [11] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. 2020.
- [12] A. Khoirunnisa, F. A. S. Wibowo, and K. Kismiantini, 'Perbandingan Analisis Agglomerative Hierarchical Clustering Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat', *Research Gate*, vol. 7, 2023, doi: 10.21831/pspmm.v7i1.273.
- [13] R. Ishak and Amiruddin, 'Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means', *Jambura Journal of Electrical and Electronics Engineering*, vol. 6, no. 1, Jan. 2024.
- [14] R. N. Puspita, 'Perbandingan Metode Centroid Dan Ward Dalam Pengelompokan Tingkat Penyelesaian Pendidikan Di Indonesia', *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 3, no. 3, Dec. 2022, doi: 10.46306/lb.v3i3.

ANALISIS SENTIMEN ULASAN PENGGUNA GOJEK DAN INDRIVE PADA GOOGLE PLAYSTORE DENGAN ALGORITMA K-NEAREST NEIGHBOR

Teguh Febriyanto^{1*}, Achmad Solichin²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ^{1*}tfebriyan10@gmail.com, ²achmad.solichin@budiluhur.ac.id

(Naskah masuk: 7 Agustus 2025, diterima untuk diterbitkan: 7 September 2025)

Abstrak

Transportasi adalah sarana atau suatu hal kendaraan yang umum digunakan sebagian masyarakat di Indonesia untuk melakukan kegiatan memindahkan barang ataupun manusia dari suatu tempat ke tempat yang lainnya. Transportasi sangat diperlukan dalam kegiatan sehari-hari namun seiring berjalannya waktu, banyak penumpukan alat transportasi di kota-kota besar seperti kota metropolitan Jakarta, tidak jarang juga ditemukan penyebab kecelakaan disebabkan terlalu banyak transportasi di suatu daerah. Maka pemerintah melakukan penghimbauan agar masyarakat terutama di daerah perkotaan padat penduduk menggunakan transportasi umum. Transportasi umum itu banyak jenisnya dari angkutan umum hingga transportasi online yang lagi trend masa kini, banyak developer maupun perusahaan dari dalam negeri maupun luar negeri. Dengan banyaknya pilihan transportasi online, dibutuhkan data ulasan pengguna yang digunakan untuk melihat tingkat kepuasan atau komentar dari pengguna dari aplikasi tersebut semisal dari gojek dan indrive. Oleh karena itu dilakukan analisis sentimen dari data ulasan pengguna aplikasi gojek dan indrive yang diambil melalui platform google playstore yang diambil pada bulan juni 2024 menggunakan google playstore scraping. Dari data kotor yang didapat, data tersebut di bersihkan melalui proses preprocessing lalu diberi label dengan menggunakan lexicon based setelah itu dilakukan split data yang menghasilkan data training dan juga data testing. Setelah di dapatkan data training lalu di hitung bobot perkata dengan TF-IDF dan nantinya data testing akan diklasifikasikan dengan metode algoritma K-Nearest Neighbor (KNN) menjadi dua label terpisah yaitu positif dan negatif. Bahwa dalam pengujian menggunakan K-Nearest Neighbor (KNN), dengan menggunakan k yang sama yaitu k=3, didapatkan akurasi pada dataset aplikasi gojek yaitu 72% dengan menggunakan rasio data 80:20 dari 860 data, sedangkan didapatkan akurasi pada dataset aplikasi indrive yaitu 85% dengan menggunakan rasio 80:20 dari 1477 data.

Kata kunci: *Gojek, Indrive, Analisis Sentimen, K-Nearest Neighbor (KNN), Text mining*

SENTIMENT ANALYSIS OF USER REVIEWS FOR GOJEK AND INDRIVE ON GOOGLE PLAYSTORE USING THE K-NEAREST NEIGHBOR ALGORITHM

Abstract

Transportation is a means or a vehicle that is commonly used by some people in Indonesia to carry out activities to move goods or people from one place to another. Transportation is very necessary in daily activities but over time, there is a lot of accumulation of transportation equipment in big cities such as the metropolitan city of Jakarta, it is not uncommon to find the cause of accidents due to too much transportation in an area. So the government makes an appeal so that people, especially in densely populated urban areas, use public transportation. There are many types of public transportation from public transportation to online transportation which is the current trend, many developers and companies from within the country and abroad. With so many online transportation options, user review data is needed which is used to see the level of satisfaction or comments from users of these applications such as gojek and indrive. Therefore, sentiment analysis is carried out from the gojek and indrive application user review data taken through the google playstore platform which was taken in June 2024 using google playstore scraping. From the dirty data obtained, the data is cleaned through a preprocessing process and then labeled using lexicon based after that split data is carried out which results in training data and also testing data. After getting the training data, the weight of each word is calculated with TF-IDF and laer the testing data will be classified by the K-Nearest Neighbor (KNN) algorithm method into two separate labels, namely positive and negative. That in testing using K-Nearest Neighbor (KNN), using the same k,

namely $k = 3$, the accuracy obtained on the gojek application dataset is 72% using a data ratio of 80: 20 of 860 data, while the accuracy obtained on the indrive application dataset is 85% using a ratio of 80: 20 of 1477 data.

Keywords: Gojek, Indrive, Sentiment Analysis, K-Nearest Neighbor (KNN), Text mining

1. PENDAHULUAN

Transportasi adalah sarana atau suatu hal kendaraan yang umum digunakan sebagian masyarakat di Indonesia untuk melakukan kegiatan memindahkan barang ataupun manusia dari suatu tempat ke tempat yang lainnya [1]. Transportasi sangat diperlukan dalam kegiatan sehari-hari namun seiring berjalannya waktu, banyak penumpukan alat transportasi di kota-kota besar seperti kota metropolitan Jakarta, tidak jarang juga ditemukan penyebab kecelakaan disebabkan terlalu banyak transportasi di suatu daerah. Maka pemerintah melakukan pengimbuhan agar masyarakat terutama di daerah perkotaan padat penduduk menggunakan transportasi umum.

Oleh sebab itu banyak dinamika yang membuat kebingungan memilih berbagai macam aplikasi ojek *online* yang dapat digunakan dalam kegiatan sehari-hari. Maka dari itu dibutuhkan suatu sistem yang dapat mengklasifikasi sentimen ulasan seseorang otomatis dalam kelas positif dan negatif yang nantinya akan menampilkan beberapa saran dari ulasan pengguna lainnya yang akan ditujukan untuk pengguna yang bingung memilih aplikasi ojek *online* apa saja. Google *play store* sebagai wadah yang menjadi tempat pengunduhan aplikasi ini memiliki fitur yang cukup bermanfaat untuk melihat ulasan-ulasan pengguna aplikasi tersebut. Fungsi awal dari adanya ulasan pengguna ini bisa dimanfaatkan sebagai tolak ukur yang efektif dan juga efisien untuk menemukan informasi yang valid pada suatu aplikasi tertentu. Ulasan yang muncul berupa saran positif dan negatif dari suatu aplikasi tersebut. Untuk melihat dan menyortir beberapa ulasan tersebut secara manual cukup sulit dilakukan. Maka ada sebuah metode yang digunakan untuk memudahkan melakukan pekerjaan tersebut.

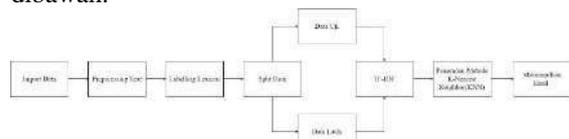
Analisis Sentimen adalah teknik yang digunakan. Ini adalah proses pengolahan data teks yang menilai apakah pendapat dalam kalimat positif atau negatif [2]. Banyak algoritma yang terkait dengan analisis sentimen, salah satunya adalah Algoritma K-Nearest Neighbor (KNN). Algoritma ini adalah metode klasifikasi yang mengelompokkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya berdasarkan jarak terdekat, atau nilai K yang biasanya disebut [3]. Studi sebelumnya tentang analisis sentimen termasuk penelitian yang mengevaluasi penggunaan algoritma K-Nearest Neighbor untuk mengategorikan ulasan pengguna dari aplikasi Digital Korlantas POLRI. Penelitian ini menggunakan 600 data ulasan pengguna, yang dibagi menjadi 70% untuk data latihan dan 30% untuk data uji. Nilai K dicari dengan *cosine similarity*, dengan nilai akurasi tertinggi $k=9$ [4]. Studi sebelumnya yang

menggunakan algoritma naive bayes untuk menganalisis sentimen ulasan pengguna untuk aplikasi MyPertamina di Google Playstore, yang menggunakan 3498 data, menemukan bahwa penelitian tersebut memiliki nilai akurasi sebesar 91%, presisi sebesar 92%, dan *recall* sebesar 100% [5].

Fokus dari penelitian ini adalah melakukan sentimen analisis tentang dua perusahaan ojek *online* yang berada di Indonesia berdasarkan ulasan pengguna yang tersedia di Google Playstore menggunakan metode K-Nearest Neighbor (KNN). Penggunaan algoritma K-Nearest Neighbor (KNN) dengan menggunakan fitur TF-IDF ini diharapkan dapat mengklasifikasikan ulasan pengguna yang berada di kolom komentar Google *Playstore* dengan baik, sehingga akan menghasilkan klasifikasi yaitu ulasan positif dan negatif. Hasil pengujian akan dilakukan menggunakan *confusion matrix*, dan nantinya sistem analisis sentimen ini akan berbentuk aplikasi berbasis web dan metode yang digunakan untuk mengumpulkan data dalam penelitian ini adalah web *scraping*.

2. METODE PENELITIAN

Dalam penelitian ini yang dikembangkan adalah sistem informasi yang menggunakan implementasi metode K-Nearest Neighbor(KNN). Dimana dalam sistem tersebut ada tahapan-tahapan yaitu import data, labeling data, preprocessing data, dan juga perhitungan TFIDF beserta perhitungan prediksi dari metode K-Nearest Neighbor. Tahapan metode penelitian tersebut dapat di lihat pada Gambar 1 dibawah.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Data ulasan yang didapatkan dari platform Google *Play Store* ini diambil dari komentar ulasan pengguna dari aplikasi Gojek dan Indrive dengan teknik *scraping* menggunakan *library google colab*. Yang dilakukan pada tahapan ini adalah menarik ulasan Gojek dan *Indrive* di *platform Google Playstore* ke dalam *file* berbentuk CSV lalu disimpan ke dalam *database* sebagai data kotor yang nantinya akan diproses dan digunakan untuk penelitian kali ini dan dapat di implementasikan ke dalam algoritma K-NN untuk analisis sentimen. Data yang didapatkan dalam ulasan Gojek sekitar 1500 data kotor.

2.2 Preprocessing Data

Text preprocessing atau prapemrosesan teks adalah langkah awal dalam melakukan text mining.

Prapemrosesan teks biasanya melibatkan penghapusan data yang tidak relevan atau transformasi data teks menjadi format yang lebih mudah di proses oleh sistem [6]. Dalam *preprocessing* ini terdapat beberapa tahapan yaitu *case folding*, *cleansing*, *slangword*, *stopword removal*, *tokenizing*, *stemming*.

- Case folding* yaitu suatu proses yang dilakukan untuk mengubah semua huruf yang berbentuk kapital dalam suatu kalimat menjadi huruf yang berbentuk kecil (*lowercase*).
- Cleansing* suatu proses yang dilakukan untuk menghilangkan beberapa karakter yang tidak dibutuhkan di dalam dokumen.
- Slangword* yaitu suatu proses yang dilakukan untuk mengubah atau menormalisasikan kata yang dilakukan dengan memperbaiki kata-kata yang disingkat menjadi kata yang memiliki arti sama, sesuai yang ada di dalam KBBI.
- Stopword removal* yaitu suatu proses yang dilakukan untuk mengambil kata-kata yang penting dan membuang kata-kata yang kurang penting dalam suatu dokumen. Pada tahapan ini teks sebelum dilabeli akan di hilangkan terlebih dahulu teks yang tidak ada hubungannya dengan analisis sentimen sehingga besarnya teks akan berkurang tanpa mengurangi isi sentimen teks [7].
- Tokenizing* yaitu suatu proses pemisahan atau pemecahan dokumen yang asalnya dari suatu kalimat menjadi potongan kata berdasarkan spasi, sehingga menghasilkan term-term yang terpisah.
- Stemming* yaitu suatu proses yang dilakukan untuk mengolah kata-kata yang memiliki imbuhan menjadi kata dasar dengan aturan-aturan tertentu [7].

2.3 Labelisasi Kamus *Lexicon*

Labeling adalah suatu proses yang memberikan klasifikasi berdasarkan karakteristik yang ada di dalam sebuah kalimat dalam suatu dokumen[8]. Pelabelan yang dilakukan atau biasa disebut pemberian polaritas pada dataset yang sudah bersih dalam tahap *preprocessing*. Pengelompokan polaritas dilakukan dengan cara otomatis menggunakan *lexicon* dengan bobot polaritas yang berbeda. Nilai positif dinyatakan 1-5 bobot, sedangkan -1 sampai dengan -5 dinyatakan sebagai nilai negatif. [9].

Dalam proses pelabelan, leksikon dengan kamus tambahan digunakan untuk menentukan kata-kata yang digunakan. Untuk menghitung polaritas, bobot kata dalam setiap komentar dijumlahkan. Hasilnya menunjukkan apakah polaritas positif atau negatif. Komentar dianggap positif jika polaritas positif, dan komentar dianggap negatif jika polaritas negatif.

2.4 TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah suatu metode pembobotan yang dilakukan untuk menghitung pada

setiap bobot kata yang terdapat dalam suatu dokumen yang tersedia. Dokumen-dokumen tersebut yang nantinya akan diubah menjadi vektor dengan jumlah kata (term) yang digunakan untuk pengklasifikasian[10]. Berikut rumus TF-IDF:

$$a. \text{ Term Frequency (TF)} \\ tf_{(ka)} = \frac{\text{jumlah frekuensi istilah } k \text{ yang muncul dalam dokumen } d}{\text{jumlah istilah dalam dokumen}} \quad (1)$$

$$b. \text{ Inverse Document Frequency (IDF)} \\ idf_{(k)} = \log \frac{N}{df_k} \quad (2)$$

$$c. \text{ TF-IDF} \\ tfidf_{(k,d)} = tf_{(k,d)} * idf_{(k)} \quad (3)$$

Keterangan :

$tf_{(ka)}$ = Frekuensi kemunculan kata (*term*) k dalam dokumen d

k = Suatu kata (*term*)

d = Suatu dokumen

$idf_{(k)}$ = Inverse Document Frequency k

N = Jumlah dokumen yang ada di dalam *database*

df_k = Jumlah dokumen yang mengandung kata (*term*) k

$tf - idf_{(k,d)}$ = Bobot kata (*term*) k terhadap dokumen d

2.5 Algoritma K-NN

Algoritma K-Nearest Neighbor (KNN) adalah suatu metode yang termasuk di dalam *supervised learning* dan melibatkan proses pembelajaran dari data *training* [4]. menjelaskan cara kerja dasar dari metode KNN adalah mencari jarak terpendek antara data yang akan direklasifikasi dengan data latih yang ada dalam lingkungan k (jarak minimum). Data yang memiliki jarak terkecil dengan kategori/kelas terbesar akan menjadi kategori/kelas baru untuk data uji (testing data). Pada penelitian ini menggunakan perhitungan jarak *Euclidean Distance* [11] yang ditunjukkan pada persamaan.

2.6 Confusion Matrix

Confusion matrix adalah sebuah metode yang seringkali dimanfaatkan dalam penelitian untuk mengetahui atau evaluasi dari hasil dan pengukuran performa dalam suatu metode klasifikasi. Dalam hal ini penggunaan *confusion matrix* sangatlah penting dikarenakan untuk mengukur sejauh mana sistem yang dibuat dapat melakukan proses klasifikasi data dengan baik. Pada *confusion matrix*, ada beberapa perhitungan yang akan dilakukan yaitu akurasi, presisi, dan *recall*. Di bawah ini adalah perhitungan akurasi, *precision* dan *recall* [12] yang dapat dilihat pada persamaan sebagai berikut:

$$accuracy = \frac{TP+TN}{TP+FN+FP+TN} X 100\% \quad \dots (5)$$

$$precision = \frac{TP}{TP+FP} X 100\% \quad \dots (6)$$

$$recall = \frac{TP}{TP+FN} X 100\% \quad \dots (7)$$

$$f1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \dots (8)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dimasukkan dalam penelitian kali ini di dapat dari ulasan aplikasi Gojek dan juga Indrive pada Google Play Store dan selanjutnya data di import kedalam database sistem. Tabel 1 menyajikan contoh dari data ulasan yang diambil pada aplikasi Gojek dan *Indrive* di *Google Play Store*.

Tabel 1. Tabel Komentar Aplikasi Indrive

No	Username	Content	Score
1	Naila Square	auto bid nya di aktifkan lagi dong	5
2	Dian Abdul Hakim	Parah malah ada peringatan padahal pake aplikasi ori dari awal juga	1
3	Hana Priyandi	gila yaaaa fungsi autobit itu buat Nerima orderan secara otomatis kn kenapa ga berfungsi sih kalo buat sistem yg bener dong makin kalah aja sama apk sebelah hadeeeeh indriver indriver kacau lu	1
...
1,993	Terry Irwansyah	Mayan,buat nambah penghasilan	5

Tabel 2. Tabel Komentar Aplikasi Gojek

No	Username	Content	Score
1	Rachel mutiara adita	Ongkirnya g ngotak mahalnya padahal dekat loh jaraknya	1
2	Ardi Budiman	kadang klo lagu hujan driver nya jauh2	2
3	Nizar Ramadhan	saldo ga balik padahal di sana tertera gagal	1
...
1,993	Tia Lovely	Go-Jek banyak diskonnya mantaap	5

3.2 Preprocessing Data

Tahap selanjutnya yaitu melakukan pembersihan data dimana data tersebut masih dalam keadaan kotor, lalu dibersihkan menjadi cleantext yang nantinya dapat di proses untuk perhitungan TF-IDF dan juga klasifikasi KNN. berikut adalah tahapan preprocessing yang dilakukan dalam penelitian kali ini:

a. Case Folding

Proses *case folding* yaitu melakukan perubahan huruf kapital menjadi huruf kecil pada seluruh data yang ada

di dalam dokumen tersebut. Proses *case folding* dapat di lihat pada tabel 3.

Tabel 3. Case Folding

Content	Casefolding
Iklan Gojek SANGAT MENGGANGGU SAAT MAIN GAME...!!!!!! TIAP MENIT IKLAN GOJEK TERUS!!!	iklan gojek sangat mengganggu saat main game...!!!!!! tiap menit iklan gojek terus!!!

b. Cleansing

Proses *cleansing* yaitu suatu tahapan yang berguna untuk menghilangkan karakter atau simbol-simbol yang tidak diperlukan pada data ulasan pengguna, termasuk tanda baca, angka, URL, dan juga simbol (,,"~&?!><#%{})([0-9]+;,:"). Proses *cleansing* dapat di lihat pada tabel 4.

Tabel 4. Cleansing

Casefolding	Cleansing
Kenapa promo cuma paylater? Dipaksa nyicil? Ngutang?	kenapa promo cuma paylater dipaksa nyicil ngutang

c. Tokenisasi

Proses *tokenisasi* yaitu suatu tahapan yang dilakukan untuk memecahkan sebuah kalimat menjadi potongan-potongan kata yang berdiri sendiri, proses ini dilakukan agar mempermudah untuk melakukan proses *slangword* dan juga *stopword*. Proses tokenisasi dapat di lihat pada tabel 5.

Tabel 5. Tokenisasi

Cleansing	Tokenisasi
kenapa promo cuma paylater? dipaksa nyicil? ngutang?	["kenapa", "promo", "cuma", "paylater", "dipaksa", "nyicil", "ngutang"]

d. Slangword

Proses *slangword* yaitu suatu tahapan yang berguna untuk mengubah atau mengganti kata-kata yang disingkat menjadi kata-kata yang terdaftar di Kamus Besar Bahasa Indonesia (KBBI) dengan arti sama tanpa merubah maksud dari kata tersebut. Proses *slangword* dapat di lihat pada tabel 6.

Tabel 6. Slangword

Tokenisasi	Slangword
aplikasinya tidak dapat dibuka, padahal sdh sy update, mhn ijin dibantu admin, trims	aplikasinya tidak dapat dibuka, padahal sudah saya update, mohon ijin dibantu admin, terimakasih

e. Stopword Removal

Stopword removal yaitu sebuah proses yang dilakukan untuk menghilangkan kata-kata yang tidak

diperlukan atau tidak penting selama proses preprocessing. Proses *slangword* dapat di lihat pada tabel 7.

Tabel 7 *Stopword Removal*

<i>Slangword</i>	<i>Stopword Removal</i>
pelayanan ramah dan sopan	pelayanan ramah sopan

f. *Stemming*

Stemming adalah sebuah proses yang dilakukan untuk mengubah kata-kata yang memiliki awalan atau akhiran yang termasuk kata sambung atau kata imbuhan menjadi kata dasar di dalam Kamus Besar Bahasa Indonesia (KBBI). Proses *slangword* dapat di lihat pada tabel 8.

Tabel 8. *Stemming*

<i>Stopword Removal</i>	<i>Stemming</i>
sangat banget <u>membantu</u>	sangat banget <u>bantu</u>

3.3 *Labelisasi Kamus Lexicon*

Pembagian label pada komentar dilakukan untuk memberikan emosi kepada komentar sehingga pesan dan kesan dapat diidentifikasi. Positif adalah sentuhan yang menunjukkan dukungan atau kesetujuan, sedangkan negatif adalah sentuhan yang menunjukkan ketidakpuasan dari komentar padangan. Sentimen yang tidak menunjukkan aspek positif atau negatif disebut *neutral*. Untuk mengetahui perasaan dari komentar, proses pelabelan ini menggunakan leksikon.

Tabel 9. *Labelisasi kamus lexicon*

<i>Document_id</i>	<i>Cleantext</i>	<i>Label</i>	<i>Skor</i>
1	terima kasih selalu beri nyaman baik	<i>Positive</i>	7
2	moga promo diskon makin banyak biaya ongkos kirim biaya lain lebih murah	<i>Positive</i>	3
3	saat sedang cari driver belum sangkut ke salah satu <i>driver</i> tidak ada pilih batal padahal sudah tunggu lama	<i>Negative</i>	-21

Pada tahapan labelisasi kali ini menggunakan kamus *lexicon*, yang didapat dari github bernama Fajri91, lalu digunakan untuk melabelisasi kalimat di dokumen tersebut adalah positif dan negatif, sehingga pada saat splitdata nantinya akan menggunakan hasil data teks yang sudah dilabelisasi oleh kamus *lexicon*. Proses *labeling* dapat dilihat pada Tabel 9.

3.4 *Split Data*

Pada tahapan pembagian data, data yang sudah di beri label dengan kamus lexicon akan dilakukan proses pembagian data menjadi data training dan juga data testing. Pada penelitian ini dilakukan pembagian data dengan rasio 80:20.

Tabel 10. *Split Data 80:20*

Jenis Data	Jumlah
<i>Data Training</i>	688
<i>Data Testing</i>	172
Jumlah	860

3.5 *Perhitungan Bobot TF-IDF*

TF-IDF adalah suatu tahapan perhitungan bobot menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Pada tahapan ini melibatkan perhitungan nilai TF untuk setiap kata yang muncul setiap di dokumen tersebut. Setelah itu nilai TF di kalikan dengan nilai IDF, yang menggambarkan pentingnya kata tersebut dalam seluruh dokumen yang tersedia. Berikut adalah simulasi perhitungan bobot TF-IDF.

Tabel 11. *Data Training*

<i>Data Ulasan</i>	<i>Label</i>
sarana transportasi bagus mahal	Positif
<i>Data Ulasan</i>	<i>Label</i>
aplikasi jelek masuk akun limit habis suruh tunggu jam	Negatif
jelek ah gocar hemat cancel melulu Gojek the best	Negatif Positif
pesan gocar jarak tidak mau ambil driver tunggu jam tidak ada ambil parah sih	Negatif

a. *Menghitung Term Frequency (TF)*

Setelah data yang sudah di preprocessing menjadi data bersih lalu langkah selanjutnya adalah perhitungan TF untuk setiap kata di dalam suatu dokumen. Perhitungan *Term Frequency* (TF) yang terdapat di Tabel 12.

b. *Menghitung Inverse Document Frequency (IDF)*

Pada perhitungan *Inverse Document Frequency* (IDF) proses ini menggunakan suatu term yang terdapat dalam seluruh dokumen. Untuk mendapatkan nilai IDF adalah dengan menghitung jumlah total dokumen (N) akan dibagi dengan jumlah DF, lalu nilai IDF di dapat dari log hasil perhitungan tersebut. Simulasi perhitungan IDF terdapat di tabel 13.

c. *TF-IDF*

Setelah dilakukan perhitungan IDF, selanjutnya adalah melakukan perhitungan TF-IDF yang didapatkan dari mengalikan nilai TF dengan nilai IDF yang telah dijelaskan sebelumnya. Simulasi perhitungan TF-IDF dapat dilihat pada tabel 14.

Tabel 16. Jarak Euclidian

Jarak Euclidian
$d_{(latih1, uji1)} = 2,3214$
$d_{(latih2, uji1)} = 2,3856$
$d_{(latih3, uji1)} = 2,3521$
$d_{(latih4, uji1)} = 2,0166$
$d_{(latih5, uji1)} = 2,7337$

c. Mencari K Tetangga Terdekat

Proses selanjutnya adalah mencari nilai K terdekat setelah melakukan perhitungan jarak antara data *training* dan juga data *testing*, menentukan jarak terdekat dengan cara diurutkan berdasarkan nilai yang terdekat dengan data *testing* yang sudah di hitung sebelumnya. Tabel berikut menunjukkan hasil klasifikasi yang diperoleh berdasarkan nilai k yang digunakan.

Tabel 17. Mencari K Tetangga

Urutan	Jarak Euclidean	Data ke -	Label Data Training
1	$d_{(latih4, uji1)} = 2,0166$	Testing 1, Training 4	Positif
2	$d_{(latih1, uji1)} = 2,3214$	Testing 1, Training 1	Positif
3	$d_{(latih3, uji1)} = 2,3521$	Testing 1, Training 3	Negatif

Di dapatkan dari k terdekat adalah *data training* 4, *training* 1, dan juga *training* 3. Jumlah perbandingan positif dan negatif yang didapatkan adalah terdapat 2 data dengan label positif dan 1 data dengan label negatif. Sehingga, hasil dari data testing pada pengujian ini memiliki tabel positif.

3.7 Confusion Matrix

Selanjutnya menampilkan Hasil yang di dapat dari klasifikasi KNN itu menghasilkan *Confusion Matrix*, pada gambar berikut adalah *confusion matrix*, akurasi, presisi, *recall*, dan *f1score* dari hasil klasifikasi KNN data gojek dan juga data *indrive*.

Tabel 18. Confusion Matrix Data Gojek

Akurasi	Presisi	Recall	F1score
72%	0%	0%	0%

Tabel 19. Confusion Matrix Data Indrive

Akurasi	Presisi	Recall	F1score
85%	0%	0%	0%

Perbandingan akurasi Indrive dan Gojek

Tabel 20. Perbandingan Hasil Akurasi, Presisi, Recall, dan f1-score

Aplika	Jumlah Data Test	Akura	Presi	Reca	F1score
Gojek	172	72%	0%	0%	0%
Indrive	295	85%	0%	0%	0%

Berdasarkan tabel 20 dapat ditarik kesimpulan bahwa dalam pengujian menggunakan *K-Nearest Neighbor* (KNN), dengan menggunakan k yang sama yaitu $k=3$, didapatkan akurasi pada dataset aplikasi gojek yaitu 72% dengan menggunakan rasio data 80:20 dari 860 data, sedangkan didapatkan akurasi pada dataset aplikasi indrive yaitu 85% dengan menggunakan rasio 80:20 dari 1477 data. Nilai presisi di kedua data tersebut didapatkan hasil 0% dikarenakan didalam perhitungan presisi harus terdapat nilai TP sedangkan TP di dalam sentimen Gojek dan *Indrive* tidak memiliki nilai klasifikasi TP. Begitu juga dengan Recall yang membutuhkan nilai TP untuk mendapatkan hasil *recall* tersebut. Jika *F1score* didapatkan dari pengalian presisi dan recall, jika didapatkan hasil presisi dan *recall* 0% maka *f1score* didapatkan 0% juga.

Didapatkan nilai akurasi dari ulasan Gojek yaitu 72% dari ulasan yang berupa komentar negatif dengan sisanya sekitar 28% berkomentar positif. Ulasan itu didapatkan dari *Google Playstore* lalu diberi labelisasi dengan kamus *lexicon*. Begitu juga dengan nilai ulasan akurasi Indrive yaitu 85% dari ulasan yang berupa komentar negatif dengan sisanya 15% berkomentar positif. Nilai akurasi diatas menjadi besar karena algoritma yang memprediksi kata tersebut sama dengan hasil labelisasi kamus *lexicon*. Hal ini menunjukkan performa sistem aplikasi semakin data yang digunakan banyak, maka semakin akurat hasil yang akan didapatkan, dan juga dalam dataset yang terdapat pada aplikasi gojek lebih banyak merujuk ke positif, dengan akurasi yang lebih kecil menandakan komentar positif masih banyak terdapat di dalam data ulasan pengguna gojek.

4. KESIMPULAN

Penelitian ini berfokus pada penerapan metode klasifikasi teks untuk menganalisis sentimen publik terhadap aplikasi Gojek dan InDrive pada platform *Google Play Store*. Data dikumpulkan dari *Google Play Store* pada bulan Juni 2024, diproses melalui pra-proses, pelabelan, pelatihan data, dan pengujian. Metode *K-Nearest Neighbor* digunakan untuk mengklasifikasikan data, mengevaluasi kinerja model klasifikasi. Matriks konfusi menunjukkan bahwa data yang digunakan lebih akurat, dengan 72% data negatif memiliki hasil lebih positif pada aplikasi Gojek dan 85% pada aplikasi *InDrive*. Penelitian ini juga menemukan bahwa jumlah data secara signifikan memengaruhi kinerja algoritma dan kinerja aplikasi. Saran yang dapat penulis berikan untuk pengembangan sistem ini lebih lanjut adalah sebagai berikut: Pengembang diharapkan untuk mempercantik kembali tampilan aplikasi agar lebih menarik. Selain itu, pengembang selanjutnya diharapkan dapat menambahkan lebih banyak metode untuk memungkinkan perbandingan tingkat performa aplikasi. Penambahan jumlah data yang sangat besar memang dapat mempengaruhi akurasi hasil, namun dengan bertambahnya data, pengembang diharapkan

juga dapat menemukan solusi agar kinerja aplikasi tidak menjadi terlalu lambat.

DAFTAR PUSTAKA

- [1] A. N. Hasanah dan B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, hal. 90–96, 2024, doi: 10.23960/jitet.v12i1.3628.
- [2] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *J. Sains dan Edukasi Sains*, vol. 5, no. 1, hal. 28–37, 2022, doi: 10.24246/juses.v5i1p28-37.
- [3] M. K. Anam, B. N. Pikir, dan M. B. Firdaus, "Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, hal. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.
- [4] E. P. Sutrisno dan S. Amini, "Implementasi Algoritma K-Nearest Neighbor Pada Analisis Sentimen Ulasan Pengguna Aplikasi Digital Korlantas Polri," *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 2, hal. 687–695, 2023.
- [5] G. Darmawan, S. Alam, dan M. I. Sulisty, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *STORAGE – J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, hal. 100–108, 2023.
- [6] utomo budiyanto marlina hidayat, "Sentimen Analisis Tentang Hilirisasi Industri Berdasarkan Opini Masyarakat Di Twitter Menggunakan Metode K-Nearest Neighbor," vol. 2, no. September, hal. 826–835, 2023, [Daring]. Tersedia pada: <http://3.8.6.95/ijcs/index.php/ijcs/article/view/3154%0Ahttp://3.8.6.95/ijcs/index.php/ijcs/article/download/3154/104>.
- [7] M. F. Rizki, W. Pramusinto, M. Hardjianto, dan S. Subandi, "Implementasi Algoritma K-Nearest Neighbors Untuk Analisis Sentimen Aplikasi Jobstreet," *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 1, hal. 267–276, 2023.
- [8] M. Priandi dan Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 2, no. 2, hal. 311–319, 2021.
- [9] S. Mujahidin, B. Prasetyo, dan M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.*, vol. 10, no. 3, hal. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [10] M. Furqan, S. Sriani, dan S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, hal. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [11] A. Yudhana, S. Sunardi, dan A. J. S. Hartanta, "Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon," *Transmisi*, vol. 22, no. 4, hal. 123–129, 2020, doi: 10.14710/transmisi.22.4.123-129.
- [12] S. Juniarsih, E. F. Ripanti, dan E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, hal. 239, 2020, doi: 10.26418/justin.v8i3.39118.

KLASTERISASI TINGKAT KEMISKINAN MENGGUNAKAN METODE K-MEANS DI DKI JAKARTA

Muhamad Khaerul Rafli^{1*}, Utomo Budiyanto²

^{1,2} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia
Email: ^{1*}2011501836@student.budiluhur.ac.id, ²utomo.budiyanto@budiluhur.ac.id

(Naskah masuk: 7 Agustus 2024, diterima untuk diterbitkan: 7 September 2024)

Abstrak

Tingkat kemiskinan dipengaruhi oleh faktor seperti kurangnya lapangan kerja yang memadai ataupun suatu pekerjaan yang tidak mampu mencukupi tanggungan keluarga serta anak-anak. Selain itu, salah satu alasan mengapa masalah kemiskinan belum terselesaikan adalah hasil survei menunjukkan bahwa bantuan pemerintah sering kali tidak sesuai dengan kebutuhan penduduk. Selain itu, dalam pendistribusian bantuan sosial yang dilakukan oleh pemerintah juga tidak tepat sasaran serta tidak merata. Hal tersebut bisa diakibatkan oleh ketidakakuratan data validasi. Tujuan dari penelitian ialah mengidentifikasi hasil *Clustering K-means* dengan data primer yang diperoleh dari Pusat Data dan Teknologi Informasi Dinas Sosial Provinsi DKI Jakarta dan mengelompokkan tingkat kemiskinan di Provinsi DKI Jakarta menggunakan metode *Clustering K-means*. Metode penelitian didasarkan pada penggunaan algoritma *K-means clustering*. Kesimpulan yang diambil ialah penggunaan metode *Clustering K-Means* dapat diterapkan untuk mengelompokkan *cluster* kemiskinan di Provinsi DKI Jakarta. Pada proses pengujian terhadap 8000 data penduduk DKI Jakarta. Didapatkan hasil 14 data. Pengujian dilakukan dengan jumlah *cluster* yang diuji $K = 4$. Menghitung jarak setiap data dengan *centroid* awal menggunakan persamaan *Euclidean Distance*. Hasil yang diperoleh *centroid* ke-1 14,97 dan *centroid* ke-2 22,05. Dari 11 iterasi yang diproses didapatkan hasil *clustering* yaitu iterasi ke-3. Dikarenakan iterasi ke-3 dipilih sebagai yang paling cocok karena beberapa alasan utama. Pertama, *centroid* tidak banyak berubah lagi setelah *iterasi* ini, menunjukkan bahwa algoritma telah mencapai stabilitas. Kedua, terdapat penurunan *inertia* yang signifikan, mengindikasikan bahwa jarak total dari titik-titik ke *centroid* mereka telah diminimalkan secara optimal. Terakhir, validasi visual menunjukkan bahwa data telah dikelompokkan dengan baik sesuai dengan *centroid* yang dihitung.

Kata kunci: *clustering, data mining, Kemiskinan DKI Jakarta, k-means, euclidean distance*

CLUSTERIZATION OF POVERTY LEVELS USING THE K-MEANS METHOD IN DKI JAKARTA

Abstract

The poverty rate is influenced by factors such as lack adequate employment opportunities or a job that is unable to support family and children. In addition, one reasons why the poverty problem has not been resolved is that survey results show that government assistance often does not match the needs population. In addition, the distribution of social assistance carried out by the government is also not on target and uneven. This can be caused by inaccurate validation data. The purpose study was to identify the results of *K-means Clustering* with primary data obtained from the Data and Information Technology Center DKI Jakarta Provincial Social Service and to group the poverty rate in DKI Jakarta Province using the *K-means Clustering* method. The research method is based on the use *K-means clustering* algorithm. The conclusion drawn is that the use *K-Means Clustering* method can be applied to group poverty clusters in DKI Jakarta Province. In the testing process on 8000 DKI Jakarta population data. The results obtained were 14 data. The test was carried out with the number of clusters tested $K = 4$. Calculating the distance of each data with the initial *centroid* using the *Euclidean Distance* equation. The results obtained *centroid* 1 14,97 and *centroid* 2 22,05. From 11 iterations processed, the clustering results obtained were iteration 3. Because iteration 3 was chosen as the most suitable for several main reasons. First, the *centroids* did not change much after this iteration, indicating that the algorithm had reached stability. Second, there was a significant decrease in *inertia*, indicating that the total distance from the points to their *centroids* had been optimally minimized. Finally, visual validation showed that the data had been well clustered according to the calculated *centroids*.

Keywords: *Clustering, Data Mining, Poverty of DKI Jakarta, k-means, euclidean distance*

1. PENDAHULUAN

Kemiskinan tetap menjadi tantangan kritis di Indonesia. Meskipun pemerintah telah

mengimplementasikan berbagai program untuk mengatasinya [1]. Masalah ini persisten dan kompleks, terutama karena berkaitan erat dengan

tingkat pendapatan individu dan rumah tangga [2]. Pendapatan yang tidak memadai menyebabkan individu hanya mampu memenuhi kebutuhan dasar, yang merupakan standar minimum untuk kehidupan yang layak. Tingkat pendapatan yang cukup untuk memenuhi kebutuhan dasar ini menjadi penentu garis kemiskinan, yaitu ambang batas yang membedakan antara kondisi miskin dan tidak miskin [3]. Meskipun upaya-upaya pemerintah terus berlangsung, penyelesaian penuh terhadap masalah kemiskinan membutuhkan pendekatan yang lebih mendalam dan berkelanjutan untuk meningkatkan pendapatan dan kualitas hidup masyarakat [4].

Tingkat kemiskinan dipengaruhi oleh faktor seperti kurangnya lapangan kerja yang memadai ataupun suatu pekerjaan yang tidak mampu mencukupi tanggungan keluarga serta anak-anak [5]. Selain itu, salah satu alasan mengapa masalah kemiskinan belum terselesaikan adalah hasil survei menunjukkan bahwa bantuan pemerintah sering kali tidak sesuai dengan kebutuhan penduduk. Selain itu, dalam pendistribusian bantuan sosial yang dilakukan oleh pemerintah juga tidak tepat sasaran serta tidak merata [6]. Hal tersebut bisa diakibatkan oleh ketidakakuratan data validasi.

Menurut studi literatur yang dilakukan memaparkan tingkat kemiskinan dipengaruhi oleh faktor seperti kurangnya lapangan pekerjaan yang memadai ataupun suatu pekerjaan yang tidak mampu mencukupi tanggungan keluarga serta anak-anak. Masalah kemiskinan yang terjadi dalam ruang lingkup masyarakat belum terselesaikan dengan merata sebab ketidaktepatan bantuan yang diberikan oleh pemerintah. Pemerintah dalam memberikan sebuah bantuan kepada masyarakat miskin belum dilakukan secara merata [7].

Pada implementasi *data mining* terdapat suatu teknik yang disebut *clustering* yang bertujuan dalam mengelompokkan beberapa obyek berdasarkan kesamaan karakteristiknya kedalam beberapa *cluster*. Terdapat dua jenis metode *clustering* yaitu *clustering* hierarkis dan *clustering* partisi. Dalam *clustering* hierarkis data dikelompokkan ke dalam struktur pohon di mana kelompok yang paling mirip digabungkan bertahap hingga semua data tergabung dalam satu *cluster* [8].

Pada penelitian terdahulu yang berjudul “Analisis *Clustering* Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means” memaparkan provinsi yang termasuk dalam *cluster* 0 atau provinsi yang memiliki tingkat kemiskinan rendah diantaranya adalah Gorontalo, Sulawesi Tengah, BTT, NTB, DI Yogyakarta, Jawa Tengah, Lampung, Bengkulu, dan Aceh. Selanjutnya provinsi yang termasuk dalam *cluster* 1 atau provinsi yang memiliki tingkat kemiskinan sedang memiliki sejumlah 21 provinsi di Indonesia. Terakhir, *cluster* 2 yaitu provinsi yang memiliki tingkat kemiskinan tinggi diantaranya adalah Papua, Papua Barat, dan Maluku [9]. Pada penelitian tersebut telah dipaparkan

sejumlah provinsi tingkat kemiskinan di seluruh Indonesia. Sedangkan kebaruan yang diambil dalam penelitian ini, peneliti menyempitkan atau mengambil salah satu provinsi di Indonesia untuk dilakukan analisis tingkat kemiskinan dengan menggunakan metode *K-means*. Adapun provinsi yang menjadi tinjauan analisis pada penulisan ini ialah DKI Jakarta.

Pada penelitian terdahulu yang dilakukan pada obyek warga DKI Jakarta dengan menggunakan algoritma K-Means hanya terbatas pada analisis kepadatan penduduk DKI Jakarta. Pengelompokan kepadatan penduduk di Provinsi DKI Jakarta dapat menghasilkan 2 *Cluster* yaitu tertinggi dan terendah. Kepadatan penduduk tertinggi terdapat pada kelurahan Kali Anyar pada tahun 2019 yang berjumlah 95676.10063 (jiwa/km), dan daerah kelurahan terendah terdapat pada kelurahan Pulau Harapan pada tahun 2018 dengan jumlah 1045.276234 (jiwa/km) [10]. Pada penelitian ini, peneliti menggunakan *clustering K-means* yang merupakan metode lebih kompleks dan detail dalam mengelompokkan data kemiskinan. Selain itu, penelitian ini berfokus pada data dari Dinas Sosial Provinsi DKI Jakarta yang belum banyak dieksplorasi sebelumnya dengan pendekatan data *mining*. *Gap* penelitian ini mencakup penerapan metode yang lebih inovatif dan spesifik dalam konteks DKI Jakarta, serta penggunaan *dataset* yang lebih akurat.

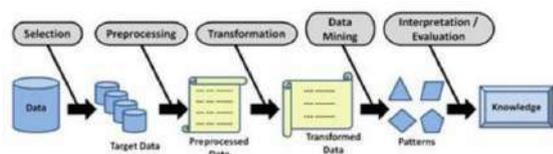
K-means adalah salah satu metode *clustering* yang sering digunakan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan karakteristik. Untuk aplikasi Klasterisasi tingkat kemiskinan, metode ini dapat sangat berguna dalam mengidentifikasi kelompok-kelompok dengan tingkat kemiskinan yang serupa. Dalam hal ini untuk memanfaatkan teknik *data mining* dan *clustering Kmeans*, dengan tujuan utama mengidentifikasi dan memetakan wilayah-wilayah di DKI Jakarta yang mengalami kondisi kemiskinan. Hasil dari penelitian ini bertujuan untuk memberikan gambaran bagi Dinas Sosial Provinsi DKI Jakarta, membantu dalam merumuskan strategi dan keputusan yang lebih informasi. Melalui pendekatan yang sistematis dan berbasis data, penelitian ini berupaya memfasilitasi peningkatan kebijakan publik di DKI Jakarta terutama dalam upaya mengurangi kemiskinan dan meningkatkan kualitas hidup warga DKI Jakarta.

Dalam penelitian ini, teknik *data mining* yang digunakan untuk mengelompokkan wilayah di DKI Jakarta adalah metode *Clustering K-means*. Metode ini memiliki keunggulan dalam hal kemudahan pemahaman dan implementasi serta proses pembelajarannya yang relatif cepat. Metode *Clustering K-means* diterapkan untuk melakukan klasterisasi wilayah di Provinsi DKI Jakarta guna mengidentifikasi daerah dengan tingkat kemiskinan tertentu [11]. Penentuan jumlah klaster yang optimal dilakukan menggunakan metode *Elbow* [12].

Berdasarkan pemaparan tersebut, penulis mengambil judul “Klasterisasi Tingkat Kemiskinan Menggunakan Metode *K-Means* Di DKI Jakarta”. Tujuan dari penelitian ialah mengidentifikasi hasil *Clustering K-means* dengan data primer yang diperoleh dari Pusat Data dan Teknologi Informasi Dinas Sosial Provinsi DKI Jakarta dan mengelompokkan tingkat kemiskinan di Provinsi DKI Jakarta menggunakan metode *Clustering K-means*.

2. METODE PENELITIAN

Metode penelitian didasarkan pada penggunaan algoritma *K-means clustering*. *K-means clustering* adalah metode dalam statistik dan *machine learning* yang digunakan untuk mengelompokkan data ke dalam beberapa grup atau kluster. Tujuannya adalah untuk membagi data yang memiliki kemiripan satu sama lain ke dalam kluster-kluster yang berbeda, sehingga data dalam satu kluster lebih mirip satu sama lain dibandingkan dengan data di kluster lainnya. Data penelitian adalah informasi yang diperoleh langsung dari Dinas Sosial Provinsi DKI Jakarta. Adapun beberapa langkah yang digunakan dalam penerapan metode *clustering k-means* diantaranya (1) menentukan jumlah *cluster*, (2) inisialisasi *centroid* awal, (3) mengelompokkan data, (4) iterasi, (5) evaluasi hasil. Gambar 1 dibawah ini memaparkan terkait dengan metode *Knowledge Discovery in Database (KDD)*, yaitu:



Gambar 1. Metode *Knowledge Discovery in Database (KDD)*

Pada tahap *selection* bertujuan untuk memastikan pemilihan variabel yang tepat dalam pengolahan data, sehingga menghindari duplikasi atau perulangan yang tidak diperlukan. Pada tahap *preprocessing* meliputi dua langkah yaitu data *cleaning* dan data *integration*. Pada tahap *transformation* melibatkan perubahan data ke dalam format yang sesuai untuk dianalisis menggunakan data mining. Pada tahap *data mining* adalah mengekstrak pengetahuan baru dari data yang telah diproses. Pada penelitian ini teknik *Clustering* yang digunakan adalah metode *K-means*. Pada tahap *evaluation* difokuskan pada pengidentifikasian pola-pola menarik dari basis data yang telah ditentukan. Pada tahap *knowledge* bertujuan agar pengetahuan baru yang diperoleh dari proses ini dapat dipahami oleh semua pihak yang terlibat dan digunakan sebagai dasar untuk pengambilan keputusan.

Selanjutnya teknik pengumpulan data yang dilakukan pada penelitian meliputi (1) wawancara

dengan KAPUSDATIN KESOS Dinas Sosial Provinsi DKI Jakarta dilakukan untuk memperoleh informasi mendalam tentang prosedur kerja, tantangan yang dihadapi dan strategi yang diterapkan dalam menangani masalah kemiskinan, (2) studi pustaka dilakukan dengan memilih, menetapkan, dan mempersiapkan daftar pustaka yang terdiri dari jurnal, buku, makalah, dan skripsi yang memiliki korelasi relevan, (3) tinjauan pustaka dilakukan untuk mengumpulkan informasi dari berbagai jurnal yang relevan.

Proses data menggunakan *Knowledge Discovery in Database (KDD)* melalui beberapa tahapan diantaranya *data selection*, *data processing*, *transformation*, serta *data mining* [13]. Selanjutnya terkait dengan rancangan basisdata meliputi *logical record structure* dan spesifikasi basis data. Rancangan pengujian sendiri terbagi menjadi beberapa *point* diantaranya metode pengujian, tahapan pengujian yang terdiri dari pengujian UI sertapengujian fungsi dasar aplikasi [14].

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Metode

Pada pengujian ini peneliti memakai data sebanyak 14 data hanya sebagai sampel untuk proses perhitungan secara manual perhitungan *Clustering K-means*. Tetapi untuk di sistem yang dibangun peneliti memakai semua data yang akan di uji yaitu sebanyak 800 data. Pada proses perhitungan *Clustering K-means* penulis melakukan perhitungan manual. Perhitungan manual dilakukan menggunakan di program yang telah dibuat. Berikut langkah-langkah penyelesaiannya:

a. Menentukan Jumlah *Cluster*

Jumlah cluster menyesuaikan dengan kebutuhan analisis dalam penelitian ini jumlah cluster yang akan digunakan sebanyak 4 *clusters*.

b. Hasil Data *Cleaning*

Proses ini dilakukan untuk membersihkan data dari elemen-elemen yang tidak diperlukan. Termasuk menghapus nilai yang tidak lengkap, mengurangi gangguan dari data yang berisik (*noise*) dan menangani konsistensi serta keterkaitan dalam data. Kemudian, dari hasil *cleaning* 8000 data diperoleh data sebanyak 14.

c. Menentukan *Centroid* Awal

Centroid awal merupakan titik pusat *cluster* pertama yang diperoleh secara acak dari data sampel dimana dari 14 data, data yang dipilih data ke-7. Pemilihan titik *centroid* awal akan sangat berpengaruh terhadap hasil *cluster*. Sebagai salah satu contoh, disini peneliti atau penulis akan melakukan perhitungan untuk data ke-1 terhadap data ke-7. Selanjutnya, Tabel 1 memaparkan terkait dengan penentuan dari *centroid* awal yang dilakukan.

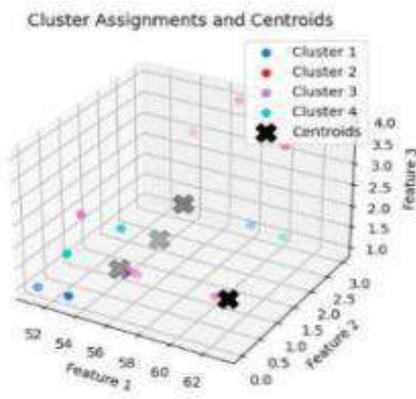
Perhitungan penentuan *centroid*
 $= (57-53)^2 + (0-0)^2 + (2-1)^2 + (3-2)^2 + (17-4)^2 + (9-3)^2 + (4-3)^2$
 $= 4^2 + 0^2 + 1^2 + 1^2 + 13^2 + 6^2 + 1^2$
 $= 16 + 0 + 1 + 1 + 169 + 36 + 1 = 224$

Tabel 1. Hasil Penentuan Centroid Awal

<i>Cen</i> <i>troi</i> <i>d</i>	<i>Fea</i> <i>ture</i> <i>1</i>	<i>Fea</i> <i>ture</i> <i>2</i>	<i>Fea</i> <i>ture</i> <i>3</i>	<i>Fea</i> <i>ture</i> <i>4</i>	<i>Fea</i> <i>ture</i> <i>5</i>	<i>Fea</i> <i>ture</i> <i>6</i>	<i>Fea</i> <i>ture</i> <i>7</i>
C1	53	0	1	2	4	3	3
C2	63	0	2	4	38	7	2
C3	54	3	3	1	25	9	1
C4	60	3	1	3	31	4	3

- d. Menghitung Jarak Data ke Titik Pusat *Cluster*
 Sebagai dua contoh, disini peneliti akan melakukan perhitungan untuk data ke-1 terhadap centroid 1 dan centroid 2.

Pada pengujian *cluster* dengan *range* nilai K dilakukan dengan jumlah *cluster* yang akan diuji adalah dari K=4. Pengujian menggunakan data penduduk diperoleh dari Dinas Sosial Provinsi DKI Jakarta periode data 2019. Dari 11 iterasi yang diproses didapatkan hasil *clustering* yaitu iterasi ke-3. Dikarenakan iterasi ke-3 dipilih sebagai yang paling cocok karena beberapa alasan utama. Pertama, centroid tidak banyak berubah lagi setelah iterasi ini, menunjukkan bahwa algoritma telah mencapai stabilitas. *Kedua*, terdapat penurunan inerti yang signifikan, mengindikasikan bahwa jarak total dari titik-titik ke *centroid* mereka telah diminimalkan secara optimal. Terakhir, validasi visual menunjukkan bahwa data telah dikelompokkan dengan baik sesuai dengan centroid yang dihitung. Kombinasi faktor-faktor ini menjadikan iterasi ke-3 sebagai pilihan terbaik dalam analisis *clustering* yang dilakukan. Pada gambar 2 memaparkan terkait dengan visualisasi hasil *clustering*.



Gambar 2. Visualisasi Hasil *Clustering*

Visualisasi hasil *clustering* dalam Gambar 2 yang dilampirkan menunjukkan representasi 3D dari pengelompokan data dengan jelas. Pada gambar tersebut, *centroid* dari setiap *cluster* ditandai dengan simbol 'X', sementara data yang dikelompokkan dalam masing-masing *cluster* ditandai dengan warna yang berbeda. Warna-warna yang berbeda ini memudahkan dalam mengidentifikasi dan membedakan kelompok data yang terbentuk. Visualisasi ini sangat membantu dalam memahami bagaimana data terdistribusi ke dalam *cluster* yang berbeda dan memberikan gambaran yang intuitif tentang struktur internal dari dataset yang dianalisis. Dengan visualisasi 3D ini, hasil *clustering* dapat dievaluasi dengan lebih baik, memastikan bahwa pengelompokan yang dilakukan telah mencerminkan pola yang ada dalam data secara efektif.

3.2 Pengujian Aplikasi

K-means adalah salah satu metode *clustering* yang sering digunakan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan karakteristik. Untuk aplikasi klasterisasi tingkat kemiskinan, metode ini dapat sangat berguna dalam mengidentifikasi kelompok-kelompok dengan tingkat kemiskinan yang serupa. Pada pengujian aplikasi yang dilakukan dengan penggunaan metode *blackbox testing* memiliki tujuan dalam menghasilkan data yang berkesesuaian antara *output* dan *input* dan pula memiliki fungsi dalam pengujian fungsionalitas [15]. Penulis juga menampilkan grafik dengan kualitas *cluster* yang terbaik menggunakan *Elbow Method*. Tabel 2 dibawah ini memaparkan terkait dengan hasil pengujian *blackbox testing dashboard* yang dihasilkan.

Tabel 2. *Blackbox Testing Dashboard*

<i>N</i> <i>o</i>	<i>Field</i>	<i>Scenario</i>	<i>Expected Result</i>	<i>Statu</i> <i>s</i>
1	Halaman <i>Dashboard</i>	Klik halaman <i>Dashboard</i>	<i>Success</i> sistem menampilkan halaman <i>Dashboard</i>	OK
2	Halaman <i>K-Means</i>	Klik halaman <i>K-Means</i>	<i>Success</i> sistem menampilkan halaman <i>K-Means Dataset</i>	OK

Selanjutnya pada pengujian *blackbox testing* pada halaman *K-Means* dihasilkan data yang tersajikan pada Tabel 3. Kemudian hasil dari pengujian *blackbox testing* pada halaman *K-Means Dataset* yang diperoleh dari penelitian disajikan dalam Tabel 4.

Tabel 3. *Blackbox Testing K-Means* dengan format

No	Field	Scenario	Expected Result	Status
1	Halaman <i>K-Means</i>	Klik <i>Choose File</i>	Success sistem menampilkan file <i>excel xlsx</i>	OK
2	Halaman <i>K-Means</i>	Pilih file <i>excel</i> dengan format selain <i>xlsx</i>	Success sistem menampilkan pop up “ <i>Uploadfile excel dengan format xlsx</i> ”	OK
3	Halaman <i>K-Means</i>	Pilih file <i>excel</i> dengan format <i>xlsx</i>	Success sistem menampilkan file <i>excel</i> dengan format <i>xlsx</i>	OK
4	Halaman <i>K-Means</i>	<i>Upload file excel</i>	Success sistem menampilkan data yang telah di <i>upload</i>	OK
5	Halaman <i>K-Means</i>	Klik tombol button <i>use this data</i>	Success sistem menampilkan halaman <i>K-Means Dataset</i>	OK

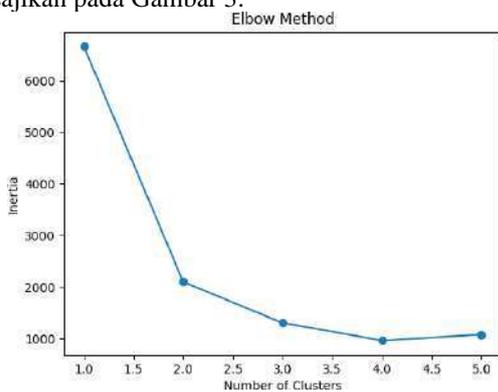
Tabel 4. *Blackbox testing K-Means Dataset*

No	Field	Scenario	Expected Result	Status
1	Halaman <i>Cluster Config</i>	Input pada <i>clusters</i> dan <i>Maximum Iteranios</i>	Success sistem input <i>clusters & Maximum Iteranios</i>	OK
2	Halaman <i>Cluster Config</i>	Klik tombol button <i>process</i> pada data yang telah di input	Success sistem menampilkan halaman <i>Process</i>	OK
3	Halaman <i>Process</i>	Klik button halaman <i>Process</i>	Success menampilkan <i>exploring the K-Means Clustering Methodology</i>	OK
4	Halaman <i>Elbow</i>	Input pada <i>MaxClusters</i>	Success sistem input <i>MaxClusters</i>	OK
5	Halaman <i>Elbow</i>	Klik tombol button <i>process</i>	Success sistem menampilkan grafik <i>ElbowMethod</i>	OK
6	Halaman <i>Clstering</i>	Klik button halaman <i>Clsturing</i>	Success sistem menampilkan <i>Clustering Result</i>	OK

Tabel 5. *Blackbox testing History*

No	Field	Scenario	Expected Result	Status
1	Menu <i>History</i>	Klik menu <i>History</i>	Success sistem menampilkan halaman <i>History</i>	OK
2	Halaman <i>K-Means</i>	Klik tombol button detail	Success sistem menampilkan halaman <i>DetailHistory K-MeansClustering</i>	OK

Pada tahap ini akan menjelaskan hasil dari pengujian hasil *Elbow Method* pada halaman *Elbow* disajikan pada Gambar 3.



Gambar 3. Hasil *Elbow Method*

Hasil dari pengujian *blackbox testing* pada menu *History*, yang diperoleh dari penelitian disajikan dalam Tabel 5.

Pengujian kualitas *cluster* diterapkan menggunakan data penduduk Provinsi DKI Jakarta memiliki jumlah 800 data, menguji kualitas *cluster*

dengan 4 *cluster* menggunakan metode *Clustering K-means* dengan kategori/label yang disajikan pada Tabel 6.

Tabel 6. *Cluster Label*

Cluster	Label
1	Tidak Miskin
2	Sedang Miskin
3	Miskin
4	Sangat Miskin

Pengujian untuk menentukan jumlah atau nilai *cluster* terbaik menggunakan metode *Clustering K-means* dengan jumlah *cluster* dimana hasil *cluster* terbaik akan menjadi acuan penggunaan *cluster* untuk penelitian ini. Langkah selanjutnya adalah menentukan *cluster* terbaik dengan label yang telah diperoleh dari *result Clustering*. Pada penduduk dengan label “Tidak Miskin” yang disajikan pada tabel 7 adalah mereka yang tidak mengalami kesulitan ekonomi. Mereka memiliki pendapatan yang cukup dalam pemenuhan kebutuhan primer. Mereka mampu hidup dengan nyaman dan tidak perlu khawatir tentang kebutuhan sehari-hari.

Tabel 7. Detail Cluster 1

Nama	Ijazah Tertinggi	Jenis Cacat	Status Pekerjaan	Jumlah Jam Kerja	Lapangan Usaha	Penyakit Kronis
Marwiyah, 60 Tahun	M. Ibtidaiyah	Tuna netra/buta	Berusaha dibantu buruh tidak tetap/ tidak dibayar	4	8	Hipertensi (tekanan darah tinggi)
Abd Manaf, 53 Tahun	SD/SDLB	Tuna daksa/ cacat tubuh	Berusaha dibantu buruh tetap/dibayar	4	3	Asma
Juhri, 51 Tahun	SD/SDLB	Tuna daksa/ cacat tubuh	Buruh/ karyawan/ pegawai swasta	3	4	Masalah jantung
Aggregasi Nilai Tertinggi/Terbanyak	SD/SDLB	Tuna daksa/ cacat tubuh	Berusaha dibantu buruh tidak tetap	4	8	Hipertensi

Tabel 8. Detail Cluster 2

Nama, Umur	Ijazah Tertinggi	Jenis Cacat	Status Pekerjaan	Jumlah Jam Kerja	Lapangan Usaha	Penyakit Kronis
IBRAHIM, 63 Tahun	SD/SDLB	Tuna netra/buta	PNS/ TNI/ Polri/ BUMN/ BUMD/ anggota legislatif	38	7	Rematik
Aggregasi Nilai Tertinggi/Terbanyak	SD/SDLB	Tuna netra/buta	PNS/ TNI/ Polri/ BUMN/ BUMD/ anggota legislatif	38	7	Rematik

Tabel 9. Detail Cluster 3

Nama, Umur	Ijazah Tertinggi	Jenis Cacat	Status Pekerjaan	Jumlah Jam Kerja	Lapangan Usaha	Penyakit Kronis
Ruslan, 57 Tahun	SD/SDLB	Tuna netra/buta	Buruh/ karyawan/ pegawai swasta	17	9	Masalah jantung
Nursaman BinAbd Kadir, 54 Tahun	SMP/ SMPLB	Tuna rungu	Berusaha dibantu buruh tidak tetap/ tidak dibayar	25	9	Hipertensi (tekanan darah tinggi)
Sarifudin, 55 Tahun	Paket A	Tuna daksa/ cacat tubuh	Berusaha dibantu buruh tetap/dibayar	20	5	Rematik
Junaidi, 60 Tahun	Paket A	Tuna daksa/ cacat tubuh	Buruh/ karyawan/ pegawai swasta	16	1	Rematik
Mat Jamin, 57 Tahun	SMP /SMPLB	Tuna wicara	Buruh/ karyawan/ pegawai swasta	25	7	Asma
Hasanah, 62 Tahun	M. Ibtidaiyah	Tuna wicara	PNS/ TNI/ Polri/ BUMN/ BUMD/ anggota legislatif	20	7	Asma
Uto Masjoyo, 54 Tahun	SD/SDLB	Tuna rungu	Berusaha dibantu buruh tetap/ dibayar	25	7	Asma
Aggregasi Nilai Tertinggi/Terbanyak	SD/SDLB	Tuna daksa/ cacat tubuh	Buruh/ karyawan/ pegawai swasta	25	9	Asma

Tabel 10. Detail Cluster 4

Nama, Umur	Ijazah Tertinggi	Jenis Cacat	Status Pekerjaan	Jumlah Jam Kerja	Lapangan Usaha	Penyakit Kronis
Rausin, 54 Tahun	Paket A	Tuna netra/buta	Berusaha dibantu buruh tidak tetap/ tidak dibayar	34	2	Masalah jantung
Sartini, 60 Tahun	SMP/SMPLB	Tuna daksa/ cacat tubuh	Buruh/ karyawan/ pegawai swasta	31	4	Asma
Husaini, 53 Tahun	SD/SDLB	Tuna netra/buta	Buruh/ karyawan/ pegawai swasta	29	5	Asma
Aggregasi Nilai Tertinggi/Terbanyak	SD/SDLB	Tuna netra/buta	Buruh/ karyawan/ pegawai swasta	34	5	Asma

Pada penduduk dengan label “Miskin” adalah *cluster* menghadapi kesulitan yang signifikan dalam

memenuhi kebutuhan dasar mereka. Mereka sering kali berada dalam kondisi ekonomi yang rentan dan

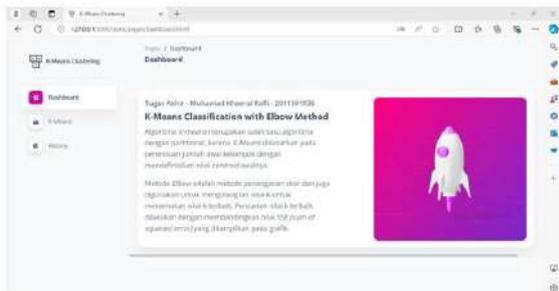
memerlukan bantuan dari luar untuk mencukupi kebutuhan hidup sehari-hari. Kesejahteraan mereka terganggu oleh pendapatan yang tidak mencukupi dan kondisi hidup yang kurang layak. Berikut *cluster* 3 dapat dilihat pada Tabel 9.

Pada penduduk dengan label “Sangat Miskin” adalah *cluster* berada dalam kondisi ekonomi yang sangat sulit. Mereka tidak mampu memenuhi kebutuhan dasar tanpa bantuan signifikan dari pemerintah atau organisasi sosial. Kehidupan mereka sangat dipengaruhi oleh kekurangan ekonomi yang ekstrem, dan mereka sering kali menghadapi masalah kesehatan, pendidikan, dan tempat tinggal yang serius. Hasil *cluster* 4 dapat dilihat pada Tabel 10.

Dari hasil pengujian kualitas *cluster*, *cluster* terbaik adalah *Cluster* 3 (Miskin). Hal ini karena *cluster* ini memberikan gambaran yang jelas mengenai penduduk yang berada dalam kondisi ekonomi yang rentan dan memerlukan perhatian khusus dari Dinas Sosial Provinsi DKI Jakarta. Penduduk dalam *cluster* ini sering kali membutuhkan intervensi yang lebih intensif untuk membantu mereka keluar dari kemiskinan dan mencapai kesejahteraan yang lebih baik. Dengan mengidentifikasi *cluster* ini sebagai *cluster* terbaik, dapat diambil tindakan yang lebih tepat sasaran dalam program bantuan sosial untuk meningkatkan kualitas hidup penduduk yang termasuk dalam kategori miskin

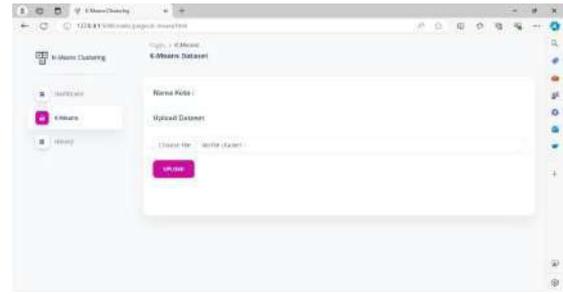
3.3 Tampilan Layar

Tampilan layar *dashboard* pada halaman *dashboard* ini terdapat dua navbar yang tersedia: *Dashboard* dan *K-Means* yang disajikan pada gambar 4.



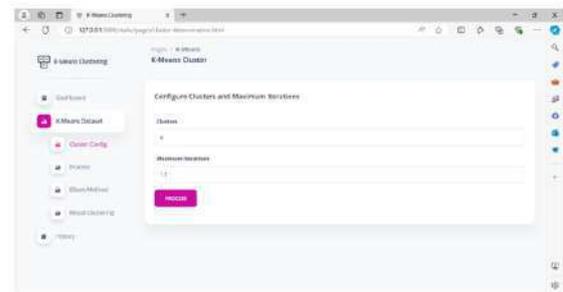
Gambar 4. Tampilan Layar *Dashboard*

Tampilan layar *K-Means* muncul setelah pengguna mengklik menu *K-Means*. Pengguna dapat mengunggah data dalam format *Excel* “.xlsx” melalui tombol “Choose File” kemudian mengklik tombol “Upload” yang disajikan pada gambar 5.



Gambar 5. Tampilan Layar *K-Means*

Tampilan layar *Cluster Config* merupakan sub menu dari menu *K-Means Dataset*. Saat user klik button *Use This Data* akan menampilkan 2 kolom yaitu *Clusters* dan *Maximum Iterations*. Langkah selanjutnya user mengisi 2 kolom tersebut untuk menentukan *Clusters* dan *Maximum Iterations* yang disajikan pada gambar 6.



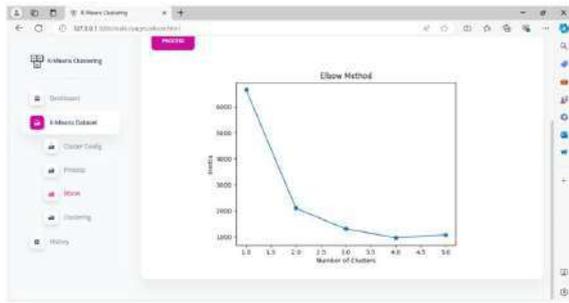
Gambar 6. Tampilan Layar *K-Means Dataset (Cluster Config)*

Tampilan layar *Process* merupakan sub-menu dari menu *K-Means*. Ketika user telah menentukan *Clusters* dan *Maximum Iterations* langkah selanjutnya user klik button *process* pada menu *Cluster Config* setelah itu akan menampilkan halaman menu *Process* pada gambar 7.

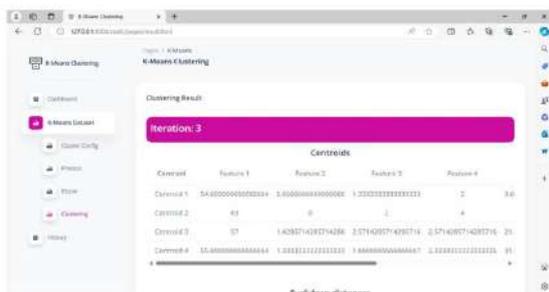


Gambar 7. Tampilan Layar *K-Means Dataset (Process)*

Tampilan layar *Elbow* merupakan sub-menu dari menu *K-Means*. Selanjutnya user klik menu *Elbow* dan mengisi kolom *Max Clusters* klik button *process*. Menampilkan grafik *Elbow* yang disajikan pada gambar 8.

Gambar 8. Tampilan Layar *K-Means Dataset (Elbow)*

Tampilan layar *Clustering* merupakan sub-menu dari menu *K-Means*. Selanjutnya user klik menu *Clustering*. Menampilkan halaman *Clustering Result* pada Gambar 9.

Gambar 9. Tampilan Layar *K-Means Dataset (Clustering)*

Tampilan layar *History* adalah hasil *clustering*. Selanjutnya user klik history. Menampilkan halaman history yang disajikan pada Gambar 10.

No	Nama File	Nama Data	Tanggal
1	data_1.csv	data_1.csv	2024-09-01 10:00:00
2	data_2.csv	data_2.csv	2024-09-01 10:00:00
3	data_3.csv	data_3.csv	2024-09-01 10:00:00
4	data_4.csv	data_4.csv	2024-09-01 10:00:00
5	data_5.csv	data_5.csv	2024-09-01 10:00:00
6	data_6.csv	data_6.csv	2024-09-01 10:00:00
7	data_7.csv	data_7.csv	2024-09-01 10:00:00
8	data_8.csv	data_8.csv	2024-09-01 10:00:00

Gambar 10. Tampilan Layar *History*

Hasil dari penelitian senada dengan penelitian terdahulu yang memaparkan metode *K-means* dapat digunakan dalam klusterisasi tingkat kemiskinan. Pada penelitian terdahulu menjelaskan provinsi yang termasuk dalam *cluster 0* atau provinsi yang memiliki tingkat kemiskinan rendah diantaranya adalah Gorontalo, Sulawesi Tengah, BTT, NTB, DI Yogyakarta, Jawa Tengah, Lampung, Bengkulu, dan Aceh. Selanjutnya provinsi yang termasuk dalam *cluster 1* atau provinsi yang memiliki tingkat kemiskinan sedang memiliki sejumlah 21 provinsi di Indonesia. Terakhir, *cluster 2* yaitu provinsi yang memiliki tingkat kemiskinan tinggi diantaranya adalah Papua, Papua Barat, dan Maluku [9].

4. KESIMPULAN

Kesimpulan yang diambil berdasarkan paparan diatas ialah penggunaan metode *Clustering K-Means* dapat diterapkan untuk mengelompokkan *cluster* kemiskinan di Provinsi DKI Jakarta. Pada proses pengujian terhadap 8000 data penduduk DKI Jakarta. Didapatkan hasil 14 data. Pengujian dilakukan dengan jumlah *cluster* yang diuji $K = 4$. Menghitung jarak setiap data dengan *centroid* awal menggunakan persamaan *Euclidean Distance*. Hasil yang diperoleh *centroid* ke-1 14,97 dan *centroid* ke-2 22,05. Dari 11 iterasi yang diproses didapatkan hasil *clustering* yaitu iterasi ke-3. Dikarenakan iterasi ke-3 dipilih sebagai yang paling cocok karena beberapa alasan utama. Pertama, *centroid* tidak banyak berubah lagi setelah iterasi ini, menunjukkan bahwa algoritma telah mencapai stabilitas. Kedua, terdapat penurunan *inertia* yang signifikan, mengindikasikan bahwa jarak total dari titik-titik ke *centroid* mereka telah diminimalkan secara optimal. Terakhir, validasi visual menunjukkan bahwa data telah dikelompokkan dengan baik sesuai dengan *centroid* yang dihitung.

DAFTAR PUSTAKA

- [1] F. D. Hardini, A. K. Joy, B. Lidya Maharani, A. Rizky Airlangga, and C. Asnanti, "Kebijakan Pajak sebagai Upaya Pengentasan Kemiskinan," *J. Krit. Stud. Huk.*, vol. 9, no. 5, pp. 203–208, 2024, [Online]. Available: <https://ojs.co.id/1/index.php/jksh/article/view/1245>.
- [2] I. N. Pratama, "Analisis Determinan Tingkat Kemiskinan Di Kabupaten Sumbawa," *J. Law Gov.*, vol. 11, no. 2, pp. 143–153, 2023, doi: 10.58406/jeb.v11i2.1314.
- [3] P. I. Lestari, B. Robiani, and Sukanto, "Kemiskinan Ekstrem, Ketimpangan Dan Pertumbuhan Ekonomi Di Indonesia," *J. Ilm. Ekon. dan Bisnis*, vol. 11, no. 2, pp. 1739–1752, 2023, [Online]. Available: <https://jurnal.unived.ac.id/index.php/er/article/view/4789>.
- [4] A. Natalia and E. N. Maulidya, "Aktualisasi Empat Pilar Sustainable Development Goals (SDGs) Di Perdesaan Kecamatan Natar Kabupaten Lampung Selatan," *JiIP J. Ilm. Ilmu Pemerintah.*, vol. 8, no. 1, pp. 21–41, 2023, doi: 10.14710/jiip.v8i1.16513.
- [5] U. Kencana, Yuswalina, and T. Eza, "Efektivitas Peraturan Daerah yang Berkesejahteraan Sosial di Kota Palembang: Studi Kasus Anak Jalanan, Gelandangan dan Pengemis di Masa Pandemi Covid-19," *Simbur Cahaya*, vol. 27, no. 2, pp. 70–97, 2021, doi: 10.28946/sc.v27i2.1039.
- [6] R. Nasution and M. Marliyah, "Analisis Program Pemerintah Dalam Penanggulangan Kemiskinan Dan Pengangguran Di Kecamatan Pulau Rakyat Kabupaten Asahan," *Jesya*, vol. 6, no. 1, pp. 810–823, 2023, doi: 10.36778/jesya.v6i1.1031.
- [7] S. Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021, doi: 10.29408/jit.v4i1.2986.
- [8] N. Sepriyanti, R. Sani Nahampun, M. H. Zikri, I.

- Ambarani, and A. Rahmadeyan, "Implementation of K-Means Clustering to Group Poverty Levels in Riau Province," *Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 59–65, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>.
- [9] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.
- [10] A. Khalif, A. N. Hasanah, M. H. Ridwan, and B. N. Sari, "Klasterisasi Tingkat Kemiskinan di Indonesia menggunakan Algoritma K-Means," *Gener. J.*, vol. 8, no. 1, pp. 54–62, 2024, doi: 10.29407/gj.v8i1.21470.
- [11] N. Oktaviany, N. Suarna, and W. Prihartono, "Implementasi Algoritma K-Means Clustering Dalam Mengelompokkan Kepadatan Penduduk Di Provinsi Dki Jakarta," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 119–126, 2024, doi: 10.36040/jati.v8i1.8241.
- [12] N. A. Maori and E. Evanita, "Metode *Elbow* dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [13] M. A. S. Fazrin, A. Voutama, and Y. Umidah, "Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1409–1415, 2023, doi: 10.36040/jati.v7i2.6810.
- [14] N. K. Ningrum, I. U. W. Mulyono, and Z. Umami, "Pengujian UI/UX dengan System Usability Scale dan Single Ease Question pada Aplikasi Pantau untuk Monitoring Perkembangan Penanaman Tanaman di Lahan Hijau," *Proceeding Sci. Eng. Natl. Semin. 7 (SENS 7)*, pp. 9–16, 2022, [Online]. Available: <https://conference.upgris.ac.id/index.php/sens/article/view/3466%0Ahttps://conference.upgris.ac.id/index.php/sens/article/download/3466/2146>.
- [15] S. Catrio, M. Rachmanto, H. A. Agatha, T. Ramdani, and A. Yusuf, "Pengujian Aplikasi Sapawarga (Jabar Super Apps) Menggunakan Metode Black Box Testing," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2319–2334, 2024.

INDEKS NAMA PENULIS

A	
Achmad Solichin	161-168
Agnes Aryasanti	112-119
Anita Diana	112-119
Arnita Piliang	131-138
B	
Bianca Vania Adeline	112-119
D	
Damar Bambang Suwarno	145-152
G	
Gufron	105-111
H	
Hafiizh Taufiqul Hakim	153-160
Hijka Listia	131-138
L	
Lauw Li Hin	120-130
M	
Mardi Hardjianto	145-152
Mohammad Anif	112-119
Mohammad Syafrullah	139-144
Muhammad Khaerul Rafli	169-177
Muhammad Rafi Pratama	120-130
P	
Putri Tasya Agustina Tampubolon	131-138
R	
Raihan Labib Hanif	139-144
S	
Sabina Wardaniah	131-138
Siti Wulandari	131-138
T	
T.Sofia Chairani	131-138
Teguh Febriyanto	161-168
U	
Utomo Budiyanto	169-177
W	
Wendi Usino	153-160

INDEKS KATA KUNCI

A	
<i>agglomerative clustering</i>	153-160
analisis data	131-138
analisis sentimen	161-168
B	
BMC	112-119
C	
cctv	105-111
<i>clustering</i>	169-177
CMS	112-119
<i>content management system</i>	120-130
crisp-dm	153-160
<i>cyber crime</i>	145-152
D	
digitalisasi bisnis	120-130
data mining	169-177
E	
E-Commerce	112-119; 120-130
G	
gojek	161-168
I	
ikn	139-144
indikator pendidikan	153-160
<i>indrive</i>	161-168
K	
<i>k-means</i>	169-177
<i>k-means clustering</i>	131-138
<i>k-nearest neighbor (knn)</i>	161-168
keamanan	131-138
keamanan cerdas	105-111
kemiskinan dki jakarta	169-177
klasterisasi	131-138
kejahatan	131-138
L	
<i>latent dirichlet allocation (LDA)</i>	139-144
M	
<i>machine learning</i>	145-152
media sosial	139-144
P	
pengembangan toko online	120-130

penyimpanan awan	105-111
<i>phishing</i>	145-152
R	
<i>random forest</i>	145-152
S	
sekolah menengah atas (SMA)	153-160
<i>smartphone</i>	105-111
T	
teknologi informasi	139-144
toko vape	120-130
<i>text mining</i>	161-168
W	
website	120-130
X	
X	139-144

JUDUL ARTIKEL BAHASA INDONESIA (maksimal 12 kata, huruf besar, times new roman, 14pt, tebal, dan rata tengah)

Penulis Satu^{1*}, **Penulis Dua**² (10pt, tebal, dan rata tengah)

¹Afiliasi Penulis Satu (10pt)

²Afiliasi Penulis Dua (10pt)

Email: ¹penulis.satu@email.ac.id, ²penulis.dua@email.ac.id (10pt)

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak (10pt, tebal, dan di tengah)

Tempatkan abstrak berbahasa Indonesia pada bagian ini. Abstrak memberikan gambaran umum tentang isi makalah dan harus ditulis dengan *Times New Roman* 10pt dalam format satu kolom. Panjang ideal sebuah abstrak adalah 150 sampai 250 kata. Jika terdapat istilah-istilah asing yang belum dibakukan ditulis *italic*.

Kata kunci: kata kunci sedapat mungkin menjelaskan isi tulisan, ditulis dengan huruf kecil kecuali singkatan, maksimum enam kata, masing-masing dipisahkan dengan koma, *Times New Roman 10pt, italic*

JUDUL ARTIKEL BAHASA INGGRIS (huruf besar, times new roman, 14pt, tebal, dan rata tengah)

Abstract (10pt, italic, tebal, dan di tengah)

Place the Indonesian abstract in this section. Abstracts provide an overview of the content of the paper and should be written in Times New Roman 10pt in a one-column format. The ideal length of an abstract is 150 to 250 words. All abstracts are written in italics.

Keywords: kata kunci sedapat mungkin menjelaskan isi tulisan, ditulis dengan huruf kecil kecuali singkatan, maksimum enam kata, masing-masing dipisahkan dengan koma, *Times New Roman 10, italic*

1. PENDAHULUAN [HEADING LEVEL 1: KAPITAL, TIMES NEW ROMAN, 11, BOLD]

Berikut ini adalah petunjuk penulisan makalah **JURNAL BIT**, Fakultas Teknologi Informasi Universitas Budi Luhur. Paper bersifat terbuka bagi masyarakat ilmiah di bidang TIK. Naskah yang ditulis untuk paper ini adalah publikasi ilmiah di bidang Teknologi Informasi dan Komputer serta aplikasinya dalam industri TIK.

Naskah yang diusulkan harus merupakan hasil pemikiran, hasil penelitian dan atau pengembangan yang bersifat asli, Naskah paper dapat ditulis dalam bahasa Indonesia atau bahasa Inggris.

Seluruh makalah yang telah lulus akan diterbitkan sesuai dengan yang dikirim oleh penulis. **Penulis bertanggung jawab sepenuhnya** terhadap isi naskah yang ditulis dan naskah merupakan tulisan yang **belum pernah dipublikasikan**. Peserta yang akan memasukkan papernya bisa mengirimkan papernya sesuai format template ini melalui laman web <https://journal.budiluhur.ac.id/index.php/bit> dengan register terlebih dahulu dan submission online lewat OJS situs jurnal bit. Seluruh makalah yang telah

lulus dari dewan redaksi BIT akan dipublikasikan dalam bentuk *online OJS*.

2. FORMAT NASKAH

2.1 Panjang Naskah [Heading Level 2: Times New Roman 10 bold]

Naskah paper ditulis pada ukuran kertas kertas A4 (21 cm x 29,7 cm) dengan total halaman 8 hingga 15 halaman termasuk tabel dan gambar. Ketika diajukan kepada Penyunting, naskah tidak perlu diberi nomor halaman, *header* dan *footer*.

Penulisan naskah menggunakan huruf Times New Roman, berukuran 10 pt, dengan batas atas, bawah, kiri dan kanan masing-masing berukuran 2,5 cm. Naskah dibuat dengan menggunakan *Microsoft Word*.

Judul, identitas penulis, abstrak dan kata kunci dibuat dalam *layout* satu kolom. Bagian utama naskah disajikan dalam *layout* dua kolom, dengan lebar setiap kolom 7,5 cm dan jarak antar kolom 1 cm. Naskah ditulis dalam spasi satu. Tambahkan satu spasi untuk setiap antar item, yaitu: antara judul dengan penulis, antara penulis dengan abstrak, antara abstrak dengan kata kunci, antara gambar dengan isi,

antara tabel dengan isi, antara persamaan matematika dengan isi.

Kecuali untuk abstrak, awal paragraf isi tulisan ditulis menjorok ke dalam (*first line indent*) sejauh 7,5 mm. tata cara penulisan telah disusun pada tulisan ini.

3. PENULISAN NASKAH

Judul harus jelas dan singkat. Nama penulis dan afiliasinya seperti yang tertulis diatas. Nama penulis ditulis secara jelas tanpa gelar. Penomoran heading dengan system Arabic dengan *sub-heading* maksimal hingga 3 tingkat.

3.1 Persamaan Matematika

Persamaan matematika dinomori dengan Angka Arab dalam kurung pada sisi kanan (rata kanan) kolom. Persamaan (1) ditulis menjorok ke dalam sejauh 7,5 mm.

Penulisan simbol matematika di dalam paragraf isi tulisan hendaknya tidak menggunakan *equation editor*, tetapi menggunakan *insert symbol*.

$$p(x_t | y_{1:t}) = \frac{p(y_t | x_t)p(x_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \dots\dots\dots(1)$$

3.2 Tabel

Tabel-tabel, dan juga grafik-grafik, harus dibuat dalam mode hitam-putih (bukan color maupun grayscale). Jika diperlukan, gambar citra dapat disajikan secara grayscale, tetapi bukan color.

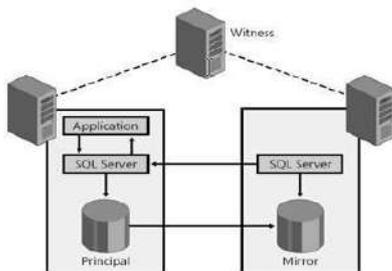
Tabel harus diberi nomor sesuai urutan presentasi (Tabel 1, dst.). Judul tabel ditulis diatas tabel dengan posisi rata kiri (*left justified*).

Tabel 1. Tabel Software dan Hardware Pendukung
[Times New Roman 9 normal center]

Product	Server	Client	Oracle Connect
Clementine	Solaris 2.X	X Windows	Server Side ODBC
Darwin	Solaris 2.X	Windows NT	Server Side ODBC
PRW	Data only	Windows NT	Client Side ODBC

3.3 Gambar

Gambar diberi nomor sesuai urutan presentasi (Gambar 1, dst.). Judul gambar yang diletakkan dibawah gambar dengan posisi tengah (*centre justified*).



Gambar 1. Database Mirroring Architecture [Judul Gambar: Times New Roman 8 italic center]

3.4 Sumber Pustaka

Sumber pustaka/rujukan sedapat mungkin merupakan pustaka-pustaka terbitan 5 tahun terakhir. Pustaka yang diutamakan adalah naskah-naskah penelitian dalam jurnal, konferensi dan/atau majalah ilmiah. Pustaka lain dapat berupa buku teks atau laporan penelitian (termasuk Skripsi/Tugas Akhir, Tesis, dan Disertasi), akan tetapi diusahakan tidak melebihi 20% dari seluruh jumlah sumber pustaka [1], [2].

Penulisan sumber pustaka dan cara mengacu menggunakan aturan IEEE. Beberapa aturan tentang penulisan sumber pustaka, yaitu: sumber pustaka yang ditulis dalam daftar pustaka sebelumnya harus pernah diacu dalam naskah, ditulis berurutan berdasarkan urutan sitasi di naskah.

Petunjuk lebih lengkap mengenai aturan penulisan sitasi dan daftar pustaka gaya IEEE dapat dibaca pada <https://ieeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf> . Sangat disarankan penggunaan perangkat lunak manajemen referensi seperti Mendeley Desktop dan Zotero. Contoh penulisan sitasi dan daftar pustaka dapat dilihat pada naskah template ini [3]–[5].

4. KESIMPULAN

Dalam kesimpulan tidak boleh ada referensi. Kesimpulan berisi fakta yang didapatkan, cukup menjawab permasalahan atau tujuan penelitian (jangan merupakan pembahasan lagi); Nyatakan kemungkinan aplikasi, implikasi dan spekulasi yang sesuai. Jika diperlukan, berikan saran untuk penelitian selanjutnya. Panduan ini telah menjelaskan bagaimana naskah Jurnal BIT Fakultas Teknologi Informasi, Universitas Budi Luhur di buat.

5. UCAPAN TERIMA KASIH

Penulis dapat menambahkan ucapan terima kasih untuk pihak-pihak yang mendukung kegiatan penelitian yang penulis lakukan.

6. DAFTAR PUSTAKA

[1] A. Solichin, A. Harjoko, and A. E. Putra, “Grid-based Histogram of Oriented Optical Flow for Analyzing Movements on Video Data,” in *2015 International Conference on Data and Software Engineering*, 2015, pp. 114–119.

[2] R. Maulunida and A. Solichin, “Optimization of LZW Compression Algorithm With Modification of Dictionary Formation,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 12, no. 1, p. 73, 2018.

[3] A. Solichin, *Pemrograman Web dengan PHP dan MySQL*. Budi Luhur Press, 2016.

[4] M. A. Romli and A. Solichin, “Pemrosesan Sinyal Digital Untuk Mengidentifikasi Akord Dasar Penyanyi Dengan Metode Mel Frequency Cepstral Coefficients (MFCC) Dan Jaringan Syaraf Tiruan Backpropagation Digital Signal Processing To Identify chords Singer Using Mel Frequency Cepstral Coef,” in *Seminar Nasional Multidisiplin Ilmu (SENMI) 2017*, 2017, no.

April, pp. 235–244.

- [5] A. Solichin, “Mengukur Kualitas Citra Hasil Steganografi,” *Achmatim.Net*, 2015. [Online]. Available: <http://achmatim.net/2015/04/16/mengukur-kualitas-citra-hasil-steganografi/>. [Accessed: 20-Jun-2016].

Peringatan

Kami sangat menghargai naskah yang dikirimkan, namun banyak kesalahan fatal dari

Author adalah tidak membaca dengan baik panduan di TEMPLATE ini, sehingga naskah yang dikirim tidak sesuai aturan template. Untuk beberapa alasan, naskah yang tidak sesuai template terkadang langsung di REJECT dan/atau minta di perbaiki jika kesalahannya minor. Untuk itu, lebih baik dibaca berulang kali, cek dan ricek sebelum submit naskah. Tujuannya untuk mempercepat proses naskah di Jurnal BIT dan secara tidak langsung Anda telah ikut membantu pengelola Jurnal.

