

Analisis Pola Permintaan Suku Cadang Menggunakan Algoritma K-Means

Nitania Yevana Hasugian^{*1}, Desi Pibriana²

^{1,2,3,4}Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

*e-mail: yevanatan@mhs.mdp.ac.id¹, desi.pibriana@mdp.ac.id²

ABSTRAK

Perkembangan teknologi informasi mendorong transformasi dalam pengelolaan data, termasuk pada manajemen persediaan barang. PT XYZ, selaku distributor suku cadang kendaraan roda dua di Palembang, menghadapi kendala dalam memahami karakteristik permintaan produk karena data historis penjualan belum dimanfaatkan secara optimal untuk keperluan analisis. Kondisi ini berpotensi menimbulkan inefisiensi operasional berupa penumpukan stok maupun kekosongan barang. Penelitian ini bertujuan untuk mengelompokkan produk berdasarkan pola permintaan menggunakan algoritma K-Means guna membantu perusahaan menyusun strategi pengelolaan stok yang lebih efektif. Metodologi yang diterapkan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk proses penambangan data dan model *Waterfall* untuk pengembangan sistem. Pengolahan data dan implementasi algoritma K-Means dilakukan menggunakan bahasa pemrograman Python dengan library *pandas* untuk pengolahan data, *scikit-learn* untuk algoritma K-Means dan evaluasi *Silhouette Score*, serta framework *Streamlit* untuk implementasi aplikasi berbasis web. Dataset mencakup 1.250 jenis produk unik dengan total 51.055 baris data transaksi periode Januari 2021 hingga November 2024. Hasil klasterisasi membagi produk ke dalam tiga kategori: 97,4% (1.217 produk) dalam kategori 'Kurang Laris', 2,3% (29 produk) kategori 'Laris', dan 0,3% (4 produk) kategori 'Sangat Laris'. Kualitas model divalidasi menggunakan *Silhouette Score* sebesar 0,939, yang menunjukkan kualitas pengelompokan yang sangat tinggi dan andal

Kata kunci : *Clustering, Data Mining, K-Means, Manajemen Stok, Suku Cadang*

ABSTRACT

The advancement of information technology is driving transformation in data management, including inventory management. PT XYZ, a distributor of two-wheeled vehicle spare parts in Palembang, faces challenges in understanding product demand characteristics due to the suboptimal use of historical sales data for analytical purposes. This condition risks operational inefficiencies such as overstock and stockout. This study aims to group products based on demand patterns using the K-Means algorithm to help the company formulate more effective stock management strategies. The applied methodology is the *Cross-Industry Standard Process for Data Mining* (CRISP-DM) for data mining and the *Waterfall* model for system development. Data processing and K-Means implementation were conducted using Python with the *pandas* library for data processing, *scikit-learn* for K-Means and *Silhouette Score* evaluation, and *Streamlit* framework for web-based application. The dataset covers 1,250 unique product types with 51,055 transaction rows from January 2021 to November 2024. Clustering results divided products into three categories: 97.4% (1,217 products) as 'Less Popular', 2.3% (29 products) as 'Popular', and 0.3% (4 products) as 'Very Popular'. Model quality was validated with a *Silhouette Score* of 0.939, indicating very high and reliable grouping accuracy.

Keywords : *Clustering, Data Mining, K-Means, Spare Parts, Stock Management*

PENDAHULUAN

Perkembangan teknologi informasi telah membawa dampak signifikan pada cara perusahaan mengelola data, terutama dalam konteks manajemen persediaan barang. Data mining sebagai salah satu cabang ilmu komputer terbukti mampu menggali pola tersembunyi dari kumpulan data berskala besar untuk menghasilkan informasi strategis yang bermanfaat bagi pengambilan keputusan bisnis (Ahmed et al., 2020; Fahmi et al., 2025; Nguyen et al., 2025). Berbagai penelitian menunjukkan bahwa penerapan data mining dalam manajemen persediaan dapat meningkatkan efisiensi operasional dengan mengidentifikasi pola permintaan produk secara akurat, sehingga perusahaan mampu menyusun strategi pengadaan yang lebih tepat sasaran (Huo & Zhan, 2024; Wijaya et al., 2025; Wang et al., 2025).

Dalam industri distribusi, khususnya distribusi suku cadang kendaraan bermotor, tantangan manajemen persediaan menjadi semakin kompleks karena tingginya variasi jenis produk dan fluktuasi pola permintaan. Kondisi ini dialami oleh PT XYZ, sebuah perusahaan distribusi suku cadang roda dua di Palembang. PT XYZ mendistribusikan lebih dari 1.250 jenis suku cadang, namun data historis penjualan yang tersimpan selama ini hanya dimanfaatkan untuk keperluan pencatatan administratif tanpa dilakukan analisis lebih lanjut. Penelitian terkini menegaskan pentingnya

analisis data historis penjualan untuk memahami pola konsumsi dan meminimalkan risiko *overstocking* maupun *stockout* (Suroso et al., 2023; Rahmania et al., 2024; Zhang et al., 2021). Berdasarkan hasil wawancara dengan direktur perusahaan, diketahui bahwa keputusan pengadaan stok masih mengandalkan intuisi dan pengalaman semata, tanpa dukungan analisis data historis, sehingga setiap produk diperlakukan secara seragam padahal masing-masing memiliki pola permintaan yang berbeda.

Permasalahan utama yang dihadapi PT XYZ adalah belum optimalnya pemanfaatan data historis penjualan untuk mendukung keputusan bisnis. Tanpa analisis berbasis data, perusahaan berisiko mengalami penumpukan stok pada produk dengan permintaan rendah sekaligus kekurangan stok pada produk dengan permintaan tinggi. Kondisi serupa telah ditemukan pada berbagai penelitian di sektor ritel dan logistik, di mana perusahaan yang tidak memanfaatkan analisis data cenderung mengalami inefisiensi biaya persediaan dan penurunan tingkat layanan kepada pelanggan (Afiasari et al., 2023; Gumilang, 2025; Manarung et al., 2025).

Untuk mengatasi permasalahan tersebut, metode *clustering* dalam *data mining* menjadi solusi yang tepat. *Clustering* berfungsi mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik, sehingga memudahkan proses analisis dan pengambilan keputusan (Annas & Wahab, 2023; Miranda & Sriani, 2025). Di antara berbagai algoritma *clustering*, K-Means merupakan yang paling banyak digunakan karena keunggulannya dalam kesederhanaan implementasi, kecepatan komputasi, dan efektivitasnya dalam menangani dataset berukuran besar (Ikotun et al., 2023; Sinaga & Yang, 2020; Falih et al., 2025). Dibandingkan algoritma seperti DBSCAN atau *Hierarchical Clustering*, K-Means lebih unggul untuk data yang memiliki distribusi permintaan yang terdefinisi dengan baik (William & Johan, 2025; Wardani et al., 2023).

Sejumlah penelitian empiris membuktikan efektivitas K-Means dalam konteks manajemen persediaan. Penelitian pada data stok *sparepart* mobil berhasil mengidentifikasi kelompok produk berdasarkan kecepatan penjualan untuk mendukung keputusan pengadaan (Haris et al., 2024). Pada sektor farmasi, K-Means berhasil membantu menentukan kebutuhan stok berdasarkan pola resep (Supriyanto et al., 2024; Dermawan et al., 2024). Penelitian pengelompokan produk ritel dan sembako menggunakan K-Means dengan metodologi CRISP-DM juga menghasilkan klaster yang mendukung efisiensi keputusan pengadaan (Setiawan et al., 2024; Sitinjak et al., 2024; Aji et al., 2023). Dalam implementasi proyek *data mining*, metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) telah menjadi standar industri yang menyediakan kerangka kerja sistematis mulai dari tahap pemahaman bisnis hingga *deployment* model (Schröer et al., 2021; Putri et al., 2025). Pengembangan sistem informasi berbasis web untuk visualisasi hasil *clustering* juga semakin banyak dilakukan menggunakan Python dan framework Streamlit (Handoto, 2025; Wikjatomiko et al., 2025).

Berdasarkan kajian literatur dan permasalahan PT XYZ, penelitian ini difokuskan pada penerapan algoritma K-Means untuk mengelompokkan produk suku cadang roda dua berdasarkan pola permintaan historis. Tujuan penelitian adalah menghasilkan sistem informasi berbasis web yang mampu mengolah data historis penjualan secara otomatis, melakukan pengelompokan produk menggunakan K-Means, menampilkan visualisasi hasil, serta memberikan rekomendasi strategi pengelolaan stok kepada manajemen PT XYZ.

METODE PENELITIAN

Penelitian ini mengintegrasikan dua pendekatan metodologis utama, yaitu CRISP-DM (**C**ross-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) sebagai kerangka kerja dalam proses pengolahan dan analisis data, serta model Waterfall sebagai metode pengembangan sistem aplikasi berbasis web. Integrasi kedua pendekatan ini bertujuan untuk menghasilkan sistem yang tidak hanya terstruktur secara teknis, tetapi juga mampu mengolah data secara sistematis dan menghasilkan informasi yang akurat serta dapat diandalkan.

Menurut Schröer dkk. [26], CRISP-DM merupakan metodologi data mining yang komprehensif, sistematis, dan telah terdokumentasi dengan baik sehingga banyak digunakan dalam berbagai penelitian dan implementasi industri. CRISP-DM menyediakan alur kerja yang jelas dan iteratif, memungkinkan peneliti

untuk melakukan perbaikan pada setiap tahap berdasarkan hasil evaluasi sebelumnya. Proses dalam CRISP-DM terdiri dari enam tahapan utama, yaitu:

1. **Business Understanding**

Tahap ini berfokus pada pemahaman tujuan penelitian atau kebutuhan bisnis yang ingin dicapai. Peneliti mengidentifikasi permasalahan, menentukan tujuan analisis, serta merumuskan kebutuhan sistem yang akan dikembangkan.

2. **Data Understanding**

Pada tahap ini dilakukan pengumpulan data awal serta eksplorasi untuk memahami karakteristik data. Kegiatan meliputi identifikasi tipe data, distribusi data, serta deteksi adanya anomali atau data yang tidak konsisten.

3. **Data Preparation**

Tahap ini merupakan proses pengolahan data sebelum dilakukan pemodelan. Kegiatan yang dilakukan meliputi pembersihan data (data cleaning), transformasi data, seleksi fitur, serta integrasi data agar siap digunakan dalam proses analisis.

4. **Modeling**

Pada tahap ini dilakukan penerapan teknik data mining atau algoritma tertentu untuk membangun model. Pemilihan metode disesuaikan dengan tujuan penelitian, misalnya klasifikasi, clustering, atau regresi.

5. **Evaluation**

Model yang telah dibangun kemudian dievaluasi untuk memastikan bahwa hasil yang diperoleh telah sesuai dengan tujuan penelitian. Evaluasi dilakukan menggunakan metrik tertentu untuk mengukur performa model.

6. **Deployment**

Tahap akhir adalah implementasi model ke dalam sistem aplikasi agar dapat digunakan secara nyata. Hasil analisis disajikan dalam bentuk yang mudah dipahami oleh pengguna, seperti dashboard atau laporan.

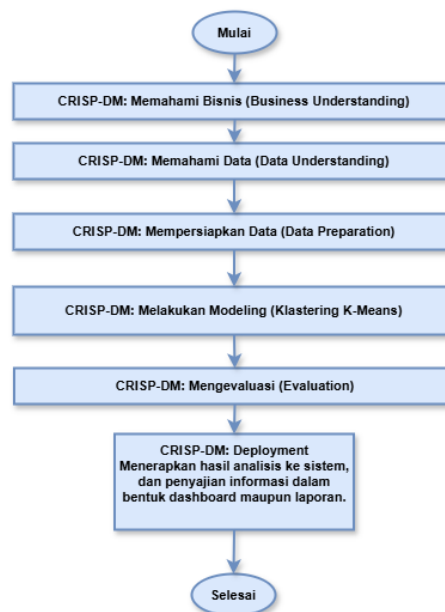
Sementara itu, model Waterfall digunakan dalam pengembangan sistem berbasis web karena menawarkan alur kerja yang linier, terstruktur, dan mudah dipahami. Model ini terdiri dari beberapa tahapan yang saling berurutan, yaitu analisis kebutuhan (requirement analysis), perancangan sistem (system design), implementasi (coding), pengujian (testing), serta pemeliharaan (maintenance). Setiap tahap harus diselesaikan terlebih dahulu sebelum melanjutkan ke tahap berikutnya, sehingga proses pengembangan menjadi lebih terkontrol dan terdokumentasi dengan baik. Pendekatan ini sangat sesuai digunakan dalam penelitian yang memiliki kebutuhan sistem yang telah terdefinisi secara jelas sejak awal, karena mampu meminimalkan perubahan yang signifikan selama proses pengembangan berlangsung.

Pada tahap analisis kebutuhan, peneliti mengidentifikasi spesifikasi sistem berdasarkan hasil dari proses business understanding pada CRISP-DM. Selanjutnya, tahap desain sistem dilakukan dengan merancang arsitektur aplikasi, struktur basis data, serta antarmuka pengguna (user interface) yang akan digunakan. Tahap implementasi merupakan proses penerjemahan desain ke dalam kode program berbasis web, sedangkan tahap pengujian bertujuan untuk memastikan bahwa sistem berjalan sesuai dengan kebutuhan dan bebas dari kesalahan. Tahap terakhir, yaitu pemeliharaan, dilakukan untuk memperbaiki bug serta melakukan pengembangan lanjutan apabila diperlukan.

Dengan mengombinasikan CRISP-DM (Cross-Industry Standard Process for Data Mining) dan model Waterfall, penelitian ini tidak hanya menekankan pada keakuratan dan ketepatan dalam proses analisis data, tetapi juga pada kualitas pengembangan sistem yang menjadi media implementasi hasil analisis tersebut. Integrasi kedua pendekatan ini menciptakan sinergi antara proses analitik dan rekayasa perangkat lunak, di mana hasil dari setiap tahapan data mining dapat langsung diakomodasi dalam tahapan pengembangan sistem.

Pendekatan terintegrasi ini juga memberikan keunggulan dalam hal keterlacakan (traceability), karena setiap hasil dari proses CRISP-DM dapat dihubungkan secara langsung dengan fitur atau modul yang dikembangkan dalam sistem berbasis web. Selain itu, dokumentasi yang dihasilkan pada setiap tahap Waterfall mendukung transparansi dan kemudahan evaluasi terhadap sistem yang dibangun. Dengan demikian, solusi yang dihasilkan tidak hanya bersifat teoritis, tetapi juga aplikatif dan implementatif.

Secara keseluruhan, kombinasi kedua metodologi ini diharapkan mampu menghasilkan solusi yang komprehensif dan berkelanjutan, mulai dari proses pengolahan dan analisis data hingga penyajian hasil dalam bentuk aplikasi berbasis web yang fungsional, mudah digunakan (user-friendly), serta mampu memberikan dukungan pengambilan keputusan yang lebih efektif dan berbasis data.



Gambar 1. Metodologi Penelitian

Business Understanding (Pemahaman Bisnis)

Pada tahap ini dilakukan identifikasi permasalahan bisnis PT XYZ secara mendalam. Berdasarkan hasil wawancara dengan direktur perusahaan, diketahui bahwa pencatatan stok bulanan yang ada belum dimanfaatkan untuk memahami pola permintaan setiap produk. Tujuan bisnis yang ditetapkan adalah mengelompokkan produk ke dalam tiga kategori berdasarkan pola permintaan: Kurang Laris (permintaan rendah), Laris (permintaan menengah), dan Sangat Laris (permintaan tinggi), agar perusahaan dapat menerapkan strategi pengadaan yang berbeda-beda sesuai karakteristik setiap kelompok.

Data Understanding (Pemahaman Data)

Data yang digunakan merupakan data primer berupa laporan stok bulanan internal PT XYZ, mencakup periode Januari 2021 hingga Desember 2024 yang disajikan dalam 48 file terpisah berformat Microsoft Excel. Proses eksplorasi awal dilakukan menggunakan Google Colaboratory dengan bahasa pemrograman Python dan library Pandas. Total data awal yang ditemukan berjumlah 51.343 baris, mencakup 1.250 jenis produk unik.

Tabel 1. Atribut dalam Dataset Awal (Mentah)

No	Atribut	Deskripsi
1	Kode Part	Kode unik internal setiap jenis suku cadang
2	Nama Part	Nama deskriptif suku cadang
3	Awal	Jumlah stok awal pada periode tersebut
4	Masuk	Jumlah stok yang masuk (pengadaan)

5	Keluar	Jumlah stok yang terjual; merepresentasikan permintaan produk
6	Akhir	Sisa stok yang tersedia di akhir periode

Sumber: Data primer penelitian PT XYZ, diolah (2021-2024)

Dari proses eksplorasi data (data exploration), teridentifikasi beberapa permasalahan kualitas data yang berpotensi memengaruhi keakuratan hasil analisis apabila tidak ditangani secara tepat. Permasalahan tersebut mencerminkan adanya ketidaksesuaian antara struktur data yang tersedia dengan kebutuhan analisis berbasis transaksi.

Pertama, ditemukan adanya baris non-transaksional yang berisi informasi administratif di bagian atas setiap file. Baris ini umumnya memuat metadata seperti judul laporan, tanggal pencetakan, atau keterangan sistem, yang tidak termasuk dalam data transaksi aktual. Keberadaan baris tersebut dapat mengganggu proses pembacaan data secara otomatis, terutama pada tahap parsing dan import data ke dalam sistem, sehingga berpotensi menyebabkan kesalahan dalam interpretasi struktur kolom maupun jumlah observasi. Oleh karena itu, diperlukan proses pembersihan awal (pre-processing) untuk menghapus baris-baris yang tidak relevan tersebut.

Kedua, ditemukan sebanyak 32.636 nilai negatif pada kolom 'Keluar', yang secara logika bisnis tidak valid karena merepresentasikan jumlah barang yang keluar atau terjual, yang seharusnya bernilai nol atau positif. Keberadaan nilai negatif ini mengindikasikan adanya kesalahan pencatatan, kesalahan input data, atau kemungkinan perbedaan definisi (misalnya penggunaan nilai negatif untuk retur atau koreksi stok). Permasalahan ini perlu ditangani secara hati-hati melalui proses validasi dan transformasi data, seperti dengan menghapus data yang tidak valid, melakukan koreksi berdasarkan aturan bisnis, atau mengklasifikasikan ulang data sesuai konteks yang benar.

Ketiga, terdapat potensi inkonsistensi tipe data dan duplikasi data yang perlu diverifikasi lebih lanjut. Inkonsistensi tipe data dapat terjadi ketika suatu kolom yang seharusnya bertipe numerik justru mengandung nilai string atau format campuran, sehingga dapat menghambat proses analisis maupun pemodelan. Sementara itu, duplikasi data dapat menyebabkan bias dalam hasil analisis karena satu transaksi dapat terhitung lebih dari satu kali. Oleh karena itu, diperlukan langkah-langkah seperti standarisasi tipe data, konversi format, serta deteksi dan penghapusan data duplikat menggunakan teknik tertentu.

Secara keseluruhan, temuan permasalahan kualitas data ini menunjukkan pentingnya tahap data preparation dalam kerangka CRISP-DM, di mana data harus melalui proses pembersihan, validasi, dan transformasi sebelum digunakan pada tahap pemodelan. Penanganan yang tepat terhadap permasalahan ini diharapkan dapat meningkatkan kualitas dataset sehingga menghasilkan analisis yang lebih akurat, reliabel, dan sesuai dengan kondisi nyata di lapangan.

Data Preparation (Persiapan Data)

Tahap ini bertujuan membersihkan, mengoreksi, dan mentransformasi data mentah menjadi dataset berkualitas tinggi yang siap dianalisis. Proses dilakukan secara otomatis menggunakan skrip Python. Tindakan pembersihan meliputi: (1) penghapusan 240 baris non-transaksional menggunakan metode deteksi header otomatis berbasis kata kunci; (2) penghapusan 48 baris tanpa Kode Part atau Nama Part yang valid; (3) koreksi 32.636 nilai negatif pada kolom 'Keluar' menggunakan fungsi nilai absolut `abs()`, mengubah nilai negatif menjadi positif yang logis (misalnya -5 menjadi 5); serta (4) validasi tipe data dan duplikasi yang mengonfirmasi tidak adanya inkonsistensi maupun pencatatan ganda dalam satu periode.

Untuk proses clustering, atribut yang dipilih adalah `kode_part`, `nama_part`, dan `stok_keluar`, karena secara langsung merepresentasikan identitas produk dan tingkat permintaannya. Atribut lain (Awal, Masuk, Akhir) tidak digunakan karena fokus penelitian adalah pada pergerakan barang keluar. Rangkuman hasil data preparation disajikan pada Tabel 2

Tabel 2. Rangkuman Hasil Data Preparation

Keterangan	Jumlah
Jumlah baris data mentah awal	51.343 baris
Baris dihapus (non-transaksional & header administratif)	240 baris
Baris dihapus (tidak memiliki Kode Part>Nama Part valid)	48 baris
Nilai negatif dikoreksi pada kolom Keluar	32.636 nilai
Jumlah baris data bersih siap analisis	51.055 baris
Jumlah produk unik	1.250 item

Modeling (Pemodelan)

Tahap pemodelan menerapkan algoritma K-Means *Clustering* terhadap dataset yang telah melalui proses persiapan data. Seluruh proses implementasi dilakukan menggunakan bahasa pemrograman Python di dalam aplikasi berbasis web yang dibangun dengan framework Streamlit (Handoto, 2025; Wikjatmiko et al., 2025).

a. Penentuan Jumlah Cluster dengan Elbow Method

Jumlah *cluster* optimal K ditentukan secara objektif menggunakan *Elbow Method*, yaitu dengan menghitung nilai *inertia* (*Sum of Squared Errors/SSE*) untuk setiap nilai K yang diuji dari K=2 hingga K=10. *Inertia* mengukur total kuadrat jarak setiap titik data ke centroid *cluster*-nya — semakin kecil nilai *inertia*, semakin padat pengelompokan yang terbentuk (Gumilang, 2025; Aji et al., 2023). Titik “siku” pada grafik *inertia*, yaitu titik di mana penurunan nilai *inertia* mulai melambat secara signifikan, menunjukkan nilai K yang paling optimal (Gumilang, 2025; Aji et al., 2023). Berdasarkan hasil pengujian *Elbow Method* yang ditampilkan pada Gambar X, titik siku terdeteksi pada K=3, sehingga jumlah *cluster* yang digunakan dalam penelitian ini adalah K=3, yang terdiri dari tiga kategori: Kurang Laris, Laris, dan Sangat Laris. Hasil ini juga konsisten dengan tujuan bisnis PT XYZ yang membutuhkan tiga kategori pengelompokan yang dapat ditindaklanjuti secara strategis.

b. Inisialisasi Centroid Awal

Setelah jumlah *cluster* ditetapkan, dipilih tiga titik data secara acak dari dataset sebagai *centroid* awal. *Centroid* merupakan nilai pusat (*center*) dari sebuah *cluster*. Inisialisasi *centroid* dilakukan secara acak sebagaimana umumnya diterapkan pada algoritma K-Means (Ikotun et al., 2023; Sinaga & Yang, 2020). Contoh *centroid* awal yang digunakan disajikan pada Tabel 3.

Tabel 3. Contoh Centroid Awal (Dipilih Secara Acak)

No.	Cluster	Kode Part	Nama Part	Stok Keluar (Unit)
1	C1	H2-35010-K25-1200	KEY SET BEAT FI	170
2	C2	YH-E76PA-NMX-1200	PAKET V-BELT	1948
3	C3	H2-131P5-KVB-1100	PAKET PISTON (1.00)	474

Sumber: Data primer penelitian PT XYZ, diolah (2021-2024)

c. Proses Iterasi K-Means

Tahap modeling dalam penelitian ini dilakukan dengan menerapkan algoritma K-Means *Clustering* terhadap dataset yang telah melalui proses data preparation. Tahap eksplorasi dan preprocessing data dilaksanakan menggunakan Google Colaboratory untuk memahami struktur, distribusi, serta karakteristik variabel yang digunakan. Sementara itu, implementasi algoritma K-Means dan seluruh proses klasterisasi dilakukan dalam aplikasi berbasis web yang dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit, sehingga memungkinkan analisis dilakukan secara interaktif dan aplikatif.. Penentuan jumlah klaster (K) ditetapkan sebanyak tiga klaster, yang disesuaikan dengan kebutuhan bisnis PT XYZ, yaitu mengelompokkan produk ke

dalam kategori Kurang Laris, Laris, dan Sangat Laris. Penetapan nilai K ini bersifat goal-oriented, sehingga hasil klasterisasi tidak hanya bersifat matematis, tetapi juga relevan secara operasional dalam mendukung pengambilan keputusan manajerial.

Secara konseptual, algoritma K-Means bekerja secara iteratif melalui dua tahapan utama, yaitu assignment dan update. Pada tahap assignment, setiap data dihitung jaraknya terhadap masing-masing centroid menggunakan metode Euclidean Distance, kemudian dialokasikan ke dalam klaster dengan jarak terdekat. Tahap ini bertujuan untuk meminimalkan jarak intra-klaster sehingga data dalam satu kelompok memiliki tingkat kemiripan yang tinggi. Selanjutnya, pada tahap update, nilai centroid diperbarui dengan menghitung rata-rata dari seluruh anggota pada masing-masing klaster, sehingga posisi centroid mencerminkan pusat distribusi data yang baru.

Proses klasterisasi diawali dengan inisialisasi centroid awal yang dipilih secara acak dari dataset. Selanjutnya, kedua tahapan tersebut dilakukan secara berulang hingga mencapai kondisi konvergensi, yaitu ketika tidak terjadi lagi perpindahan data antar klaster dan posisi centroid telah stabil. Proses iteratif ini memastikan bahwa algoritma mencapai solusi yang relatif optimal dalam meminimalkan variasi dalam klaster (within-cluster variance). Dengan pendekatan tersebut, diharapkan terbentuk struktur pengelompokan data yang tidak hanya optimal secara matematis, tetapi juga mampu merepresentasikan pola permintaan produk secara lebih akurat. Hal ini menjadi dasar penting dalam mendukung analisis lanjutan serta perumusan strategi pengelolaan persediaan yang lebih efektif dan berbasis data.

Evaluation (Evaluasi)

Evaluasi kualitas model clustering dilakukan menggunakan Silhouette Score. Metrik ini mengukur seberapa baik sebuah objek dikelompokkan ke dalam clusternya sendiri dibandingkan dengan cluster lain. Nilai Silhouette Score berkisar dari -1 hingga +1; nilai yang mendekati +1 menunjukkan pengelompokan yang sangat baik dan kohesif [30]. Pemilihan Silhouette Score sebagai satu-satunya metrik evaluasi didasarkan pada kemampuannya mengukur kualitas clustering secara komprehensif tanpa memerlukan label kelas yang diketahui sebelumnya.

Deployment (Penyebaran)

Streamlit serta sistem manajemen basis data MySQL. Implementasi ini bertujuan untuk mentransformasikan hasil analisis data menjadi sebuah sistem yang interaktif, mudah digunakan, dan dapat mendukung proses pengambilan keputusan secara langsung.

Aplikasi yang dikembangkan dirancang dengan konsep role-based access control, yaitu menyediakan dua tingkat hak akses pengguna yang memiliki fungsi dan tanggung jawab berbeda. Pertama, Admin berperan dalam pengelolaan data, termasuk proses impor data historis penjualan, validasi data, serta pemeliharaan database. Admin juga bertanggung jawab memastikan bahwa data yang digunakan dalam proses clustering selalu terbaru dan memiliki kualitas yang baik, sehingga hasil analisis tetap akurat dan relevan.

Kedua, Direktur memiliki akses terhadap fitur analisis yang lebih strategis. Pada level ini, pengguna dapat melakukan clustering secara interaktif, melihat hasil pengelompokan produk dalam bentuk visualisasi yang informatif, serta memperoleh rekomendasi strategi pengelolaan stok secara otomatis berdasarkan hasil clustering. Visualisasi yang disajikan membantu dalam memahami pola penjualan, seperti identifikasi produk dengan tingkat permintaan rendah hingga tinggi, sehingga memudahkan dalam menentukan kebijakan bisnis.

Selain itu, sistem ini juga dirancang untuk mendukung pengambilan keputusan berbasis data (data-driven decision making), di mana hasil clustering tidak hanya ditampilkan sebagai output analisis, tetapi juga diterjemahkan menjadi insight yang aplikatif. Misalnya, produk dalam kategori Kurang Laris dapat direkomendasikan untuk dilakukan promosi atau pengurangan stok, sedangkan produk Sangat Laris dapat diprioritaskan dalam pengadaan dan distribusi.

Dengan adanya integrasi antara proses data mining dan sistem berbasis web ini, hasil clustering tidak berhenti pada tahap analisis saja, tetapi dapat langsung dimanfaatkan dalam operasional bisnis sehari-hari. Hal ini menjadikan sistem yang dikembangkan tidak hanya bersifat analitis, tetapi juga implementatif, adaptif, dan memiliki nilai praktis yang tinggi bagi pengambilan keputusan strategis perusahaan.

HASIL DAN PEMBAHASAN**Hasil Penelitian****Karakteristik Sampel Penelitian****Tabel 4.1 Karakteristik Sampel Penelitian**

Kode Part	Nama Part	Stok Keluar	Bulan	Tahun
12-27517-ATR01	TIRE 27517 TRAIL	2427	Januari	2021
12-27518-AT701X	TIRE 275-18	15	Januari	2021
12-2753001400000	TUBE 275/300- 14	810	Januari	2021
12-2753001800000	TUBE, TYRE 275/300-18	404	Januari	2021
...	...	...	...	...

Sampel penelitian penelitian terdiri dari data stok keluar suku cadang periode Januari 2021 hingga Desember 2024 yang diperoleh dari sistem persediaan perusahaan. Jenis part yang dianalisis meliputi suku cadang kendaraan (*sparepart*) ban, oli / pelumas, aksesoris kendaraan untuk berbagai ukuran.

Variasi volume permintaan antar part cukup signifikan, yang menjadi dasar pengelompokan menggunakan algoritma K-Means untuk mengidentifikasi kategori part berdasarkan tingkat permintaannya (tinggi, sedang, rendah) selama periode empat tahun tersebut.

Hasil

Penelitian ini menghasilkan pengelompokan 1.250 jenis produk suku cadang PT XYZ ke dalam tiga kategori berdasarkan pola permintaan historis periode 2021–2024. Proses clustering diimplementasikan pada aplikasi berbasis web dengan framework Streamlit. Berikut dipaparkan hasil dari setiap tahapan pemodelan.

Elbow Method dan Penentuan K Optimal

Pengujian Elbow Method dilakukan terhadap nilai K=2 hingga K=10. Hasil perhitungan inertia untuk setiap nilai K disajikan pada Tabel 4.

Tabel 4. Nilai Inertia per Jumlah Cluster (K)

K	Inertia (SSE)
2	365,5074
3	171,7132
4	71,8926
5	40,0797
6	20,7029
7	12,8918
8	8,8608
9	6,7358
10	5,4236

Sumber: Data primer penelitian PT XYZ, diolah (2021-2024)

Centroid Akhir dan Konvergensi

Algoritma K-Means mencapai konvergensi setelah 12 iterasi, ditandai dengan tidak adanya perpindahan anggota antar cluster dan posisi centroid yang stabil. Posisi centroid akhir yang menentukan batas setiap kategori disajikan pada Tabel 4.

Tabel 3. Nilai Centroid Akhir dan Kategori Cluster

Cluster	Kategori	Nilai Centroid Akhir (Stok Keluar)
Cluster 1 (C1)	Kurang Laris	2.390,90
Cluster 2 (C2)	Laris	94.516,10
Cluster 3 (C3)	Sangat Laris	538.430,25

Nilai centroid akhir mencerminkan rata-rata total permintaan (stok keluar) selama periode analisis. Cluster 1 (Kurang Laris) memiliki centroid 2.390,90 unit, Cluster 2 (Laris) sebesar 94.516,10 unit, dan Cluster 3 (Sangat Laris) sebesar 538.430,25 unit.

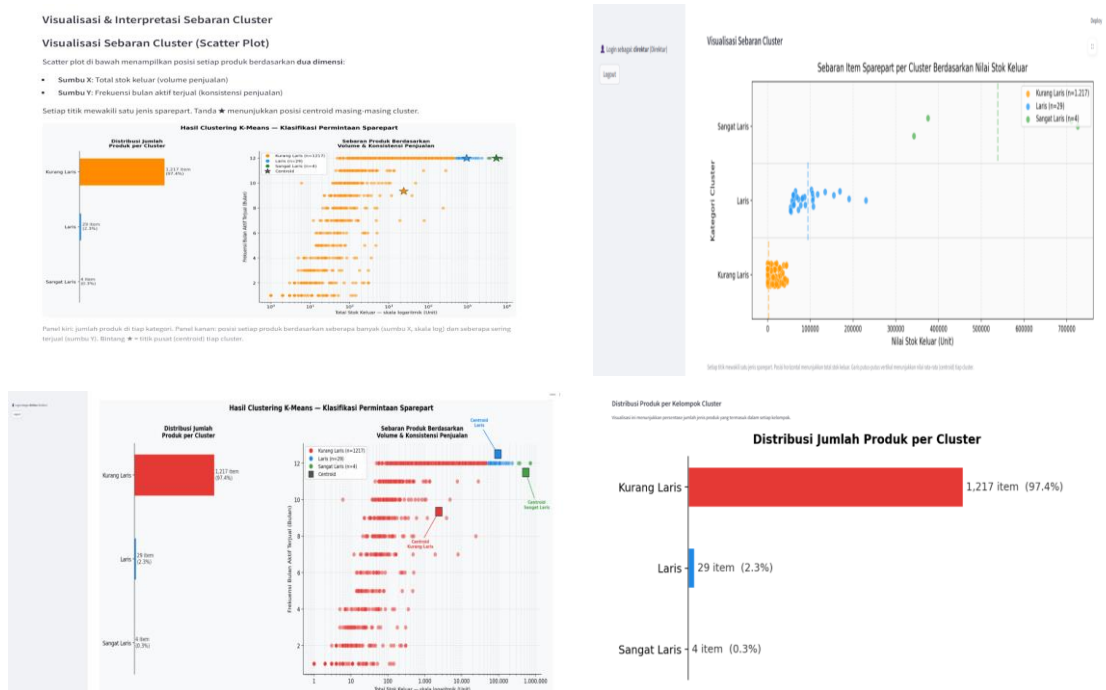
Distribusi Produk per Cluster

Distribusi produk ke dalam masing-masing cluster disajikan pada Tabel 5. Tampilan visualisasi distribusi produk pada aplikasi web ditunjukkan pada Gambar 1.

Tabel 4. Distribusi Produk per Cluster

Kategori Cluster	Jumlah Produk	Persentase (%)
Kurang Laris	1.217	97,4%
Laris	29	2,3%
Sangat Laris	4	0,3%
Total	1.250	100%

Sumber: Data primer penelitian PT XYZ, diolah (2021-2024)



Gambar 1. Visualisasi Distribusi Produk per Cluster pada Aplikasi Web

Hasil distribusi menunjukkan bahwa portofolio produk PT XYZ memiliki komposisi permintaan yang sangat tidak seimbang. Mayoritas absolut produk, yaitu 1.217 item (97,4%), masuk ke dalam kategori 'Kurang Laris'. Hanya 29 produk (2,3%) yang masuk kategori 'Laris' dan 4 produk (0,3%) yang masuk kategori 'Sangat Laris'. Dengan demikian, hanya 2,6% dari total produk yang tergolong aktif bergerak, sementara 97,4% sisanya merupakan produk slow-moving.

Evaluasi Kualitas Clustering

Evaluasi kualitas model clustering dilakukan menggunakan Silhouette Score, yang ditampilkan pada tab analitik direktur dalam aplikasi web. Hasil evaluasi menunjukkan nilai sebesar 0,939, yang mendekati +1. Menurut Hasan [30], nilai tersebut mengindikasikan bahwa setiap cluster memiliki tingkat kohesi yang tinggi dan terpisah dengan jelas dari cluster lainnya. Dengan demikian, model clustering yang dihasilkan dinilai sangat baik dan layak digunakan sebagai dasar dalam pengambilan keputusan strategis.

Pembahasan

Hasil penelitian menunjukkan bahwa penerapan algoritma K-Means mampu mengelompokkan 1.250 produk suku cadang PT XYZ ke dalam tiga kategori utama yang memiliki makna strategis dalam pengelolaan bisnis. Tingginya kualitas klasterisasi yang ditunjukkan oleh nilai Silhouette Score sebesar 0,939 mengindikasikan bahwa pemisahan antar cluster sangat jelas (well-separated) dan kohesi dalam cluster sangat baik, sehingga hasil pengelompokan dapat dikategorikan sangat reliabel untuk mendukung pengambilan keputusan berbasis data.

Distribusi produk yang sangat tidak merata, dengan 97,4% produk berada pada kategori kurang laris, menjadi temuan krusial. Kondisi ini mencerminkan adanya ketidakseimbangan dalam portofolio produk, di mana sebagian besar item memiliki perputaran rendah (slow-moving). Secara manajerial, hal ini berimplikasi pada tingginya potensi overstock dan modal kerja yang terikat, serta meningkatnya biaya penyimpanan. Di sisi lain, proporsi produk sangat laris yang sangat kecil namun memiliki volume permintaan tinggi menunjukkan adanya ketergantungan bisnis pada sejumlah kecil produk kunci (high-demand items).

Analisis per cluster memberikan wawasan yang lebih operasional. Cluster kurang laris (C1) yang mendominasi menunjukkan pola permintaan yang tidak stabil dengan rentang yang luas, sehingga memerlukan pendekatan pengendalian stok yang adaptif, seperti penerapan sistem just-in-time, pembatasan stok minimum, serta strategi pemasaran seperti diskon dan bundling untuk mempercepat perputaran barang. Cluster laris (C2) merepresentasikan produk dengan permintaan yang relatif stabil, sehingga strategi yang tepat adalah menjaga keseimbangan stok melalui pemantauan berkala dan optimalisasi promosi musiman. Sementara itu, cluster sangat laris (C3) meskipun jumlahnya sangat terbatas, memiliki kontribusi signifikan terhadap volume penjualan, sehingga harus menjadi prioritas utama dalam perencanaan pengadaan, dengan pengawasan ketat terhadap ketersediaan untuk menghindari stockout.

Dibandingkan dengan penelitian sebelumnya, nilai Silhouette Score yang diperoleh dalam penelitian ini lebih tinggi, yang menunjukkan kualitas segmentasi data yang lebih baik. Hal ini mengindikasikan bahwa pola permintaan produk pada PT XYZ memiliki karakteristik yang tersegmentasi secara alami, sehingga metode clustering menjadi sangat efektif dalam mengidentifikasi struktur data tersebut.

Selain itu, implementasi sistem berbasis web menggunakan Python dan framework Streamlit memberikan nilai tambah praktis, karena memungkinkan proses analisis dilakukan secara interaktif dan real-time oleh pihak manajemen. Hal ini tidak hanya meningkatkan efisiensi analisis data, tetapi juga memperkuat penerapan data-driven decision making dalam pengelolaan persediaan dan strategi bisnis.

SIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan 1.250 produk suku cadang roda dua PT XYZ ke dalam tiga kategori berdasarkan pola permintaan, yaitu Kurang Laris (97,4%), Laris (2,3%), dan Sangat Laris (0,3%). Kualitas model tergolong sangat baik dengan Silhouette Score sebesar 0,939, yang menunjukkan pengelompokan yang akurat dan sesuai dengan karakteristik data.

Hasil analisis mengungkap distribusi permintaan yang tidak seimbang, di mana mayoritas produk termasuk kategori slow-moving. Temuan ini memberikan dasar penting bagi perusahaan dalam merancang strategi manajemen inventori yang lebih efisien dan tepat sasaran.

Implementasi hasil penelitian diwujudkan dalam aplikasi berbasis web menggunakan Python dan framework Streamlit, sehingga memudahkan manajemen dalam melakukan analisis, visualisasi, serta memperoleh rekomendasi strategi stok. Ke depan, disarankan pengembangan model prediktif untuk mendukung pengambilan keputusan yang lebih proaktif.

SARAN

Penelitian ini menghasilkan beberapa saran yang dapat dijadikan tindak lanjut. Bagi PT XYZ, hasil clustering dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam manajemen persediaan, di mana produk kategori *Sangat Laris* perlu diprioritaskan untuk menjaga ketersediaan stok, sementara produk *Kurang Laris* dapat dievaluasi melalui strategi seperti pengurangan stok, promosi, atau penghentian distribusi. Selain itu, sistem berbasis web yang telah dikembangkan menggunakan Python dan framework Streamlit sebaiknya diintegrasikan ke dalam operasional perusahaan agar analisis dapat dilakukan secara berkelanjutan. Dari sisi pengembangan, sistem masih dapat ditingkatkan melalui penambahan fitur pembaruan data secara real-time, integrasi dengan sistem lain seperti ERP, serta peningkatan kualitas visualisasi. Untuk penelitian selanjutnya, disarankan pengembangan model prediktif (forecasting) guna memperkirakan permintaan di masa depan serta eksplorasi metode clustering lain untuk membandingkan performa model. Selain itu, penambahan variabel seperti faktor musiman atau tren pasar juga dapat dilakukan agar menghasilkan analisis yang lebih komprehensif dan mendalam.

DAFTAR PUSTAKA

- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The K-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1–12. <https://doi.org/10.3390/electronics9081295>
- Fahmi, S. A., Ikhwan, A., & Sibarani, F. H. (2025). Implementasi data mining dengan menggunakan metode K-Means clustering untuk menentukan penjualan sparepart mobil. *Jurnal Komputer Teknologi Informasi Sistem Informasi*, 4(1), 204–209. <https://doi.org/10.62712/juktisi.v4i1.388>
- Nguyen, D. V. H., Nguyen, T. X. H., & Nguyen, H. G. (2025). K-means clustering and machine learning-based forecasting model for the automotive spare parts industry. *Lecture Notes in Networks and Systems*, 1255, 473–484. https://doi.org/10.1007/978-981-96-1747-0_39
- Huo, Z., & Zhan, X. (2024). Revolutionizing inventory management: The role of data mining in Industry 4.0. *Advances in Economics, Management and Political Sciences*, 109(1), 67–72. <https://doi.org/10.54254/2754-1169/109/2024bj0121>
- Wijaya, H. R., Khairil, K., & Lianda, D. (2025). Implementasi data mining menggunakan algoritma apriori dalam menentukan persediaan barang. *Digital Transformation Technology*, 5(1), 427–433. <https://doi.org/10.47709/digitech.v5i1.6381>
- Wang, Z., Xue, H., & Hu, J. (2025). Research on supply chain inventory demand forecasting and optimization model based on spatio-temporal data mining methods. *International Journal of Housing Science and Its Applications*, 46(3), 5774–5784. <https://doi.org/10.70517/ijhsa463497>
- Suroso, F., Rahmah, G. M., & Permana, D. R. A. (2023). Implementasi sistem peramalan kebutuhan spare part mobil dengan WMA. *Jurnal Teknologi dan Manajemen*, 21(2), 113–122. <https://doi.org/10.52330/jtm.v21i2.136>

- Rahmania, R. A., Indriani, N., & Wulandari, C. P. (2024). Optimizing spare parts inventory management of truck dealer services using forecasting methods and continuous review system approach. *Journal of Advanced Technology and Multidisciplinary*, 3(2), 60–65. <https://doi.org/10.20473/jatm.v3i2.66530>
- Zhang, S., Huang, K., & Yuan, Y. (2021). Spare parts inventory management: A literature review. *Sustainability*, 13(5), 1–23. <https://doi.org/10.3390/su13052460>
- Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi data mining transaksi penjualan menggunakan algoritma clustering dengan metode K-Means. *Jurnal SAINTEKOM*, 13(1), 100–110. <https://doi.org/10.33020/saintekom.v13i1.402>
- Gumilang, M. C. (2025). Optimasi manajemen stok melalui pengelompokan produk ritel berdasarkan data penjualan dengan K-Means. *Jurnal Algoritma*, 22(2). <https://doi.org/10.33364/algoritma/v.22-2.2526>
- Manarung, R. I., Widodo, E., & Rifai, A. M. (2025). Sales data clustering using the K-Means algorithm to determine retail product needs. *International Journal of Software Engineering and Computer Science*, 5(1), 226–234. <https://doi.org/10.35870/ijsecs.v5i1.4090>
- Annas, M., & Wahab, S. N. (2023). Data mining methods: K-Means clustering algorithms. *International Journal of Cyber and IT Service Management*, 3(1), 40–47.
- Miranda, M., & Sriani, S. (2025). Implementation of K-Means clustering in grouping sales data at Zura Mart. *Journal of Applied Informatics and Computing*, 9(2), 547–555. <https://doi.org/10.30871/jaic.v9i2.9160>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Falih, A. R. F., Kurniawan, R., Wijaya, Y. A., & Anwar, S. (2025). Algoritma K-Means untuk optimalisasi model clustering data penjualan toko online di TikTok Shop dalam strategi pemasaran. *Jurnal Sistem Informasi Kaputama*, 9(1), 1–11. <https://doi.org/10.59697/jsik.v9i1.929>
- William, & Johan, M. E. (2025). Implementation of customer segmentation model using K-Means and DBSCAN for fashion industry product transaction. *International Journal of Informatics and Visualization*, 9(6), 2559–2568. <https://doi.org/10.62527/joiv.9.6.2978>
- Wardani, S. D. K., Ariyanto, A. S., Umroh, M., & Rolliawati, D. (2023). Perbandingan hasil metode clustering K-Means, DBSCAN & hierarchical untuk analisa segmentasi pasar. *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 191. <https://doi.org/10.26798/jiko.v7i2.796>
- Haris, Y., Friadi, J., Frederick, A. E. S., Huda, D. N., & Romdoni, M. R. (2024). Clustering data stok penjualan sparepart mobil Toyota Bengkel Multi Topindo menggunakan K-Means. *Jurnal Ilmu Siber dan Teknologi Digital*, 2(2), 109–121. <https://doi.org/10.35912/jisted.v2i2.3308>
- Supriyanto, H., et al. (2024). Klasterisasi data obat farmasi berdasarkan jumlah persediaan dengan menggunakan metode K-Means. *Teknika*, 13(3), 361–369. <https://doi.org/10.34148/teknika.v13i3.987>

- Dermawan, A. A., et al. (2024). Optimizing inventory management: Data-driven insights from K-Means clustering analysis of prescription patterns. *Scientific Journal of Informatics*, 11(3), 811–820. <https://doi.org/10.15294/sji.v11i3.8690>
- Setiawan, D. Y. D., Hadikristanto, W., & Edora. (2024). Grouping production goods requirements using the K-Means clustering method. *International Journal of Software Engineering and Computer Science*, 4(2), 418–429. <https://doi.org/10.35870/ijsecs.v4i2.2863>
- Sitinjak, J. M., Sari, K., & Yetri, M. (2024). Penerapan data mining dalam penjualan sparepart motor dengan menggunakan metode K-Means clustering. *Jurnal Sistem Informasi Triguna Dharma*, 3(6), 862–871.
- Aji, N. S., Natsir, F., & Istianah, S. (2023). Penentuan penjualan barang berdasarkan pengelompokan produk dengan K-Means clustering metode CRISP-DM pada CV Sembako Dina. *Jurnal Zetroem*, 5(2), 119–126. <https://doi.org/10.36526/ztr.v5i2.3041>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Putri, A. E., MZ, Y., & Bororing, J. E. (2025). Implementasi K-Means clustering dan model CRISP-DM untuk pengelompokan daerah rawan tindak pidana narkoba di DIY. *Jurnal Informatika Teknologi dan Sains*, 7(2), 606–615. <https://doi.org/10.51401/jinteks.v7i2.5690>
- Handoto, H. C. (2025). Perancangan website interaktif untuk analisis disparitas kabupaten/kota di Indonesia. *[Nama jurnal tidak lengkap]*, 20(2), 193–204.
- Wikjatmiko, Z. B., Setyati, E., & Kristian, Y. (2025). Clustering dan visualisasi data ASN dalam penunjang analisis kecukupan data di perangkat daerah pemerintah Provinsi Jawa Timur. *Jurnal Informatic Engineering*, 10(2), 183–193.
- Hasan, Y. (2024). Pengukuran silhouette score dan Davies-Bouldin index pada hasil cluster. *Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, 6(1), 60–74.
- Lestari, E. D. N., Rezky, S. F., & Alhafiz, A. (2025). Penerapan data mining untuk pengelompokan data stok barang dengan metode K-Means clustering. *Jurnal Sistem Informasi Triguna Dharma*, 4(3), 528–539. <https://doi.org/10.53513/jursi.v4i3.8515>
- Miranda, M., & Sriani, S. (2025). Implementation of K-Means clustering in grouping sales data at Zura Mart. *Journal of Applied Informatics and Computing*, 9(2), 547–555. <https://doi.org/10.30871/jaic.v9i2.9160>