

DASHBOARD VISUALISASI PERFORMA PENJUALAN MENGGUNAKAN METODE KNOWLEDGE DISCOVERY IN DATABASE (KDD) DENGAN ALGORITMA K-MEANS DAN REGRESI LINEAR

Deah Ayu Safitri¹, Andri^{2*}
Sistem Informasi^{1,2}, Universitas Bina Darma^{1,2}
deahayusafitri@gmail.com¹, andri@binadarma.ac.id²

*Corresponding Author: andri@binadarma.ac.id

Abstrak

Pemanfaatan data penjualan secara optimal menjadi kebutuhan penting dalam mendukung pengambilan keputusan bisnis berbasis data. Kesuma Motor Sport sebagai toko variasi dan aksesoris motor memiliki volume data transaksi yang besar, namun pemanfaatannya masih terbatas pada laporan statis sehingga belum mampu memberikan gambaran performa penjualan secara komprehensif. Penelitian ini bertujuan untuk menganalisis performa penjualan dan perilaku pelanggan melalui penerapan metode Knowledge Discovery in Database (KDD) dengan algoritma K-Means dan Regresi Linear, serta mengembangkan dashboard visualisasi interaktif berbasis Microsoft Power BI. Algoritma K-Means digunakan untuk mengelompokkan data penjualan berdasarkan pola dan karakteristik pelanggan, sedangkan Regresi Linear diterapkan untuk menganalisis hubungan antar variabel dan memprediksi tren penjualan di masa mendatang. Hasil penelitian menunjukkan terbentuknya beberapa cluster penjualan dengan karakteristik performa yang berbeda serta pola tren penjualan yang meningkat, menurun, dan fluktuatif. Dashboard yang dikembangkan mampu menyajikan informasi secara visual dan interaktif, sehingga membantu manajemen dalam melakukan evaluasi kinerja dan pengambilan keputusan strategis berbasis data.

Kata kunci: Knowledge Discovery in Database (KDD); K-Means Clustering; Regresi Linear; Dashboard Power BI; Analisis Penjualan.

Abstract

Data plays a crucial role in supporting business decision-making through systematic processing that transforms raw data into valuable information. This study aims to develop a sales performance analysis and visualization system for Kesuma Motor Sport using the Knowledge Discovery in Database (KDD) approach. The research applies K-Means clustering to segment customer behavior and sales performance patterns, and Linear Regression to analyze relationships among sales variables and predict future sales trends. The processed data are visualized through an interactive dashboard developed using Microsoft Power BI to enhance data interpretation and managerial insight. The results demonstrate that the implementation of KDD successfully identifies meaningful sales patterns, customer segments, and trend behaviors that were previously difficult to interpret using conventional reports. The dashboard provides an informative and user-friendly visualization, supporting faster and more data-driven decision-making. This research contributes to the application of data mining and business intelligence in the automotive retail sector and supports digital transformation in sales performance monitoring.

Keywords : Knowledge Discovery in Database (KDD); K-Means Clustering; Linear Regression; Power BI Dashboard; Sales Analysis.

1. Pendahuluan

Berbagai sektor kehidupan memanfaatkan data sebagai instrumen utama untuk menghasilkan informasi yang bermakna dan bernilai guna, serta menjadi dasar dalam pengambilan keputusan dan pengembangan inovasi. Salah satu pendekatan yang banyak digunakan dalam mengolah data menjadi informasi bernilai adalah data mining. *Data mining* merupakan proses penemuan pola, hubungan, dan pengetahuan tersembunyi dari kumpulan data berukuran besar dengan memanfaatkan metode statistik, algoritma, dan kecerdasan buatan[1]. Salah satu metode yang ada di dalam data mining adalah *Knowledge Discovery in Database* (KDD). KDD merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan berguna[2].

Salah satu algoritma yang sering digunakan dalam *data mining* adalah *K-Means Clustering*, yaitu algoritma yang berfungsi untuk mengelompokkan data berdasarkan tingkat kemiripan atributnya. Dalam konteks bisnis, algoritma ini dapat dimanfaatkan untuk mendukung keakuratan strategi pemasaran, pengelolaan stok, promosi, hingga analisis performa pada platform digital seperti media sosial dan e-commerce[1]

Saat ini, Kesuma Motor Sport telah menggunakan sistem akuntansi internal bernama Sevensoft, yaitu aplikasi yang berfungsi untuk mencatat dan menganalisis berbagai transaksi bisnis. Meskipun fitur yang tersedia tergolong lengkap, sistem ini masih memiliki keterbatasan dalam hal visualisasi data. Hasil analisis yang dihasilkan masih berupa laporan dalam format Excel sederhana, yang bersifat statis dan kurang interaktif. Kondisi ini membuat proses interpretasi data menjadi kurang efisien, terutama ketika pihak manajemen membutuhkan gambaran performa penjualan secara cepat dan menyeluruh.

Oleh karena itu penelitian ini bertujuan untuk menerapkan metode *Knowledge Discovery in Database* (KDD) dalam proses pengolahan dan analisis data penjualan, mengelompokkan perilaku pelanggan berdasarkan cluster dengan algoritma K-Means, menganalisis tren penjualan dengan regresi linear serta sebagai output yaitu perancangan dashboard visualisasi penjualan dengan 4 dashboard utama, sehingga penelitian ini dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang sistem informasi serta sebagai pemberi masukan tambahan bagi pihak terkait dalam pengambilan keputusan dan evaluasi penjualan.

2. Kajian Pustaka Dan Pengembangan Hipotesis

2.1 Konsep Dasar Knowledge Discovery in Database (KDD) dan Data Mining

Knowledge Discovery in Databases (KDD) adalah proses secara umum yang melibatkan beberapa tahapan atau langkah[3]. Knowledge Discovery in Database (KDD) merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan bermanfaat[2]. Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data[4]. Di dalam *Knowledge Discovery in Database* terdapat 5 tahapan utama yaitu seleksi data, pra-proses data, transformasi data, data mining, dan yang terakhir interpretasi dan evaluasi. Data mining merupakan salah satu tahap keseluruhan dari proses KDD atau bisa dikatakan sebagai inti dari proses KDD.

2.2 Algoritma K-Means dan Regresi Linear dalam Analisa data serta prediksi penjualan

K-Means merupakan metode pengelompokan data non-hierarkis yang bekerja dengan cara membagi data ke dalam sejumlah klaster berdasarkan tingkat kemiripan atribut, di mana setiap klaster direpresentasikan oleh titik pusat (centroid). Metode K-Means merupakan salah satu algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan karakteristik antar data[5].

Selain pengelompokan, analisis tren dan hubungan antar variabel juga menjadi aspek penting dalam analisis data. Regresi Linear merupakan metode statistik yang digunakan untuk memodelkan hubungan linier antara variabel independen dan variabel dependen, serta digunakan untuk melakukan prediksi nilai di masa mendatang. Dalam konteks bisnis, regresi linear sering dimanfaatkan untuk menganalisis tren penjualan dan memprediksi perkembangan kinerja berdasarkan data historis[6].

Kombinasi algoritma K-Means dan Regresi Linear memungkinkan analisis data yang lebih komprehensif, di mana K-Means berperan dalam mengelompokkan data berdasarkan karakteristik tertentu, sementara Regresi Linear digunakan untuk memahami pola tren dan melakukan prediksi. Integrasi kedua metode ini mendukung pengambilan keputusan yang lebih efektif dan berbasis data.

3. Metode Penelitian

Penelitian ini merupakan penelitian terapan (applied research) dengan pendekatan kuantitatif dan deskriptif yang berfokus pada penerapan metode Knowledge Discovery in Database (KDD) menggunakan algoritma K-Means dan Regresi Linear untuk menganalisis serta memvisualisasikan data penjualan Kesuma Motor Sport melalui Microsoft Power BI. Data penelitian diperoleh dari dokumen transaksi penjualan yang tercatat dalam sistem internal Seventsoft, serta didukung oleh wawancara dengan pihak manajemen dan staf administrasi guna memahami kebutuhan bisnis dan indikator kinerja utama (KPI) yang relevan. Proses analisis data dilakukan melalui tahapan KDD yang meliputi pemilihan atribut penjualan yang relevan, pembersihan dan transformasi data, penerapan K-Means untuk pengelompokan pola transaksi, serta Regresi Linear untuk analisis hubungan variabel dan prediksi tren penjualan. Hasil analisis kemudian disajikan dalam bentuk dashboard interaktif yang menampilkan informasi penjualan secara informatif dan mendukung pengambilan keputusan manajerial berbasis data.

4. Hasil dan Pembahasan

4.1 Penerapan Algoritma K-Means Dalam Segmentasi Pelanggan

Berdasarkan ketentuan dalam penerapan analisis K-Means menggunakan aplikasi RapidMiner, data penjualan mentah perlu diolah terlebih dahulu dengan menggunakan pendekatan RFM (Recency, Frequency, Monetary) seperti Tabel 1.

Tabel 1. Transformasi data dengan pendekatan RFM

CUSTOMER	Recency	Frequency	Monetary
PELANGGAN A	11	794	Rp 120.220.056,00
PELANGGAN B	5	669	Rp 49.090.500,00
PELANGGAN C	5	447	Rp 25.109.000,00
PELANGGAN D	5	343	Rp 48.187.500,00
PELANGGAN E	1	326	Rp 51.826.156,00
PELANGGAN F	2	276	Rp 55.885.500,00
PELANGGAN G	22	275	Rp 19.719.000,00

RapidMiner adalah platform open source untuk data mining dan analisis bisnis yang memungkinkan pengguna mengekstraksi pengetahuan dari data dan membangun model prediktif tanpa memerlukan pengetahuan pemrograman[7].

Berdasarkan transformasi data tersebut, langkah selanjutnya yaitu melakukan normalisasi data dengan metode **Z-Score Normalization** dengan rumus:

$$Z = \frac{(X - \mu)}{\sigma} \dots\dots\dots (1)$$

Keterangan :

X = Nilai RFM asli

μ = Rata-rata variabel

σ = Standar deviasi variable

Tabel 2. Tabel Normalisasi RFM

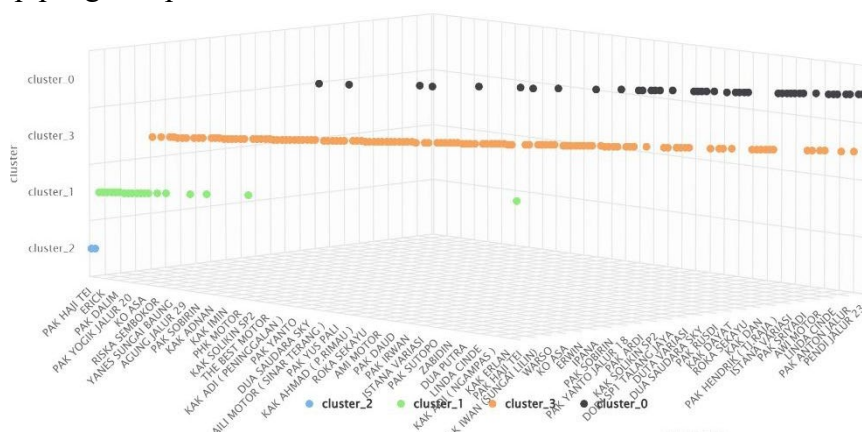
CUSTOMER	Z_Recency	Z_Frequency	Z_Monitory
PELANGGAN A	-0,7629093946	7,246887416	6,649904162
PELANGGAN B	-0,8380228794	5,996827581	2,374741659
PELANGGAN C	-0,8380228794	3,776721315	0,933360392
PELANGGAN D	-0,8380228794	2,736671532	2,320467853
PELANGGAN E	-0,8880985360	2,566663395	2,539165207
PELANGGAN F	-0,8755796218	2,066639461	2,783147543
PELANGGAN G	-0,6252013391	2,056638982	0,609400463
...

Penentuan jumlah cluster tidak dilakukan secara manual, melainkan menggunakan proses otomatis dengan blok Multiplay, Set Parameters, Loop Parameters, dan Clustering Performance (Davies-Bouldin Index). Berdasarkan hasil perhitungan Davies-Bouldin Index, setiap nilai K menghasilkan DBI sebagai berikut:

Table 3. Tabel Penentuan Jumlah Cluster berdasarkan DBI

Jumlah Cluster (K)	Davies-Bouldin Index
2	0.819
3	0.767
4	0.691
5	0.759

Nilai DBI yang lebih kecil menunjukkan cluster yang lebih terpisah dengan baik dan lebih homogen dalam masing-masing kelompok[8]. Semakin kecil nilai DBI menunjukkan kualitas cluster terbaik karena DBI mempertimbangkan jarak antar cluster dan kedekatan dalam cluster dan semakin kecil nilai DBI maka semakin terpisah dan kompak clusternya. Dari table diatas terlihat nilai DBI yang paling kecil ditunjukkan pada K=4 dengan nilai Davies Bouldin Index (DBI) sebesar 0.691 sehingga jumlah cluster terbaik yang digunakan dalam tahap pengelompokkan adalah 4 Cluster.



Gambar 1. Hasil clustering

Clustering ini dilakukan untuk melihat Behavioral Segmentation atau perilaku pelanggan. Cluster 2 – Pelanggan Bernilai Tinggi (High Value Customer), Cluster 1 – Pelanggan Potensial (Potential Growth Customer), Cluster 3 – Pelanggan Aktif dengan Nilai Transaksi sedang (Mid-Value Customer), Cluster 0 – Pelanggan dengan Aktivitas Terendah (Low Value Customer).

4.2 Penerapan Regresi Linear Dalam Prediksi Tren Penjualan

Prediksi/forecasting Merupakan salah satu metode yang digunakan untuk memprediksi atau meramalkan suatu hal yang belum pernah terjadi[9]. Metode regresi linear dipilih karena mampu menggambarkan hubungan antara variabel waktu (bulan) sebagai variabel independent (X) dengan total penjualan sebagai variabel dependen (Y).

Table 4. Ringkasan Regresi Linear Penjualan Online

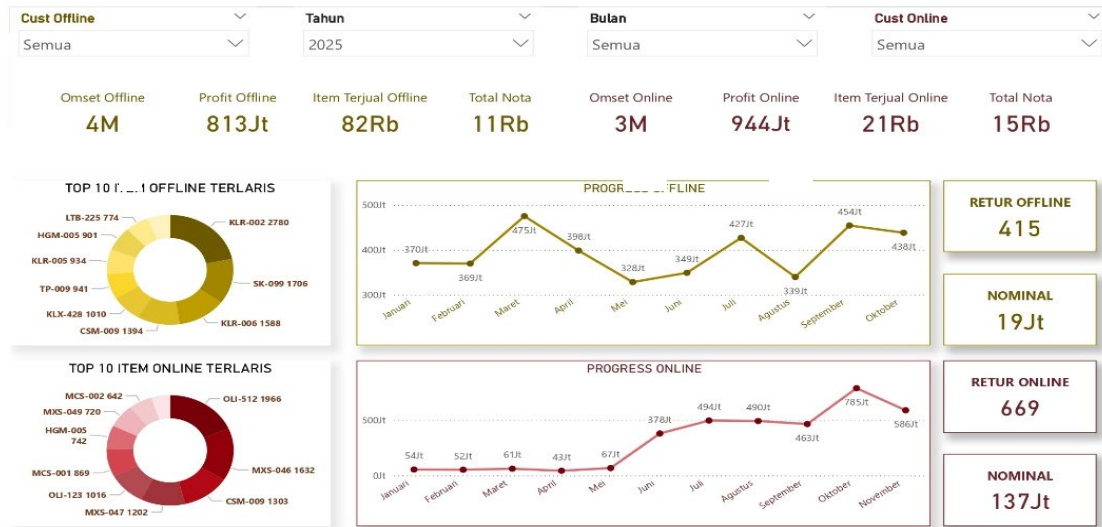
KODE ITEM	SLOPE (b_1)	INTERCEPT (b_0)	R-SQUARE	KET
AC-001	3.550.714	-19.815.771	0,47	Naik
AC-002	4.787.271	-29.118.724	0,65	Naik
AC-003	486.186	898.471	0,31	Naik
AC-006	29.429	1.403.190	0,00	Naik
AC-016	147.657	243.181	0,17	Naik
BD-003	33.543	-138.590	0,21	Naik
BD-018	-8.500	111.233	0,75	Turun
BD-019	-13.026	346.202	0,05	Turun
...

Table 5. Ringkasan Regresi Linear Penjualan Offline

KODE ITEM	SLOPE (b_1)	INTERCEPT (b_0)	R-SQUARE	KET
AC-001	144.021	176.598	0,00	Naik
AC-004	-72.964	1.329.600	0,00	Turun
AC-250	260.000	-200.000	1,00	Naik
AC-260	41.227	187.733	1,00	Naik
BD-003	-87.085	877.740	0,00	Turun
BD-006	-191.482	2.308.042	0,00	Turun
BD-018	15.116	57.521	0,00	Naik
BD-019	52.939	915.866	0,00	Naik

Analisis regresi yang dilakukan pada data online maupun offline baik per item maupun per kategori menunjukkan adanya kecenderungan tren naik, stabil, maupun menurun. Nilai slope dan R^2 yang diperoleh mengonfirmasi kekuatan hubungan antara waktu dan volume penjualan. Hasil regresi memberikan gambaran tren jangka panjang yang dapat digunakan untuk memprediksi performa penjualan di masa depan. Misalnya, beberapa kategori seperti oli, ban, atau produk aftermarket tertentu menunjukkan tren positif, sedangkan beberapa item lainnya mengalami penurunan konsisten. Temuan ini membuktikan bahwa regresi linear berhasil digunakan untuk memodelkan tren penjualan dan mendukung tujuan penelitian secara kuantitatif.

4.3 Implementasi Dashboard Visualisasi Power BI



Gambar 2. Dashboard Visualisasi Menu Overview

Visualisasi memegang peranan penting dalam analisis data. Power BI menawarkan beragam opsi visualisasi yang dapat disesuaikan, seperti grafik batang, diagram garis, dan peta geografis, yang memungkinkan user untuk mengidentifikasi tren penjualan, preferensi konsumen, dan pola geografis dengan lebih mudah [10]. Berdasarkan tampilan overview dashboard penjualan, dapat diperoleh gambaran umum mengenai kinerja penjualan Kesuma Motor Sport baik pada kanal offline maupun online selama periode pengamatan tahun 2025. Dashboard ini menyajikan indikator utama (Key Performance Indicators/KPI) yang mencerminkan performa penjualan secara agregat, sehingga memudahkan manajemen dalam memahami kondisi bisnis secara cepat dan komprehensif.

Dari sisi tren penjualan bulanan, grafik progress menunjukkan bahwa baik penjualan offline maupun online mengalami fluktuasi sepanjang tahun. Dashboard juga menampilkan Top 10 item terlaris baik untuk penjualan offline maupun online.

Visualisasi ini menunjukkan bahwa kontribusi penjualan didominasi oleh sejumlah item tertentu, yang mengindikasikan adanya konsentrasi permintaan pada produk-produk unggulan.



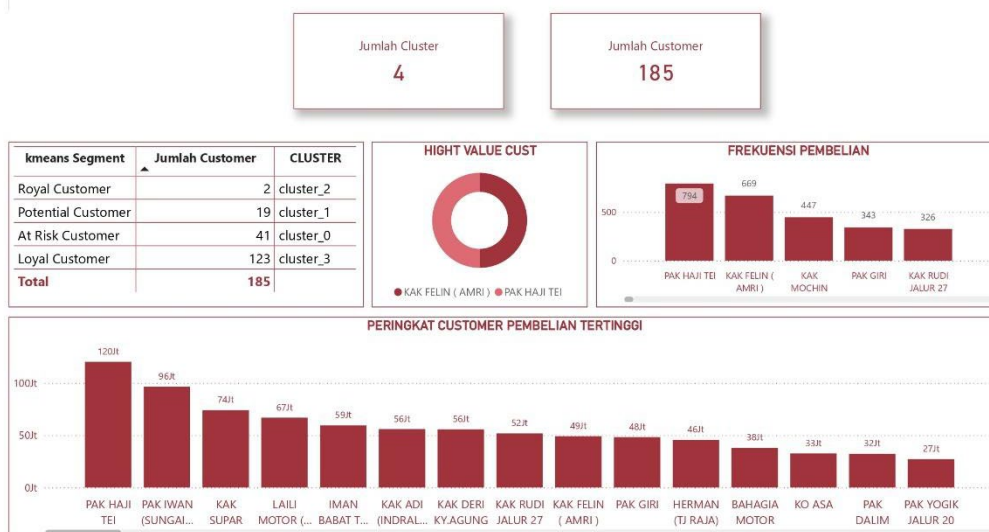
Gambar 3. Dashboard Visualisasi Penjualan Offline

Dashboard penjualan offline menyajikan gambaran kinerja penjualan secara komprehensif melalui indikator utama berupa omzet, jumlah item terjual, total transaksi (nota), grafik pertumbuhan penjualan, 10 item terlaris, 10 item paling jarang dibeli, kontribusi produk terjual, tampilan kartu (card) serta grafik tingkat retur barang serta nominal rata-rata penjualan maupun pembelian barang.



Gambar 4. Overview Dashboard visualisasi penjualan online

Dashboard penjualan online menampilkan performa penjualan kanal digital yang diukur melalui indikator omzet, jumlah item terjual, total nota, tren penjualan bulanan, top 10 item terlaris, kontribusi kanal penjualan, serta retur online. Analisis kontribusi kanal penjualan online menunjukkan bahwa marketplace seperti Shopee Evo, Shopee Samurái, dan Shopee Race memberikan kontribusi omzet terbesar dibandingkan kanal online lainnya.



Gambar 5. Dashboard Analisa K-Means

Berdasarkan hasil penerapan algoritma K-Means Clustering terhadap data Recency, Frequency, dan Monetary (RFM), pelanggan berhasil dikelompokkan ke dalam empat cluster utama yang merepresentasikan perbedaan perilaku pembelian konsumen. Secara keseluruhan,

proses clustering menghasilkan 185 pelanggan yang terdistribusi ke dalam empat segmen dengan karakteristik yang berbeda.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Secara keseluruhan, penelitian ini membuktikan bahwa integrasi metode KDD, algoritma K-Means, regresi linear, dan dashboard visualisasi mampu memberikan insight yang komprehensif terkait kinerja penjualan dan perilaku pelanggan. Meskipun demikian, efektivitas dashboard sebagai alat pengambilan keputusan masih bergantung pada pemahaman pengguna dalam menginterpretasikan hasil analisis serta konsistensi pembaruan data yang digunakan.

5.2 Saran

Penelitian selanjutnya disarankan untuk memperluas cakupan data dengan menambahkan periode waktu yang lebih panjang serta mengintegrasikan variabel lain, seperti data promosi, musiman, dan preferensi pelanggan, guna meningkatkan akurasi hasil analisis. Selain itu, penggunaan algoritma data mining alternatif atau metode prediksi yang lebih kompleks, seperti time series atau machine learning, dapat dipertimbangkan sebagai pembandingan untuk memperoleh model yang lebih optimal. Dari sisi implementasi sistem, pengembangan dashboard visualisasi diharapkan dapat dilengkapi dengan fitur evaluasi performa secara real-time dan indikator kinerja utama (KPI) yang lebih komprehensif agar mendukung pengambilan keputusan yang lebih efektif.

Referensi

- [1] A. T. Suseno, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024.
- [2] D. A. Fakhri, S. Defit, and Sumijan, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering," *J. Inf. dan Teknol.*, vol. 3, pp. 160–166, 2021, doi: 10.37034/jidt.v3i3.137.
- [3] I. G. I. Sudipa, I. G. M. Darmawiguna, I. M. Dendi, and ..., *Buku Ajar Data Mining. PT*, no. April. 2024.
- [4] & S. Tarigan, Hardinata, Qurniawan, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *ADA J. Inf. Syst. Res.*, vol. 1, no. 1, pp. 1–8, 2023, doi: 10.64366/adajisr.v1i1.7.
- [5] H. Adam, E. Novalia, and A. L. Hananto, "Bulletin Of Computer Science Research Prediksi Penjualan Barang Menggunakan Metode K-Means dan Regresi Linear," vol. 5, no. 4, 2025, doi: 10.47065/bulletinsr.v5i4.541.
- [6] I. Ardhanur *et al.*, "Implementasi Algoritma Regresi Linear Untuk Model Prediksi Penjualan Di Toko," vol. 13, no. 2, 2025.
- [7] S. R. Agustin, I. Purnamasari, and B. N. Sari, "Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis Web," vol. 5, no. 3, pp. 167–176, 2025, doi: 10.47065/jimat.v5i3.610.
- [8] B. D. Oktavian, I. Lumintu, and S. Amar, "Optimalisasi Strategi Product Bundling melalui Pemetaan Pola Peminatan dan Pola Penjualan Produk menggunakan K-Means Clustering dan Apriori Optimizing Product Bundling Strategy through Mapping Product Interest Patterns and Sales Patterns using K-Means Clustering and Apriori," vol. 8, no. 1, pp. 26–41, 2025.
- [9] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [10] J. Barat, "Visualisasi Prediksi Penjualan Game Di Dunia," vol. 13, no. 2, 2025.