

Analisis Pola Penjualan Toko Ritel NisaMart Menggunakan Algoritma K-Means Klasterisasi

Muhamad Hanif Kurnia Hadi¹, Dimas Andreas Mustofa²,
Muhammad Rizky Perdana³, Ina Sholihah Widiati⁴

¹²³⁴Prodi S1 Informatika, STMIK Amikom Surakarta

¹²³⁴Sukoharjo Indonesia

Email: ¹muh.130538@mhs.amikomsolo.ac.id,

²dimas130518@mhs.amikomsolo.ac.id, ³muh.130537@mhs.amikomsolo.ac.id,

⁴inasw@dosen.amikomsolo.ac.id

Abstract

Utilizing transaction data in retail stores is essential for understanding customer purchasing patterns and supporting effective operational decision-making. This study aims to analyze product sales patterns at NisaMart Wonogiri using the K-Means clustering algorithm. The dataset consists of sales transactions recorded on August 5, 2025, and is processed through several preprocessing stages, including data cleaning, normalization, and categorization of transaction time. The clustering process is carried out using the WEKA software with the number of clusters set to three. The results indicate that the K-Means algorithm successfully groups products into three main clusters, namely products with high sales and lower prices, products with moderate sales and medium prices, and products with low sales and higher prices. These clusters provide a clear representation of product characteristics based on their sales movement levels. The findings can be utilized to optimize stock management, determine promotional strategies, and identify products with low turnover that require additional managerial attention. Overall, the K-Means algorithm proves effective in uncovering sales patterns in retail environments and can serve as a valuable tool to assist in operational decision-making.

Keywords: K-Means, clustering, data mining, retail sales, WEKA

Abstraksi

Pemanfaatan data transaksi penjualan pada toko ritel dapat memberikan informasi penting untuk memahami pola pembelian konsumen dan mendukung pengambilan keputusan operasional. Penelitian ini bertujuan untuk menganalisis pola penjualan produk pada ritel NisaMart Wonogiri menggunakan algoritma K-Means klasterisasi. Data yang digunakan berasal dari transaksi penjualan tanggal 5 Agustus 2025 dan diproses melalui tahapan pra-pemrosesan, meliputi pembersihan data, normalisasi, serta pengelompokan waktu transaksi. Proses klasterisasi dilakukan menggunakan perangkat lunak WEKA dengan jumlah klaster ditetapkan sebanyak tiga. Hasil penelitian menunjukkan bahwa K-Means berhasil mengelompokkan produk ke dalam tiga klaster utama, yaitu klaster dengan penjualan tinggi dan harga rendah, klaster dengan penjualan sedang dan harga menengah, serta klaster dengan penjualan rendah dan harga lebih tinggi. Pembagian klaster ini memberikan gambaran yang jelas tentang karakteristik produk berdasarkan tingkat pergerakan penjualannya. Temuan ini dapat dimanfaatkan untuk mengoptimalkan pengelolaan stok, menyusun strategi promosi, serta mengidentifikasi

produk dengan tingkat perputaran rendah yang memerlukan perhatian khusus. Secara keseluruhan, algoritma K-Means terbukti efektif dalam mengungkap pola penjualan dan dapat digunakan sebagai alat pendukung pengambilan keputusan pada lingkungan ritel.

Kata Kunci: K-Means, klasterisasi, data mining, penjualan ritel, WEKA

1. PENDAHULUAN

Industri ritel di Indonesia berkembang pesat sehingga pemanfaatan data transaksi menjadi bagian penting dalam mendukung pengambilan keputusan operasional dan strategi bisnis[1]. Setiap transaksi penjualan menghasilkan data dalam jumlah besar yang dapat dimanfaatkan untuk memahami pola pembelian pelanggan serta mengidentifikasi peluang peningkatan penjualan[2]. Pengolahan data tersebut memerlukan pendekatan analitis yang mampu menemukan pola tersembunyi, salah satunya melalui teknik data mining yang berfokus pada pencarian informasi bermakna dari kumpulan data besar[3].

Dalam konteks ritel, data mining digunakan untuk memetakan perilaku pembelian, mengelompokkan produk, dan mengidentifikasi kecenderungan penjualan berdasarkan karakteristik tertentu[4]. Salah satu metode yang banyak diterapkan adalah algoritma K-Means, yaitu teknik klasterisasi yang membagi data ke dalam kelompok berdasarkan kemiripan atribut untuk menghasilkan segmentasi yang jelas[5]. Efektivitas algoritma ini ditunjukkan pada berbagai penelitian, misalnya pada pengelompokan penjualan produk frozen food dan identifikasi kategori permintaan yang berbeda[6]. Metode ini juga digunakan pada segmentasi pelanggan untuk memetakan perilaku pembelian dan menentukan strategi pemasaran yang tepat sasaran[7].

K-Means semakin luas digunakan karena mampu memberikan pemetaan data yang terstruktur, termasuk pada penerapan di lingkungan akademik seperti pengelompokan dosen berdasarkan jabatan fungsional menggunakan perangkat lunak pendukung analisis data[8]. Implementasi metode ini juga diperkuat oleh perangkat lunak WEKA yang menyediakan fitur klasterisasi secara cepat dan akurat sehingga memudahkan proses analisis tanpa memerlukan implementasi algoritma secara manual[9]. Penelitian terkait pola penjualan makanan menunjukkan bahwa kombinasi K-Means dan WEKA mampu menghasilkan pengelompokan produk yang membantu manajemen dalam merancang strategi stok dan promosi yang efektif[10].

Berdasarkan pentingnya pemanfaatan data transaksi serta bukti efektivitas algoritma K-Means pada berbagai penelitian sebelumnya, studi ini berfokus pada analisis pola penjualan produk di ritel NisaMart Wonogiri. Data transaksi yang digunakan berasal dari penjualan pada tanggal 5 Agustus 2025 dan selanjutnya diolah melalui tahapan pra-pemrosesan, proses klasterisasi dengan algoritma K-Means menggunakan WEKA, serta analisis terhadap hasil pengelompokan. Melalui proses tersebut, penelitian ini diharapkan mampu menghasilkan klaster produk berdasarkan tingkat penjualan yang dapat mendukung perencanaan strategi pengelolaan stok dan pemasaran secara lebih tepat.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan proses pengolahan data dalam jumlah besar untuk menemukan pola dan informasi penting yang dapat digunakan dalam pengambilan keputusan bisnis. Data mining menggabungkan konsep statistik dan machine learning untuk membuka pola tersembunyi dari data transaksi penjualan[3]. Melalui penerapan teknik ini, pelaku usaha dapat mengidentifikasi peluang peningkatan penjualan serta memahami perilaku konsumen secara lebih mendalam[2].

2.2. Algoritma K-Means

Algoritma K-Means merupakan salah satu teknik klasterisasi paling umum digunakan karena mampu membagi data ke dalam kelompok berdasarkan tingkat kemiripan atribut [5]. Algoritma ini bekerja dengan menentukan centroid pada tiap klaster serta menghitung jarak data terhadap pusat tersebut sehingga kelompok terbentuk dengan karakteristik yang serupa. Penggunaan K-Means terbukti efektif dalam mengelompokkan produk maupun pelanggan berdasarkan aktivitas transaksi sehingga dapat membantu strategi pemasaran dan pengelolaan stok [7].

2.3. Penerapan K-Means dalam Analisis Penjualan

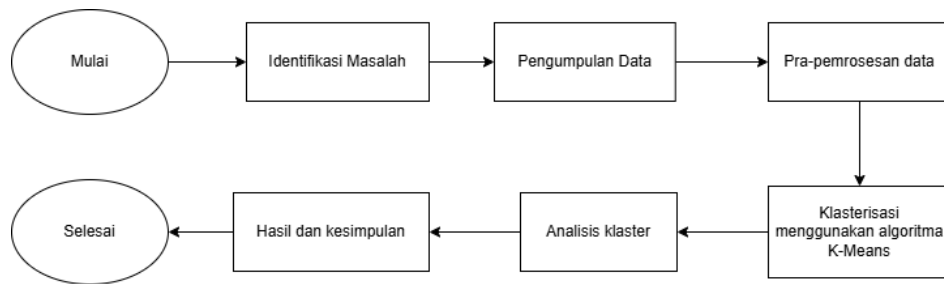
Penelitian terdahulu menunjukkan bahwa K-Means dapat digunakan untuk mengidentifikasi kelompok produk berdasarkan tingkat penjualan pada berbagai jenis ritel [4]. Hasil klasterisasi membantu toko menentukan kategori produk terlaris, penjualan sedang, dan produk dengan permintaan rendah sehingga strategi promosi dan pengelolaan stok dapat dilakukan secara lebih tepat sasaran [10]. Efektivitas metode ini menjadikan K-Means banyak digunakan dalam analisis penjualan dan perilaku konsumen.

2.4. Weka

Weka merupakan perangkat lunak open-source yang menyediakan berbagai algoritma data mining, termasuk K-Means, dan memungkinkan proses analisis dilakukan secara visual tanpa perlu pemrograman kompleks [8]. Penggunaan WEKA mempermudah proses klasterisasi karena antarmukanya yang intuitif dan kemampuannya dalam mengolah data secara cepat dan akurat [9]. Dukungan fitur-fitur ini menjadikan WEKA alat yang tepat untuk penelitian terkait pengelompokan data penjualan.

3. METODE PENELITIAN

Penelitian ini menggunakan algoritma K-Means untuk menganalisis pola penjualan produk pada ritel. Alur penelitian mencakup pengumpulan data, pra-pemrosesan, klasterisasi menggunakan WEKA, serta analisis klaster untuk menghasilkan kelompok produk berdasarkan tingkat penjualan. Proses klasterisasi membagi data ke dalam tiga kelompok produk berdasarkan tingkat penjualan, sehingga diperoleh informasi yang berguna untuk pengelolaan stok dan strategi pemasaran.



Gambar 1. Flowchart alur penelitian

3.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini berasal dari data transaksi penjualan di ritel NisaMart, sebuah toko ritel yang berlokasi di Wonogiri, Indonesia. Data diperoleh langsung dari sistem kasir toko, sehingga merepresentasikan kondisi penjualan nyata di lapangan. Pengambilan data dilakukan pada tanggal 5 Agustus 2025 selama satu hari penuh untuk memperoleh gambaran pola transaksi harian. Dataset awal berjumlah 170 transaksi, kemudian dilakukan proses seleksi untuk menghapus duplikasi dan kesalahan pencatatan hingga tersisa 100 data yang layak digunakan sebagai data uji.

3.2. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menyiapkan dataset agar sesuai dengan kebutuhan analisis. Beberapa langkah yang diterapkan meliputi:

- Pembersihan data, yaitu menghapus entri duplikat dan data yang tidak valid.
- Transformasi atribut waktu, dari format jam (HH:MM:SS) ke kategori waktu:
 - 1) Pagi (07:00–10:59)
 - 2) Siang (11:00–14:59)
 - 3) Sore (15:00–19:00)

Tahap ini bertujuan memastikan dataset memiliki struktur yang seragam dan siap dianalisis.

3.3. Proses Menggunakan Metode K-Means Klusterisasi

Setelah pra-pemrosesan, data dianalisis menggunakan algoritma K-Means untuk mengelompokkan produk berdasarkan kemiripan atribut penjualan, seperti jumlah penjualan dan harga produk. Jumlah kluster ditentukan sebanyak tiga untuk membedakan kategori produk dengan tingkat penjualan tinggi, sedang, dan rendah. Proses klusterisasi dilakukan menggunakan perangkat lunak WEKA sehingga perhitungan centroid, pengelompokan, dan visualisasi kluster diperoleh secara otomatis. Untuk memperjelas proses perhitungan jarak pada algoritma K-Means, rumus dasar yang digunakan ditunjukkan pada persamaan berikut:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(f)} - c_j\|^2$$

Euclidean distance adalah metode yang paling umum untuk mengukur jarak antara pengamatan, termasuk ketika pengamatan variabel kontinu:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

3.4. Analisis Hasil Klaster

Tahapan analisis dilakukan dengan mengevaluasi karakteristik tiap klaster berdasarkan nilai centroid yang dihasilkan. Hasil klasterisasi ditinjau dari rata-rata jumlah penjualan, rata-rata harga produk, serta kecenderungan waktu pembelian. Analisis ini digunakan untuk menginterpretasikan kategori produk yang masuk dalam klaster penjualan tinggi, sedang, maupun rendah sehingga dapat memberikan rekomendasi pada pengelolaan stok dan strategi pemasaran.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan Data

Penelitian ini menggunakan data transaksi penjualan yang diperoleh dari sistem kasir ritel NisaMart Wonogiri pada tanggal 5 Agustus 2025. Dari total 170 transaksi awal, dilakukan proses pembersihan data untuk menghapus duplikasi dan entri yang tidak valid, sehingga diperoleh 100 data akhir yang siap dianalisis. Data ini berisi atribut jumlah penjualan produk, harga, dan waktu transaksi.

	A	B	C	D
1	ID Transaksi	Nama Barang	Jumlah Terjual (pcs)	Harga
2	T-20250805-8754;09:09:32	Indomie Ayam Bawang	5	2900
3	T-20250805-2122;07:11:43	Teh Kotak 200ml	8	2500
4	T-20250805-3367;07:18:12	Roma Kelapa 300gr	6	2250
5	T-20250805-5647;08:23:45	Aqua Botol 600ml	10	3000
6	T-20250805-4487;08:46:05	Biskuit Oreo Double Stuff	4	2500
7	T-20250805-6578;09:32:45	Kopi Kapal Api Mix	7	1800
8	T-20250805-6574;07:22:33	Cat Rambut Miranda	1	12500
9	T-20250805-8876;07:48:56	Minyak Goreng Fortune	3	19500

Gambar 2. Data Uji

4.2. Hasil Pra-pemrosesan Data

Pada tahap pra-pemrosesan, atribut waktu yang awalnya dalam format jam (HH:MM:SS) dikonversi menjadi tiga kategori waktu: pagi, siang, dan sore. Perubahan ini bertujuan memudahkan analisis kecenderungan waktu pembelian. Selain itu, dilakukan normalisasi variabel numerik agar skala data lebih seragam sebelum proses klasterisasi.

	A	B	C	D
1	Nama Barang	Jumlah Terjual	Harga	Waktu Terjual
2	Indomie Ayam B	5	2900	pagi
3	Teh Kotak 200m	8	2500	pagi
4	Roma Kelapa 30	6	2250	pagi
5	Aqua Botol 600r	10	3000	pagi
6	Biskuit Oreo Do	4	2500	pagi
7	Kopi Kapal Api M	7	1800	pagi
8	Cat Rambut Mir	1	12500	pagi
9	Minyak Goreng	3	19500	pagi
10	Sabun Mandi Lif	5	2600	pagi
11	Sabun Cuci Rins	4	7000	pagi
12	Shampo Head &	3	8500	pagi
13	Sabun Mandi Bi	6	3000	pagi
14	Sabun Cuci Att	2	10500	pagi
15	Shampo Sunsilk	5	4500	pagi
16	Handbody Vase	4	6300	pagi
17	Gel Rambut Gat	3	8000	pagi
18	Cat Rambut Mir	2	12500	pagi
19	Bedak Marcks C	5	3600	pagi
20	Handbody Mari	6	3500	pagi
21	Indomie Kari Ay	6	2900	pagi
22	Aqua Botol 600r	12	3000	pagi

Gambar 3. Potongan data set yang telah diubah atribut waktu

4.3. Hasil Klasterisasi Menggunakan K-Means

Pada tahap ini dilakukan proses klasterisasi terhadap data penjualan menggunakan algoritma K-Means melalui perangkat lunak WEKA. Hasil klasterisasi menunjukkan pembagian data ke dalam tiga klaster berdasarkan kemiripan atribut penjualan. Output klasterisasi dari WEKA ditampilkan

Missing values globally replaced with mean/mode				
Final cluster centroids:				
Attribute	Full Data	Cluster#		
	(100.0)	0	1	2
		(34.0)	(44.0)	(22.0)
=====				
Jumlah Terjual (pcs)	4.79	6.6765	4.4545	2.5455
	+/-2.6488	+/-2.446	+/-2.3273	+/-1.1434
Harga	6694.5	3216.1765	7179.5455	11100
	+/-4885.972	+/-1162.7893	+/-5265.0815	+/-3638.9428
Waktu Terjual	pagi	pagi	sore	pagi
pagi	39.0 (39%)	24.0 (70%)	0.0 (0%)	15.0 (68%)
siang	27.0 (27%)	10.0 (29%)	10.0 (22%)	7.0 (31%)
sore	34.0 (34%)	0.0 (0%)	34.0 (77%)	0.0 (0%)
Time taken to build model (full training data) : 0 seconds				
=== Model and evaluation on training set ===				
Clustered Instances				

Gambar 4. Hasil Weka menampilkan masing masing cluster

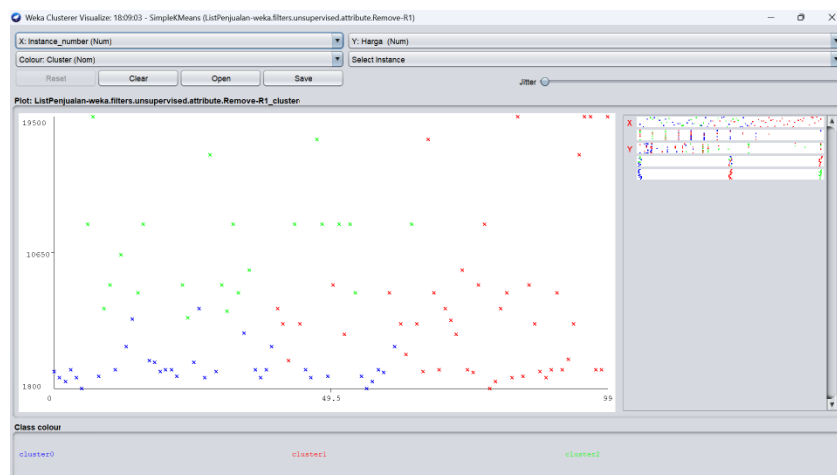
Cluster 0: 'Shampo Sunsilk Black Shine 170ml',5,4500,pagi
 Cluster 1: 'Sabun Mandi Biore 80gr',4,3000,sore
 Cluster 2: 'Sabun Cuci Rinso Liquid 750ml',4,7000,pagi

Gambar 5. Salah satu barang dari masing masing cluster

WEKA menghasilkan nilai centroid untuk masing-masing klaster sebagai berikut:

Cluster	Rata rata terjual	Harga rata rata	Persentase Data	Kategori
Cluster 0	6,67 unit	Rp 3.216	34%	Tinggi
Cluster 1	4,45 unit	Rp 7.179	44%	Sedang
Cluster 2	2,54 unit	Rp 11.000	22%	Rendah

Sehingga dapat diketahui hasil pengelompokan Weka dapat dilihat pada gambar berikut :



Gambar 6. Gambar visual pengelompokan data

Visualisasi hasil klasterisasi (Gambar 6) memperlihatkan masing-masing data produk yang masuk dalam tiga kelompok tersebut.

4.4. Pembahasan Pembagian Klaster

1. Klaster 0 – Produk dengan Tingkat Penjualan Tinggi

Klaster 0 memiliki rata-rata jumlah penjualan 6,67 unit, tertinggi di antara semua klaster. Harga produk di klaster ini relatif rendah (Rp 3.216) sehingga mendorong konsumen untuk membeli dalam frekuensi yang lebih sering. Produk dalam klaster ini cenderung merupakan produk kebutuhan harian, sehingga tingkat rotasinya cepat. Pada Gambar 5, beberapa contoh produk dalam klaster ini menunjukkan konsistensi pembelian terutama pada waktu pagi.

2. Klaster 1 – Produk dengan Tingkat Penjualan Sedang

Kluster 1 mencakup produk dengan rata-rata terjual 4,45 unit per hari pada harga menengah (Rp 7.179). Produk pada kelompok ini tidak secepat klaster 0 dalam rotasi stok, namun penjualannya cukup stabil. Waktu pembelian didominasi pada periode sore, yang mengindikasikan bahwa produk ini sering dibeli setelah aktivitas harian pelanggan selesai. Secara visual, Gambar 4 menunjukkan sebaran kluster ini berada di antara klaster 0 dan 2 sehingga menandakan transisi kategori produk.

3. Klaster 2 – Produk dengan Tingkat Penjualan Rendah

Klaster 2 merupakan kelompok dengan rata-rata penjualan terendah yaitu 2,54 unit per hari namun memiliki harga tertinggi (Rp 11.100). Produk pada klaster ini umumnya termasuk produk kebutuhan tidak rutin seperti barang rumah tangga atau produk berdaya pakai lama. Contohnya pada dataset adalah Rinso Liquid, yang dibeli hanya ketika pelanggan membutuhkan pengganti barang lama. Hal ini menjelaskan mengapa klaster 2 memiliki volume penjualan rendah meskipun harganya tinggi.

4.5. Interpretasi Hasil Klasterisasi

Hasil klasterisasi memberikan pemahaman bahwa: produk berharga rendah cenderung memiliki tingkat penjualan tinggi, produk berharga menengah memiliki pola pembelian stabil dan membutuhkan strategi promosi tambahan untuk meningkatkan penjualan, produk berharga tinggi memiliki tingkat pembelian rendah dan bersifat musiman atau kebutuhan jangka panjang. Informasi ini membantu ritel dalam: menentukan prioritas pengadaan stok, menyusun strategi promosi berdasarkan tingkat penjualan, mengevaluasi produk kurang laku untuk meminimalkan potensi dead stock.

5. KESIMPULAN

Berdasarkan hasil klasterisasi menggunakan algoritma K-Means terhadap data penjualan ritel, dapat disimpulkan bahwa: proses klasterisasi berhasil membagi produk ke dalam tiga kelompok yang merepresentasikan tingkat penjualan tinggi, sedang, dan rendah sehingga pola penjualan dapat dikenali secara lebih jelas; dan hasil pengelompokan tersebut memberikan informasi yang dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pengelolaan stok, penentuan strategi promosi, serta evaluasi produk dengan tingkat perputaran rendah agar strategi operasional ritel dapat dilakukan secara lebih efektif.

DAFTAR PUSTAKA

- [1] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 1–7, 2022.
- [2] M. Sulaiman, R. Yudistira, R. Narasati, and R. Herdiana, "Penerapan Data Mining dengan Metode Clustering untuk menentukan Strategi Peningkatan Penjualan Berdasarkan Data Transaksi," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 307–314, 2024.
- [3] S. Aulia, "Klusterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno: Jurnal Teknologi Informasi*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [4] S. Lestari, Muhatri, A. R. Fachrezi, M. Agung Sutrisno, and M. Geovany, "Implementasi Algoritma K-Means Clustering pada Penjualan Sepatu Futsal Merk Specs," vol. 2, no. 3, pp. 50–61, 2024.
- [5] L. Azzahra and Amru Yasir, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 3, no. 1, pp. 1–10, 2024, doi: 10.70340/jirsi.v3i1.88.
- [6] H. O. Siti Andry Yani Nasution, Poningsih, "Penerapan Algoritma K-Means Pada Penjualan Frozen Food Pada UD Soise Sosis Pematangsiantar," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 2, pp. 171–177, 2021, doi: 10.30865/json.v2i2.2768.
- [7] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 8, no. 3, 2022, doi: 10.33197/jitter.vol8.iss3.2022.883.
- [8] N. Noviyanto and P. Ekasari, "Data Mining Menggunakan Algoritma K-Means dengan Weka Interface untuk Klusterisasi Dosen Yang Memiliki Jabatan Fungsional pada Perguruan Tinggi Swasta dilingkungan LLDikti Wilayah III," *Paradigma - Jurnal Komputer dan Informatika*, vol. 24, no. 1, pp. 103–107, 2022, doi: 10.31294/paradigma.v24i1.1112.
- [9] Pangestu, "Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka," *JUST IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 11, no. 3, pp. 67–71, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/article/view/11591>
- [10] D. Rusvinasari and L. H. Annisa, "Klusterisasi Pola Penjualan Menu Makanan pada Rumah Makan menggunakan Metode K-Means Clustering," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 2, pp. 398–409, 2025, doi: 10.30591/jpit.v10i2.8511.