

CLUSTERING DATA BARANG DARI TRANSAKSI CLOTHING BENGKULU MENGGUNAKAN ALGORITMA K-MEANS

Muhammad Dafa Sadikin, Muhammad Husni Rifqo

Teknik Informatika, Universitas Muhammadiyah Bengkulu
Jalan Bali, Kampung Bali, Kota Bengkulu, Bengkulu, Indonesia
dafasusiana@gmail.com

ABSTRAK

Perkembangan usaha ritel menuntut pengelolaan data penjualan yang akurat guna mendukung pengambilan keputusan berbasis data. Toko Clothing Bengkulu menghadapi kendala dalam mengidentifikasi produk yang tergolong laris, kurang laris, dan tidak laris berdasarkan data transaksi yang tersedia. Penelitian ini bertujuan untuk mengelompokkan produk berdasarkan tingkat penjualan menggunakan algoritma K-Means dengan pendekatan Knowledge Discovery in Databases (KDD). Data yang dianalisis berjumlah 115 item produk selama periode Januari–Mei 2025 melalui tahapan seleksi data, pembersihan, transformasi, proses clustering, dan evaluasi. Jumlah kluster ditetapkan sebanyak tiga kategori. Hasil pengelompokan menunjukkan 33 produk termasuk kategori laris, 62 produk kurang laris, dan 20 produk tidak laris. Evaluasi performa menghasilkan nilai average within centroid distance sebesar 966,181 dan Davies Bouldin Index sebesar 0,724 yang menunjukkan kualitas kluster yang cukup baik. Hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam perencanaan strategi stok dan promosi guna meningkatkan efektivitas pengelolaan persediaan dan penjualan.

Kata kunci : data mining, clustering, K-Means, penjualan, RapidMiner, KDD, Toko Clothing, evaluasi performa

1. PENDAHULUAN

Clothing Bengkulu merupakan sebuah usaha yang bergerak di bidang penjualan berbagai aksesoris pria, seperti baju, celana, jam tangan, skateboard, cincin, dan produk lainnya. Dalam menghadapi persaingan bisnis yang semakin ketat serta perubahan tren konsumen yang dinamis, toko dituntut untuk mampu menyediakan produk yang sesuai dengan kebutuhan dan minat pelanggan. Keberhasilan dalam mempertahankan daya saing tersebut sangat bergantung pada kemampuan pengelola dalam memahami pola penjualan berdasarkan data historis yang dimiliki.

Seiring meningkatnya jumlah data penjualan, pengelolaan data secara konvensional menjadi kurang efektif dan menyulitkan pengelola dalam mengidentifikasi produk yang memiliki tingkat penjualan tinggi, rendah, maupun yang tidak diminati konsumen. Kondisi ini menyebabkan pengambilan keputusan terkait pengelolaan stok dan strategi pemasaran belum sepenuhnya didasarkan pada analisis data yang optimal. Oleh karena itu, diperlukan suatu pendekatan analisis data yang komprehensif untuk menggali informasi dan pola tersembunyi dari data penjualan yang tersedia, sehingga dapat digunakan sebagai dasar dalam perumusan strategi bisnis yang lebih efektif [1].

Salah satu teknik dalam data mining yang dapat digunakan untuk mengatasi permasalahan tersebut adalah clustering. Clustering merupakan teknik pengelompokan data berdasarkan tingkat kemiripan karakteristik, di mana data dalam satu kelompok memiliki kesamaan yang lebih tinggi dibandingkan dengan data di kelompok lainnya [2]. Teknik ini banyak dimanfaatkan untuk menyelesaikan

permasalahan kompleks di berbagai bidang, termasuk dalam analisis penjualan dan pengelolaan inventaris.

Metode K-Means Clustering merupakan salah satu algoritma clustering yang paling umum digunakan karena kemampuannya dalam mengelompokkan data numerik yang bersifat nyata dan terukur. Algoritma ini bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan kesamaan karakteristik, sehingga setiap kluster berisi data dengan tingkat kemiripan yang tinggi [3]. Keunggulan metode K-Means terletak pada sifatnya yang adaptif, karena jumlah kluster dapat ditentukan sesuai dengan kebutuhan analisis, sehingga hasil pengelompokan dapat disesuaikan dengan tujuan penelitian [4].

Penelitian ini bertujuan untuk menganalisis data penjualan pada Toko Clothing Bengkulu dengan menerapkan metode K-Means Clustering guna mengelompokkan produk berdasarkan tingkat penjualannya, seperti produk yang sangat laku, kurang laku, dan tidak laku. Hasil pengelompokan tersebut diharapkan dapat memberikan informasi yang bermanfaat bagi pengelola toko dalam menentukan strategi pemasaran dan pengelolaan stok yang lebih tepat sasaran.

Proses penyelesaian masalah dalam penelitian ini mengacu pada konsep Knowledge Discovery in Database (KDD), yang meliputi tahapan pemilihan data, pembersihan data, transformasi data, proses data mining, serta interpretasi dan evaluasi hasil [5].

Untuk mendukung proses analisis, digunakan perangkat lunak RapidMiner Studio karena kemudahan penggunaan dan kompatibilitasnya dengan berbagai format data [6]. Dengan penerapan metode K-Means Clustering, diharapkan Toko Clothing Bengkulu dapat meningkatkan efektivitas pengambilan keputusan serta mengoptimalkan strategi

penjualan guna memenuhi kebutuhan pelanggan secara lebih optimal.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu digunakan untuk memberikan gambaran mengenai penerapan algoritma K-Means dalam

Penelitian yang ditulis oleh Triyandana, Putri, dan Umaidah berjudul *“Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means”* pada tahun 2022 meneliti data transaksi penjualan menu makanan dan minuman dengan tujuan mengelompokkan menu berdasarkan tingkat penjualannya. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu membentuk kluster menu laris dan kurang laris, namun penelitian ini belum dilengkapi dengan evaluasi kualitas kluster secara kuantitatif sehingga tingkat keoptimalan hasil pengelompokan belum dapat diukur secara objektif [1].

Penelitian yang ditulis oleh Nugraha, Nurdian, dan Dwilestari berjudul *“Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport”* pada tahun 2022 meneliti data penjualan produk olahraga berdasarkan jumlah penjualan dan stok barang. Penelitian ini menunjukkan bahwa K-Means dapat membantu pengambilan keputusan pengadaan barang, namun objek penelitian masih terbatas pada produk dengan karakteristik yang relatif homogen [2].

Penelitian yang ditulis oleh Afiasari, Suarna, dan Rahaningsi berjudul *“Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means”* pada tahun 2023 meneliti data transaksi penjualan numerik untuk mengungkap pola penjualan tersembunyi. Hasil penelitian menunjukkan efektivitas algoritma K-Means dalam proses clustering, tetapi belum membahas evaluasi performa kluster menggunakan indeks validitas tertentu [3].

Penelitian yang ditulis oleh Fajar, Adji, dan Dwilestari berjudul *“Analisis Data Transaksi Penjualan Barang Menggunakan Teknik K-Means Clustering”* pada tahun 2025 meneliti data transaksi penjualan barang untuk mendukung manajemen stok. Penelitian ini membuktikan bahwa clustering mampu mengelompokkan produk berdasarkan tingkat penjualan, namun belum mengaitkan hasil pengelompokan dengan pengujian kualitas kluster secara mendalam [4].

Penelitian yang ditulis oleh Rahmawati dan Prihartono berjudul *“Optimasi Stok dengan Clustering Data Menggunakan K-Means”* pada tahun 2025 meneliti data stok dan penjualan untuk tujuan optimalisasi persediaan barang. Clustering digunakan untuk mengelompokkan barang berdasarkan pergerakan stok, tetapi belum secara spesifik

mengklasifikasikan produk ke dalam kategori laris, kurang laris, dan tidak laris [5].

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma K-Means efektif digunakan dalam analisis data penjualan. Namun, masih terdapat keterbatasan pada aspek evaluasi kualitas kluster dan penerapan pada usaha ritel clothing dengan variasi produk yang tinggi. Oleh karena itu, penelitian ini difokuskan pada pengelompokan produk clothing berdasarkan tingkat penjualan dengan disertai evaluasi performa kluster agar hasil analisis lebih komprehensif dan aplikatif.

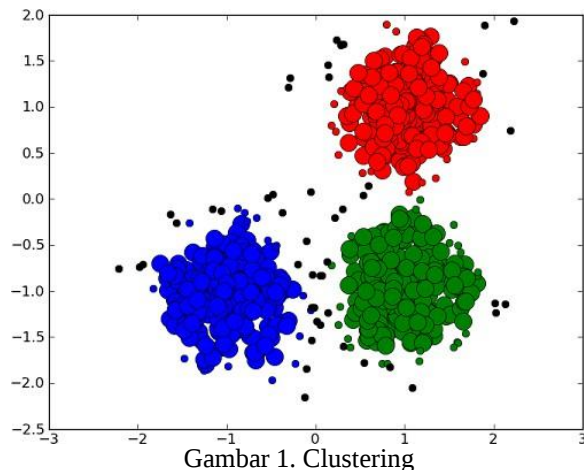
2.2. Data Mining

Data mining merupakan suatu proses untuk memecah kumpulan data yang besar dan kompleks menjadi informasi yang bernilai. Informasi yang dihasilkan biasanya tidak langsung terlihat secara jelas, karena sebelumnya belum diketahui atau tersembunyi di dalam data mentah yang tersedia, proses ini juga dapat dijelaskan sebagai kegiatan menggali dan menganalisis data dalam skala besar menggunakan bantuan teknologi, baik yang bersifat otomatis maupun setengah otomatis [7]. Tujuan utamanya adalah untuk menemukan pola-pola tertentu yang memiliki makna penting dan dapat dimanfaatkan lebih lanjut, dengan teknik dan metode tertentu, data mining memungkinkan pengguna untuk mengungkap keterkaitan antar data yang tidak langsung tampak. Pola-pola tersembunyi yang ditemukan ini sering menjadi dasar dalam merumuskan keputusan strategis yang lebih akurat dan berdasarkan data, secara keseluruhan, data mining menjadi bagian dari proses yang lebih luas yang disebut sebagai *Knowledge Discovery in Databases (KDD)*, yaitu sebuah tahapan sistematis dalam mengekstraksi pengetahuan dari kumpulan data yang tersimpan di dalam sistem basis data.

2.3. Clustering

Clustering adalah salah satu pendekatan dalam data mining yang digunakan untuk mengelompokkan data ke dalam sejumlah grup atau kluster tertentu [8]. Tujuan utama dari teknik ini adalah menyusun data ke dalam kelompok berdasarkan tingkat kemiripan antar data tersebut, Setiap data yang memiliki atribut atau ciri-ciri yang mirip akan dikelompokkan bersama dalam satu kluster, sedangkan data yang berbeda sifat atau karakternya akan dipisahkan ke dalam kluster lainnya. Proses ini memungkinkan penyusunan data yang lebih terstruktur dan mudah dianalisis, pengelompokan ini didasarkan pada kesamaan karakteristik yang dimiliki masing-masing data, sehingga pola dalam data dapat teridentifikasi dengan lebih jelas. Dengan cara ini, informasi tersembunyi dalam data dapat lebih mudah diungkapkan, terdapat berbagai teknik yang bisa diterapkan dalam proses clustering, dan salah satu metode yang cukup populer serta banyak digunakan adalah algoritma K-Means,

yang dikenal efektif dalam mengelola data berjumlah besar dengan efisien [9].



Gambar 1. Clustering

2.4. Algoritma K-means

Algoritma K-Means dikenal sebagai salah satu metode pengelompokan data yang paling awal dikembangkan dan masih sering diterapkan hingga kini. Popularitasnya bertahan karena kemudahan dalam proses penerapan, terutama untuk kebutuhan aplikasi berskala kecil hingga menengah, metode ini digunakan secara luas di berbagai bidang karena mampu memberikan hasil yang cukup baik dengan proses yang tidak terlalu kompleks. Sifatnya yang sederhana menjadikannya pilihan utama bagi banyak praktisi dalam melakukan analisis kluster, konsep utama dari K-Means adalah mencari cara untuk mengelompokkan data sehingga perbedaan antar data dalam satu kelompok seminimal mungkin [10]. Dengan kata lain, algoritma ini bertujuan untuk meminimalkan jarak antar data terhadap pusat klusternya, proses dalam algoritma ini dilakukan melalui empat tahapan utama, dengan tujuan akhir untuk mengurangi jumlah kesalahan kuadrat atau *Sum of Squared Error (SSE)* antara setiap titik data dengan pusat kelompoknya masing-masing [2].

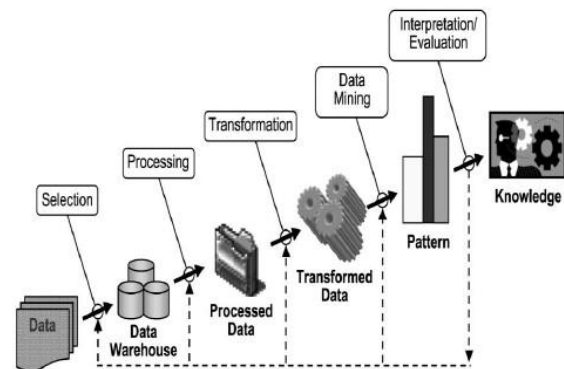
3. METODE PENELITIAN

Penelitian ini memanfaatkan informasi dari aktivitas penjualan sebagai bahan utama analisis, jumlah keseluruhan produk yang dijadikan objek pengamatan dalam studi ini mencapai 115 item. Data tersebut mencerminkan catatan penjualan yang terjadi selama lima bulan penuh dalam kurun waktu 2025, Periode pengumpulan data berlangsung mulai bulan Januari hingga akhir Mei di tahun yang sama.

Teknik *clustering* K-Means merupakan salah satu pendekatan dalam analisis data yang termasuk dalam metode *data mining*. Metode ini bekerja secara *unsupervised* atau tanpa pengawasan langsung dalam proses pemodelan data, pendekatan ini bertujuan untuk mengelompokkan data berdasarkan kesamaan karakteristik dengan cara membagi data ke dalam beberapa kelompok atau klaster melalui sistem partisi, dalam proses pengolahan dan analisis data, penerapan

metode ini mengikuti tahapan kerja yang dikenal sebagai *Knowledge Discovery in Databases (KDD)*, sebuah kerangka sistematis dalam penemuan pengetahuan dari data.

Langkah-langkah dalam proses KDD mencakup pemilihan data (*selection*), pembersihan data (*cleaning*), transformasi data, penerapan teknik *data mining*, serta evaluasi terhadap hasil yang diperoleh guna menemukan informasi yang bermakna[3].



Gambar 2. Tahapan kdd

Proses tersebut tersusun atas beberapa tahapan utama sebagai berikut:

a. Data Selection (pemilihan data)

Dalam praktik pengolahan data, tidak seluruh data yang tersedia akan dimanfaatkan dalam analisis, hal ini disebabkan karena sebagian informasi yang terkumpul mungkin tidak relevan atau tidak berkaitan dengan tujuan penelitian oleh sebab itu, diperlukan proses seleksi data agar hanya informasi yang sesuai dengan kebutuhan yang dipilih untuk dianalisis lebih lanjut. data yang telah disesuaikan tersebut nantinya akan diambil secara spesifik dari sumber database yang tersedia.

b. Pembersihan data dan proses data (*cleaning* and *processing*)

Langkah ini diterapkan untuk menyaring informasi yang bersifat statis dan tidak mengalami perubahan, tujuan utamanya adalah menghilangkan gangguan data atau *noise* yang dapat memengaruhi kualitas informasi secara keseluruhan. gangguan tersebut umumnya muncul dari berbagai sumber data yang memiliki format atau sistem penyimpanan yang berbeda-beda, setelah proses pembersihan selesai, seluruh data yang telah diselaraskan kemudian digabungkan ke dalam satu basis data terpadu yang dikenal dengan sebutan *data warehouse*.

c. Transformasi data (*transformation*)

Informasi yang tersimpan di dalam basis data akan diproses ulang menggunakan beragam metode transformasi, transformasi ini dilakukan untuk mengolah data menjadi bentuk yang lebih sesuai dengan kebutuhan analisis lanjutan, tahap seleksi data menjadi penting agar hanya informasi yang relevan saja yang digunakan dalam proses analisis, Dengan melakukan pemilihan data yang tepat, hasil yang diperoleh akan lebih akurat serta efisien

dalam hal waktu pemrosesan, terutama untuk data dalam jumlah besar.

d. Penambangan data (data mining)

Data yang sudah melalui tahap pemilihan dan transformasi selanjutnya diolah menggunakan berbagai metode penambangan, tahapan ini dikenal sebagai proses *data mining*, yang berfungsi untuk menggali informasi penting dari kumpulan data yang telah disiapkan, tujuan utama dari proses ini adalah menemukan pola-pola tersembunyi atau pengetahuan baru dengan memanfaatkan fungsi-fungsi tertentu sesuai kebutuhan analisis, penentuan algoritma yang digunakan sangat dipengaruhi oleh arah tujuan analisis serta keseluruhan proses pencarian wawasan dari data tersebut.

e. Evaluasi pola dan presentasi pengetahuan (knowledge extraction)

Langkah ini bertujuan untuk mengevaluasi apakah hasil pola yang ditemukan sejalan atau justru bertentangan dengan fakta maupun dugaan awal yang telah ditentukan sebelumnya. Selain itu, tahap ini juga mencakup proses visualisasi data guna memperjelas hasil yang telah diperoleh selama analisis, Informasi yang dihasilkan kemudian disusun dan disajikan dalam bentuk pengetahuan yang menunjukkan cara atau pendekatan yang digunakan untuk memperoleh informasi tersebut, Akhirnya, pengetahuan tersebut dituangkan dalam format yang mudah dimengerti oleh pengguna agar dapat dimanfaatkan secara maksimal.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Informasi yang digunakan berasal dari pencatatan aktivitas penjualan yang mencakup sebanyak 115 jenis produk, rentang waktu pengambilan data dimulai dari bulan Januari hingga Mei tahun 2025, Sumber data yang digunakan merupakan data primer, yang berarti bahwa data diperoleh secara langsung dari lokasi usaha, yaitu Toko Clothing Bengkulu, detail dari data transaksi tersebut dapat ditemukan pada Tabel 1.

Tabel 1. Data selection

Kode	Barang	Stok Awal	Stok Terjual	Stok Akhir
B01	Skateboard	90	58	32
B02	Sepatu Adidas Samba	40	33	7
B03	Kacamata Style Hitam	35	12	23
B04	Kaos Kerah Marun	70	47	23
B05	Jeans Wash Blue	58	28	30
B06	Tali pinggang kulit coklat	48	45	3
B07	Celana Gangster	36	32	4

Kode	Barang	Stok Awal	Stok Terjual	Stok Akhir
B08	Kacamata Vincent	144	87	57
B09	Baju lakers	48	12	36
B10	Baju Gws	36	23	13
B11	Celana Blackhawk	132	97	35
B12	Laofers	40	36	4
B13	Sepatu Vans	38	29	9
B14	Kaos Barca	24	15	9
B15	Tas Style Kulit	45	40	5
B16	Handbag Black	120	78	42
B17	Sepatu Docmart	50	47	3
B18	Gantungan HP	60	27	33
B19	Chinos Khaki	58	15	43
B20	Ripped Jeans Wash Black	36	35	1
B21	Maternal Shirt	48	15	33
B22	Varsity	36	17	19
B23	Jaket Parasut	180	115	65
B24	SmartWatch	65	27	38
B25	Cincin Romawi	75	27	48
B26	Sendal Crocs	96	39	57
B27	Hoodie Oversize	80	50	30
B28	Topi Snapback	60	25	35
B29	Dompot Kulit Pria	55	32	23
B30	Kemeja Flanel Hitam Merah	70	44	26
B31	Sepatu Converse High	45	28	17
B32	Celana Cargo Abu-Abu	48	36	12
B33	Rompi Jeans	30	14	16
B34	Tas Selempang Kanvas	64	27	37
B35	Kaos Oversize Putih Polos	92	58	34
B36	Gelang Kulit Hitam	40	19	21
B37	Sweater Rajut	75	40	35
B38	Jaket Hoodie	32	32	0
B39	Kaos Raglan	113	17	96
B40	Celana Pendek Pria	78	5	73
B41	Sepatu Slip-On	108	107	1
B42	Tas Punggung	116	68	48
B43	Kemeja Batik	94	29	65
B44	Dompot Kecil	101	61	40
B45	Topi Fedora	138	91	47
B46	Jam Tangan Analog	38	11	27
B47	Kaos Lengan Panjang	62	12	50
B48	Celana Jogger	72	24	48
B49	Rompi Outdoor	74	50	24
B50	Jaket Denim	117	29	88
B51	Baju Oversize	111	7	104
B52	Sepatu Lari	71	63	8
B53	Kaos Jersey	142	44	98
B54	Sweater Polos	46	17	29
B55	Sendal Jepit	140	136	4

Kode	Barang	Stok Awal	Stok Terjual	Stok Akhir
B56	Sling Bag Kulit	113	97	16
B57	Gelang Manik	147	27	120
B58	Kaos Polo Pria	140	7	133
B59	Handbag Wanita	94	69	25
B60	Tas Bahu Remaja	145	67	78
B61	Celana Legging	91	81	10
B62	Long Coat	41	22	19
B63	Sepatu Pantofel	86	59	27
B64	Kaos Komunitas	93	56	37
B65	Cardigan	61	25	36
B66	Sarung Tangan Motor	36	21	15
B67	Scarf Rajut	148	32	116
B68	Topi Bucket	141	18	123
B69	Jam Digital	123	75	48
B70	Sandal Gunung	98	33	65
B71	Kemeja Linen	105	93	12
B72	Jas Pria	93	18	75
B73	Kaos Sablon	37	31	6
B74	Kacamata Fashion	69	58	11
B75	Tas Laptop	95	19	76
B76	Tas Waist Bag	70	25	45
B77	Rompi Tactical	79	27	52
B78	Jaket Kulit	78	67	11
B79	Sepatu Casual	66	38	28
B80	Celana Bahan Slimfit	74	49	25
B81	Kemeja Oxford	70	46	24
B82	Tas Travel	129	75	54
B83	Gelang Stainless	103	58	45
B84	Sendal Stylish	82	74	8
B85	Tote Bag	104	86	18
B86	Kacamata Anti Radiasi	67	4	63
B87	Sweater Rajut	44	4	40
B88	Jaket Hoodie	45	33	12
B89	Kaos Raglan	84	1	83
B90	Celana Pendek Pria	55	20	35
B91	Sepatu Slip-On	71	12	59
B92	Tas Punggung	124	18	106
B93	Kemeja Batik	144	142	2
B94	Dompot Kecil	113	107	6
B95	Topi Fedora	123	3	120
B96	Jam Tangan Analog	33	33	0
B97	Kaos Lengan Panjang	43	17	26
B98	Celana Jogger	37	36	1
B99	Rompi Outdoor	113	8	105
B100	Jaket Denim	81	69	12
B101	Baju Oversize	75	47	28
B102	Sepatu Lari	131	89	42
B103	Kaos Jersey	98	10	88
B104	Sweater Polos	54	6	48
B105	Sendal Jepit	100	83	17

Kode	Barang	Stok Awal	Stok Terjual	Stok Akhir
B106	Sling Bag Kulit	37	6	31
B107	Gelang Manik	30	25	5
B108	Kaos Polo Pria	132	94	38
B109	Handbag Wanita	91	53	38
B110	Tas Bahu Remaja	113	109	4
B111	Celana Legging	124	82	42
B112	Long Coat	57	7	50
B113	Sepatu Pantofel	53	50	3
B114	Kaos Komunitas	33	22	11
B115	Cardigan	79	74	5

4.2. Data Cleaning dan Processing

Langkah awal yang dilakukan melibatkan proses dasar, salah satunya adalah pembersihan data dari gangguan atau noise.

Tahapan ini penting untuk memastikan bahwa data yang digunakan dalam analisis berada dalam kondisi optimal, Proses pembersihan ini termasuk dalam bagian dari tahap pra-pemrosesan data. Data yang digunakan merupakan hasil transaksi penjualan yang kemudian diolah menggunakan perangkat lunak RapidMiner.

Tabel 2. Data cleaning

Kode	Barang	Saldo Awal	Saldo Terjual	Saldo Akhir
B01	Skateboard	90	58	32
B02	Sepatu Adidas Samba	40	33	7
B03	Kacamata Style Hitam	35	12	23
B04	Kaos Kerah Marun	70	47	23
B05	Jeans Wash Blue	58	28	30
B06	Tali pinggang kulit coklat	48	45	3
B07	Celana Gangster	36	32	4
B08	Kacamata Vincent	144	87	57
B09	Baju lakers	48	12	36
B10	Baju Gws	36	23	13
B11	Celana Blackhawk	132	97	35
B12	Laofers	40	36	4
B13	Sepatu Vans	38	29	9
B14	Kaos Barca	24	15	9
B15	Tas Style Kulit	45	40	5
B16	Handbag Black	120	78	42
B17	Sepatu Docmart	50	47	3
B18	Gantungan HP	60	27	33
B19	Chinos Khaki	58	15	43
B20	Ripped Jeans Wash Black	36	35	1
B21	Maternal Shirt	48	15	33

Kode	Barang	Saldo Awal	Saldo Terjual	Saldo Akhir
B22	Varsity	36	17	19
B23	Jaket Parasut	180	115	65
B24	SmartWatch	65	27	38
B25	Cincin Romawi	75	27	48
B26	Sendal Crocs	96	39	57
B27	Hoodie Oversize	80	50	30
B28	Topi Snapback	60	25	35
B29	Dompot Kulit Pria	55	32	23
B30	Kemeja Flanel Hitam Merah	70	44	26
B31	Sepatu Converse High	45	28	17
B32	Celana Cargo Abu-Abu	48	36	12
B33	Rompi Jeans	30	14	16
B34	Tas Selempang Kanvas	64	27	37
B35	Kaos Oversize Putih Polos	92	58	34
B36	Gelang Kulit Hitam	40	19	21
B37	Sweater Rajut	75	40	35
B38	Jaket Hoodie	32	32	0
B39	Kaos Raglan	113	17	96
B40	Celana Pendek Pria	78	5	73
B41	Sepatu Slip-On	108	107	1
B42	Tas Punggung	116	68	48
B43	Kemeja Batik	94	29	65
B44	Dompot Kecil	101	61	40
B45	Topi Fedora	138	91	47
B46	Jam Tangan Analog	38	11	27
B47	Kaos Lengan Panjang	62	12	50
B48	Celana Jogger	72	24	48
B49	Rompi Outdoor	74	50	24
B50	Jaket Denim	117	29	88
B51	Baju Oversize	111	7	104
B52	Sepatu Lari	71	63	8
B53	Kaos Jersey	142	44	98
B54	Sweater Polos	46	17	29
B55	Sendal Jepit	140	136	4
B56	Sling Bag Kulit	113	97	16
B57	Gelang Manik	147	27	120
B58	Kaos Polo Pria	140	7	133
B59	Handbag Wanita	94	69	25
B60	Tas Bahu Remaja	145	67	78
B61	Celana Legging	91	81	10
B62	Long Coat	41	22	19
B63	Sepatu Pantofel	86	59	27
B64	Kaos Komunitas	93	56	37
B65	Cardigan	61	25	36

Kode	Barang	Saldo Awal	Saldo Terjual	Saldo Akhir
B66	Sarung Tangan Motor	36	21	15
B67	Scarf Rajut	148	32	116
B68	Topi Bucket	141	18	123
B69	Jam Digital	123	75	48
B70	Sandal Gunung	98	33	65
B71	Kemeja Linen	105	93	12
B72	Jas Pria	93	18	75
B73	Kaos Sablon	37	31	6
B74	Kacamata Fashion	69	58	11
B75	Tas Laptop	95	19	76
B76	Tas Waist Bag	70	25	45
B77	Rompi Tactical	79	27	52
B78	Jaket Kulit	78	67	11
B79	Sepatu Casual	66	38	28
B80	Celana Bahan Slimfit	74	49	25
B81	Kemeja Oxford	70	46	24
B82	Tas Travel	129	75	54
B83	Gelang Stainless	103	58	45
B84	Sendal Stylish	82	74	8
B85	Tote Bag	104	86	18
B86	Kacamata Anti Radiasi	67	4	63
B87	Sweater Rajut	44	4	40
B88	Jaket Hoodie	45	33	12
B89	Kaos Raglan	84	1	83
B90	Celana Pendek Pria	55	20	35
B91	Sepatu Slip-On	71	12	59
B92	Tas Punggung	124	18	106
B93	Kemeja Batik	144	142	2
B94	Dompot Kecil	113	107	6
B95	Topi Fedora	123	3	120
B96	Jam Tangan Analog	33	33	0
B97	Kaos Lengan Panjang	43	17	26
B98	Celana Jogger	37	36	1
B99	Rompi Outdoor	113	8	105
B100	Jaket Denim	81	69	12
B101	Baju Oversize	75	47	28
B102	Sepatu Lari	131	89	42
B103	Kaos Jersey	98	10	88
B104	Sweater Polos	54	6	48
B105	Sendal Jepit	100	83	17
B106	Sling Bag Kulit	37	6	31
B107	Gelang Manik	30	25	5
B108	Kaos Polo Pria	132	94	38
B109	Handbag Wanita	91	53	38
B110	Tas Bahu Remaja	113	109	4
B111	Celana Legging	124	82	42
B112	Long Coat	57	7	50

Kode	Barang	Saldo Awal	Saldo Terjual	Saldo Akhir
B113	Sepatu Pantofel	53	50	3
B114	Kaos Komunitas	33	22	11
B115	Cardigan	79	74	5

Data penjualan yang telah diseleksi selanjutnya akan diproses menggunakan aplikasi RapidMiner. Data tersebut kemudian dijalankan ke dalam sistem untuk tahap analisis lebih lanjut. Sebelum pemrosesan, dilakukan pemeriksaan kualitas data untuk memastikan validitas dan konsistensinya. Hasil pengecekan menunjukkan bahwa data ini bersih dan tidak mengandung noise, sehingga dapat langsung digunakan dalam proses analisis.

4.3. Data Transformation

Proses transformasi data dilakukan untuk menyiapkan dataset agar sesuai dengan kebutuhan dalam analisis data lanjutan. Tahapan ini bertujuan untuk mengubah struktur data mentah menjadi format yang lebih rapi dan siap digunakan. Setiap elemen data diperiksa dan disesuaikan dengan tipe data yang relevan, agar kompatibel dengan algoritma dalam proses data mining. Data tersebut dapat dilihat pada Gambar 3.

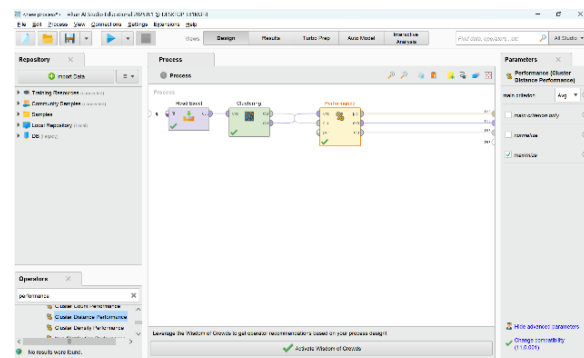
KD	Barang	Stok Awal	ST	SAK
polynomial 12	polynomial 12	integer	integer	integer
1 B01	Stanoboard	50	58	32
2 B02	Sepatu Adidas Samba	40	33	7
3 B03	Kacamata Style Hitam	35	12	23
4 B04	Kaos Kerah Marun	70	47	23
5 B05	Jens Wash Blue	55	28	30
6 B06	Tali pinggang kulit co.	48	45	3
7 B07	Celana Gangster	36	32	4
8 B08	Kacamata Vincent	144	87	57
9 B09	Baju lakers	48	12	36
10 B10	Baju Gens	36	23	13
11 B11	Celana Blackhawk	132	97	35
12 B12	Lodders	40	36	4

Gambar 3. Pemilihan atribut yang unik

4.4. Data Mining

Data mining merupakan suatu tahapan dalam pengolahan data yang bertujuan untuk menemukan pola tertentu atau informasi penting dari data yang telah dipilih, dengan memanfaatkan algoritma *k-means* sebagai metode pengelompokan, pada Gambar 4 diperlihatkan bahwa proses analisis diawali dengan penggunaan operator *Retrieve* untuk mengambil dataset yang akan dianalisis dalam penelitian ini. Setelah data berhasil diakses, dilakukan pemodelan menggunakan teknik *k-means clustering*. Evaluasi terhadap hasil pengelompokan dilakukan dengan operator *Cluster Distance Performance* untuk menilai sejauh mana kelompok yang terbentuk memiliki perbedaan yang signifikan. Dalam proses ini, ditetapkan parameter *k* sebesar 3, yang artinya data dikelompokkan ke dalam tiga segmen, yaitu produk

yang laris, produk dengan penjualan kurang laris, dan produk yang tidak laris. Selain itu, jumlah maksimal iterasi yang digunakan dalam pengelompokan ditentukan sebanyak 10 kali atau *max run = 10*.



Gambar 4. Model Algoritma K-means

4.5. Evaluasi

Berdasarkan hasil evaluasi performa dari algoritma K-Means, diketahui bahwa Cluster_0 memiliki nilai *centroid distance* sebesar 1338.544, yang menunjukkan kelompok produk dengan tingkat penjualan tertinggi dan oleh karena itu dikategorikan sebagai produk laris. Selanjutnya, Cluster_1 memiliki nilai 690.725, mencerminkan tingkat penjualan yang rendah dalam kelompok ini umumnya memiliki stok yang tidak terlalu banyak, namun tetap tidak mampu mencapai angka penjualan yang signifikan. Hal ini mengindikasikan bahwa produk dalam cluster ini kurang diminati. Oleh karena itu, dikategorikan sebagai produk kurang laris. Sementara itu, Cluster_2 dengan nilai 1205.695 berada di antara keduanya namun menunjukkan stok yang tinggi dengan penjualan lebih rendah, sehingga dapat dikategorikan sebagai produk tidak laris. *davies bouldin* dengan hasil 0.724 dapat dilihat pada gambar 5.

Performance
PerformanceVector:
Avg. within centroid distance: 946.183
Avg. within centroid distance_cluster_0: 1338.544
Avg. within centroid distance_cluster_1: 690.725
Avg. within centroid distance_cluster_2: 1205.695
Davies Bouldin: 0.724

Gambar 5. Data Performance

4.6. Pembahasan

Berdasarkan hasil dari nilai performance diatas yang mengkategorikan sebagai Cluster_0 dengan nilai 1338.544 dikategorikan sebagai Laris, Cluster_1 dengan nilai 690.725 dikategorikan sebagai Kurang Laris, sedangkan Cluster_2 dengan nilai 1205.695 dikategorikan sebagai Tidak Laris.

Berikut Tabel 3 Cluster_0 dengan kategori Laris.

Tabel 3. Cluster_0

NO	KD	Barang	Hasil Cluster	Keterangan
1	B01	Skateboard	cluster_0	Laris
2	B08	Kacamata Vincent	cluster_0	Laris
3	B11	Celana Blackhawk	cluster_0	Laris
4	B16	Handbag Black	cluster_0	Laris
5	B23	Jaket Parasut	cluster_0	Laris
6	B35	Kaos Oversize Putih Polos	cluster_0	Laris
7	B41	Sepatu Slip-On	cluster_0	Laris
8	B42	Tas Punggung	cluster_0	Laris
9	B44	Dompot Kecil	cluster_0	Laris
10	B45	Topi Fedora	cluster_0	Laris
11	B55	Sendal Jepit	cluster_0	Laris
12	B56	Sling Bag Kulit	cluster_0	Laris
13	B59	Handbag Wanita	cluster_0	Laris
14	B61	Celana Legging	cluster_0	Laris
15	B63	Sepatu Pantofel	cluster_0	Laris
16	B64	Kaos Komunitas	cluster_0	Laris
17	B69	Jam Digital	cluster_0	Laris
18	B71	Kemeja Linen	cluster_0	Laris
19	B78	Jaket Kulit	cluster_0	Laris
20	B82	Tas Travel	cluster_0	Laris
21	B83	Gelang Stainless	cluster_0	Laris
22	B84	Sendal Stylish	cluster_0	Laris
23	B85	Tote Bag	cluster_0	Laris
24	B93	Kemeja Batik	cluster_0	Laris
25	B94	Dompot Kecil	cluster_0	Laris
26	B100	Jaket Denim	cluster_0	Laris
27	B102	Sepatu Lari	cluster_0	Laris
28	B105	Sendal Jepit	cluster_0	Laris
29	B108	Kaos Polo Pria	cluster_0	Laris
30	B109	Handbag Wanita	cluster_0	Laris
31	B110	Tas Bahu Remaja	cluster_0	Laris
32	B111	Celana Legging	cluster_0	Laris
33	B115	Cardigan	cluster_0	Laris

Selanjutnya Tabel 4 Cluster_1 dengan kategori Kurang laris.

Tabel 4. Cluster_1

NO	KD	Barang	Hasil Cluster	Keterangan
1	B02	Sepatu Adidas Samba	cluster_1	Kurang Laris
2	B03	Kacamata Style Hitam	cluster_1	Kurang Laris
3	B04	Kaos Kerah Marun	cluster_1	Kurang Laris

NO	KD	Barang	Hasil Cluster	Keterangan
4	B05	Jeans Wash Blue	cluster_1	Kurang Laris
5	B06	Tali pinggang kulit coklat	cluster_1	Kurang Laris
6	B07	Celana Gangster	cluster_1	Kurang Laris
7	B09	Baju lakers	cluster_1	Kurang Laris
8	B10	Baju Gws	cluster_1	Kurang Laris
9	B12	Laofers	cluster_1	Kurang Laris
10	B13	Sepatu Vans	cluster_1	Kurang Laris
11	B14	Kaos Barca	cluster_1	Kurang Laris
12	B15	Tas Style Kulit	cluster_1	Kurang Laris
13	B17	Sepatu Docmart	cluster_1	Kurang Laris
14	B18	Gantungan HP	cluster_1	Kurang Laris
15	B19	Chinos Khaki	cluster_1	Kurang Laris
16	B20	Ripped Jeans Wash Black	cluster_1	Kurang Laris
17	B21	Maternal Shirt	cluster_1	Kurang Laris
18	B22	Varsity	cluster_1	Kurang Laris
19	B24	SmartWatch	cluster_1	Kurang Laris
20	B25	Cincin Romawi	cluster_1	Kurang Laris
21	B27	Hoodie Oversize	cluster_1	Kurang Laris
22	B28	Topi Snapback	cluster_1	Kurang Laris
23	B29	Dompot Kulit Pria	cluster_1	Kurang Laris
24	B30	Kemeja Flanel Hitam Merah	cluster_1	Kurang Laris
25	B31	Sepatu Converse High	cluster_1	Kurang Laris
26	B32	Celana Cargo Abu-Abu	cluster_1	Kurang Laris
27	B33	Rompi Jeans	cluster_1	Kurang Laris
28	B34	Tas Selempang Kanvas	cluster_1	Kurang Laris
29	B36	Gelang Kulit Hitam	cluster_1	Kurang Laris
30	B37	Sweater Rajut	cluster_1	Kurang Laris
31	B38	Jaket Hoodie	cluster_1	Kurang Laris
32	B46	Jam Tangan Analog	cluster_1	Kurang Laris
33	B47	Kaos Lengan Panjang	cluster_1	Kurang Laris
34	B48	Celana Jogger	cluster_1	Kurang Laris

NO	KD	Barang	Hasil Cluster	Keterangan
35	B49	Rompi Outdoor	cluster_1	Kurang Laris
36	B52	Sepatu Lari	cluster_1	Kurang Laris
37	B54	Sweater Polos	cluster_1	Kurang Laris
38	B62	Long Coat	cluster_1	Kurang Laris
39	B65	Cardigan	cluster_1	Kurang Laris
40	B66	Sarung Tangan Motor	cluster_1	Kurang Laris
41	B73	Kaos Sablon	cluster_1	Kurang Laris
42	B74	Kacamata Fashion	cluster_1	Kurang Laris
43	B76	Tas Waist Bag	cluster_1	Kurang Laris
44	B77	Rompi Tactical	cluster_1	Kurang Laris
45	B79	Sepatu Casual	cluster_1	Kurang Laris
46	B80	Celana Bahan Slimfit	cluster_1	Kurang Laris
47	B81	Kemeja Oxford	cluster_1	Kurang Laris
48	B86	Kacamata Anti Radiasi	cluster_1	Kurang Laris
49	B87	Sweater Rajut	cluster_1	Kurang Laris
50	B88	Jaket Hoodie	cluster_1	Kurang Laris
51	B91	Sepatu Slip-On	cluster_1	Kurang Laris
52	B96	Jam Tangan Analog	cluster_1	Kurang Laris
53	B97	Kaos Lengan Panjang	cluster_1	Kurang Laris
54	B98	Celana Jogger	cluster_1	Kurang Laris
55	B99	Rompi Outdoor	cluster_1	Kurang Laris
56	B101	Baju Oversize	cluster_1	Kurang Laris
57	B104	Sweater Polos	cluster_1	Kurang Laris
58	B106	Sling Bag Kulit	cluster_1	Kurang Laris
59	B107	Gelang Manik	cluster_1	Kurang Laris
60	B112	Long Coat	cluster_1	Kurang Laris
61	B113	Sepatu Pantofel	cluster_1	Kurang Laris
62	B114	Kaos Komunitas	cluster_1	Kurang Laris

Selanjutnya Tabel 5 Cluster_2 dengan kategori Tidak Laris

Tabel 5. Cluster_2

NO	KD	Barang	Hasil Cluster	Keterangan
1	B26	Sendal Crocs	cluster_2	Tidak Laris
2	B39	Kaos Raglan	cluster_2	Tidak Laris
3	B40	Celana Pendek Pria	cluster_2	Tidak Laris
4	B43	Kemeja Batik	cluster_2	Tidak Laris
5	B50	Jaket Denim	cluster_2	Tidak Laris
6	B51	Baju Oversize	cluster_2	Tidak Laris
7	B53	Kaos Jersey	cluster_2	Tidak Laris
8	B57	Gelang Manik	cluster_2	Tidak Laris
9	B58	Kaos Polo Pria	cluster_2	Tidak Laris
10	B60	Tas Bahu Remaja	cluster_2	Tidak Laris
11	B67	Scarf Rajut	cluster_2	Tidak Laris
12	B68	Topi Bucket	cluster_2	Tidak Laris
13	B70	Sandal Gunung	cluster_2	Tidak Laris
14	B72	Jas Pria	cluster_2	Tidak Laris
15	B75	Tas Laptop	cluster_2	Tidak Laris
16	B89	Kaos Raglan	cluster_2	Tidak Laris
17	B92	Tas Punggung	cluster_2	Tidak Laris
18	B95	Topi Fedora	cluster_2	Tidak Laris
19	B99	Rompi Outdoor	cluster_2	Tidak Laris
20	B103	Kaos Jersey	cluster_2	Tidak Laris

Berdasarkan hasil penelitian, penerapan algoritma K-Means pada data penjualan berhasil mengelompokkan produk ke dalam tiga kategori utama. Terdapat sebanyak 33 item produk yang termasuk dalam kategori *laris*, 62 item produk masuk ke dalam kelompok *kurang laris*, dan 20 item produk tergolong dalam kategori *tidak laris*. Evaluasi hasil klusterisasi dilakukan untuk membandingkan performa antara dua alat bantu analisis, yaitu RapidMiner dan Microsoft Excel. Hasil evaluasi menunjukkan bahwa RapidMiner menghasilkan nilai akurasi, recall, dan precision yang lebih unggul dibandingkan dengan Excel. Perhitungan yang dilakukan menggunakan RapidMiner menunjukkan keandalan yang lebih tinggi, terutama dalam mengelompokkan data secara tepat sesuai dengan karakteristik penjualannya. Hal ini membuktikan

bahwa RapidMiner lebih efektif dalam proses analisis clustering dibandingkan dengan metode manual menggunakan Excel. Dengan mempertimbangkan hasil evaluasi yang diperoleh, peneliti merekomendasikan penggunaan RapidMiner sebagai alat utama dalam pengelompokan data penjualan. Hal ini bertujuan untuk mengurangi risiko kesalahan dalam perhitungan serta memperoleh hasil yang lebih akurat dalam menentukan produk laris, kurang laris, dan tidak laris.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan terhadap data penjualan Toko Clothing Bengkulu yang beralamat di Jalan Bumi Ayu 5 Kota Bengkulu, Provinsi Bengkulu, penerapan algoritma K-Means berhasil mengelompokkan 115 produk ke dalam tiga kategori, yaitu 33 produk tergolong *laris*, 62 produk *kurang laris*, dan 20 produk *tidak laris*. Evaluasi performa model menunjukkan nilai *average within centroid distance* sebesar 966.181, dengan rincian: *Cluster 0* (*laris*) sebesar 1338.544, *Cluster 1* (*kurang laris*) sebesar 690.725, dan *Cluster 2* (*tidak laris*) sebesar 1205.695. Selain itu, nilai *Davies Bouldin Index* sebesar 0.724 menunjukkan kualitas pengelompokan yang cukup baik dan dapat dijadikan dasar pengambilan keputusan. Hasil evaluasi juga memperlihatkan bahwa RapidMiner memiliki performa yang lebih unggul dibandingkan Microsoft Excel dalam aspek akurasi, recall, dan precision, sehingga sangat direkomendasikan sebagai alat utama dalam proses analisis data penjualan. Penelitian ini menyarankan agar pada tahap selanjutnya dilakukan pengembangan dengan memasukkan variabel tambahan seperti waktu penjualan, musim, dan tren konsumen agar hasil analisis semakin akurat. Secara khusus, peneliti juga menyarankan kepada pemilik Toko Clothing Bengkulu untuk memanfaatkan hasil pengelompokan ini sebagai dasar dalam menyusun strategi promosi dan pengadaan stok, misalnya dengan lebih fokus memasarkan produk *laris*, memberikan diskon untuk produk *kurang laris*, serta mengevaluasi keberadaan produk *tidak laris* agar efisiensi operasional toko dapat ditingkatkan secara optimal.

DAFTAR PUSTAKA

- [1] G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 40–46, 2022, doi: 10.30871/jaic.v6i1.3824.
- [2] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [3] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.
- [4] M. Fajar, M. Adji, and G. Dwilestari, "ANALISIS DATA TRANSAKSI PENJUALAN BARANG MENGGUNAKAN TEKNIK K-MEANS CLUSTERING," vol. 9, no. 1, pp. 619–625, 2025.
- [5] R. Rahmawati, W. Prihartono, and K. Cirebon, "OPTIMASI STOK DENGAN CLUSTERING DATA MENGGUNAKAN," vol. 13, no. 2, 2025.
- [6] C. L. Angraeni, D. F., Rahaningsih, N., Dana, R. D., & Rohmat, "PEMANFAATAN ALGORITMA K-MEANS DALAM ANALISIS DATA PENJUALAN TOKO BUYUNG UPIK JS DI LAZADA," vol. 13, no. 2, pp. 1297–1306, 2025.
- [7] M. Studi, K. Kfc, M. T. Haryono, K. Nadhilah, and A. Mulyani, "ANALISIS PENGELOMPOKKAN MENU TERLARIS DI RESTORAN CEPAT SAJI MENGGUNAKAN METODE K-," vol. 9, no. 2, pp. 674–684, 2025, doi: 10.52362/jisamar.v9i2.1808.
- [8] J. Nasional, S. Informasi, E. Paskasius, and E. Magdalena, "Penerapan Metode Clustering Pada Penjualan Makanan Dan Minuman Dalam Menentukan Strategi Marketing," vol. 1, pp. 48–56, 2025.
- [9] W. A. Wahyuni and S. Saepudin, "Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform. Univ. Nusa Putra)*, vol. 1, no. 1, pp. 306–313, 2021.
- [10] N. Rahaningsih, I. Ali, K. Akuntansi, and S. I. Cirebon, "SEGMENTASI DATA TRANSAKSI PENJUALAN TOKO ONLINE VASTYLE UNTUK MANAJEMEN STOK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," vol. 9, no. 3, pp. 5093–5099, 2025.