

Application of Data Mining to Predict Distro Clothing Sales Using the K-Means Clustering Method

[Penerapan Data Mining Untuk Prediksi Penjualan Baju Distro Menggunakan Metode *K-Means Clustering*]

Shafa Arrizqa Az Zahroh¹⁾, Nuril Lutvi Azizah²⁾, Novia Ariyanti³⁾, Irwan Alnarus Kautsar⁴⁾

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

³⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

⁴⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

Abstract. *The rapid development of technology in the modern era has brought significant changes to various aspects of society, including the world of business and commerce. This research aims to analyze the data to identify product groups based on their sales levels. The results of this analysis are expected to help Aldi Store in developing more effective inventory and marketing strategies, thereby increasing sales and the store's competitiveness. During the implementation stage, data processing was carried out using Google Colaboratory as the primary tool to execute the K-Means algorithm. In this process, the K-Means algorithm was not only applied independently but also combined with two relevant evaluation techniques: the Silhouette Coefficient and the Davies-Bouldin Index. This study demonstrates that the K-Means Clustering method can be effectively used to group sales patterns of distro clothing. With an optimal cluster, the evaluation results indicate high cluster quality based on the Silhouette Coefficient (0.576), Calinski-Harabasz Index (19.125), and Davies-Bouldin Index (0.308). These findings suggest that the formed clusters have high cohesion and clear separation, enabling more accurate customer segmentation.*

Keywords - *Data Mining, K-Means Clustering, Sales Prediction, Customer Segmentation*

Abstrak. *Pesatnya perkembangan teknologi di era modern telah membawa perubahan besar dalam berbagai aspek kehidupan masyarakat, termasuk dalam dunia bisnis dan perdagangan. penelitian ini bertujuan untuk menganalisis data tersebut guna mengidentifikasi kelompok produk berdasarkan tingkat penjualannya. Hasil analisis ini diharapkan dapat membantu Toko Aldi dalam menyusun strategi persediaan dan pemasaran yang lebih efektif, sehingga dapat meningkatkan penjualan dan daya saing toko. Pada tahap implementasi, dilakukan pemrosesan data dengan memanfaatkan platform Google Colaboratory sebagai alat utama untuk menjalankan algoritma k-means. Dalam proses ini, algoritma k-means tidak hanya diterapkan secara mandiri, tetapi juga dikombinasikan dengan dua teknik evaluasi yang relevan, yaitu Silhouette Coefficient dan Davies-Bouldin Index. Penelitian ini menunjukkan bahwa metode K-Means Clustering dapat digunakan secara efektif untuk mengelompokkan pola penjualan baju distro. Dengan klaster optimal, hasil evaluasi menunjukkan kualitas klaster yang sangat baik berdasarkan nilai Silhouette Coefficient (0.576), Calinski-Harabasz Index (19.125), dan Davies-Bouldin Index (0.308). Hasil ini mengindikasikan bahwa klaster yang terbentuk memiliki kohesi yang tinggi dan pemisahan yang jelas, sehingga mampu memberikan segmentasi pelanggan yang lebih akurat.*

Kata Kunci - *Data Mining, K-Means Clustering, Prediksi Penjualan, Segmentasi Pelanggan*

I. PENDAHULUAN

Pesatnya perkembangan teknologi pada era modern sudah menunjukkan transformasi besar untuk sejumlah bidang kehidupan masyarakat, mencakup untuk bidang bisnis dan perdagangan [1].. Teknologi informasi sudah sebagai salah satu komponen penting untuk mendukung proses bisnis, mulai dari pengelolaan inventaris hingga strategi pemasaran [2]. Pemanfaatan teknologi tidak lagi terbatas pada pengumpulan data, tetapi juga pada analisis data untuk mendapatkan wawasan yang bermanfaat bagi pengambilan keputusan. Dalam dunia perdagangan yang semakin kompetitif, pemanfaatan data menjadi sangat penting, khususnya bagi industri yang bergerak dalam penjualan barang, seperti baju distro [3].

Industri baju distro, yang dikenal dengan kreativitas dan inovasi desainnya, menghadapi tantangan besar dalam mempertahankan daya saing [4]. Persaingan yang semakin ketat menuntut para pengusaha untuk terus beradaptasi dan mencari cara yang lebih efektif dalam mengelola bisnis mereka [5]. Salah satu aspek penting dalam industri ini adalah kemampuan untuk memenuhi kebutuhan konsumen secara tepat waktu dan efisien. Untuk mencapai tujuan tersebut, para pelaku usaha perlu memahami pola penjualan dan tren pasar sehingga dapat mengambil langkah strategis yang tepat dalam pengelolaan stok dan pemasaran produk [6].

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

Berdasarkan tinjauan pustaka, penelitian-penelitian sebelumnya menunjukkan bahwa analisis data penjualan dapat memberikan dampak signifikan terhadap peningkatan efisiensi dan efektivitas pengelolaan bisnis [7]. Sebagai contoh, algoritma Apriori telah digunakan untuk menganalisis pola kombinasi item dengan frekuensi tinggi, yang bermanfaat dalam pengambilan keputusan terkait persediaan barang [8]. Namun, penelitian-penelitian ini masih memiliki keterbatasan, seperti kesulitan dalam memperoleh data yang terkini dan terbatasnya pemanfaatan data secara maksimal. Kelemahan ini menunjukkan adanya peluang untuk mengembangkan pendekatan baru yang lebih sesuai dengan kebutuhan industri tertentu, seperti baju distro [9].

Data mining, sebagai salah satu cabang dari kecerdasan buatan, menawarkan berbagai teknik untuk menggali pola dan informasi penting dari data dalam jumlah besar [10]. Data mining memungkinkan pengusaha dalam menganalisis data historis serta mengidentifikasi pola yang relevan, seperti tren musiman, preferensi konsumen, atau kategori produk yang paling diminati. Dalam konteks penjualan baju distro, data mining dapat digunakan untuk mengoptimalkan pengelolaan stok, memperkirakan permintaan, dan menyusun strategi pemasaran yang lebih efektif [11].

K-Means Clustering adalah salah satu metode pada data mining dengan memiliki potensi besar dalam suatu metode. Algoritma ini dimanfaatkan dalam mengklasifikasikan data pada beberapa kelompok menurut kemiripan karakteristik [12]. Dengan menggunakan K-Means Clustering, produk-produk dalam toko distro dapat dikelompokkan berdasarkan pola penjualan yang serupa, misalnya produk dengan sangat laku, cukup laku, serta kurang laku. Informasi ini dapat digunakan sebagai dasar untuk memprediksi permintaan produk di masa depan dan mengoptimalkan strategi persediaan. Selain itu, metode ini memungkinkan toko untuk mengidentifikasi tren musiman atau pola penjualan tertentu yang tidak mudah terlihat secara langsung dari data mentah.

Penelitian ini dilatar belakangi oleh kelemahan yang ada pada penelitian sebelumnya, di mana belum ada implementasi sistem pendukung keputusan berbasis K-Means Clustering yang secara khusus diterapkan pada industri baju distro. Penelitian ini akan fokus pada data penjualan dari Toko Aldi, yang meliputi variabel-variabel seperti tanggal transaksi, status pembayaran, jumlah barang, harga satuan, dan total harga. Dengan menerapkan metode K-Means Clustering, penelitian ini terdapat tujuan untuk menganalisis data tersebut guna mengidentifikasi kelompok produk berdasarkan tingkat penjualannya. Hasil analisis ini diharapkan dapat membantu Toko Aldi dalam menyusun strategi persediaan dan pemasaran secara lebih efektif, dengan demikian mampu meningkatkan penjualan serta daya saing toko.

Dengan demikian, penelitian ini berupaya untuk menjawab kebutuhan akan sistem pendukung keputusan yang berbasis data mining dalam konteks penjualan baju distro.

II. METODE

2.1 Pengumpulan Data

1. Lokasi dan Waktu Penelitian

Riset dilaksanakan di Toko Aldi Baju Distro yang terletak di Perumahan Mega Asri Larangan Kabupaten Sidoarjo dan waktu pelaksanaan dilakukan tanggal 1 Juni 2024.

2. Studi Literatur

Di tahap ini, referensi dikumpulkan melalui membaca jurnal ilmiah dan melakukan penelusuran internet terkait penerapan data mining, khususnya K-Means Clustering, dalam prediksi penjualan.

3. Wawancara

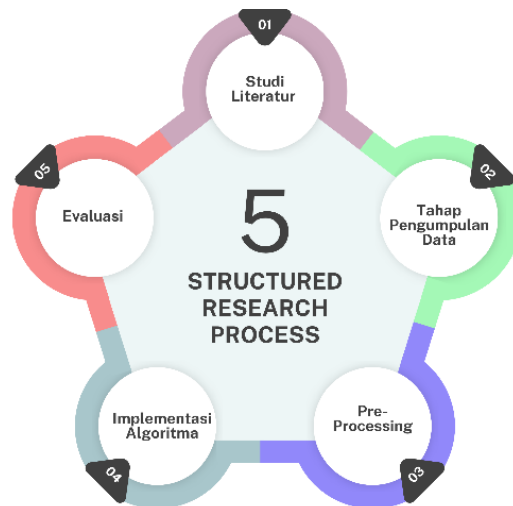
Data diperoleh setelah melakukan wawancara langsung dengan pihak yang mempunyai Toko Aldi Baju Distro sehingga diketahui data transaksi penjualan baju yang dijual.

4. Pengambilan data langsung

Pengambilan data untuk penelitian ini dilaksanakan dengan berbagai tahapan, diantaranya wawancara langsung dengan pemilik Toko Aldi Baju Distro untuk mendapatkan informasi mengenai transaksi penjualan baju. Kemudian pengambilan dataset diambil dari bulan April-Juli 2024 sebanyak 393 data yang di dapat dari toko tersebut. Penelitian ini juga didukung oleh studi literatur yang mencakup berbagai referensi terkait penerapan metode data mining, utamanya K-Means Clustering, yang dimanfaatkan dalam menganalisis dan memprediksi pola penjualan.

2.2 Tahapan Penelitian

Tahapan penelitian adalah langkah yang akan diambil untuk menuntaskan penelitian dari rancangan awal hingga pengujian. Dimulai dari tahap pengumpulan data kemudian di lanjut ke tahap pre-processing kemudian di lanjut ke tahap pemrosesan data hingga ke tahap validasi sebagaimana terlampir dalam rangkaian berikut pada Gambar 1 :



Gambar 1. Alur Penelitian

2.2.1 Studi Literatur

Tahapan studi literatur merupakan langkah penting yang dilakukan untuk mengumpulkan berbagai referensi yang relevan dan mendukung penelitian. Proses ini mencakup penelaahan sumber-sumber informasi yang berasal dari beragam media, termasuk buku, jurnal ilmiah, artikel penelitian, serta materi audio-visual seperti video [13]. Penelaahan ini bertujuan untuk memperkaya landasan teori yang menjadi dasar dalam merancang program yang lebih efektif dan inovatif. Dalam konteks penelitian ini, tahap studi literatur bertujuan untuk memperdalam pemahaman mengenai konsep dasar data mining serta penerapannya yang dapat mendukung pengelolaan data di Toko Baju Distro. Maka dari itu, ditujukan hasil penelitian mampu menunjukkan kontribusi yang signifikan untuk pengembangan program secara lebih bermanfaat serta aplikatif.

2.2.2 Pengumpulan Data

Dalam tahapan ini ini, proses pengumpulan data mulai dilaksanakan di Toko Aldi sebagai langkah strategis untuk mendukung jalannya penelitian ini. Data yang berhasil dihimpun berjumlah sebanyak 393 entri, yang dikumpulkan selama periode waktu bulan April hingga Juli tahun 2024. Informasi tersebut diperoleh melalui analisis hasil rekapan penjualan yang dilakukan secara cermat dan sistematis selama dua bulan tersebut, sehingga diharapkan dapat memberikan gambaran yang komprehensif terkait aktivitas penjualan di toko tersebut.

2.2.3 Pre-Processing

Pre-processing data merupakan salah satu langkah dengan sangat penting dan mendasar yang harus dilakukan sebelum data dapat dianalisis menggunakan metode clustering seperti K-Means [14]. Tahapan ini terdapat tujuan yaitu dalam menjamin mengenai data berada dalam keadaan terbaik agar analisis yang dilakukan menghasilkan tingkat akurasi dan validitas yang tinggi. Beberapa langkah utama yang biasanya dilakukan dalam tahap pre-processing data meliputi :

a. Pembersihan Data (Data Cleaning)

Eliminasi data yang tidak lengkap atau hilang, penanganan masalah yang berkaitan dengan duplikasi data, serta penerapan prosedur untuk memastikan konsistensi pada data yang akan digunakan dalam analisis atau pengambilan keputusan.

b. Normalisasi Data

Meningkatkan tingkat akurasi dalam proses pengelompokan data, dilakukan transformasi data ke dalam skala yang seragam melalui penerapan metode seperti min-max scaling atau standardisasi. Proses transformasi ini bertujuan untuk mereduksi ketidakseimbangan yang mungkin terjadi akibat perbedaan skala di antara variabel, sehingga variabel dengan rentang nilai yang lebih besar, seperti harga, tidak secara tidak proporsional memengaruhi hasil pengelompokan dibandingkan variabel lain yang memiliki skala lebih kecil, seperti jumlah atau kuantitas penjualan. Dengan demikian, analisis dapat dilakukan secara lebih adil dan representatif terhadap keseluruhan variabel yang terlibat.

c. Pemilihan Fitur (Feature Selection) :

Dalam rangka memastikan keakuratan dan relevansi hasil dari proses clustering, langkah awal yang sangat penting adalah melakukan seleksi terhadap atribut-atribut yang dianggap relevan. Atribut-atribut ini, seperti kuantitas penjualan dan harga barang, dipilih berdasarkan pertimbangan teoritis maupun empiris untuk mewakili karakteristik utama yang menjadi fokus analisis. Proses seleksi ini bertujuan untuk mengurangi noise atau informasi yang tidak diperlukan, sehingga algoritma clustering dapat bekerja secara lebih efisien dan menghasilkan pengelompokan data yang lebih bermakna.

2.2.4 Implementasi Algoritma

Pada tahap implementasi, dilakukan pemrosesan data dengan memanfaatkan platform Google Colaboratory sebagai alat utama untuk menjalankan algoritma k-means. Dalam proses ini, algoritma k-means tidak hanya diterapkan secara mandiri, tetapi juga dikombinasikan dengan dua teknik evaluasi yang relevan, yaitu Silhouette Coefficient dan Davies-Bouldin Index. Kombinasi kedua teknik tersebut bertujuan untuk mengevaluasi kinerja model clustering secara lebih komprehensif, sehingga memungkinkan penentuan jumlah kluster optimal serta memastikan bahwa pengelompokan yang dihasilkan terdapat kualitas secara tinggi juga sejalan terhadap karakteristik data yang dianalisis.

2.2.5 Evaluasi

1. Silhouette Score

Metode ini mengevaluasi kedekatan data terhadap klasternya serta seberapa jauh kluster-kluster terpisah. Skor Silhouette bernilai dalam kisaran -1 sampai dengan 1, yang mana nilai secara lebih besar membuktikan kualitas pengelompokan secara lebih baik.

$$S(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \dots 1$$

2. Davies-Bouldin Index (DBI)

Indeks ini berfungsi dalam menilai kualitas kluster, di mana nilai DBI secara lebih rendah mengindikasikan bahwa hasil pengelompokan semakin optimal.

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{s_i + s_j}{d(c_i, c_j)} \right) \dots 2$$

III. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dikumpulkan berupa data penjualan baju distro di Toko Aldi Sidoarjo dengan format .csv (microsoft excel) sebanyak 393 baris data dan memiliki 6 atribut utama yang dijadikan variabel untuk mendukung proses prediksi yaitu Tanggal, Status Pembayaran, Type Barang, Quantity, Harga, Jumlah. Berikut merupakan sampel data yang disajikan pada tabel 1 :

Tabel 1 Sampel Data

No	Tanggal	Status Pembayaran	Type Barang	Quantity	Harga	Jumlah
1.	01-Apr-24	Transfer	PU	1	Rp.45.000	Rp.45.000
2.	01-Apr-24	Cash	KU	2	Rp.30.000	Rp.60.000
3.	01-Apr-24	Cash	PH-WL	1	Rp.65.000	Rp.65.000
4.	01-Apr-24	Cash	KU	1	Rp.30.000	Rp.30.000
...
390.	26-Jul-24	Transfer	PH-WP	2	Rp.65.000	Rp.130.000
391.	27-Jul-24	Transfer	PH-WL	1	Rp.65.000	Rp.65.000
392.	28-Jul-24	Transfer	PJU	1	Rp.50.000	Rp.50.000
393.	29-Jul-24	Transfer	PU	1	Rp.45.000	Rp.45.000

Pada tahap selanjutnya pengumpulan data ini di proses dengan menggunakan Google Colaboratory sebagai alat utama untuk menjalankan algoritma k-means. Dalam proses ini, maka data tersebut di lakukan dengan cara kategorikal. Kategorikal merupakan proses mengubah data agar lebih mudah diolah. Karena beberapa atribut yang digunakan berbentuk non-numerik, maka perlu dilakukan konversi data tersebut ke dalam bentuk numerik. Data yang dirubah ke kategorikal yaitu di tunjukkan oleh tabel 2

Tabel 2. Data Dirubah Ke Kategorikal

No.	Nama	Keterangan
1.	Status Pembayaran	Kategorikal {1, 2}
2.	Type Barang	Kategorikal {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}

3.2 Pre-Processing

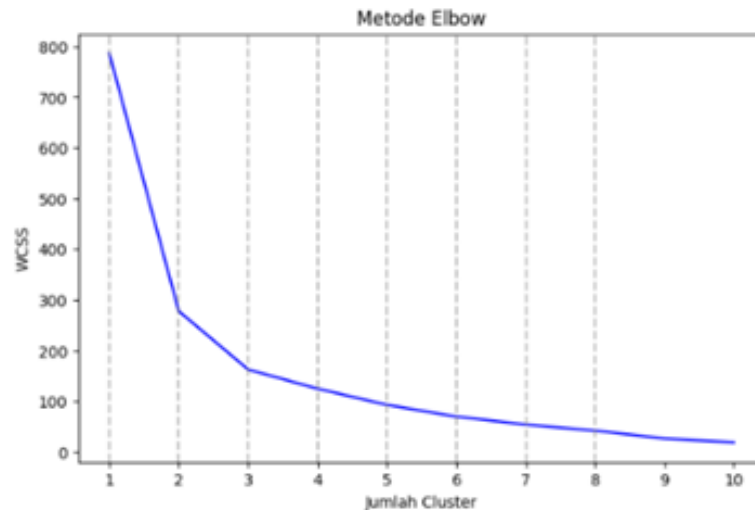
Proses pengolahan data awal dilaksanakan dengan tujuan dalam memperbaiki kualitas data serta membuat data siap untuk diproses lebih lanjut [8]. Tahapan yang dilakukan meliputi pengambilan atribut yang dibutuhkan yaitu atribut “Type Barang”, “Quantity”, dan “Jumlah” dan menghilangkan atribut yang tidak digunakan. Selanjutnya menghilangkan simbol “Rp” dan tanda baca “titik (.)” pada isi atribut biaya sewa per hari dari yang awalnya “Rp. 35.000” menjadi “35000”. Sampel data yang sudah melalui proses Pre-Processing data sebelumnya akan ditunjukkan oleh tabel 3 :

Tabel 3. Data Setelah *Pre-Processing*

	Type Barang	Quantity	Jumlah
1	1	1	45000
2	2	2	60000
3	3	1	65000
4	2	1	30000
...
390.	6	2	130000
391.	3	1	65000
392.	4	1	50000
393.	1	1	45000

3.3 Implementasi Algoritma

K-Means Clustering adalah salah satu metode pada data mining dengan terdapat tujuan untuk mengklasifikasikan data menurut kesamaan karakteristik tertentu. Pada penelitian ini, metode K-Means dimanfaatkan dalam menganalisis serta mengelompokkan pola penjualan baju distro dengan mempertimbangkan atribut seperti jumlah penjualan, kategori produk, dan periode waktu penjualan. Jumlah cluster yang paling optimal ditentukan menggunakan metode Elbow, yang memanfaatkan grafik Within-Cluster Sum of Squares (WCSS) [15]. Dalam grafik tersebut, jumlah cluster optimal dapat diidentifikasi pada titik siku, yang menunjukkan jumlah cluster yang memberikan keseimbangan antara variasi dalam cluster dan kompleksitas model. Berdasarkan hasil analisis, titik siku teridentifikasi pada $k = 2$ atau $k = 3$, yang mengindikasikan jumlah cluster optimal berada dalam rentang tersebut. Berikut hasil perhitungan dari metode elbow ditunjukkan oleh gambar 2 :



Gambar 2. Grafik Metode Elbow

Setelah dilakukan perhitungan menggunakan metode Elbow dengan mengambil sampel cluster dengan range 1-10 cluster terjadi sudut lengkungan pada cluster 2 dan 3. Hal tersebut membuktikan mengenai cluster optimal ada dalam cluster 2 dan 3. Namun, untuk menguji keakuratan cluster perlu menggunakan matriks untuk menghitungnya.

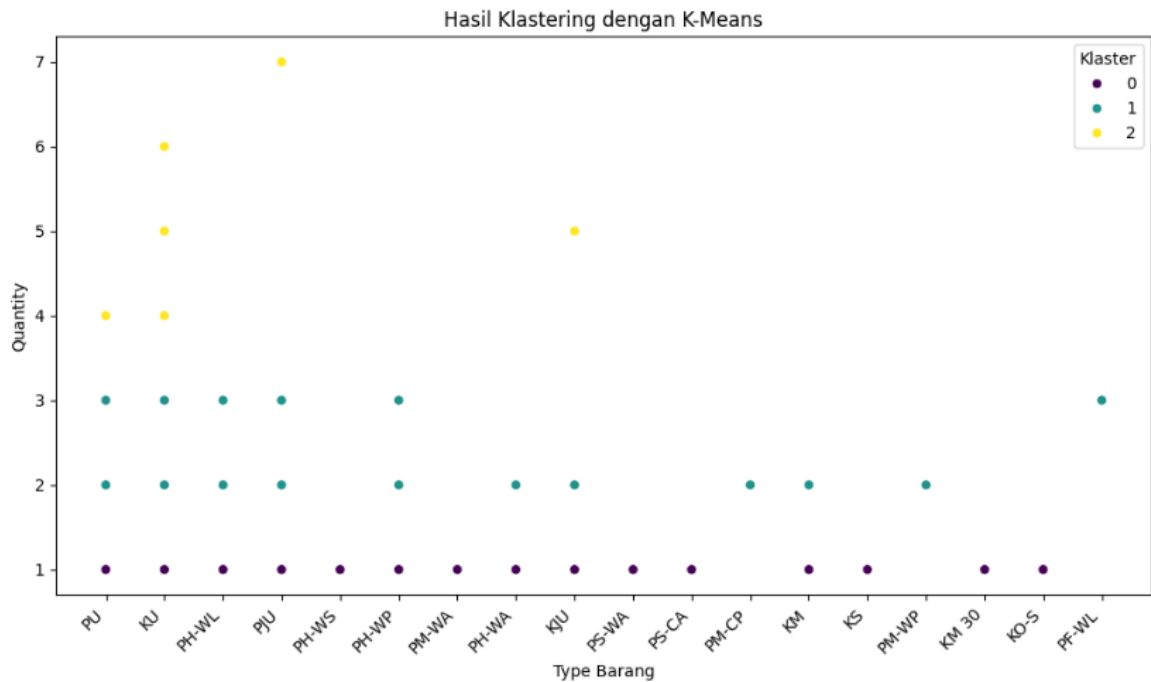
Setelah jumlah cluster optimal ditentukan, algoritma K-Means diterapkan melalui serangkaian tahapan. Proses ini dimulai dengan inisialisasi awal pusat cluster yang ditentukan dengan acak. Selanjutnya, setiap data dihitung jarak Euclidean terhadap pusat cluster yang sudah ditetapkan, selanjutnya dikelompokkan ke dalam cluster terdekat. Setelah itu, pusat cluster diperbarui berdasarkan rata-rata nilai dalam masing-masing cluster. Proses ini berlanjut secara iteratif hingga tidak terjadi perubahan signifikan dalam pusat cluster. Dari hasil pengelompokan, diperoleh tiga kategori utama yang menggambarkan karakteristik penjualan, yaitu produk dalam tingkatan penjualan tinggi produk dalam tingkatan penjualan sedang, serta produk dalam tingkatan penjualan rendah.

Hasil dari penerapan metode ini menunjukkan bahwa produk dengan tingkat penjualan tinggi dapat dijadikan prioritas utama dalam strategi pemasaran, sementara produk dengan tingkat penjualan sedang memerlukan upaya promosi tambahan untuk meningkatkan daya tariknya. Adapun produk dengan tingkat penjualan rendah memerlukan evaluasi lebih lanjut untuk menentukan kelayakan keberlanjutannya atau strategi inovasi yang dapat diterapkan untuk meningkatkan minat konsumen. Wawasan yang dihasilkan dari pengelompokan ini memberikan manfaat signifikan bagi industri distro dalam menyusun strategi pemasaran yang lebih efektif. Dengan pemahaman yang lebih baik terhadap pola penjualan, pemilik bisnis dapat mengalokasikan sumber daya secara lebih efisien, mengoptimalkan strategi promosi, serta meningkatkan daya saing pada pasar secara kompetitif.

Secara keseluruhan, metode K-Means Clustering terbukti sebagai pendekatan yang efektif untuk mengelompokkan pola penjualan baju distro dan memberikan wawasan yang berharga bagi pengambilan keputusan bisnis. Dengan pendekatan berbasis data ini, strategi pemasaran dapat lebih terfokus dan sesuai dengan karakteristik pasar, sehingga meningkatkan peluang keberhasilan dalam industri distro yang semakin dinamis.

3.4 Evaluasi Hasil

Hasil clustering yang divisualisasikan dalam diagram scatter mengindikasikan bahwa data tersegmentasi ke dalam sejumlah kelompok yang dibedakan berdasarkan warna, di mana setiap kelompok merepresentasikan satu klaster yang dibentuk oleh algoritma K-Means..



Gambar 3. Grafik Hasil Clustering

Dari hasil clustering yang ditampilkan pada gambar 3, dapat diidentifikasi bahwa terdapat beberapa kluster dengan karakteristik yang berbeda, mulai dari kluster dengan quantity rendah dan pendapatan rendah, kluster dengan quantity sedang dan pendapatan sedang, hingga kluster dengan quantity tinggi dan pendapatan tinggi. Kluster dengan quantity rendah dan pendapatan rendah menunjukkan produk yang terjual dalam jumlah sedikit dan menghasilkan pendapatan kecil. Kluster dengan quantity sedang dan pendapatan sedang menunjukkan produk dengan tingkat penjualan moderat dan pendapatan yang lebih besar dibandingkan kluster sebelumnya. Kluster dengan quantity tinggi dan pendapatan tinggi menandakan produk unggulan yang memiliki tingkat penjualan tertinggi serta menyumbang pendapatan terbesar bagi distro.

Distribusi data pada grafik juga menunjukkan adanya pola tertentu di mana produk dengan jumlah penjualan rendah cenderung penjualan terendah dalam pengelompokan kluster 0 dengan jumlah quantity barang yang terjual sebanyak 1, sementara produk dengan jumlah penjualan sedang lebih cenderung penjualan terendah dalam pengelompokan kluster 1 dengan jumlah quantity barang yang terjual sebanyak 2-3, sedangkan produk dengan jumlah penjualan tinggi lebih cenderung penjualan terbanyak dalam pengelompokan kluster 2 dengan jumlah quantity barang yang terjual sebanyak 4-7. Berdasarkan hasil clustering, ditemukan beberapa pola penjualan yang dapat dianalisis lebih lanjut. Produk dengan jumlah quantity penjualan rendah namun pendapatan dapat mengindikasikan bahwa produk tersebut memiliki total keseluruhan terjual sebanyak 214 barang dengan total harga keseluruhan Rp.10.267.000, sementara produk dengan jumlah quantity penjualan sedang namun pendapatan dapat mengindikasikan bahwa produk tersebut memiliki total keseluruhan terjual sebanyak 317 barang dengan total harga keseluruhan Rp.13.555.000. Produk dengan jumlah penjualan tinggi namun pendapatan sedang kemungkinan memiliki harga jual yang lebih rendah memiliki total keseluruhan terjual sebanyak 206 barang dengan total harga keseluruhan Rp.13.160.000, namun sangat diminati oleh pelanggan sehingga volume penjualannya tinggi tetapi tidak memberikan margin keuntungan yang besar. Produk dengan jumlah penjualan tinggi dan pendapatan tinggi merupakan produk unggulan yang berkontribusi besar terhadap total pendapatan sehingga dapat dijadikan fokus dalam strategi pemasaran dan produksi agar terus meningkatkan penjualan. Berikut merupakan jumlah anggota untuk setiap cluster ditunjukkan dalam tabel 4 :

Tabel 4. Data Anggota Cluster

c	n	x	nx	hx
0	216	PU : Polo Urgan	77	3.465.000
		KU : Kaos Urgan	43	1.290.000
		PH-WL : Polo Hold/W/List	26	1.690.000
		PJU : Polo Jumbo Urgan	13	650.000
		KJU : Kaos Jumbo Urgan	10	350.000
		PH-WS : Polo Hold/W/Salur	9	657.000
		PH-WA : Polo Hold/W/Apl	8	520.000
		PH-WP : Polo Hold/W/Polos	6	390.000
		PM-WA : Polo Monkl/W/Apl	6	390.000
		PS-WA : Polo Six/W/Apl	4	280.000
		PS-CA : Polo Six/C/Apl	3	195.000
		KM : Kaos Mirror	3	150.000
		KO-S : Kaos Oastok/Slur	3	90.000
		KM 30 : Kaos Mirror 30'S	2	100.000
		KS : Kaos Salur	1	50.000
1	317	KU : Kaos Urgan	138	4.140.000
		PU : Polo Urgan	97	4.365.000
		PJU : Polo Jumbo Urgan	18	900.000
		KJU : Kaos Jumbo Urgan	4	140.000
		PH-WP : Polo Hold/W/Polos	18	1.170.000
		PF-WL : Polo Feold/W/List	9	675.000
		PH-WL : Polo Hold/W/List	13	845.000
		PH-WA : Polo Hold/W/Apl	6	390.000
		PM-WP : Polo Monkl/W/Print	6	438.000
PM-CP : Polo Monkl/C/Polos	4	292.000		
KM : Kaos Mirror	4	200.000		
2	206	KU : Kaos Urgan	152	4.560.000
		PU : Polo Urgan	20	900.000
		KJU : Kaos Jumbo Urgan	20	700.000
		PJU : Polo Jumbo Urgan	14	700.000

Keterangan :

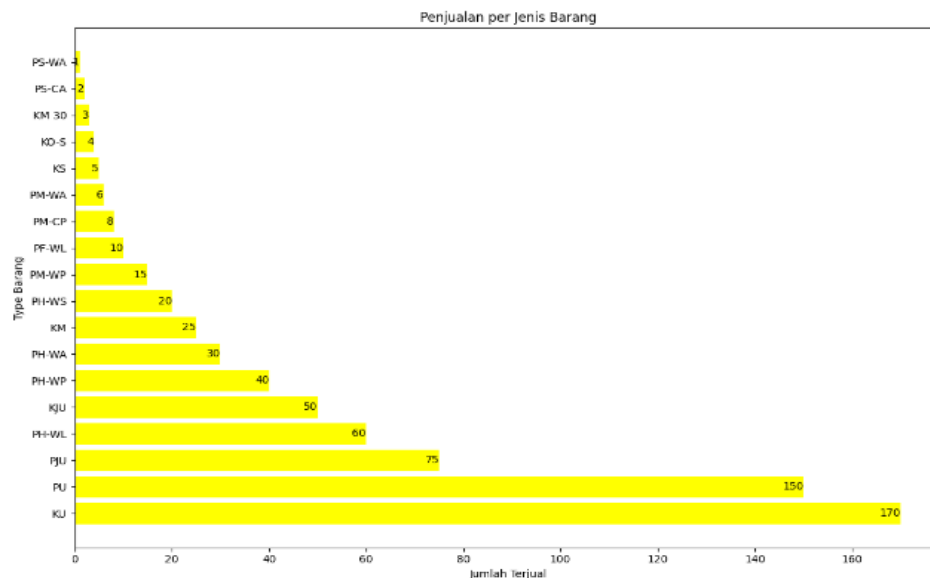
- c : Cluster
- n : Jumlah total terjual keseluruhan type barang
- x : Jenis type barang
- nx : Jumlah total terjual pada masing-masing type barang
- hx : Total harga keseluruhan pada masing-masing type barang

Pada tabel 4 data anggota cluster menunjukkan pembagian data pada masing-masing cluster. Cluster 0 terdapat 216 jumlah total terjual type barang dengan 15 type barang, cluster 1 terdapat 317 jumlah total terjual type barang dengan 11 type barang, dan cluster 2 terdapat 206 jumlah total terjual type barang dengan 4 type barang.

Dari hasil clustering dan analisis pola penjualan, adanya berbagai strategi yang mampu dimanfaatkan dari pihak distro untuk meningkatkan efektivitas penjualan. Optimalisasi stok produk perlu dilakukan dengan memahami kelompok produk dengan terdapat tingkat penjualan tinggi dan rendah maka manajemen mampu mengatur stok dengan lebih efektif agar tidak terjadi overstock atau kehabisan produk yang populer. Strategi harga dan promosi dapat dioptimalkan, di mana produk dalam klaster dengan quantity rendah tetapi pendapatan tinggi dapat didukung dengan strategi pemasaran yang lebih eksklusif, sedangkan produk dalam klaster quantity tinggi dengan pendapatan sedang dapat ditingkatkan profitabilitasnya melalui promosi diskon atau bundling. Diversifikasi produk menjadi salah satu langkah penting untuk mengurangi ketergantungan pada beberapa produk unggulan saja, sehingga distribusi penjualan lebih merata dan risiko bisnis dapat diminimalisir.

Hasil dari penerapan K-Means Clustering dalam penelitian ini menunjukkan bahwa metode ini efektif dalam mengelompokkan produk berdasarkan pola penjualan. Dengan memahami karakteristik setiap klaster, pihak distro dapat merancang strategi penjualan yang lebih akurat dan berbasis data sehingga meningkatkan efisiensi

operasional dan profitabilitas bisnis secara keseluruhan. Penerapan metode ini juga memberikan wawasan yang lebih dalam terkait perilaku pelanggan dan preferensi produk, yang mampu dimanfaatkan menjadi landasan ketika pengambilan keputusan bisnis di masa mendatang.



Gambar 4. Tabel Jumlah Penjualan Per Jenis Barang

Analisis data penjualan merupakan tahap yang sangat krusial dalam menentukan pola serta tren yang terjadi di dalam pasar. Berdasarkan visualisasi data yang disajikan dalam diagram batang horizontal, dapat diidentifikasi perbedaan signifikan dalam jumlah penjualan tiap kategori produk. Produk dengan type baju KU, PU, dan PJU menunjukkan tingkat penjualan yang jauh lebih tinggi (laku) dengan jumlah terjual sebesar 170, 150 dan 75 barang, kemudian dengan type baju PH-WL, KJU, PH-WP, dan PH-WA tingkat penjualan menengah (cukup laku) dengan jumlah terjual sebesar 60, 50, 40 dan 30 barang. Sebaliknya, beberapa kategori barang seperti KS, KO-S, KM 30, PS-CA dan PS-WA memiliki jumlah penjualan yang relatif rendah (kurang laku), dengan angka penjualan yang berkisar antara 1 hingga 5 barang. Disparitas ini mengindikasikan bahwa terdapat produk tertentu yang lebih diminati oleh konsumen, sementara produk lainnya mengalami permintaan yang lebih rendah.

Dalam penelitian ini, metode K-Means Clustering dimanfaatkan untuk mengklasifikasikan produk berdasarkan tingkat penjualannya guna memperoleh pola yang lebih jelas mengenai distribusi permintaan. Metode ini bekerja dengan mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan pola penjualan, dengan demikian dapat membantu untuk merumuskan strategi bisnis yang lebih tepat sasaran. Berdasarkan hasil clustering, produk dengan tingkat penjualan yang tinggi dikelompokkan dalam satu cluster sebagai kategori dengan terdapat daya tarik pasar yang besar. Produk dengan tingkat penjualan sedang dikelompokkan dalam cluster lain yang mencerminkan kategori dengan tingkat penerimaan pasar yang stabil, sedangkan produk dengan tingkat penjualan rendah dikelompokkan dalam cluster yang mencerminkan kategori dengan daya saing yang lemah di pasar.

Penerapan metode ini memberikan wawasan mendalam terhadap strategi pengelolaan bisnis yang dapat dioptimalkan berdasarkan pola yang telah ditemukan. Salah satu implikasi utama dari hasil clustering ini adalah optimalisasi stok barang. Produk dengan tingkat penjualan tinggi perlu diprioritaskan dalam produksi dan distribusi guna memenuhi permintaan pasar, sementara produk dengan tingkat penjualan rendah perlu dievaluasi ulang terkait kelayakan produksinya. Dengan demikian, bisnis dapat menghindari terjadinya kelebihan stok pada produk yang kurang diminati dan memastikan ketersediaan produk yang memiliki permintaan tinggi.

Berdasarkan hasil analisis, algoritma K-Means berhasil membentuk dua klaster utama. Evaluasi terhadap kualitas klaster dilakukan menggunakan metrik Silhouette Coefficient, Calinski-Harabasz Index, dan Davies-Bouldin Index untuk menilai sejauh mana hasil klasterisasi memberikan pemisahan dan kohesi yang optimal.

3.5 Hasil Klasterisasi

Hasil evaluasi menunjukkan bahwa nilai Silhouette Coefficient sebesar 0.576, yang menunjukkan bahwa klaster yang terbentuk memiliki kualitas yang sangat baik. Yang mengindikasikan bahwa data dalam satu klaster memiliki kemiripan yang tinggi, sementara antar klaster memiliki perbedaan yang signifikan. Nilai Calinski-Harabasz Index sebesar 19.125 mengindikasikan bahwa rasio antara kepadatan intra-klaster dan keterpisahan antar-klaster cukup tinggi, sehingga mendukung bahwa hasil klasterisasi ini memiliki struktur yang jelas.

Sementara itu, nilai Davies-Bouldin Index sebesar 0.308 menunjukkan bahwa rata-rata rasio kesalahan antara jarak antar-klaster dan penyebaran dalam satu klaster sangat kecil, yang menandakan klaster yang terbentuk memiliki batasan yang jelas dan tidak terjadi banyak tumpang tindih antar kelompok.

Berdasarkan hasil ini, dapat disimpulkan bahwa segmentasi penjualan baju distro menggunakan K-Means Clustering memberikan hasil yang sangat baik dalam mengelompokkan pola pembelian pelanggan. Keberhasilan algoritma ini dalam membentuk klaster yang optimal memiliki implikasi yang signifikan terhadap strategi bisnis, terutama dalam hal personalisasi strategi pemasaran dan optimasi manajemen stok. Dengan mengetahui karakteristik masing-masing klaster, pelaku usaha dapat menyusun strategi pemasaran yang lebih efektif dengan menargetkan pelanggan sesuai dengan pola pembelian yang telah teridentifikasi. Tidak hanya itu, hasil klasterisasi tersebut mampu dimanfaatkan untuk optimalisasi inventaris, di mana stok barang dapat disesuaikan dengan kebutuhan berdasarkan segmentasi pelanggan yang telah terbentuk, sehingga dapat mengurangi risiko overstock maupun understock.

Lebih lanjut, hasil klasterisasi juga dapat digunakan sebagai dasar dalam analisis perilaku pelanggan, yang memungkinkan peritel untuk memahami preferensi pelanggan secara lebih mendalam. Informasi ini dapat dimanfaatkan untuk merancang program loyalitas yang lebih tepat sasaran, meningkatkan retensi pelanggan, serta mengidentifikasi peluang untuk memperluas pasar berdasarkan karakteristik demografi dan pola pembelian dari masing-masing klaster. Dengan demikian, penerapan K-Means Clustering dalam prediksi penjualan baju distro bukan hanya berfungsi sebagai alat analisis data, tetapi juga sebagai dasar dalam pengambilan keputusan strategis yang lebih berbasis data dalam industri fashion.

```

→ Estimated number of clusters: 3
Silhouette Coefficient: 0.576
Calinski_harabasz Index: 19.125
Davies-Bouldin Index: 0.308

```

Gambar 5. Hasil Silhoutte dan DBI

Menurut hasil serta pembahasan yang sudah diuraikan, mampu diperoleh kesimpulan mengenai metode K-Means Clustering mampu mengelompokkan pola penjualan baju distro dengan sangat baik. Dengan jumlah klaster optimal sebanyak dua, serta hasil evaluasi yang menunjukkan kualitas klaster yang sangat baik berdasarkan nilai Silhouette Coefficient (0.576), Calinski-Harabasz Index (19.125), dan Davies-Bouldin Index (0.308), metode ini dapat dijadikan sebagai alat analisis yang efektif dalam memahami tren penjualan dan perilaku pelanggan. Dengan demikian, penelitian ini memberikan kontribusi dalam penerapan data mining dalam industri fashion, khususnya dalam upaya optimalisasi strategi pemasaran dan manajemen stok yang lebih berbasis data dan prediktif.

V. SIMPULAN

Berdasarkan penelitian yang dilakukan, metode K-Means Clustering terbukti mampu mengelompokkan pola penjualan baju distro di Toko Aldi secara efektif berdasarkan tingkat penjualannya. Dari 393 transaksi yang dianalisis, hasil klasterisasi mengidentifikasi tiga kategori utama, yakni produk dengan tingkat penjualan tinggi (laku) sebanyak 170, 150 dan 75 barang, tingkat penjualan menengah (cukup laku) sebanyak 60, 50, 40 dan 30 barang, dan tingkat penjualan rendah (kurang laku) sebanyak 1 hingga 5 barang. Evaluasi menggunakan Silhouette Coefficient (0.576), Calinski-Harabasz Index (19.125), dan Davies-Bouldin Index (0.308) membuktikan mengenai klaster yang terbentuk terdapat tingkat kohesi secara baik serta pemisahan yang jelas. Temuan ini memberikan wawasan bagi pemilik usaha dalam merancang strategi pemasaran yang lebih optimal serta meningkatkan efisiensi pengelolaan stok, sehingga dapat meningkatkan kinerja operasional dan daya saing bisnis di industri distro.

REFERENSI

- [1] R. G. Solechati and A. Jananto, "Penerapan Algoritma K-Means Clustering Pada Data Brain Stroke Untuk Pengelompokan Profile Pasien," *semantik*, vol. 9, no. 1, p. 39, 2023, doi: 10.55679/semantik.v9i1.29446.
- [2] R. Maulana, "OPTIMISASI PENGGUNAAN ALGORITMA MACHINE LEARNING," vol. 1, no. 6, pp. 1–16, 2024.
- [3] C. Purnama, W. Witanti, and P. Nurul Sabrina, "Klasterisasi Penjualan Pakaian untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," *J. Inf. Technol.*, vol. 4, no. 1, pp. 35–38, 2022, doi: 10.47292/joint.v4i1.79.
- [4] A. T. Suseno, A. R. Naufal, and M. Al Amin, "Market Based Analysis Sebagai Peningkatan Penjualan Produk

- Menggunakan Algoritma K-Medoids Dan Fp-Growth,” *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 301, 2022, doi: 10.37600/tekinkom.v5i2.646.
- [5] P. Putra, “PENGEMBANGAN MODEL PREDIKSI RISIKO KREDIT,” vol. 1, no. 6, pp. 1–18, 2024.
- [6] O. B. Ginting, A. Anita, and E. Y. Tumanggor, “Penerapan Metode Trend Moment Untuk Memprediksi Jumlah Penjualan Dan Stok Kopi Pada Omilen Coffee,” *J. Tekinkom (Teknik Inf. dan Komputer)*, vol. 7, no. 1, pp. 395–401, 2024, doi: 10.37600/tekinkom.v7i1.1332.
- [7] I. B. Perkasa and I. Komputer, “STRATEGI DATA MINING UNTUK IDENTIFIKASI POLA,” vol. 1, no. 6, pp. 1–16, 2024.
- [8] R. Dwi Putra, “Klasifikasi Penjualan Produk Customer Relationship Management dengan Algoritma K-Nearest Neighbors,” *J. Comput. Scine Inf. Technol.*, vol. 8, pp. 48–55, 2022, doi: 10.35134/jcsitech.v8i2.34.
- [9] Ismai, “Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur Dengan Algoritma Naive Bayes Pada Sistem JURAGAN Berbasis Komunitas,” vol. 03, no. 01, pp. 24–31, 2020.
- [10] N. Farida, M. T. Chulkamdi, and Z. Wulansari, “Application of Data Mining By Using a Priori Algorithm To Improve Customer Purchasing Decisions At Mikamart Blitar Store,” *Int. J. Multidiscip. Res. Lit.*, vol. 1, no. 5, pp. 526–534, 2022, doi: 10.53067/ijomral.v1i5.58.
- [11] R. Komansilan, V. Tarigan, and A. Yusupa, “Analisis Perbandingan Metode Trend Moment dan Regresi Linear Untuk Meramal Harga Saham Bank BRI,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 7, no. 1, p. 24, 2024, doi: 10.53513/jsk.v7i1.9474.
- [12] L. P. Dalova, Nurmawanti, N. E. Faizah, and S. B. Syahputro, “Efektifitas Penerapan Customer Relationship Management Pada Usaha Jasa Desain Iklan Citra Karya Setia (Advertising & Digital Printing) Melalui Pemasaran Electronic Word of Mouth (E-Wom),” *Neraca Manajemen, Akunt. Ekon.*, vol. 1, no. 3, pp. 1–17, 2023.
- [13] A. Nugraha, O. Nurdiawan, and G. Dwilestari, “PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT,” 2022.
- [14] N. Luh, P. P. Dewi, I. Nyoman Purnama, and N. W. Utami, “Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara),” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 16, no. 2, 2022.
- [15] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, I. Setiawan Mangku Negara, I. Ahmad Ashari, P. Studi Teknologi Informasi, F. Sains dan Teknologi, and U. Harapan Bangsa, “Terakreditasi SINTA Peringkat 4 Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means,” 2021.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.