

Analisis Perilaku Belanja Konsumen UMKM Pondok Gede Menggunakan Algoritma K-Means Clustering

Alfriska Fefviana Sigit Saputra^{1*}, Satriyo Wicaksono², Aughestin Neizza Frisenta³, Ferdian Ady Sudrajat⁴

¹Teknik Informatika/Fakultas
Ilmu Komputer (Universitas Duta
Bangsa Surakarta)

^{1*}230103087@mhs.udb.ac.id

²Teknik Informatika/Fakultas
Ilmu Komputer (Universitas Duta
Bangsa Surakarta)

²230103118@mhs.udb.ac.id

³Teknik Informatika/Fakultas
Ilmu Komputer (Universitas Duta
Bangsa Surakarta)

³230103093@mhs.udb.ac.id

⁴Teknik Informatika/Fakultas
Ilmu Komputer (Universitas Duta
Bangsa Surakarta)

⁴230103101@mhs.udb.ac.id

Abstrak— Penelitian ini bertujuan untuk menganalisis perilaku belanja konsumen pada UMKM Pondok Gede menggunakan algoritma K-Means Clustering berdasarkan data transaksi penjualan. Analisis ini dilakukan untuk mengidentifikasi pola transaksi konsumen sehingga dapat menjadi dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran. Dataset yang digunakan terdiri atas 7.233 data transaksi dengan atribut Jumlah Terjual, Harga Satuan, Total Penjualan, Kategori Produk, Metode Pembayaran, dan Channel Penjualan. Tahapan penelitian meliputi proses data preprocessing yang mencakup pembersihan data, penanganan nilai yang tidak valid, transformasi data kategorikal, serta normalisasi menggunakan metode Min-Max Scaling agar seluruh atribut berada pada skala yang sama. Selanjutnya, jumlah cluster optimal ditentukan menggunakan Elbow Method dan Silhouette Score sebelum dilakukan proses pengelompokan menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster dengan nilai Silhouette Score sebesar 0,5143, yang mengindikasikan kualitas pengelompokan berada pada kategori cukup baik. Setiap cluster memiliki karakteristik transaksi yang berbeda berdasarkan nilai pembelian, jenis produk, metode pembayaran, dan saluran penjualan. Hasil pengelompokan menghasilkan tiga segmen konsumen yang terdiri atas Cluster 0 sebanyak 1.555 transaksi, Cluster 1 sebanyak 2.939 transaksi, dan Cluster 2 sebanyak 2.739 transaksi. Masing-masing cluster menunjukkan karakteristik transaksi yang berbeda berdasarkan jumlah pembelian, nilai transaksi, kategori produk, metode pembayaran, dan channel penjualan sehingga dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran.

Kata kunci— Data Mining, K-Means Clustering, UMKM, Perilaku Konsumen, Segmentasi Konsumen.

Abstract— This study aims to analyze consumer shopping behavior at Pondok Gede UMKM using the K-Means Clustering algorithm based on transaction data. The analysis was conducted to identify consumer transaction patterns that can serve as a basis for developing more targeted marketing strategies. The dataset consisted of 7,233 transaction records with the attributes of Quantity Sold, Unit Price, Total Sales, Product Category, Payment Method, and Sales Channel. The research process included data preprocessing, which involved data cleaning, handling invalid values, transforming categorical data, and normalizing the dataset using the Min-Max Scaling method to ensure that all attributes were on the same scale. Furthermore, the optimal number of clusters was determined using the Elbow Method and Silhouette Score before applying the K-Means clustering algorithm. The results showed that the optimal number of clusters was three, with a Silhouette Score of 0.5143, indicating a moderately good clustering quality. Each cluster exhibited distinct transaction characteristics based on purchase value, product category, payment method, and sales channel. The clustering process resulted in three consumer segments, consisting of 1,555 transactions in Cluster 0, 2,939 transactions in Cluster 1, and 2,739 transactions in Cluster 2. Each cluster exhibited distinct transaction characteristics in terms of purchase quantity, transaction value, product category, payment method, and sales channel, providing a basis for developing more targeted marketing strategies.

Keywords— Data Mining, K-Means Clustering, UMKM, Consumer Behavior, Consumer Segmentation.

I. PENDAHULUAN

Perilaku konsumen merupakan salah satu faktor penting yang perlu dipahami oleh pelaku usaha dalam menentukan strategi pemasaran dan meningkatkan kepuasan pelanggan. Pelanggan merupakan kunci sukses dalam bisnis retail.

Pembagian atau pengelompokan pelanggan menjadi dasar dari analisis kebutuhan dari berbagai kebutuhan yang berbeda[1].

Berdasarkan hasil sensus ekonomi BPS tahun 2016, jumlah UMKM di Kota Bekasi cukup banyak yaitu kurang lebih 203,000 unit UMKM,

sebanyak 12,186 merupakan jenis usaha perdagangan pengolahan makanan dan minuman dengan serapan tenaga kerja sebesar 410,000 orang pekerja pada tahun 2016. Jumlah UMKM yang sudah terdaftar sebagai binaan Dinas Koperasi dan Usaha Kecil Menengah Kota Bekasi sebanyak 2,186 UMKM dari 12,988 UMKM dan mampu menyerap tenaga kerja sebanyak 6,558 orang. Adapun jenis UMKM yang akan diteliti adalah UMKM kategori pengolahan makanan di Kota Bekasi.

Salah satu tantangan yang dihadapi UMKM adalah memahami karakteristik konsumennya berdasarkan data transaksi yang dimiliki. Banyak UMKM masih mengalami kendala dalam menganalisis pola perilaku konsumen serta dalam mengelompokkan produk secara efektif[2]. Beberapa penelitian sebelumnya telah menerapkan metode clustering, khususnya algoritma K-Means, dalam berbagai konteks seperti evaluasi pembelajaran[3]. Hasil penelitian tersebut menunjukkan bahwa algoritma K-Means Efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik, dan dapat dievaluasi menggunakan metrik seperti Silhouette Score, Sum of Squared Error (SSE), serta David Bouldin Index (DBI)[4].

Kemampuan untuk memperkirakan volume penjualan secara akurat pada setiap item yang dijual di toko sangat penting guna kelangsungan hidup pada bisnis[5]. Oleh karena itu, penelitian ini bertujuan untuk menganalisis perilaku belanja konsumen pada UMKM Pondok Gede menggunakan algoritma K-Means Clustering berdasarkan data transaksi yang meliputi total transaksi, jumlah transaksi, kategori barang, dan metode pembayaran. Hasil pengelompokan yang diperoleh diharapkan dapat membantu mengidentifikasi karakteristik setiap kelompok konsumen sehingga dapat menjadi dasar dalam penyusunan strategi pemasaran yang lebih efektif dan tepat sasaran.

II. METODOLOGI PENELITIAN

Metode pendekatan yang digunakan dalam penelitian ini adalah kuantitatif. Metode kuantitatif adalah suatu pendekatan penelitian yang menggunakan angka dan statistik dalam

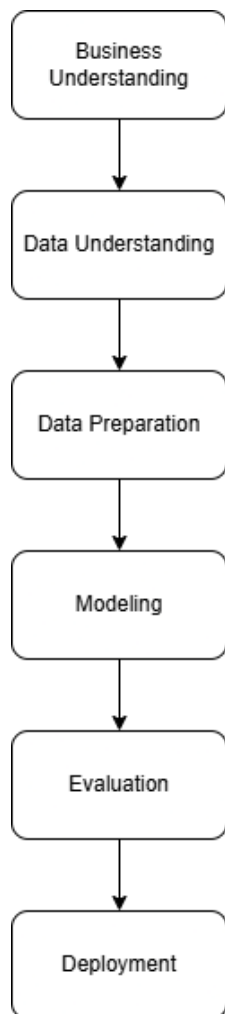
pengumpulan serta analisis data yang dapat diukur[6]. Dalam konteks ini, penelitian menggunakan pendekatan kuantitatif eksploratif untuk mengidentifikasi pola perilaku belanja konsumen berdasarkan data transaksi. Metode yang diterapkan adalah data mining dengan teknik unsupervised learning melalui algoritma K-Means Clustering, yang digunakan untuk mengelompokkan konsumen berdasarkan tingkat kemiripan karakteristik transaksi yang dimiliki.

Pendekatan ini memungkinkan pengelompokan produk secara otomatis berdasarkan kesamaan pola penjualannya tanpa memerlukan label data sebelumnya, sehingga cocok untuk membantu UMKM dalam menganalisis dan memahami segmentasi produk mereka secara lebih efektif. Melalui pendekatan ini, mereka dapat menggunakan informasi yang diperoleh untuk merancang strategi pengembangan UMKM yang lebih efektif, menyesuaikan alokasi sumber daya, serta merumuskan kebijakan yang lebih tepat[7].

Penerapan algoritma K-Means melibatkan pengelompokan data transaksi penjualan berdasarkan kesamaan karakteristik[8]. Algoritma ini menentukan sejumlah *centroid* secara acak, lalu secara iteratif mengukur jarak terdekat setiap data transaksi ke pusat kluster hingga mencapai kondisi stabil (*converged*). Proses ini memastikan terbentuknya kelompok dengan kemiripan perilaku belanja yang tinggi di dalam kluster (*high intra-cluster similarity*) dan perbedaan yang jelas antar-kluster (*low inter-cluster similarity*), sehingga menghasilkan segmentasi data mentah yang akurat dan objektif.

Penelitian ini menggunakan metodologi CRISP-DM sebagai kerangka kerja dalam proses data mining. Metodologi ini dipilih karena menyediakan tahapan yang sistematis, mulai dari pemahaman masalah bisnis hingga evaluasi hasil, sehingga mampu mendukung proses penggalian informasi dari data secara terstruktur. Dalam penelitian ini, pendekatan analisis cluster diterapkan untuk mengelompokkan konsumen berdasarkan karakteristik transaksi yang terdapat pada Dataset Penjualan UMKM Pondok Gede. Data transaksi yang meliputi total penjualan, jumlah transaksi, kategori produk, dan metode pembayaran diolah menjadi informasi yang dapat digunakan untuk

mengidentifikasi pola perilaku belanja konsumen. Informasi tersebut diharapkan dapat memberikan wawasan yang bermanfaat dalam mendukung pengambilan keputusan bisnis yang lebih tepat sasaran [9]. Penelitian ini terdiri atas enam tahapan sesuai dengan metodologi CRISP-DM, yang dijelaskan pada Gambar 1.



Gambar 1. Tahapan Penelitian Crisp-DM

A. Pemahaman Bisnis (*Business Understanding*)

Tahap pemahaman bisnis bertujuan untuk mengidentifikasi permasalahan yang dihadapi pelaku UMKM dalam menganalisis data transaksi penjualan. Banyak pelaku UMKM masih mengalami kesulitan dalam mengetahui karakteristik transaksi yang terjadi sehingga strategi pemasaran dan pengelolaan produk sering kali belum dilakukan berdasarkan data. Oleh karena itu, penelitian ini

bertujuan menerapkan algoritma K-Means Clustering untuk mengelompokkan transaksi berdasarkan karakteristik penjualannya. Algoritma K-Means sendiri memiliki mekanisme yang secara berulang mengelompokkan dataset menjadi beberapa kluster yang telah ditentukan pada awal proses, serta dapat dikustomisasi dengan relatif cepat dan mudah [10]. Melalui karakteristik tersebut, hasil pengelompokan diharapkan dapat membantu pelaku UMKM memahami pola transaksi secara efisien dan mendukung pengambilan keputusan dalam menentukan strategi penjualan yang lebih tepat.

B. Pemahaman Data (*Data Understanding*)

Tahap ini dilakukan untuk memahami karakteristik *dataset* yang digunakan dalam penelitian. Data yang digunakan merupakan data sekunder yang diperoleh dari Dataset Penjualan UMKM Pondok Gede yang tersedia pada platform Kaggle[11]. Dataset terdiri dari 7.233 data transaksi dengan atribut Invoice_ID, Tanggal, Nama_Produk, Kategori, Jumlah_Terjual, Harga_Satuan, Total_Penjualan, Metode_Pembayaran, Channel_Penjualan, Nama_Pelanggan, dan Wilayah. Dari seluruh atribut tersebut, atribut Jumlah_Terjual dan Harga_Satuan digunakan sebagai variabel utama dalam proses clustering, sedangkan atribut lainnya dimanfaatkan untuk menganalisis karakteristik setiap cluster yang terbentuk.

C. Persiapan Data (*Data Preparation*)

Pada tahap persiapan data dilakukan pemeriksaan terhadap dataset transaksi UMKM yang berjumlah 7.233 data. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki nilai kosong maupun data duplikat sehingga seluruh data dapat digunakan pada proses clustering. Selanjutnya dilakukan seleksi atribut dengan menggunakan atribut Jumlah_Terjual, Harga_Satuan, Kategori, Metode_Pembayaran, dan Channel_Penjualan. Atribut Harga_Satuan kemudian ditransformasikan menggunakan log transform, atribut kategorikal dikonversi menggunakan One-Hot Encoding, sedangkan atribut numerik dinormalisasi menggunakan Min-Max Scaling agar seluruh fitur memiliki rentang nilai yang sama dan tidak mendominasi proses clustering[12].

Tabel 1. Hasil Pemeriksaan Kualitas Dataset

Pemeriksaan	Hasil
Jumlah Data	7233
Jumlah Atribut	11
Missing Value	0
Data Duplikat	0

D. Pemodelan (Modeling)

Tahap pemodelan dilakukan menggunakan algoritma K-Means Clustering. Sebelum proses clustering dilakukan, jumlah cluster optimal ditentukan menggunakan Elbow Method dan divalidasi menggunakan Silhouette Score.

Persamaan Fungsi Objektif K-Means:

$$J = \sum_{i=1}^K \sum_{x \in S_i} \|x - \mu_i\|^2$$

Keterangan:

J = nilai fungsi objektif K-Means

K = jumlah cluster

S_i = himpunan data pada cluster ke- i

x = data ke- x

μ_i = centroid cluster ke- i

Berdasarkan hasil pengujian menggunakan Elbow Method dan Silhouette Score, diperoleh jumlah cluster optimal sebanyak 3 cluster ($K = 3$). Selanjutnya algoritma K-Means diterapkan pada data hasil preprocessing hingga menghasilkan tiga kelompok transaksi berdasarkan kemiripan karakteristik data.

Rumus Silhouette Score:

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

Keterangan:

$S(i)$ = nilai Silhouette data ke- i

$a(i)$ = rata-rata jarak data ke anggota cluster yang sama

$b(i)$ = rata-rata jarak data ke cluster terdekat

E. Evaluasi (Evaluation)

Tahap evaluasi bertujuan untuk menilai kualitas hasil clustering yang diperoleh. Pengujian data dilakukan menggunakan metode Silhouette serta

Elbow Method untuk melihat seberapa baiknya clustering yang sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data[13]. Pengujian ini mengukur tingkat kemiripan data dalam satu cluster dibandingkan dengan cluster lainnya, di mana nilai yang diperoleh digunakan sebagai dasar dalam menentukan jumlah cluster terbaik yang optimal [14]. Selain itu, hasil clustering dianalisis berdasarkan rata-rata atribut Jumlah_Terjual dan Harga_Satuan, serta distribusi Kategori, Metode_Pembayaran, dan Channel_Penjualan pada setiap cluster untuk mengetahui karakteristik masing-masing kelompok transaksi.

F. Penerapan (Deployment)

Pada era digital, pengelolaan data yang efisien menjadi kunci pengambilan keputusan yang tepat[15]. Sejalan dengan hal tersebut, tahap deployment dilakukan dengan mengimplementasikan hasil clustering ke dalam aplikasi berbasis web sebagai media visualisasi hasil analisis. Aplikasi ini memungkinkan pengguna memilih rentang waktu transaksi yang ingin dianalisis, kemudian sistem akan menampilkan hasil pengelompokan beserta karakteristik setiap cluster dalam bentuk tabel dan grafik. Implementasi ini diharapkan dapat membantu pelaku UMKM dalam memahami pola transaksi penjualan secara efisien, serta mendukung pengambilan keputusan strategis yang lebih tepat sasaran berdasarkan hasil analisis clustering.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi setiap tahapan dalam metodologi CRISP-DM, yang meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Setiap tahapan menghasilkan proses pengelompokan data transaksi UMKM menggunakan algoritma K-Means, mulai dari penentuan jumlah cluster optimal, pembentukan cluster, analisis karakteristik setiap cluster, hingga implementasi hasil clustering ke dalam aplikasi berbasis web sebagai media visualisasi dan analisis data.

Tabel 2. Nilai Centroid Setiap Cluster

Cluster	Jumlah Terjual	Harga Sat	Total Penjualan
0	10.00	21110.93	211109.34
1	5.52	21054.68	115926.27
2	1.00	21068.03	21068.03

Berdasarkan nilai centroid, Cluster 0 memiliki rata-rata jumlah transaksi tertinggi, Cluster 1 didominasi oleh transaksi dengan harga satuan tertinggi, sedangkan Cluster 2 memiliki rata-rata jumlah transaksi dan total penjualan yang paling rendah. Perbedaan nilai centroid menunjukkan bahwa setiap cluster memiliki karakteristik transaksi yang berbeda sehingga dapat digunakan sebagai dasar segmentasi konsumen.

A. Hasil Pemahaman Bisnis (Business Understanding)

Pada tahap ini ditetapkan tujuan utama penelitian, yaitu mengelompokkan transaksi penjualan UMKM berdasarkan karakteristik transaksi sehingga dapat memberikan informasi mengenai pola penjualan pada setiap kelompok. Hasil clustering diharapkan dapat membantu pelaku UMKM dalam memahami karakteristik transaksi serta mendukung pengambilan keputusan terkait strategi pemasaran dan pengelolaan produk.

B. Hasil Pemahaman (Data Understanding)

Tahap ini menghasilkan informasi mengenai karakteristik dataset yang digunakan dalam penelitian. Dataset terdiri dari 7.233 data transaksi dengan atribut Invoice_ID, Tanggal, Nama_Produk, Kategori, Jumlah_Terjual, Harga_Satuan, Total_Penjualan, Metode_Pembayaran, Channel_Penjualan, Nama_Pelanggan, dan Wilayah. Berdasarkan eksplorasi data diketahui bahwa atribut Jumlah_Terjual dan Harga_Satuan memiliki rentang nilai yang berbeda sehingga diperlukan proses normalisasi sebelum dilakukan clustering.

C. Hasil Persiapan Data (Data Preparation)

Pada tahap persiapan data dilakukan proses data cleaning untuk memastikan data tidak mengandung nilai kosong maupun data duplikat. Selanjutnya, atribut Harga_Satuan ditransformasikan menggunakan log transform, atribut kategorikal dikonversi dengan One-Hot Encoding, dan atribut

numerik dinormalisasi menggunakan QuantileTransformer. Hasil preprocessing tersebut kemudian digunakan sebagai masukan pada proses clustering menggunakan algoritma K-Means.

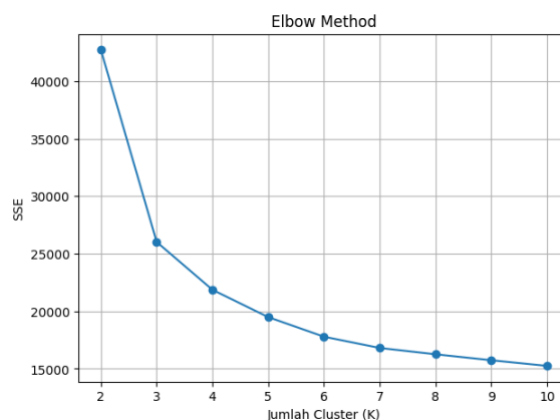
Tahap berikutnya dilakukan transformasi logaritmik pada atribut numerik untuk mengurangi kemencengan distribusi data, kemudian dilakukan normalisasi menggunakan logaritmik sehingga data numerik berada pada skala yang lebih seragam. Contoh data hasil preprocessing ditunjukkan pada Tabel 3.

Tabel 3. Contoh Hasil Data PreProcessing

No	Jmlh Trjual	Hrg Sat	Total	Kategori
1	0.214	-1.128	9.632	Makanan
2	0.357	-1.386	10.273	Snack
3	0.214	-1.023	9.695	Minuman
4	0.928	0.842	12.321	Minuman
5	0.928	0.635	12.259	Minuman

D. Hasil Pemodelan Data (Modeling)

Tahap pemodelan dilakukan menggunakan algoritma K-Means Clustering. Sebelum proses clustering dilakukan, jumlah cluster optimal ditentukan menggunakan Elbow Method dan Silhouette Score.



Gambar 2. Elbow Method

Berdasarkan hasil Elbow Method, penurunan nilai SSE mulai melandai pada $K = 3$, sehingga jumlah cluster yang digunakan pada penelitian ini adalah tiga cluster.

E. Hasil Evaluasi (Evaluation)

Tabel 4. Hasil Silhouette Score

Jumlah Cluster (K)	Silhouette Score
2	0.4950
3	0.5143
4	0.2516
5	0.2444
6	0.2743
7	0.2598
8	0.2391
9	0.2112
10	0.1769

Berdasarkan nilai Silhouette Score, nilai tertinggi diperoleh pada K = 3 sebesar 0,5143, sehingga jumlah cluster tersebut dipilih sebagai jumlah cluster optimal.

Tabel 4. Pengelompokan Data Hasil Clustering

Cluster	Jumlah Transaksi	Karakteristik
0	1555	Harga satuan tinggi dan total penjualan tinggi
1	2939	Jumlah terjual tinggi dengan total penjualan tinggi
2	2739	Jumlah terjual dan total penjualan relatif rendah

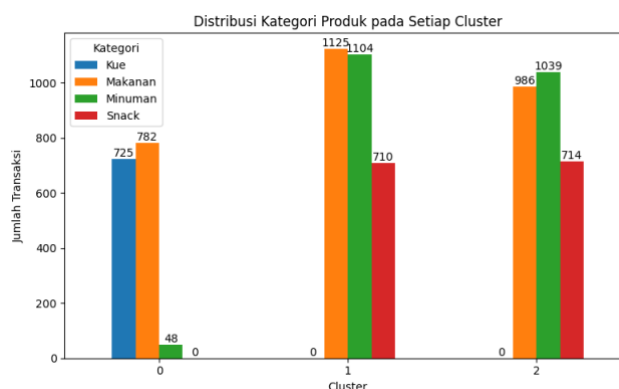
- Cluster 0 memiliki harga satuan transaksi paling tinggi.
- Cluster 1 memiliki jumlah transaksi paling tinggi.
- Cluster 2 memiliki jumlah transaksi paling sedikit.

Karakteristik masing-masing cluster dianalisis berdasarkan nilai rata-rata atribut Jumlah_Terjual, Harga_Satuan, dan Total_Penjualan. Hasil perhitungan rata-rata setiap atribut pada masing-masing cluster ditunjukkan pada Gambar 3. Perbedaan nilai rata-rata tersebut digunakan sebagai dasar untuk mengidentifikasi karakteristik utama dari setiap cluster.



Gambar 3. Hasil Rata-Rata Setiap Atribut

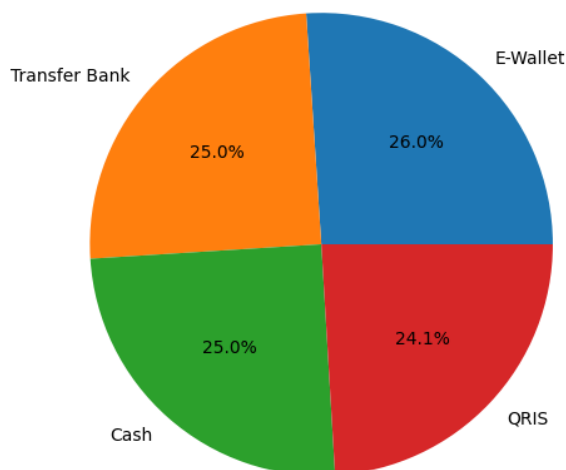
Selain berdasarkan atribut numerik, karakteristik cluster juga dianalisis berdasarkan distribusi kategori produk. Distribusi kategori pada setiap cluster ditampilkan pada Gambar 4 untuk mengetahui jenis produk yang mendominasi pada masing-masing kelompok transaksi.



Gambar 4. Distribusi Kategori Produk

Distribusi metode pembayaran dianalisis untuk mengetahui kecenderungan penggunaan metode pembayaran pada setiap cluster. Hasil distribusi tersebut disajikan pada Gambar 5 sehingga dapat diketahui apakah terdapat perbedaan pola pembayaran antarcluster.

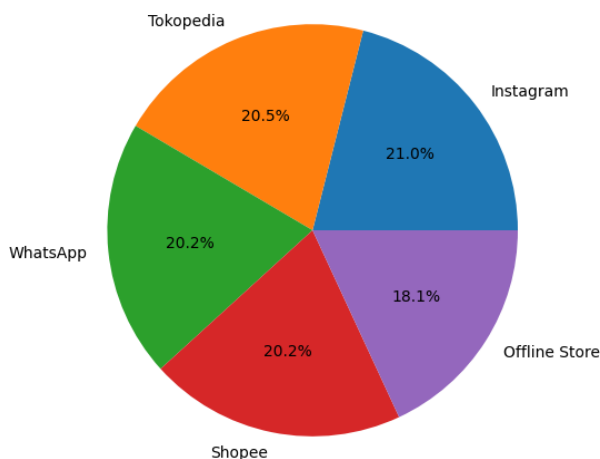
Cluster Berdasarkan Metode Pembayaran



Gambar 5. Distribusi Media Pembayaran

Distribusi channel penjualan pada setiap cluster dianalisis untuk melihat kecenderungan media penjualan yang digunakan dalam setiap kelompok transaksi. Hasil analisis ditampilkan pada Gambar 6. sebagai informasi pendukung dalam mengidentifikasi karakteristik masing-masing cluster.

Cluster Berdasarkan Channel Penjualan

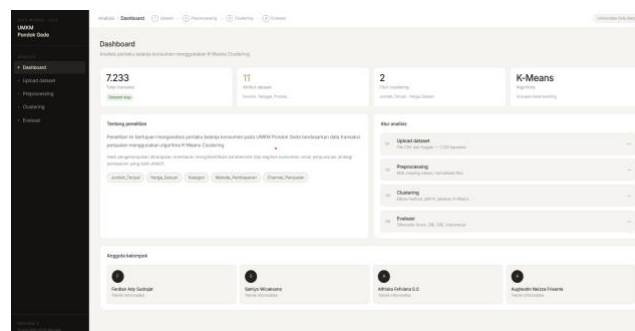


Gambar 6. Distribusi Channel Pembayaran

F. Hasil Fase Penerapan Sistem (Deployment)

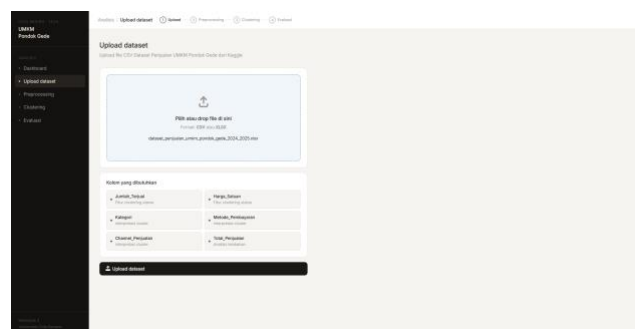
Pada tahap deployment, hasil analisis clustering yang telah diperoleh diimplementasikan ke dalam

sebuah prototipe aplikasi berbasis web menggunakan framework Flask (Python) sebagai media visualisasi bagi pelaku UMKM Pondok Gede. Aplikasi ini dirancang agar pengguna dapat menjalankan keseluruhan alur analisis secara interaktif, mulai dari unggah dataset transaksi, tahap preprocessing, penentuan jumlah cluster, hingga menampilkan hasil evaluasi clustering. Untuk memudahkan pengguna memahami alur kerja sistem, antarmuka aplikasi dibagi ke dalam lima halaman utama, yaitu dashboard, upload dataset, preprocessing, clustering, dan hasil evaluasi.



Gambar 7. Antarmuka Dashboard Sistem

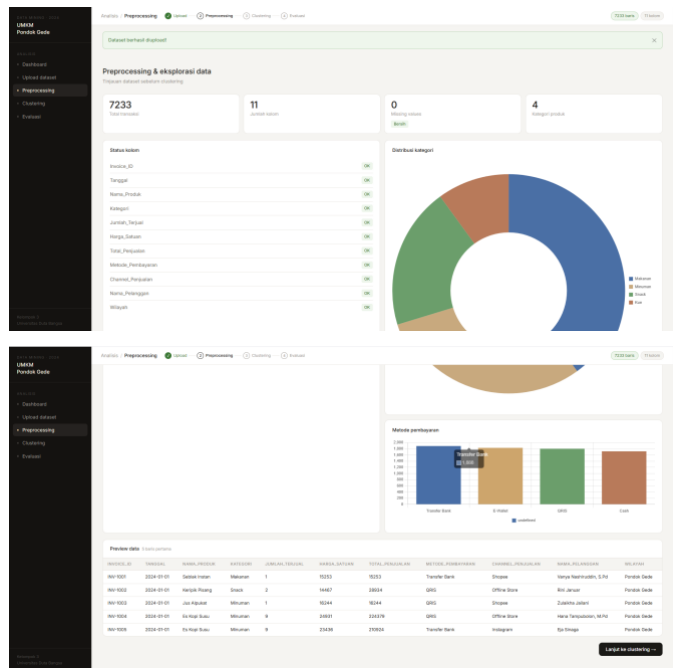
Seperti ditunjukkan pada Gambar 7, halaman dashboard menampilkan ringkasan umum penelitian, meliputi jumlah total transaksi (7.233 data), jumlah atribut dataset, dua fitur utama yang digunakan dalam proses clustering (Jumlah_Terjual dan Harga_Satuan), serta tahapan alur analisis mulai dari upload dataset hingga evaluasi. Halaman ini berfungsi sebagai titik awal bagi pengguna untuk memahami gambaran keseluruhan sistem sebelum melakukan analisis.



Gambar 8. Antarmuka Upload Dataset

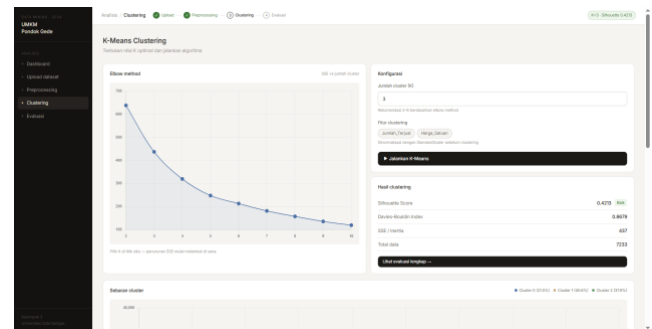
Gambar 8 memperlihatkan halaman upload dataset, tempat pengguna dapat mengunggah berkas transaksi dalam format CSV atau XLSX melalui fitur

drag-and-drop maupun pemilihan berkas secara manual. Sistem menampilkan daftar kolom yang wajib tersedia pada dataset, yaitu Jumlah_Terjual dan Harga_Satuan sebagai fitur utama clustering, serta Kategori, Metode_Pembayaran, Channel_Penjualan, dan Total_Penjualan sebagai atribut pendukung analisis karakteristik cluster.



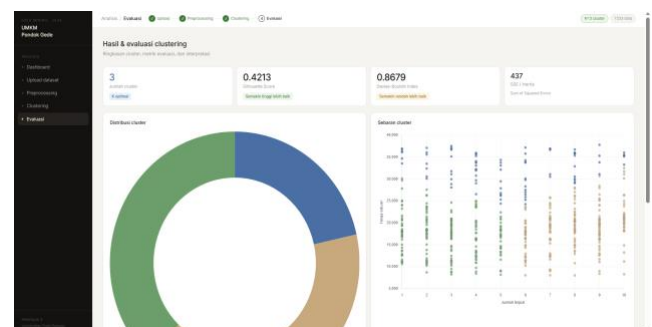
Gambar 9. Antarmuka Preprocessing dan Eklporasi Data

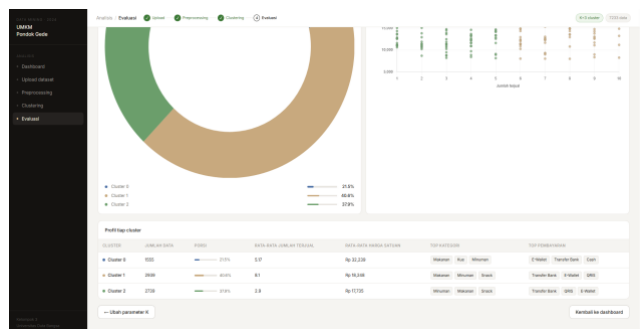
Setelah dataset berhasil diunggah, sistem secara otomatis menampilkan hasil eksplorasi data pada halaman preprocessing, sebagaimana terlihat pada Gambar 9. Halaman ini menyajikan informasi jumlah baris dan kolom dataset, status missing values pada setiap kolom, distribusi kategori produk, distribusi metode pembayaran, serta pratinjau lima baris pertama data. Informasi ini digunakan untuk memastikan bahwa data telah bersih dan siap digunakan pada tahap clustering.



Gambar 10. Antarmuka Clustering (Elbow Method dan Konfigurasi K)

Pada halaman clustering yang ditampilkan pada Gambar 10, pengguna dapat melihat grafik Elbow Method (SSE terhadap jumlah cluster) untuk membantu menentukan nilai K optimal, kemudian memilih jumlah cluster yang diinginkan serta menjalankan algoritma K-Means dengan menekan tombol jalankan. Setelah proses clustering selesai, sistem langsung menampilkan hasil evaluasi berupa Silhouette Score, Davies-Bouldin Index, dan SSE, beserta sebaran proporsi setiap cluster pada halaman yang sama.





Gambar 11. Antarmuka Hasil dan Evaluasi Clustering

Gambar 11 menunjukkan halaman hasil dan evaluasi clustering yang menyajikan ringkasan akhir berupa jumlah cluster, nilai Silhouette Score, Davies-Bouldin Index, dan SSE, dilengkapi dengan grafik distribusi cluster. Selain itu, halaman ini menampilkan tabel profil tiap cluster yang berisi jumlah data, porsi persentase, rata-rata Jumlah_Terjual, rata-rata Harga_Satuan, kategori produk dominan, serta metode pembayaran dominan pada masing-masing cluster.

Implementasi prototipe aplikasi web ini menjembatani hasil analisis data mining dengan kebutuhan praktis pelaku UMKM Pondok Gede, sehingga proses eksplorasi data, penentuan jumlah cluster, hingga interpretasi karakteristik konsumen dapat dilakukan secara interaktif tanpa memerlukan keahlian teknis dalam pemrograman maupun analisis statistik.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma K-Means berhasil mengelompokkan 7.233 data transaksi UMKM ke dalam 3 cluster berdasarkan karakteristik Jumlah Terjual, Harga Satuan, dan Total Penjualan. Penentuan jumlah cluster dilakukan menggunakan Elbow Method dan divalidasi dengan Silhouette Score, yang menunjukkan nilai terbaik pada $K = 3$ dengan nilai 0,4213.

Hasil clustering menunjukkan bahwa setiap cluster memiliki karakteristik yang berbeda. Cluster 0 didominasi oleh transaksi dengan harga satuan dan total penjualan tinggi, Cluster 1 memiliki jumlah terjual tertinggi, sedangkan Cluster 2 merupakan kelompok transaksi dengan jumlah terjual dan total penjualan relatif rendah. Analisis terhadap kategori

produk, metode pembayaran, dan channel penjualan menunjukkan bahwa perbedaan utama antarcluster lebih dipengaruhi oleh karakteristik nilai transaksi dibandingkan atribut pendukung lainnya.

Hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam memahami pola transaksi UMKM serta membantu pelaku usaha dalam menyusun strategi pemasaran, pengelolaan produk, dan pengambilan keputusan berdasarkan karakteristik masing-masing kelompok transaksi.

REFERENSI

- [1] Wu, Z., Jin, L., Zhao, J., Jing, L., & Chen, L. (2022). Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm. *Computational Intelligence and Neuroscience*, 2022(1), 9930613.
- [2] Amalina, T., Pramana, D. B. A., & Sari, B. N. (2022). Metode K-Means clustering dalam pengelompokan penjualan produk frozen food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574-583.
- [3] Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46-56.
- [4] Happy Maulana, M. K. (2024). *PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT* (Doctoral dissertation, UPN Veteran Jawa Timur).
- [5] Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 109-113.
- [6] Sulianta, N. R. N. F. Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja menggunakan Metode K-Means Clustering.
- [7] Azzam, A., Purnamasari, A. I., & Ali, I. (2024). Implementasi algoritma K-Means clustering untuk analisis persebaran UMKM di Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3062-3070.
- [8] Bui, M. A., & Bahtiar, A. (2024). Implementasi metode algoritma K-Means Clustering untuk mengelompokkan transaksi penjualan barang di Toko Arino. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1451-1456.
- [9] Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. *Jurnal Saintekom: Sains, Teknologi, Komputer dan Manajemen*, 13(1), 100-110.
- [10] Al Rasyid, H., Soebari, B. F. K., & Kartika, D. S. Y. (2022, September). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi. In *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi* (Vol. 2, No. 1, pp. 242-248).
- [11] H. Saputro, "Dataset Penjualan,UMKM" Kaggle, 2023. [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/hermawansaputro1/dataset-penjualan>. [Diakses: 23-Jun-2026].
- [12] Yulistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20-28.
- [13] Ana, S. (2022). Pengklasteran risiko covid-19 di riau menggunakan teknik one hot encoding dan algoritma k-means clustering. *Jurnal Informasi dan Komputer*.
- [14] Hartama, D., Wanayumini, W., & Damanik, I. S. (2024). Pengelompokan algoritma k-means dan k-medoid berdasarkan lokasi daerah rawan bencana di Indonesia dengan optimasi elbow, DBI, dan silhouette. *Building of Informatics, Technology and Science (BITS)*, 6(2).
- [15] Aradhana, M. A. M., Orisa, M., & Wahyuni, F. S. (2025). PENERAPAN METODE K-MEANS DALAM PENGELOMPOKAN PELANGGAN PADA TOKO PERCETAKAN BERBASIS WEB. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(6), 10760-10769.