

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK SEGMENTASI PERFORMA PENJUALAN PRODUK EUNOLA COFFEE SPACE

Abu Hanifah¹.

Program Studi Teknik Informatika ¹

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
abiehanifah01@gmail.com

(*) Corresponding Author : abiehanifah01@gmail.com

Published : 25 Juli 2026

Abstract—This research was motivated by the need of Eunola Coffee Space to utilize sales data as a more objective basis for business decision-making. The wide variety of products requires management to identify groups of products with high, medium, and low performance so that stock management, promotion, bundling, and menu evaluation strategies can be carried out more effectively. This study aims to identify the characteristics of product sales data, apply the K-Means Clustering algorithm, and explain the results of product performance segmentation as a basis for business strategy recommendations. The research method applies a data mining approach using the CRISP-DM workflow, consisting of business understanding, data understanding, data preparation, modeling, evaluation, and interpretation. The data were obtained from Eunola Coffee Space internal sales reports for the November–December 2024 and January–February 2025 periods in PDF format. The data were extracted using Python and pdfplumber, then processed through data cleaning, column standardization, numerical value conversion, period merging, product aggregation, data validation, and normalization using StandardScaler. The clustering features used were quantity sold, sales, and gross profit. The results show that the initial data consisted of 59 rows and, after aggregation, produced 48 unique products, with a total of 17,244 units sold, total sales of Rp410,741,500, and total gross profit of Rp410,255,500. K-Means modeling with K=3 produced three clusters: Best Seller with 1 product, Middle Performer with 15 products, and Low Performer with 32 products. Internal evaluation produced a Silhouette Score of 0.6010, Davies-Bouldin Index of 0.4368, and Calinski-Harabasz Index of 111.58, indicating sufficiently good cluster quality. The findings show that Kopi Susu Eunola + is the most dominant product, while the Espresso Based category contributes the largest sales proportion. The segmentation results can be used as a business recommendation basis to maintain leading products, strengthen supporting products, and evaluate low-performing products.

Keywords: Data Mining, Eunola Coffee Space, K-Means Clustering, Product Sales, Product Segmentation

Abstrak—Penelitian ini dilatarbelakangi oleh kebutuhan Eunola Coffee Space dalam memanfaatkan data penjualan sebagai dasar pengambilan keputusan bisnis yang lebih objektif. Banyaknya variasi produk menyebabkan manajemen perlu mengetahui kelompok produk dengan performa tinggi, menengah, dan rendah agar strategi stok, promosi, bundling, serta evaluasi menu dapat dilakukan secara lebih terarah. Tujuan penelitian ini adalah mengetahui karakteristik data penjualan produk, menerapkan algoritma K-Means Clustering, serta menguraikan hasil segmentasi performa produk sebagai dasar rekomendasi strategi bisnis. Metode penelitian menggunakan pendekatan data mining dengan alur CRISP-DM, dimulai dari business understanding, data understanding, data preparation, modeling, evaluation, hingga interpretation. Data yang digunakan berasal dari laporan penjualan internal Eunola Coffee Space periode November–Desember 2024 dan Januari–Februari 2025 dalam format PDF. Data diekstraksi menggunakan Python dan pdfplumber, kemudian diproses melalui pembersihan data, standarisasi kolom, konversi nilai numerik, penggabungan periode, agregasi produk, validasi data, dan normalisasi menggunakan StandardScaler. Fitur yang digunakan dalam clustering adalah jumlah terjual, penjualan, dan laba kotor. Hasil penelitian menunjukkan bahwa data awal berjumlah 59 baris dan setelah agregasi menghasilkan 48 produk unik, dengan total 17.244 unit terjual, total penjualan Rp410.741.500, dan total laba kotor Rp410.255.500. Pemodelan K-Means dengan K=3 menghasilkan tiga cluster, yaitu Best Seller sebanyak 1 produk, Middle Performer sebanyak 15 produk, dan Low Performer sebanyak 32 produk. Evaluasi internal menghasilkan Silhouette Score 0,6010, Davies-Bouldin Index 0,4368, dan Calinski-Harabasz Index 111,58,

yang menunjukkan kualitas cluster cukup baik. Temuan penelitian menunjukkan bahwa Kopi Susu Eunola + menjadi produk paling dominan, sedangkan kategori Espresso Based memberikan kontribusi penjualan terbesar. Hasil segmentasi dapat digunakan sebagai dasar rekomendasi bisnis untuk mempertahankan produk unggulan, memperkuat produk penopang, dan mengevaluasi produk berperforma rendah.

Kata Kunci : Data Mining, Eunola Coffee Space, K-Means Clustering, Penjualan Produk, Segmentasi Produk

INTRODUCTION

Perkembangan teknologi informasi telah mendorong pelaku usaha untuk memanfaatkan data sebagai dasar dalam proses pengambilan keputusan bisnis. Data transaksi penjualan yang sebelumnya hanya digunakan sebagai arsip administratif kini dapat diolah menjadi informasi strategis untuk memahami pola permintaan, kontribusi produk, dan perilaku pasar. Dalam konteks bisnis ritel dan layanan makanan-minuman, pemanfaatan data penjualan menjadi penting karena setiap produk memiliki tingkat kontribusi yang berbeda terhadap pendapatan dan keberlanjutan usaha. Pendekatan berbasis data dinilai mampu membantu perusahaan menyusun strategi pemasaran yang lebih tepat sasaran melalui proses segmentasi yang objektif dan sistematis [1]. Selain itu, integrasi teknik kecerdasan buatan dan clustering dalam strategi pemasaran ritel juga terbukti dapat membantu perusahaan mengelompokkan karakteristik transaksi untuk mendukung optimalisasi strategi bisnis [2].

Pada usaha kafe, variasi menu yang beragam sering kali menjadi daya tarik utama bagi konsumen, tetapi di sisi lain juga menimbulkan tantangan dalam pengelolaan produk. Tidak semua menu memiliki tingkat penjualan yang sama, sehingga manajemen perlu mengetahui produk mana yang memiliki performa tinggi, sedang, dan rendah. Apabila keputusan stok, promosi, dan evaluasi menu hanya didasarkan pada intuisi, maka risiko kelebihan bahan baku, rendahnya perputaran produk, dan tidak optimalnya strategi promosi dapat meningkat. Pengelompokan data penjualan menggunakan algoritma K-Means dapat membantu pelaku usaha dalam mengidentifikasi pola kebutuhan produk berdasarkan data historis, sehingga keputusan operasional dapat dilakukan secara lebih terarah [3]. Dalam konteks penjualan produk kopi, penerapan K-Means juga dapat digunakan untuk membantu simulasi pengelolaan barang masuk berdasarkan pola penjualan yang terjadi [4].

Eunola Coffee Space merupakan usaha kafe yang memiliki berbagai jenis produk, seperti minuman berbasis kopi, makanan ringan, makanan utama, mocktail, pastry, dan beberapa produk pendukung lainnya. Banyaknya variasi

produk tersebut menuntut adanya analisis yang mampu menggambarkan performa setiap produk secara lebih objektif. Berdasarkan data penjualan periode November–Desember 2024 dan Januari–Februari 2025, terlihat bahwa terdapat perbedaan kontribusi penjualan antarproduk, di mana sebagian produk memiliki kontribusi tinggi sedangkan sebagian lainnya memiliki performa yang relatif rendah. Kondisi tersebut menunjukkan perlunya proses segmentasi performa produk agar manajemen dapat memahami produk yang menjadi unggulan, produk yang masih stabil, dan produk yang perlu dievaluasi. Segmentasi produk berbasis data penjualan dinilai relevan karena dapat mendukung pengambilan keputusan dalam pengelolaan produk, promosi, serta efisiensi operasional ritel [5]. Penelitian lain juga menunjukkan bahwa K-Means dapat diterapkan untuk mengelompokkan produk berdasarkan data penjualan sehingga hasilnya dapat digunakan sebagai dasar analisis performa produk [6].

Permasalahan yang muncul pada Eunola Coffee Space adalah belum adanya pengelompokan produk secara terukur berdasarkan performa penjualan. Laporan penjualan yang tersedia masih perlu diolah lebih lanjut agar dapat menghasilkan informasi yang lebih mudah dipahami oleh manajemen. Jika data hanya dilihat berdasarkan total penjualan secara umum, maka produk dengan performa rendah dapat luput dari perhatian, sementara produk yang sebenarnya memiliki potensi untuk dikembangkan belum tentu memperoleh strategi promosi yang tepat. Segmentasi menggunakan pendekatan data mining dapat membantu membagi objek ke dalam beberapa kelompok berdasarkan kemiripan karakteristik, sehingga pola yang tersembunyi dalam data dapat terlihat lebih jelas [1]. Dalam konteks ritel, segmentasi berbasis atribut transaksi juga dapat membantu perusahaan memahami karakteristik kelompok secara lebih detail sehingga strategi yang dihasilkan menjadi lebih efektif [7].

Metode yang digunakan dalam penelitian ini adalah data mining dengan teknik clustering menggunakan algoritma K-Means. Metode ini dipilih karena data penjualan produk Eunola Coffee Space tidak memiliki label kelas sebelumnya, sehingga pendekatan yang digunakan bukan klasifikasi, melainkan pengelompokan berdasarkan kesamaan karakteristik data. K-Means merupakan

algoritma clustering yang bekerja dengan cara membagi data ke dalam sejumlah cluster berdasarkan kedekatan data terhadap centroid. Dalam penelitian ini, fitur yang digunakan meliputi jumlah terjual, penjualan, dan laba kotor karena ketiga atribut tersebut dapat merepresentasikan performa produk dari sisi volume permintaan dan kontribusi finansial. Penggunaan K-Means dalam segmentasi berbasis transaksi dinilai efektif karena mampu menghasilkan kelompok data yang dapat diinterpretasikan untuk kebutuhan strategi bisnis [2]. Selain itu, penggunaan K-Means pada data penjualan ritel juga dapat membantu mengidentifikasi kebutuhan produk berdasarkan perbedaan karakteristik transaksi [3].

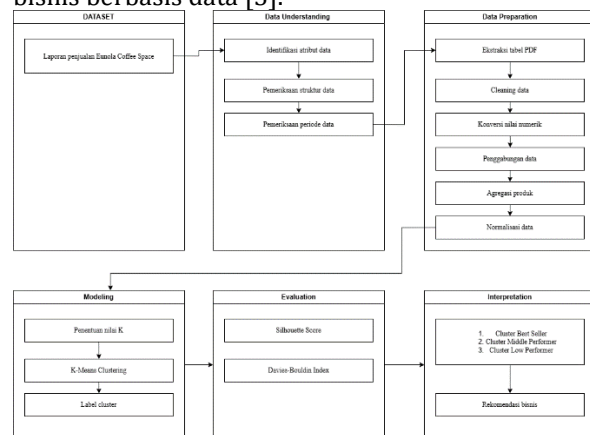
Berdasarkan uraian tersebut, penelitian mengenai implementasi algoritma K-Means Clustering untuk segmentasi performa penjualan produk pada Eunola Coffee Space penting untuk dilakukan. Secara praktis, penelitian ini dapat membantu pihak manajemen dalam menentukan produk best seller, produk dengan performa menengah, dan produk dengan performa rendah sebagai dasar pengambilan keputusan terkait stok, promosi, bundling, dan evaluasi menu. Secara teoritis, penelitian ini memberikan kontribusi dalam penerapan data mining pada studi kasus bisnis kafe lokal, khususnya dalam mengolah data penjualan menjadi informasi segmentasi produk yang lebih bermakna. Penerapan clustering pada data penjualan telah terbukti dapat mendukung pengambilan keputusan berbasis data dalam bisnis ritel dan pengelolaan produk [5]. Oleh karena itu, penelitian ini diarahkan untuk menjawab bagaimana algoritma K-Means dapat diterapkan dalam mengelompokkan performa penjualan produk Eunola Coffee Space serta bagaimana hasil segmentasi tersebut dapat dimanfaatkan sebagai rekomendasi strategi bisnis [4].

MATERIALS AND METHODS

Prosedur penelitian merupakan tahapan sistematis yang digunakan untuk menjelaskan alur pelaksanaan penelitian dari awal sampai akhir. Pada penelitian ini, prosedur penelitian disusun berdasarkan pendekatan CRISP-DM karena penelitian berfokus pada proses data mining, mulai dari pemahaman kebutuhan penelitian, pemahaman data, persiapan data, pemodelan, evaluasi, hingga interpretasi hasil. CRISP-DM relevan digunakan dalam penelitian data mining karena menyediakan kerangka

kerja yang sistematis dan dapat diterapkan pada berbagai jenis proyek analisis data [8].

Penelitian ini menggunakan data penjualan produk Eunola Coffee Space sebagai input utama, kemudian data tersebut diproses melalui tahapan data understanding, data preparation, modeling, evaluation, dan interpretation. Setiap tahapan memiliki peran yang saling berkaitan untuk menghasilkan segmentasi performa produk menggunakan algoritma K-Means Clustering. Penggunaan K-Means pada data penjualan dinilai relevan karena metode ini dapat mengelompokkan data berdasarkan kemiripan karakteristik sehingga dapat membantu proses pengambilan keputusan bisnis berbasis data [3].



Gambar 1 Prosedur Penelitian

Business Understanding

Tahap Business Understanding dilakukan untuk memahami permasalahan utama yang menjadi dasar penelitian. Pada penelitian ini, permasalahan yang diangkat adalah kebutuhan untuk mengetahui segmentasi performa produk Eunola Coffee Space berdasarkan data penjualan historis. Data penjualan produk perlu dianalisis agar manajemen dapat memperoleh gambaran mengenai kelompok produk yang memiliki performa tinggi, menengah, dan rendah. Pemanfaatan data mining dalam konteks bisnis dapat membantu perusahaan menyusun strategi berbasis data, sehingga keputusan tidak hanya bergantung pada intuisi atau penilaian subjektif [1].

Pada tahap ini, peneliti menentukan tujuan penelitian, yaitu mengetahui karakteristik data penjualan produk, menerapkan algoritma K-Means Clustering, dan menghasilkan segmentasi performa produk sebagai dasar rekomendasi strategi bisnis. Output dari tahap ini berupa identifikasi masalah bisnis, kebutuhan segmentasi produk, dan arah rekomendasi yang berkaitan dengan pengelolaan stok, promosi, bundling menu, serta evaluasi produk. Clustering dalam bidang ritel dapat digunakan untuk membentuk kelompok data yang

lebih mudah dipahami sehingga strategi bisnis dapat disesuaikan dengan karakteristik setiap kelompok [2].

Data Understanding

Tahap Data Understanding dilakukan untuk memahami struktur dan karakteristik data yang digunakan dalam penelitian. Data yang digunakan berasal dari laporan penjualan produk Eunola Coffee Space dalam format PDF. Data tersebut memuat beberapa atribut, seperti produk, SKU, kategori, jumlah terjual, penjualan, jumlah refund, refund, dan laba kotor. Pemahaman awal terhadap data penting dilakukan agar peneliti dapat mengetahui atribut yang tersedia, periode data, serta kesesuaian data dengan tujuan penelitian. Data penjualan historis dapat digunakan untuk memahami pola permintaan dan performa produk dalam periode tertentu [4].

Pada tahap ini, peneliti melakukan identifikasi atribut data, pemeriksaan struktur data, dan pemeriksaan periode data sebagaimana ditunjukkan pada Gambar 3.1. Atribut yang menjadi perhatian utama adalah jumlah terjual, penjualan, dan laba kotor karena ketiga atribut tersebut dapat merepresentasikan performa produk dari sisi volume penjualan dan kontribusi finansial. Pemahaman terhadap atribut data penjualan menjadi langkah penting sebelum proses clustering karena kualitas pemilihan atribut akan memengaruhi hasil pengelompokan data [3].

Data Preparation

Tahap Data Preparation merupakan proses persiapan data sebelum dilakukan pemodelan. Pada penelitian ini, data awal berupa laporan penjualan dalam format PDF sehingga perlu diekstraksi terlebih dahulu menjadi data tabular. Setelah proses ekstraksi, data melalui tahapan cleaning data, konversi nilai numerik, penggabungan data, agregasi produk, dan normalisasi data. Tahap persiapan data menjadi bagian penting dalam CRISP-DM karena data yang belum bersih dan belum terstruktur dapat memengaruhi kualitas hasil analisis data mining [8].

Pada tahap ini, data dari periode November–Desember 2024 dan Januari–Februari 2025 digabungkan untuk memperoleh dataset yang lebih utuh. Setelah itu, data diagregasi berdasarkan produk agar setiap produk memiliki nilai total yang dapat digunakan dalam proses clustering. Normalisasi data dilakukan karena atribut jumlah terjual, penjualan, dan

laba kotor memiliki skala yang berbeda. K-Means merupakan algoritma yang berbasis jarak, sehingga proses normalisasi penting dilakukan agar setiap atribut memiliki kontribusi yang lebih seimbang dalam pembentukan cluster [9].

Modeling

Tahap Modeling merupakan proses penerapan algoritma K-Means Clustering pada dataset yang telah melalui tahap data preparation. Pada tahap ini, data diproses menggunakan fitur jumlah terjual, penjualan, dan laba kotor. Proses modeling diawali dengan penentuan nilai K, kemudian dilanjutkan dengan penerapan algoritma K-Means Clustering dan pemberian label cluster. K-Means banyak digunakan dalam segmentasi karena mampu mengelompokkan data numerik berdasarkan kedekatan data terhadap centroid [9].

Dalam penelitian ini, label cluster diarahkan untuk memudahkan interpretasi hasil segmentasi produk. Label tersebut digunakan untuk menjelaskan kelompok produk berdasarkan performa penjualan, yaitu Best Seller, Middle Performer, dan Low Performer. Pada bagian ini, hasil detail cluster belum disajikan karena pembahasan mengenai anggota cluster, karakteristik cluster, dan hasil segmentasi akan dijelaskan pada BAB IV. Penggunaan K-Means pada data penjualan produk dapat membantu mengelompokkan produk berdasarkan karakteristik penjualan sehingga hasilnya dapat dimanfaatkan untuk mendukung strategi persediaan dan promosi [6].

Evaluation

Tahap Evaluation dilakukan untuk menilai kualitas hasil clustering yang terbentuk. Karena penelitian ini menggunakan metode clustering, maka evaluasi tidak menggunakan metrik klasifikasi seperti accuracy, precision, recall, atau F1-score. Evaluasi dilakukan menggunakan metrik internal clustering, yaitu Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Metrik internal digunakan karena clustering tidak memiliki label kelas awal, sehingga kualitas kelompok dinilai berdasarkan kedekatan data dalam cluster dan pemisahan antar cluster [10].

Pada tahap ini, hasil evaluasi digunakan untuk memastikan bahwa cluster yang terbentuk memiliki kualitas yang dapat dipertanggungjawabkan secara ilmiah. Silhouette Score digunakan untuk menilai kesesuaian data terhadap cluster-nya, Davies-Bouldin Index digunakan untuk menilai rasio kepadatan dan pemisahan cluster, sedangkan Calinski-Harabasz Index digunakan untuk melihat perbandingan variasi antar cluster dan dalam

cluster. Evaluasi menggunakan metrik internal seperti Silhouette Score relevan digunakan dalam penelitian K-Means karena dapat membantu menilai kualitas pembentukan cluster tanpa membutuhkan label kelas sebenarnya [11].

Interpretation

Tahap Interpretation dilakukan untuk menafsirkan hasil clustering menjadi informasi bisnis yang mudah dipahami. Pada tahap ini, cluster yang terbentuk dianalisis berdasarkan karakteristik jumlah terjual, penjualan, dan laba kotor. Interpretasi dilakukan untuk menjelaskan makna setiap cluster dalam konteks bisnis Eunola Coffee Space. Hasil clustering tidak hanya dipandang sebagai output teknis, tetapi juga sebagai dasar untuk memahami performa produk dan menyusun rekomendasi bisnis [5].

Pada penelitian ini, interpretasi diarahkan pada tiga kelompok produk, yaitu Best Seller, Middle Performer, dan Low Performer, sebagaimana ditunjukkan pada alur workflow. Rekomendasi bisnis kemudian disusun berdasarkan karakteristik masing-masing kelompok produk, seperti pengelolaan stok, promosi, bundling menu, dan evaluasi produk. Namun, rincian hasil cluster dan rekomendasi detail tidak disajikan pada bagian ini karena akan dibahas secara khusus pada BAB IV. Pengelompokan produk berbasis data penjualan dapat membantu perusahaan memahami performa produk secara lebih terarah dan mendukung strategi pengelolaan produk [6].

RESULTS AND DISCUSSION

Parameter pemodelan yang digunakan dalam penelitian ini ditampilkan pada Tabel 4.10. Tabel tersebut menunjukkan komponen utama dalam proses K-Means Clustering, mulai dari dataset, fitur yang digunakan, teknik normalisasi, jumlah cluster, hingga label bisnis yang diberikan pada hasil cluster.

Tabel 1 Parameter Pemodelan K-Means

N o	Komponen	Nilai	Keterangan
1	Dataset	48 produk unik	Dataset final hasil agregasi produk dari dua periode penjualan
2	Fitur clustering	Jumlah terjual, penjualan, laba kotor	Fitur utama yang merepresentasikan

			performa produk
3	Teknik normalisasi	StandardScaler	Menyamakan skala fitur agar proses perhitungan jarak lebih seimbang
4	Algoritma	K-Means Clustering	Algoritma yang digunakan untuk mengelompokkan produk
5	Jumlah cluster	K=3	Jumlah cluster yang digunakan dalam pemodelan
6	random_state	42	Menjaga hasil pemodelan agar dapat direplikasi
7	n_init	10	Jumlah inisialisasi centroid untuk memilih hasil terbaik
8	Label cluster	Best Seller, Middle Performer, Low Performer	Label bisnis untuk memudahkan interpretasi hasil clustering

Hasil pemodelan K-Means Clustering menghasilkan tiga kelompok produk, yaitu Best Seller, Middle Performer, dan Low Performer. Pemberian label ini dilakukan agar hasil clustering lebih mudah dipahami dalam konteks bisnis. Cluster Best Seller digunakan untuk menunjukkan produk dengan performa penjualan paling tinggi. Cluster Middle Performer digunakan untuk menunjukkan produk dengan performa menengah atau stabil. Sementara itu, cluster Low Performer digunakan untuk menunjukkan produk dengan performa penjualan rendah yang perlu dievaluasi lebih lanjut.

Berdasarkan hasil pemodelan, cluster Best Seller terdiri dari 1 produk atau 2,08% dari total produk. Cluster Middle Performer terdiri dari 15 produk atau 31,25% dari total produk. Adapun cluster Low Performer terdiri dari 32 produk atau 66,67% dari total produk. Distribusi ini menunjukkan bahwa sebagian besar produk berada pada kelompok Low Performer, sedangkan produk dengan performa paling tinggi hanya terdiri dari satu produk. Hasil

distribusi jumlah produk per cluster ditampilkan pada Tabel 2.

Tabel 2 Distribusi Jumlah Produk Percluster

No	Cluster	Jumlah Produk	Persentase Produk	Keterangan
1	Best Seller	1	2,08%	Produk dengan performa penjualan paling tinggi
2	Middle Performer	15	31,25%	Produk dengan performa menengah atau stabil
3	Low Performer	32	66,67%	Produk dengan performa rendah dan perlu dievaluasi

Hasil pemodelan ini menunjukkan bahwa algoritma K-Means Clustering berhasil membagi produk Eunola Coffee Space ke dalam tiga kelompok performa. Pembagian cluster tersebut menjadi dasar untuk tahap pembahasan berikutnya, yaitu evaluasi hasil clustering dan interpretasi karakteristik masing-masing cluster. Dengan adanya label Best Seller, Middle Performer, dan Low Performer, hasil clustering tidak hanya berbentuk kelompok numerik, tetapi juga dapat dipahami sebagai informasi bisnis yang relevan untuk mendukung pengambilan keputusan manajemen Eunola Coffee Space.

Evaluasi Hasil Clustering

Evaluasi hasil clustering dilakukan untuk mengetahui kualitas kelompok produk yang terbentuk setelah proses pemodelan K-Means Clustering. Karena penelitian ini menggunakan metode clustering yang termasuk ke dalam unsupervised learning, maka evaluasi tidak menggunakan istilah akurasi seperti pada metode klasifikasi. Evaluasi dilakukan menggunakan metrik internal clustering, yaitu Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Ketiga metrik tersebut digunakan untuk menilai kualitas cluster

berdasarkan kedekatan data dalam satu cluster dan pemisahan antar cluster.

Silhouette Score digunakan untuk mengukur seberapa dekat suatu data dengan cluster tempat data tersebut berada dibandingkan dengan cluster lain. Nilai Silhouette Score berada pada rentang -1 sampai 1. Semakin mendekati 1, maka kualitas cluster semakin baik karena data dalam satu cluster memiliki kedekatan yang tinggi dan memiliki jarak yang cukup baik terhadap cluster lain. Berdasarkan hasil evaluasi, nilai Silhouette Score yang diperoleh adalah 0,6010. Nilai tersebut menunjukkan bahwa hasil clustering memiliki pemisahan cluster yang cukup baik dan data dalam setiap kelompok memiliki tingkat kesesuaian yang memadai.

Davies-Bouldin Index digunakan untuk mengukur rasio antara kepadatan data dalam cluster dan jarak antar cluster. Berbeda dengan Silhouette Score, nilai Davies-Bouldin Index yang semakin kecil menunjukkan kualitas cluster yang semakin baik. Berdasarkan hasil evaluasi, nilai Davies-Bouldin Index yang diperoleh adalah 0,4368. Nilai ini menunjukkan bahwa rasio penyebaran data dalam cluster terhadap jarak antar cluster tergolong rendah, sehingga kelompok yang terbentuk dapat dikatakan memiliki pemisahan yang cukup baik.

Calinski-Harabasz Index digunakan untuk menilai kualitas cluster berdasarkan perbandingan antara variasi antar cluster dan variasi dalam cluster. Semakin tinggi nilai Calinski-Harabasz Index, maka semakin baik kualitas cluster yang terbentuk. Berdasarkan hasil evaluasi, nilai Calinski-Harabasz Index yang diperoleh adalah 111,58. Nilai tersebut menunjukkan bahwa terdapat pemisahan antar cluster yang baik dibandingkan dengan variasi data di dalam cluster.

Hasil evaluasi clustering menunjukkan bahwa segmentasi produk Eunola Coffee Space memiliki kualitas cluster yang dapat digunakan untuk tahap interpretasi bisnis. Nilai Silhouette Score yang berada di atas 0,6 menunjukkan bahwa cluster memiliki tingkat kohesi dan separasi yang cukup baik. Nilai Davies-Bouldin Index yang rendah juga mendukung bahwa jarak antar cluster cukup terpisah, sedangkan nilai Calinski-Harabasz Index menunjukkan adanya perbedaan karakteristik antar cluster. Dengan demikian, hasil clustering layak digunakan sebagai dasar untuk menginterpretasikan kelompok produk Best Seller, Middle Performer, dan Low Performer.

Tabel 3 Hasil Evaluasi Clustering

No	Metrik Evaluasi	Nilai	Interpretasi
1	Silhouette Score	0,6010	Nilai mendekati 1

			menunjukkan bahwa cluster memiliki pemisahan yang cukup baik.
2	Davies-Bouldin Index	0,4368	Nilai mendekati 0 menunjukkan bahwa cluster memiliki pemisahan yang baik dan penyebaran dalam cluster relatif rendah.
3	Calinski-Harabasz Index	111,58	Semakin tinggi nilainya, semakin baik kualitas pemisahan antar cluster.

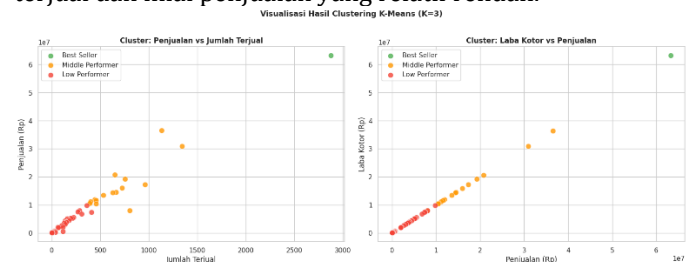
Berdasarkan Tabel 3 hasil evaluasi menunjukkan bahwa pembentukan tiga cluster pada data penjualan Eunola Coffee Space memiliki kualitas yang baik untuk dianalisis lebih lanjut. Oleh karena itu, hasil clustering dapat digunakan sebagai dasar interpretasi performa produk dan penyusunan rekomendasi strategi bisnis. Pembahasan selanjutnya akan menjelaskan karakteristik masing-masing cluster, yaitu Best Seller, Middle Performer, dan Low Performer.

Interpretasi Hasil Segmentasi Produk

Interpretasi hasil segmentasi produk dilakukan untuk menjelaskan makna dari cluster yang terbentuk melalui algoritma K-Means Clustering. Tahap ini menjadi bagian penting karena hasil clustering tidak hanya berupa pembagian kelompok secara matematis, tetapi juga perlu diterjemahkan menjadi informasi bisnis yang dapat dipahami oleh pihak manajemen Eunola Coffee Space. Berdasarkan hasil pemodelan, produk dikelompokkan ke dalam tiga cluster, yaitu Best Seller, Middle Performer, dan Low Performer. Ketiga cluster tersebut dibentuk berdasarkan fitur jumlah terjual, penjualan, dan laba kotor.

Berdasarkan visualisasi hasil clustering K-Means pada Gambar 2, terlihat bahwa cluster

Best Seller berada jauh dari kelompok produk lainnya. Pada grafik hubungan antara penjualan dan jumlah terjual, cluster Best Seller ditunjukkan sebagai titik yang berada pada posisi paling tinggi dan paling kanan. Hal ini menunjukkan bahwa produk pada cluster tersebut memiliki jumlah terjual dan nilai penjualan paling tinggi dibandingkan produk lain. Sementara itu, cluster Middle Performer berada pada area menengah dengan nilai penjualan dan jumlah terjual yang relatif stabil. Cluster Low Performer terkonsentrasi pada area bawah dan kiri, yang menunjukkan bahwa produk dalam kelompok ini memiliki jumlah terjual dan nilai penjualan yang relatif rendah.



Gambar 2 Hasil Cluster

Pada visualisasi hubungan antara laba kotor dan penjualan, pola data terlihat membentuk garis yang hampir linear. Hal ini menunjukkan bahwa nilai laba kotor pada sebagian besar produk memiliki hubungan yang sangat dekat dengan nilai penjualan. Perbedaan kecil antara penjualan dan laba kotor terjadi karena adanya refund pada beberapa produk. Visualisasi tersebut juga memperlihatkan bahwa cluster Best Seller memiliki nilai penjualan dan laba kotor paling tinggi, sedangkan cluster Middle Performer berada pada tingkat menengah, dan cluster Low Performer berada pada tingkat rendah. Dengan demikian, visualisasi clustering memperkuat hasil segmentasi bahwa produk Eunola Coffee Space dapat dikelompokkan secara jelas berdasarkan performa penjualannya.

Ringkasan karakteristik setiap cluster ditampilkan pada Tabel 4. Berdasarkan tabel tersebut, cluster Best Seller terdiri dari 1 produk dengan total jumlah terjual sebanyak 2.877 unit, total penjualan sebesar Rp63.294.000, dan total laba kotor sebesar Rp63.294.000. Cluster Middle Performer terdiri dari 15 produk dengan total jumlah terjual sebanyak 10.322 unit, total penjualan sebesar Rp246.743.500, dan total laba kotor sebesar Rp246.395.500. Adapun cluster Low Performer terdiri dari 32 produk dengan total jumlah terjual sebanyak 4.045 unit, total penjualan sebesar Rp100.704.000, dan total laba kotor sebesar Rp100.566.000.

Tabel 4 Ringkasan Karakteristik Cluster

N o	Cluster	Jumlah Produk	Total Terjual	Total Penjualan	Total Laba Kotor	Karakteristik
1	Best Seller	1	2.877	Rp63.294.000	Rp63.294.000	Produk dengan performa paling tinggi dan menjadi produk utama penopang penjualan.
2	Middle Performer	15	10.322	Rp24.674.500	Rp24.639.500	Produk dengan performa menengah, stabil, dan menjadi penopang penjualan harian.
3	Low Performer	32	4.045	Rp10.070.000	Rp10.056.000	Produk dengan performa rendah yang perlu ditingkatkan.

luasi dari sisi strategi penjualan, promosi, dan stok.

Cluster Best Seller merupakan kelompok produk dengan performa paling tinggi. Cluster ini hanya terdiri dari satu produk, yaitu Kopi Susu Eunola +. Produk tersebut memiliki jumlah terjual sebanyak 2.877 unit, penjualan sebesar Rp63.294.000, dan laba kotor sebesar Rp63.294.000. Berdasarkan hasil tersebut, Kopi Susu Eunola + dapat diinterpretasikan sebagai produk unggulan utama atau signature product Eunola Coffee Space. Produk ini perlu dijaga kualitasnya, dipastikan ketersediaan bahan bakunya, dan dipertahankan sebagai produk utama karena memiliki kontribusi yang sangat besar terhadap performa penjualan.

Tabel 5 Daftar Produk Cluster Best Seller

N o	Produk	Kategori	Jumlah Terjual	Penjualan	Lab Kotor
1	Kopi Susu Eunola +	Espresso Based	2.877	Rp63.294.000	Rp63.294.000

Cluster Middle Performer merupakan kelompok produk dengan performa menengah dan stabil. Cluster ini terdiri dari 15 produk yang secara keseluruhan menghasilkan total penjualan paling besar dibandingkan cluster lainnya. Hal ini menunjukkan bahwa meskipun tidak ada satu produk pada cluster ini yang melampaui dominasi Kopi Susu Eunola +, gabungan produk dalam cluster Middle Performer memberikan kontribusi yang sangat penting bagi penjualan Eunola Coffee Space. Produk dalam cluster ini dapat dianggap sebagai produk penopang operasional harian karena memiliki permintaan yang cukup baik dan kontribusi finansial yang stabil.

Tabel 6 Daftar Produk Cluster Middle Performer

N o	Produk	Kategori	Jumlah Terjual	Penjualan	Lab Kotor
-----	--------	----------	----------------	-----------	-----------

1	Mix Platter	Snacks	1.13 2	Rp36.48 4.000	Rp36.41 6.000	Performer memiliki jumlah produk lebih banyak tetapi kontribusi penjualannya lebih rendah. Hal ini menunjukkan bahwa sebagian besar produk dalam cluster Low Performer memiliki tingkat permintaan yang relatif rendah. Oleh karena itu, produk pada cluster ini perlu menjadi perhatian dalam evaluasi menu, strategi promosi, pengaturan stok, dan pertimbangan efisiensi bahan baku.
2	Caffe Latte	Espresso Based	1.34 3	Rp30.88 9.000	Rp30.86 6.000	
3	Chicken Butter	Main Courses	649	Rp20.76 8.000	Rp20.64 0.000	
4	Salted Caramel	Espresso Based	754	Rp19.22 7.000	Rp19.22 7.000	CONCLUSION Berdasarkan hasil penelitian mengenai implementasi algoritma K-Means Clustering untuk segmentasi performa penjualan produk pada Eunola Coffee Space, maka dapat diperoleh beberapa kesimpulan sebagai berikut: Karakteristik data penjualan produk Eunola Coffee Space dapat diketahui melalui proses pengumpulan data, ekstraksi, preprocessing, validasi, dan Exploratory Data Analysis. Data yang digunakan berasal dari laporan penjualan internal periode November–Desember 2024 dan Januari–Februari 2025 dalam format PDF. Data awal berjumlah 59 baris, kemudian setelah dilakukan proses penggabungan dan agregasi berdasarkan produk menghasilkan 48 produk unik. Berdasarkan hasil analisis, total jumlah produk terjual selama periode penelitian adalah 17.244 unit, total penjualan sebesar Rp410.741.500, total laba kotor sebesar Rp410.255.500, dan total refund sebesar Rp486.000. Produk dengan performa paling dominan adalah Kopi Susu Eunola +, sedangkan kategori dengan kontribusi penjualan terbesar adalah Espresso Based dengan kontribusi sebesar 40,71% dari total penjualan. Hal ini menunjukkan bahwa data penjualan Eunola Coffee Space memiliki pola performa produk yang tidak merata, sehingga perlu dilakukan segmentasi produk secara lebih objektif. Algoritma K-Means Clustering berhasil diterapkan untuk melakukan segmentasi performa penjualan produk Eunola Coffee Space. Proses pemodelan dilakukan menggunakan tiga fitur utama, yaitu jumlah terjual, penjualan, dan laba kotor. Sebelum proses clustering dilakukan, data dinormalisasi menggunakan StandardScaler agar setiap fitur memiliki skala yang seimbang dalam proses perhitungan jarak. Penentuan jumlah cluster dilakukan dengan membandingkan beberapa nilai K dan mempertimbangkan kebutuhan interpretasi bisnis. Berdasarkan hasil evaluasi dan kebutuhan manajemen, jumlah cluster yang digunakan adalah K=3. Pemodelan dilakukan dengan parameter K=3, random_state=42, dan n_init=10. Hasil evaluasi clustering menunjukkan nilai Silhouette Score sebesar 0,6010, Davies-Bouldin Index sebesar 0,4368, dan Calinski-Harabasz Index sebesar 111,58. Nilai tersebut menunjukkan bahwa hasil
5	Pisang Goreng	Snacks	961	Rp17.29 8.000	Rp17.28 0.000	
6	French Toast	Snacks	725	Rp15.95 0.000	Rp15.88 4.000	
7	French Fries	Snacks	661	Rp14.54 2.000	Rp14.52 0.000	
8	Cappuccino	Espresso Based	626	Rp14.39 8.000	Rp14.39 8.000	
9	Caramel Latte	Espresso Based	530	Rp13.51 5.000	Rp13.51 5.000	
10	Matcha Latte	Matcha Series	442	Rp11.93 4.000	Rp11.93 4.000	
11	Hazelnut Latte	Espresso Based	455	Rp11.60 2.500	Rp11.60 2.500	
12	Rose Berry	Mocktails	398	Rp11.14 4.000	Rp11.14 4.000	
13	Berry Vanilla	Milk Based	390	Rp10.53 0.000	Rp10.53 0.000	
14	Mushroom Crispy	Snacks	454	Rp10.44 2.000	Rp10.41 9.000	
15	Mineral Cleo	Mineral	802	Rp8.020 .000	Rp8.020 .000	

Cluster Low Performer merupakan kelompok dengan jumlah produk paling banyak, yaitu 32 produk. Meskipun jumlah produknya paling besar, total penjualan cluster ini hanya sebesar Rp100.704.000 dengan total laba kotor sebesar Rp100.566.000. Jika dibandingkan dengan cluster Middle Performer, cluster Low

clustering memiliki kualitas cluster yang cukup baik dari sisi kohesi dan separasi antarcluster.

Hasil segmentasi performa produk Eunola Coffee Space menghasilkan tiga kelompok, yaitu Best Seller, Middle Performer, dan Low Performer. Cluster Best Seller terdiri dari 1 produk, yaitu Kopi Susu Eunola +, dengan total jumlah terjual 2.877 unit, penjualan Rp63.294.000, dan laba kotor Rp63.294.000. Cluster Middle Performer terdiri dari 15 produk dengan total jumlah terjual 10.322 unit, total penjualan Rp246.743.500, dan total laba kotor Rp246.395.500. Cluster Low Performer terdiri dari 32 produk dengan total jumlah terjual 4.045 unit, total penjualan Rp100.704.000, dan total laba kotor Rp100.566.000. Berdasarkan hasil tersebut, dapat disimpulkan bahwa jumlah produk yang banyak tidak selalu memberikan kontribusi penjualan yang besar. Oleh karena itu, hasil segmentasi ini dapat digunakan sebagai dasar rekomendasi strategi bisnis, seperti menjaga kualitas dan stok produk Best Seller, memperkuat promosi dan bundling produk Middle Performer, serta mengevaluasi produk Low Performer dari sisi harga, rasa, tampilan menu, promosi, dan pengelolaan stok.

REFERENCE

- [1] N. Gautam and N. Kumar, "Customer segmentation using K-Means clustering for developing sustainable marketing strategies," *Bus. Informatics*, vol. 16, no. 1, pp. 72–82, 2022, doi: 10.17323/2587-814X.2022.172.82.
- [2] Y. Putri, D. Aldo, and W. Ilham, "Retail marketing strategy optimization: Customer segmentation with artificial intelligence integration and K-Means clustering," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 4, pp. 2155–2163, 2024, doi: 10.33395/sinkron.v8i4.14000.
- [3] R. I. Manarung, E. Widodo, and A. M. Rifai, "Sales data clustering using the K-Means algorithm to determine retail product needs," *Int. J. Softw. Eng. Comput. Sci.*, vol. 5, no. 1, pp. 226–234, 2025, doi: 10.35870/ijsecs.v5i1.4090.
- [4] M. A. Triansyah, B. Wijayanto, and A. A. Paramestuti, "Implementation of K-Means on packaged coffee sales data for simulating goods entry in sole proprietorship businesses," *J. Tek. Inform.*, vol. 6, no. 4, pp. 2392–2404, 2025, doi: 10.52436/1.jutif.2025.6.4.5245.
- [5] B. K. Kristanto, S. W. P. Listio, M. Amien, and P. I. Baskoro, "Visit recommendation model: Recursive K-Means clustering analysis of retail sales data," *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 221–225, 2024, doi: 10.30871/jaic.v8i1.8138.
- [6] C. Chairunnita, A. Handayanto, and F. Dewanto, "Penerapan algoritma K-Means Clustering dalam analisis pengelompokan produk toko oleh-oleh berdasarkan data penjualan," *J. Inf. Syst. Res.*, vol. 6, no. 4, pp. 1816–1825, 2025, doi: 10.47065/josh.v6i4.7832.
- [7] D. A. Imanuel and G. Alfian, "Visualisasi segmentasi pelanggan berdasarkan atribut RFM menggunakan algoritma K-Means untuk memahami karakteristik pelanggan pada toko retail online," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 145–156, 2025, doi: 10.25126/jtiik.2025128619.
- [8] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [9] K. Tabianan, S. Velu, and V. Ravi, "K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022, doi: 10.3390/su14127243.
- [10] I. F. Ashari, E. D. Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 95–103, 2023, doi: 10.30871/jaic.v7i1.4947.
- [11] B. N. Yuliasih, Herman, Sunardi, and H. Yuliansyah, "Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation," *Ilk. J. Ilm.*, vol. 16, no. 3, pp. 330–342, 2024, doi: 10.33096/ilkom.v16i3.2325.330-342.