

PENGELOMPOKAN PRODUK PENJUALAN MENGGUNAKAN K-MEANS SEBAGAI PENDUKUNG STRATEGI BISNIS KAFE OMAH KOPI

Ahmad Dhani^{1*}, Retno Wardhani², Ayu Ismi Hanifah³

¹²³ Prodi Teknik Informatika, Universitas Islam Lamongan, Jl Veteran Nomor 53A, Jetis, Kecamatan Lamongan, Kabupaten Lamongan, Jawa Timur, 62211.

Keywords:

sistem pendukung keputusan, *k-Means clustering*, penjualan, Google Colab, Omah Kopi.

Correspondent Email:

fdhani9@gmail.com

Abstrak. Kafe Omah Kopi by NSN merupakan usaha kuliner yang menyediakan berbagai jenis minuman kopi dan makanan pelengkap. Permasalahan utama yang dihadapi adalah kesulitan dalam mengidentifikasi produk yang paling laris serta menentukan strategi penjualan yang tepat. Selama ini, proses analisis data penjualan masih dilakukan secara manual dan belum berbasis data historis. Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan sistem pendukung keputusan (SPK) berbasis *k-Means clustering* untuk mengelompokkan produk berdasarkan jumlah penjualan dan total pendapatan. Hasil pengelompokan diklasifikasikan menjadi dua kategori, yaitu “Penjualan Tinggi” dan “Penjualan Rendah”. Sistem ini dikembangkan menggunakan bahasa pemrograman Python di lingkungan Google Colab, dengan antarmuka berbasis kode yang disusun secara terstruktur. Evaluasi sistem dilakukan melalui metode *black box testing*, *Elbow Method*, dan *Silhouette Coefficient* untuk mengukur kualitas pengelompokan. Hasil evaluasi menunjukkan bahwa sistem mampu mengelompokkan produk dengan akurasi dan konsistensi yang tinggi, sehingga dapat membantu pemilik kafe dalam merancang strategi promosi dan pengelolaan stok secara lebih efisien dan berbasis data.



Copyright © [JPI](#) (Jurnal Profesi Insinyur Universitas Lampung).

Abstract. *Omah Kopi by NSN is a culinary business that offers a variety of coffee beverages and complementary foods. One of the main problems faced is the difficulty in identifying best-selling products and determining effective sales strategies. To date, sales data analysis has been carried out manually and has not been data-driven. To address this issue, this study develops a decision support system (DSS) using the k-Means clustering algorithm to group products based on sales volume and total revenue. The clustering results are classified into two categories: "High Sales" and "Low Sales". The system is developed using the Python programming language within the Google Colab environment, with a structured code-based interface. System evaluation is conducted using black box testing, the Elbow Method, and the Silhouette Coefficient to assess clustering quality. The evaluation results show that the system effectively clusters products with high accuracy and consistency, thus helping business owners to design more targeted promotion strategies and efficient inventory management based on historical sales data.*

1. PENDAHULUAN

Kemajuan teknologi saat ini telah menjadi elemen penting dalam membantu pemilik usaha dalam mengambil keputusan yang lebih efektif dan strategis. Dengan memanfaatkan teknologi, pengelolaan bisnis dapat menjadi lebih efisien,

khususnya dalam mengatasi masalah yang berkaitan dengan stok produk dan strategi penjualan[1]. Kegagalan dalam mengelola stok dengan baik dapat menyebabkan dua skenario kerugian utama, yaitu kelebihan stok yang meningkatkan biaya operasional atau

kekurangan stok yang menyebabkan pelanggan beralih ke kompetitor[2].

Salah satu usaha yang menghadapi tantangan ini adalah Kafe Omah Kopi by NSN, sebuah usaha kuliner lokal yang berdiri sejak tahun 2016. Kafe ini awalnya dikenal sebagai penjual makanan cepat saji *fried chicken*, dan kemudian melakukan rebranding menjadi Kafe Omah Kopi by NSN untuk memperluas jangkauan pasar dan menyesuaikan diri dengan tren konsumsi kopi dan gaya hidup kekinian. Transformasi ini juga membawa perubahan dalam jenis produk yang ditawarkan, dari hanya menjual *fried chicken* menjadi menyajikan berbagai varian kopi dan makanan pendamping. Pemilihan Kafe Omah Kopi sebagai lokasi penelitian bukan tanpa alasan. Kafe ini mengalami proses rebranding yang cukup menarik, awalnya hanya menjual makanan cepat saji *fried chicken*, menjadi sebuah kafe dengan konsep yang lebih modern bernama Omah Kopi by NSN. Perubahan ini tentu memunculkan tantangan baru dalam pengelolaan bisnis, terutama dalam hal penjualan dan promosi. Meskipun secara tampilan dan menu sudah berkembang, sistem pengambilan keputusan di kafe ini masih dilakukan secara manual, tanpa bantuan data yang terstruktur. Hal ini menyulitkan pemilik usaha dalam memantau produk-produk mana yang paling diminati konsumen, serta menghambat efektivitas strategi penjualan. Keputusan yang diambil pun seringkali hanya berdasarkan pengamatan dan perkiraan, bukan dari analisis data yang objektif.

Demi mengatasi permasalahan tersebut, diperlukan sistem berbasis teknologi yang mampu membantu manajemen dalam mengolah dan menganalisis data penjualan secara sistematis. Salah satu pendekatan yang relevan adalah penerapan Sistem Pendukung Keputusan (SPK) dengan algoritma *k*-Means Clustering, yang berfungsi untuk mengelompokkan produk berdasarkan pola penjualannya. Dengan adanya segmentasi ini, perusahaan jadi lebih mudah untuk mengetahui mana produk yang laris dan mana yang kurang diminati. Hal ini tentu sangat berguna untuk menyusun strategi promosi dan pengaturan stok barang supaya lebih tepat sasaran dan efisien[3].

Menurut penelitian yang dimuat dalam Jurnal Teknologi dan Informasi (JATI) tahun 2023, segmentasi pelanggan pada transaksi

online retail dilakukan menggunakan algoritma *k*-Means dengan pendekatan *Elbow Method* dan eksplorasi data (EDA). Hasil penelitian tersebut menunjukkan bahwa metode ini cukup efektif dalam mengelompokkan data pelanggan, sehingga bisa membantu pelaku usaha, termasuk UMKM, untuk lebih memahami segmentasi pasar dan menyusun strategi pemasaran yang lebih tepat sasaran. Pemanfaatan big data bisa meningkatkan efektivitas dalam operasional bisnis dan membantu pengambilan keputusan yang lebih tepat, terutama dalam menghadapi perubahan pasar yang cepat[4]. Oleh karena itu, penerapan SPK berbasis *k*-Means pada Kafe Omah Kopi tidak hanya relevan, tetapi juga sangat potensial dalam mendukung peningkatan daya saing dan kepuasan pelanggan. Selain itu, penelitian ini juga bertujuan untuk mengembangkan sistem pendukung keputusan berbasis *k*-Means Clustering yang dapat membantu Kafe Omah Kopi dalam mengidentifikasi produk-produk terlaris dan merancang strategi penjualan berbasis data. Implementasi sistem ini diharapkan dapat memberikan rekomendasi yang akurat, mendukung pengelolaan persediaan secara efisien, serta menjadi landasan yang kuat dalam merumuskan kebijakan promosi untuk mendukung pertumbuhan bisnis kafe ke depannya.

Beberapa penelitian terdahulu menunjukkan bahwa pendekatan serupa telah berhasil diimplementasikan dalam konteks bisnis kuliner. Misalnya, penelitian oleh Syahputra juga melakukan pengelompokan menu F&B dengan algoritma *k*-Means menggunakan data penjualan tahunan. Hasilnya terbukti efektif untuk memisahkan menu yang populer dan kurang populer, sehingga memudahkan pelaku usaha dalam memprioritaskan stok yang tepat[5]. Penelitian lain mengembangkan SPK pada usaha kedai kopi menggunakan *k*-Means berbasis harga dan kuantitas penjualan, yang terbukti mampu meningkatkan efisiensi distribusi bahan baku[6]. Selain itu, studi yang menggabungkan *k*-Means dengan visualisasi dashboard interaktif juga mampu membantu manajer restoran mengambil keputusan cepat terhadap fluktuasi penjualan harian[7]. Pada ranah yang berbeda namun metodologinya serupa, menerapkan *k*-means dalam analisis klasifikasi data industri rumahan berbasis Web-SIG dan CRISP-DM, menghasilkan tiga

klaster dengan validitas tinggi (DBI = 0,184) untuk pengelompokan kategori usaha (pemula, berkembang, maju)[8].

2. TINJAUAN PUSTAKA

2.1 Strategi Penjualan

Strategi penjualan adalah serangkaian tindakan yang dirancang untuk meningkatkan hasil penjualan serta memaksimalkan kepuasan pelanggan. Strategi ini meliputi kegiatan seperti promosi, penyesuaian harga, hingga pengelolaan stok barang secara optimal. Pemanfaatan data analitik bisa bantu perusahaan lebih paham sama kebiasaan dan kebutuhan pelanggan. Hal ini bikin strategi pemasaran jadi lebih tepat sasaran dan keputusan bisnis bisa lebih cepat menyesuaikan diri dengan perubahan pasar[9]. Algoritma *k*-Means berhasil mengelompokkan pasar laptop di Jabodetabek menjadi enam segmen berbeda berdasarkan karakteristik demografis dan perilaku pembelian. Dengan begitu, produsen bisa menentukan produk yang tepat, menetapkan harga yang pas, dan menjalankan strategi penjualan promosi spesifik untuk tiap segmen guna meningkatkan efektivitas strategi pemasaran mereka[10].

2.2 Sistem Pendukung Keputusan (SPK)

Sistem Pendukung Keputusan (SPK) adalah sistem berbasis komputer yang dirancang untuk membantu pengambilan keputusan yang kompleks dan tidak terstruktur. SPK mengintegrasikan data, model analitis, dan antarmuka pengguna untuk menghasilkan rekomendasi yang relevan bagi pengambil keputusan. Komponen SPK meliputi:

1. Subsistem Data: Mengelola informasi dan data historis.
2. Subsistem Model: Menggunakan algoritma matematis untuk analisis data.
3. Subsistem Antarmuka: Menyajikan informasi hasil analisis dalam format yang mudah dipahami oleh pengguna.

SPK berbasis *k*-Means dapat digunakan untuk mengelompokkan data penjualan untuk mendukung keputusan promosi dan manajemen stok[11].

2.3 *k*-Means Clustering

k-Means adalah algoritma clustering yang gunanya buat membagi data ke dalam beberapa kelompok (*k*) berdasarkan kemiripan

karakteristik. Metode ini banyak dipakai karena gampang diimplementasi dan cukup cepat, terutama saat dipakai untuk dataset besar[12]. Langkah kerja algoritma *k*-Means meliputi:

1. Penentuan Jumlah *Cluster* (*k*)

Nilai *k* dapat ditentukan menggunakan Rumus 1:

$$k = \sqrt{\frac{n}{2}} \quad (1)$$

Di mana *n* adalah jumlah data[13].

2. Inisialisasi *Centroid*

Centroid awal dipilih secara acak sebanyak *k*.

3. Penghitungan Jarak

Jarak antara setiap titik data ke *centroid* dihitung menggunakan rumus *Euclidean Distance*:

$$d(i, j) = \sqrt{\sum_{m=1}^N (x_{im} - c_{jm})^2} \quad (2)$$

Keterangan untuk Rumus 2:

$d(i, j)$: Jarak antara data *i* dan *centroid* *j*.

x_{im} : Koordinat data *i*.

c_{jm} : Koordinat *centroid* *j*.

4. Pengelompokan Data.

Data dikelompokkan ke *centroid* dengan jarak terdekat.

5. Pembaruan *Centroid*

Centroid dihitung ulang berdasarkan rata-rata posisi data dalam *cluster*:

$$c_j = \left(\frac{1}{|C_j|} \right) \sum_{x \in C_j} x \quad (3)$$

Keterangan untuk Rumus 2.3:

c_j : *Centroid* baru untuk *cluster* *j*.

$|C_j|$: Jumlah data dalam *cluster* *j*.

6. Iterasi

Proses diulangi hingga posisi *centroid* stabil[14].

7. Evaluasi kualitas cluster dilakukan menggunakan *Silhouette Coefficient*, yang dihitung dengan rumus:

$$S = \frac{b - a}{\max(a, b)} \quad (4)$$

Keterangan untuk Rumus 4:

$a(i)$: Jarak rata-rata dalam *cluster*.

$b(i)$: Jarak rata-rata ke *cluster* terdekat lainnya.

Selain menggunakan *Silhouette Coefficient*, proses evaluasi dalam menentukan jumlah *cluster* yang tepat pada algoritma *k*-Means

juga dilakukan dengan metode *Elbow*. *Elbow Method* digunakan untuk menentukan jumlah *cluster* (k) yang optimal berdasarkan nilai *inertia* (jumlah kuadrat jarak antara data dengan *centroid*-nya).

$$Inertia = \sum_{i=1}^n \|x_i - \mu_c(i)\|^2 \quad (5)$$

Keterangan untuk Rumus 5:

x_i : titik data ke- i .

$\mu_c(i)$: *centroid* dari *cluster* tempat data ke- i berada.

Semakin kecil *inertia*, semakin baik pengelompokan. Namun, menambah jumlah *cluster* terus-menerus tidak selalu meningkatkan performa secara signifikan. Oleh karena itu, kita mencari "tekukan" (*Elbow*) pada grafik *inertia* vs jumlah *cluster*.

2.4 Python

Python adalah bahasa pemrograman tingkat tinggi yang bersifat *interpreted*, *open-source*, dan mendukung berbagai paradigma pemrograman, termasuk imperatif, objektif, dan fungsional. Python dikenal luas karena sintaksnya yang sederhana dan mudah dipahami, sehingga sangat populer dalam bidang analisis data, kecerdasan buatan, pembelajaran mesin, dan data mining [15].

Python memiliki banyak pustaka (*library*) yang memudahkan pengolahan data, seperti:

- *pandas* untuk manipulasi data tabular,
- *numpy* untuk komputasi numerik,
- *scikit-learn* untuk algoritma machine learning,
- serta *matplotlib* dan *seaborn* untuk visualisasi data.

Penggunaan Python dalam penelitian ini berperan penting dalam proses preprocessing data, penerapan algoritma *k*-Means, evaluasi *clustering*, hingga visualisasi hasil. Karena sifatnya yang fleksibel dan didukung komunitas yang luas, Python menjadi pilihan utama untuk implementasi sistem pendukung keputusan berbasis data.

2.5 Google Colaboratory (Google Colab)

Google Colaboratory, atau yang lebih dikenal dengan Google Colab, adalah platform berbasis cloud yang disediakan oleh Google untuk menjalankan kode Python secara interaktif

melalui antarmuka Jupyter Notebook. Platform ini memungkinkan pengguna menjalankan pemrograman Python langsung di browser tanpa harus menginstal software atau konfigurasi sistem tambahan [16].

Keunggulan Google Colab meliputi:

- Gratis dan berbasis cloud, sehingga tidak memerlukan spesifikasi komputer tinggi,
- Mendukung penyimpanan dan pembacaan file dari Google Drive,
- Terintegrasi dengan pustaka Python populer seperti *scikit-learn*, *tensorflow*, dan *matplotlib*
- Cocok untuk pengolahan data berskala besar dan eksperimen machine learning secara kolaboratif.

Dalam penelitian ini, Google Colab digunakan sebagai lingkungan pengembangan utama untuk memproses data penjualan, menerapkan algoritma *k*-Means *Clustering*, dan menghasilkan strategi penjualan berbasis hasil pengelompokan. Dengan memanfaatkan Colab, seluruh proses penelitian dapat dilakukan secara efisien, transparan, dan dapat diakses kapan saja selama terhubung dengan internet.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen sistem berbasis data, yang bertujuan untuk mengembangkan dan menguji sistem pendukung keputusan (SPK) menggunakan algoritma *k*-Means *Clustering*. Metode ini dipilih karena mampu mengelompokkan data penjualan berdasarkan kemiripan nilai, sehingga dapat memberikan informasi yang objektif dan terstruktur kepada pemilik usaha dalam mengambil keputusan bisnis.

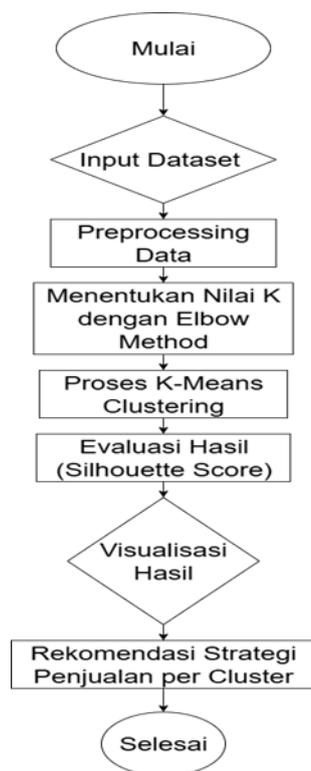
Jenis penelitian yang digunakan termasuk dalam rekayasa perangkat lunak terapan, di mana sistem dibangun untuk menyelesaikan permasalahan nyata di lapangan, khususnya dalam mengelompokkan produk berdasarkan tingkat penjualan di Kafe Omah Kopi by NSN. Penelitian dilakukan secara bertahap, mulai dari pengumpulan data, pengolahan data (*preprocessing*), pengembangan aplikasi, hingga evaluasi performa sistem menggunakan *Elbow Method* dan *Silhouette Coefficient*.

Melalui metode ini, peneliti berupaya untuk tidak hanya membangun sistem berbasis *k*-Means, tetapi juga mengevaluasi efektivitasnya

dalam mendukung strategi penjualan berbasis data historis penjualan yang telah dihimpun selama tiga tahun (2021–2024).

3.1 Perancangan Proses

Karena penelitian ini tidak menggunakan sistem basis data (database) maupun antarmuka pengguna grafis (UI), maka perancangan proses sistem digambarkan dalam bentuk flowchart algoritma, yang mewakili urutan logika kerja sistem dari input hingga output.



Gambar 1. Flow Chart

Penjelasan alur proses pada Gambar 1:

1. Mulai: Proses dimulai saat pengguna menjalankan skrip analisis di Google Colab.
2. *Input Dataset*: Dataset penjualan diunggah dalam format Excel (.xlsx) berisi kolom qty dan total_harga.
3. Preprocessing: Data dibersihkan dari nilai kosong, lalu dilakukan seleksi kolom dan normalisasi jika diperlukan.
4. Menentukan *k*: Sistem menggunakan metode *Elbow Curve* untuk menentukan jumlah *cluster* (*k*) optimal berdasarkan nilai WCSS.
5. Proses *Clustering*: Data diproses menggunakan algoritma *k*-Means untuk

mengelompokkan produk berdasarkan pola penjualan.

6. Evaluasi: Hasil *clustering* dievaluasi menggunakan *Silhouette Score* untuk mengukur *Validitas* dan pemisahan antar *cluster*.
7. Visualisasi dan Interpretasi: Sistem menampilkan grafik hasil *clustering* dan tabel kategori seperti *Penjualan Tinggi* dan *Rendah*.
8. Rekomendasi Strategi: Setiap *cluster* diberi strategi bisnis, misalnya: stok ulang, promosi, atau bundling produk.
9. Selesai: Sistem selesai dan dapat digunakan sebagai dasar pengambilan keputusan.

3.2 Perancangan Sistem

Perancangan sistem dilakukan dengan menyesuaikan alur logika pemrosesan data dalam format Google Colab. Metode pengembangan sistem yang digunakan mengadopsi pendekatan *Rapid Application Development* (RAD), yang terdiri dari tiga tahap utama:

1. Requirement Planning

Identifikasi kebutuhan utama, yaitu sistem yang mampu membaca data penjualan, mengelompokkan produk menggunakan *k*-Means, dan memberikan strategi berdasarkan pola hasil *clustering*.

2. Design Workshop

Merancang alur kerja sistem, algoritma yang digunakan, dan cara penyajian hasil (grafik *scatter plot*, tabel *cluster*, strategi rekomendasi) dalam lingkungan pemrograman Google Colab.

3. Implementation

Implementasi dilakukan menggunakan bahasa Python dengan memanfaatkan library seperti *pandas*, *numpy*, *sklearn*, *matplotlib*, dan *seaborn*. Seluruh proses dilakukan di Google Colab untuk efisiensi dan kemudahan integrasi.

3.3 Perancangan Dataset

1. Preprocessing Data

Preprocessing data adalah tahap awal yang penting dalam proses analisis data untuk memastikan data yang digunakan Valid dan siap diolah. Pada tahap ini dilakukan beberapa langkah, antara lain:

1. Data Mentah (Data Awal)

Data awal yang diperoleh dari Kafe Omah Kopi merupakan hasil rekap transaksi penjualan selama periode November 2021 sampai November 2024. Data ini terdiri atas berbagai atribut, seperti tanggal penjualan, nomor faktur, nama produk, jumlah barang terjual, harga satuan, dan total harga, dan berjumlah lebih dari 150.000 baris transaksi.

2. Data Cleaning

Langkah pertama dalam preprocessing adalah membersihkan data dari:

- Nilai kosong atau null;
- Duplikasi (baris data yang sama);
- Format yang tidak konsisten (misalnya format tanggal atau penulisan nama barang yang berbeda tapi merujuk ke produk yang sama).

3. Agregasi Data

Setelah data penjualan berhasil dibersihkan dari duplikasi, nilai kosong, dan format yang tidak sesuai, langkah selanjutnya adalah melakukan proses agregasi data. Agregasi ini dilakukan untuk menyederhanakan data penjualan berdasarkan nama barang. Dalam hal ini, setiap nama produk yang muncul berulang kali dalam transaksi penjualan akan digabungkan dan dijumlahkan nilai *quantity* serta *total_harga*.

Tujuan dari agregasi ini adalah agar sistem dapat menganalisis performa penjualan setiap produk secara keseluruhan, bukan per transaksi. Dengan cara ini, data yang semula terdiri dari banyak baris transaksi menjadi lebih ringkas dan terstruktur, sehingga lebih mudah untuk dianalisis pada tahap *clustering* selanjutnya. Agregasi dilakukan dengan menjumlahkan total kuantitas yang terjual dan total pendapatan yang dihasilkan untuk masing-masing nama barang, dan hasilnya disimpan dalam bentuk data baru yang telah diringkas.

Langkah ini penting karena membantu menghindari bias dalam pengelompokan data. Misalnya, jika satu produk dijual ratusan kali dengan jumlah kecil, tanpa agregasi, ia bisa terlihat sama dengan produk lain yang hanya terjual sekali namun dalam jumlah besar. Dengan melakukan agregasi terlebih dahulu, sistem akan memberikan hasil *clustering* yang lebih akurat dan representatif berdasarkan performa penjualan setiap produk,

4. Penyusunan Dataset Akhir

Dataset yang telah dibersihkan disusun kembali dengan format yang lebih sederhana

dan siap digunakan untuk proses *clustering*. Fokus utama dari pengelompokan adalah pada dua atribut, yaitu Quantity (jumlah produk terjual) dan Total Harga, karena kedua atribut tersebut merepresentasikan tingkat penjualan, berikut adalah dataset hasil agregasi pada Tabel 1:

Table 1. Dataset

No	Jumlah Terjual (QTY)	Total Harga (Rp)
1	5	25000
2	2	16000
3	1	7000
4	1	7000
5	5	60000
6	1	8000
7	1	10000
8	5	50000
9	1	10000
10	3	30000
11	1	12000
12	1	10000
13	1	17000
14	2	16000
15	1	12000
16	6	48000
17	2	10000

2. Proses Clustering

Setelah melalui tahap preprocessing data, dataset yang digunakan untuk proses clustering harus terstruktur dengan baik dan memiliki kesinambungan antar atributnya. Dataset yang telah diproses melalui tahap normalisasi digunakan sebagai input utama. Atribut seperti QTY (jumlah barang yang terjual), Harga (harga satuan barang), dan Total Harga (hasil perhitungan $QTY \times \text{Harga}$) telah dipastikan memiliki keterkaitan yang logis. Hal ini penting agar pola yang terbentuk dalam proses *clustering* dapat mencerminkan realitas penjualan.

3. Evaluasi Clustering

Untuk mengevaluasi hasil pengelompokan menggunakan algoritma *k*-Means, dilakukan dua metode evaluasi yaitu *Elbow Method* dan *Silhouette Coefficient*. Tujuannya adalah untuk mengetahui seberapa baik data dikelompokkan, serta menentukan jumlah cluster (*k*) yang paling optimal.

4. HASIL DAN PEMBAHASAN

4.1 Implementasi

Tahap implementasi dilakukan setelah proses perancangan sistem selesai disusun. Sistem dibangun menggunakan bahasa pemrograman Python dan dijalankan melalui platform Google Colaboratory (Colab). Penggunaan Google Colab dipilih karena memungkinkan pengguna untuk menjalankan program langsung di browser tanpa harus menginstal software tambahan.

Beberapa library Python yang digunakan dalam implementasi sistem ini antara lain:

- pandas untuk membaca dan mengolah data penjualan,
- numpy untuk membantu perhitungan matematis,
- matplotlib dan seaborn untuk menampilkan visualisasi grafik,
- scikit-learn untuk menerapkan algoritma *k*-Means *Clustering*, termasuk proses evaluasi menggunakan *Silhouette Score* dan penentuan jumlah *cluster* optimal dengan *Elbow Method*.

Sistem ini tidak menggunakan database karena seluruh data penjualan diinput melalui file Excel (.xlsx) yang diunggah secara manual oleh pengguna. Setelah data diunggah, sistem akan melakukan preprocessing, seperti pengecekan data kosong, normalisasi, dan pemilihan kolom yang relevan.

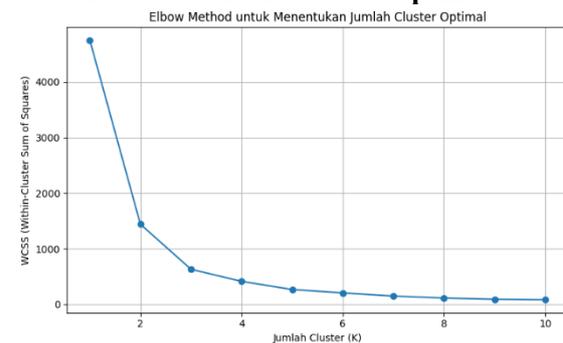
4.2 Proses Input Data Menjadi Dataset

Merupakan bagian dari proses awal dalam sistem, yaitu menggabungkan dan menyiapkan data penjualan sebelum dilakukan pengelompokan (*clustering*). Di dalam fungsi *load_and_prepare_data()*, sistem membaca tiga file Excel yang masing-masing berisi data penjualan, kemudian digabungkan menjadi satu data utama menggunakan perintah *pd.concat()*. Hal ini dilakukan agar semua data dari berbagai periode atau cabang dapat dianalisis secara menyeluruh.

Setelah itu, sistem hanya mengambil kolom yang dibutuhkan, yaitu nama barang, jumlah yang terjual (*quantity*), dan total pendapatan (*total_harga*). Data yang kosong atau tidak lengkap langsung dibuang dengan perintah *dropna()* supaya tidak mengganggu proses analisis. Langkah selanjutnya adalah mengelompokkan (menggabungkan) data berdasarkan nama barang. Jadi, jika satu produk muncul beberapa kali di data penjualan, semua

jumlah dan pendapatannya akan dijumlahkan agar lebih rapi dan mudah dianalisis. Hasil akhirnya adalah data yang sudah bersih dan siap digunakan untuk proses *clustering* menggunakan algoritma *k*-Means. Penggunaan Google Colab dipilih karena memungkinkan pengguna untuk menjalankan program langsung di browser tanpa harus menginstal software tambahan.

4.3 Proses Mencari *Cluster* Optimal

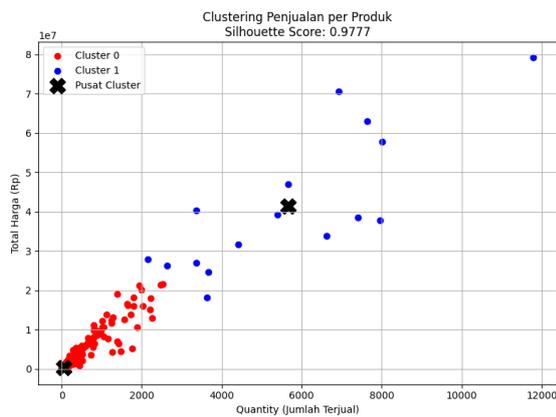


Gambar 2. Elbow Method

Gambar 2 adalah visualisasi dari *Elbow Method*, yaitu salah satu cara yang digunakan untuk menentukan jumlah *cluster* yang paling cocok dalam proses pengelompokan data menggunakan algoritma *k*-Means *Clustering*. Grafik ini menampilkan hubungan antara jumlah *cluster* (ditunjukkan pada sumbu X) dengan nilai WCSS atau *Within-Cluster Sum of Squares* (ditunjukkan pada sumbu Y). WCSS sendiri adalah nilai yang menunjukkan seberapa rapat data dalam satu *cluster*; semakin kecil nilainya, berarti data dalam *cluster* tersebut semakin mirip atau saling mendekati.

Dari grafik terlihat bahwa saat jumlah *cluster* bertambah dari 1 sampai 10, nilai WCSS semakin menurun. Namun, penurunan yang paling tajam terjadi di titik $k = 2$, setelah itu grafik mulai melandai. Titik melandainya grafik ini disebut sebagai "siku" atau *Elbow point*, dan menjadi acuan untuk menentukan jumlah *cluster* yang optimal. Berdasarkan grafik ini, jumlah *cluster* yang ideal untuk data penjualan di penelitian ini adalah 2 *cluster*, karena setelah angka tersebut penurunan WCSS tidak terlalu signifikan lagi. Dengan memilih 2 *cluster*, sistem dapat mengelompokkan produk penjualan secara lebih efisien dan relevan untuk dianalisis lebih lanjut.

4.4 Hasil Grafik *Clustering* Produk



Gambar 3. Hasil *Clustering* Produk

Gambar 3 menunjukkan hasil pengelompokan (*clustering*) produk berdasarkan data penjualan menggunakan algoritma *k*-Means. Dalam grafik ini, sumbu X menampilkan jumlah produk yang terjual (*quantity*), sedangkan sumbu Y menunjukkan total pendapatan atau total harga (*total_harga*) dari setiap produk. Titik-titik dalam grafik mewakili masing-masing produk, yang telah dikelompokkan ke dalam dua *cluster* berbeda. *Cluster 0* ditandai dengan warna merah, sedangkan *Cluster 1* berwarna biru. Setiap *cluster* memiliki ciri khas tertentu berdasarkan jumlah dan nilai penjualannya.

Tanda silang hitam besar di tengah-tengah masing-masing kelompok menandakan pusat *cluster*, yaitu titik rata-rata dari data dalam *cluster* tersebut. Kehadiran pusat ini membantu kita memahami letak konsentrasi data dari masing-masing kelompok. Selain itu, pada bagian atas grafik juga tertera nilai *Silhouette Score* sebesar 0,9777, yang merupakan indikator kualitas hasil *clustering*. Semakin mendekati angka 1, berarti pembagian *cluster* yang dihasilkan semakin baik dan setiap data sudah berada di kelompok yang paling sesuai. Dengan hasil seperti ini, sistem dapat membantu pemilik usaha untuk menyusun strategi penjualan yang lebih terarah, seperti menentukan produk mana yang layak dipromosikan, ditambah stoknya, atau bahkan dievaluasi ulang.

4.5 Hasil Penjualan Per *Cluster*

Gambar 4. Hasil Strategi Penjualan Per *Cluster*

=== Ringkasan Hasil Penjualan per *Cluster* ===

Cluster 0 (Total 2359 produk):

Rata-rata kuantitas: 43.87

Rata-rata harga: Rp393,320.35

Contoh karakteristik produk:

quantity	total_harga
1.0	7000.0
1.0	8000.0
2.0	14000.0

Cluster 1 (Total 16 produk):

Rata-rata kuantitas: 5660.62

Rata-rata harga: Rp41,444,470.31

Contoh karakteristik produk:

quantity	total_harga
2626.0	26258000.0
4407.0	31679000.0
3658.0	24576000.0

Gambar 5. Ringkasan Hasil *Clustering*

Gambar 4 merupakan ringkasan hasil pengelompokan produk berdasarkan data penjualan menggunakan metode *k*-Means *Clustering*. Sistem berhasil membagi seluruh produk menjadi dua kelompok (*cluster*), yaitu *Cluster 0* dan *Cluster 1*, dengan karakteristik yang sangat berbeda. Pada *Cluster 0*, terdapat sebanyak 2.359 produk, yang rata-rata hanya terjual sebanyak 43,87 unit per produk, dengan rata-rata total pendapatan sebesar Rp 393.320,35. Data ini menunjukkan bahwa produk-produk di *Cluster 0* termasuk dalam kategori penjualan rendah, karena kuantitas dan nilai penjualannya tergolong kecil. Contoh data dari *cluster* ini terlihat dari beberapa produk yang hanya terjual 1–2 kali, dengan total harga penjualan di bawah Rp 15.000.

Sementara itu, *Cluster 1* hanya terdiri dari 16 produk, namun masing-masing produk memiliki rata-rata kuantitas penjualan yang sangat tinggi, yaitu 5.660,62 unit, dan menghasilkan rata-rata total pendapatan sebesar Rp 41.444.470,31. Produk-produk dalam *cluster* ini bisa dikatakan termasuk dalam kategori produk unggulan atau best seller, karena tingkat penjualannya jauh lebih tinggi dibandingkan produk lain. Beberapa contohnya bahkan mencapai penjualan ribuan unit dengan nilai pendapatan di atas Rp 20.000.000,00. Dari ringkasan ini, sistem memberikan gambaran

yang jelas mengenai performa masing-masing produk, sehingga sangat membantu dalam menentukan strategi penjualan seperti peningkatan stok, promosi, atau evaluasi produk

4.6 Hasil Strategi Penjualan Per Cluster

=== Strategi Penjualan per Cluster ===

Cluster	Jumlah Produk	Rata-rata Kuantitas	Rata-rata Harga	Total Penjualan	Strategi Penjualan
0	2359	43.87	Rp393,320.35	Rp927,842,694.00	Penjualan Rendah - Perlu evaluasi lebih lanjut
1	16	5660.62	Rp41,444,470.31	Rp663,111,525.00	Penjualan Tinggi - Fokuskan promosi dan stok

Gambar 5 merupakan tabel strategi penjualan berdasarkan hasil *clustering* data penjualan produk. Tabel ini disusun setelah data dianalisis menggunakan algoritma *k*-Means, yang menghasilkan dua kelompok utama: *Cluster 0* dan *Cluster 1*. Pada *Cluster 0*, terdapat 2.359 produk yang secara rata-rata hanya terjual sekitar 43 unit dengan rata-rata harga sekitar Rp 393.000,00 per produk. Total penjualan dari seluruh produk dalam *cluster* ini mencapai lebih dari Rp 927.000.000,00, namun karena jumlah produknya sangat banyak, rata-rata performa tiap produk tergolong rendah. Maka dari itu, sistem memberikan strategi “Penjualan Rendah – Perlu evaluasi lebih lanjut,” artinya produk-produk dalam kelompok ini perlu ditinjau kembali apakah layak untuk dipertahankan, diganti, atau ditingkatkan melalui promosi.

Sementara itu, *Cluster 1* hanya berisi 16 produk, namun masing-masing produk memiliki rata-rata penjualan yang sangat tinggi, yaitu sekitar 5.660 unit dengan pendapatan rata-rata lebih dari Rp 41.000.000,00. Total penjualan dari *cluster* ini mencapai lebih dari Rp 663.000.000,00, meskipun produknya sedikit. Artinya, produk-produk dalam *cluster* ini memiliki potensi besar dan terbukti diminati konsumen. Oleh karena itu, strategi yang disarankan untuk *cluster* ini adalah “Penjualan Tinggi – Fokuskan promosi dan stok,” yang berarti produk-produk dalam *cluster* ini sebaiknya diprioritaskan dalam strategi penjualan, seperti diperbanyak stoknya atau lebih sering dipromosikan. Dengan analisis seperti ini, pemilik usaha dapat mengambil keputusan yang lebih terarah dan berdasarkan data nyata dari performa penjualan.

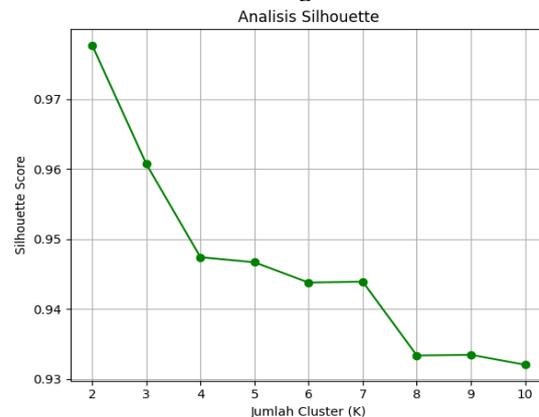
4.7 Ekspor Laporan

=== Download Laporan ===
[Download strategi_penjualan.csv](#)
[Download data_clustering.csv](#)

Gambar 6. Ekspor Laporan

Gambar 6 menunjukkan fitur unduh laporan (download report) yang disediakan oleh sistem setelah proses pengelompokan data penjualan selesai dilakukan. Dalam tampilan tersebut, pengguna diberikan dua tautan unduhan berupa file dengan format .csv, yaitu: *strategi_penjualan.csv* dan *data_clustering.csv*.

4.8 Evaluasi Clustering Silhouette Score



Gambar 7. Grafik Silhouette Score

Gambar 4.6 menampilkan hasil analisis menggunakan *Silhouette Score* untuk menentukan jumlah *cluster* yang paling optimal dalam proses pengelompokan data penjualan. *Silhouette Score* merupakan ukuran yang digunakan untuk menilai kualitas pengelompokan, di mana nilai yang mendekati angka 1 menunjukkan bahwa data dalam suatu *cluster* memiliki kesamaan tinggi dan cukup jauh jaraknya dari *cluster* lain. Pada grafik tersebut, sumbu horizontal menunjukkan jumlah *cluster* (*k*), sementara sumbu vertikal menunjukkan nilai *Silhouette Score* untuk masing-masing nilai *k*. Dari grafik dapat dilihat bahwa nilai tertinggi terdapat pada saat jumlah *cluster* adalah 2, dengan skor mendekati 0,98. Hal ini menunjukkan bahwa pembagian data ke dalam dua *cluster* memberikan hasil pengelompokan yang paling baik dan jelas, karena jarak antar *cluster* cukup jauh dan data di dalam *cluster* lebih seragam. Oleh karena itu, berdasarkan analisis ini, jumlah *cluster* yang paling optimal untuk data penjualan adalah dua.

5. KESIMPULAN

1. Kesimpulan

Kesimpulan berdasarkan penelitian yang telah dilakukan adalah sebagai berikut:

1. Berdasarkan penelitian ini, sistem pendukung keputusan (SPK) berbasis algoritma *k-Means Clustering* berhasil dibangun untuk mengelompokkan produk berdasarkan jumlah penjualan dan total pendapatan.
2. Sistem berhasil memproses dan mengelompokkan data penjualan dari Kafe Omah Kopi by NSN ke dalam dua kategori, yaitu “Penjualan Tinggi” dan “Penjualan Rendah”.
3. Evaluasi menggunakan metode *Silhouette Coefficient* menunjukkan bahwa sistem menghasilkan pemisahan cluster yang baik, dengan nilai silhouette tertinggi saat jumlah cluster (K) adalah dua.
4. Sistem menyediakan visualisasi hasil clustering dan rekomendasi strategi bisnis dalam bentuk tabel dan grafik, serta dapat mengeksplor laporan dalam format Excel.
5. Berdasarkan hasil evaluasi, sistem menghasilkan *Silhouette Score* sebesar 0,9777, yang mengindikasikan bahwa pembentukan cluster sangat baik dan representatif terhadap pola distribusi data penjualan.

2. Saran

Saran yang diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Sistem sebaiknya dikembangkan menggunakan database agar dapat menyimpan riwayat hasil clustering dan strategi penjualan secara berkelanjutan.
2. Visualisasi hasil clustering dapat ditingkatkan dengan menambahkan fitur filter berdasarkan periode waktu atau kategori produk agar memudahkan pengguna dalam analisis.
3. Jumlah cluster sebaiknya divariasikan secara dinamis dan dibandingkan lebih

lanjut menggunakan berbagai metrik evaluasi, seperti *Davies-Bouldin Index* atau *Calinski-Harabasz Score*.

4. Penelitian selanjutnya dapat mengintegrasikan algoritma clustering lain, seperti DBSCAN atau Hierarchical Clustering, sebagai pembanding terhadap hasil dari k-Means.
5. Pengembangan sistem dengan antarmuka web interaktif (GUI) berbasis Streamlit atau framework web lainnya dapat membantu pengguna non-teknis dalam mengakses dan mengoperasikan sistem secara lebih praktis.

DAFTAR PUSTAKA

- [1] M. Darmansyah, T., Aidin, W., Hadi, F., Husnah, M. A. S., Novaliza, A., & Ayumi, “Peran Teknologi dalam Meningkatkan Efisiensi Manajemen Kinerja,” *J. Penelit. Ilm. Multidisiplin.*, 2024.
- [2] K. S. R. Sugiartama and N. L. P. S. P. Pradnyani, “ANALISIS KETERSEDIAAN STOK BARANG TERHADAP VOLUME PENJUALAN DI CANGGU SHOP,” *Semin. Ilm. Nas. Teknol. Sains, dan Sos. Hum.*, vol. 5, no. 0 SE-Articles, Jan. 2023, [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/sintesa/article/view/2268>
- [3] E. F. L. Awalina and W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail,” *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [4] Alya Hastin Nafisah et al, “MEMAHAMI PENGARUH BIG DATA DALAM BISNIS MODERN DENGAN PENGAMBILAN KEPUTUSAN BERBASIS DATA Alya Hastin Nafisah, Jesslyn Olyviane, Rusdi Hidayat Nugroho, Indah Respati Kusumasari UPN “Veteran” Jawa Timur,” vol. 5, no. 4, 2024.
- [5] H. Syahputra, “Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means,” *J. KomtekInfo*, vol. 9, pp. 29–33, 2022, doi: 10.35134/komtekinfo.v9i1.274.
- [6] G. Triyandana, L. A. Putri, and Y. Umidah, “Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Triyandana, G., Putri, L. A., & Umidah, Y. (2022). Penerapan Data Mining

- Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means. Jou,” *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 40–46, 2022, doi: 10.30871/jaic.v6i1.3824.
- [7] H. A. Siregar, A. Azlan, and N. Y. Lumban Gaol, “Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 5, p. 750, 2023, doi: 10.53513/jursi.v2i5.8955.
- [8] H. Santoso, I. Rochmad, and S. Budiyanto, “Implementasi Etika Profesi Pelaksanaan Pemetaan Cerdas Lokasi Industri Rumahan Berbasis WEB-SIG di Dinas P3ACSKB Provinsi Kepulauan Bangka Belitung,” *J. Profesi Ins. Univ. Lampung*, vol. 6, no. 1 SE-ARTIKEL, Jun. 2025, doi: 10.23960/jpi.v6n1.149.
- [9] B. Erik Sibarani, S. Setiawan, T. Hadi, T. Williams, and T. Mkhize, “Penggunaan Data Analistik dalam Strategi Pemasaran untuk Mempertahankan Loyalitas Pelanggan,” vol. 3, no. 1, pp. 30–39, 2024, [Online]. Available: <https://doi.org/10.33050/mentari.v3i1>
- [10] T. C. Saputra, S. M. Fadhilah, and S. U. Mangkuto, “Segmentation , targeting and positioning analysis using k- means clustering model : A case study of the laptop market in Indonesia,” vol. 12, no. 2, pp. 195–203, 2024.
- [11] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*. Prentice Hall, 2011. [Online]. Available: <https://books.google.co.id/books?id=TtE0QgAACAAJ>
- [12] “K Means Clustering,” 2005, doi: 10.1201/9781420034912.ch3.
- [13] 2012 Madhulatha, “PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U) Kajian Multi Disiplin Ilmu untuk Mewujudkan Poros Maritim dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-ME,” pp. 978–979, 2015, [Online]. Available: https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/3333
- [14] M. . Rani Rotul Muhima, S.Si. *et al.*, *Kupas Tuntas Algoritma Clustering Konsep, Perhitungan Manual dan Program*. 2018.
- [15] W. McKinney, *Python for data analysis : data wrangling with Pandas, NumPy, and IPython*, Second edi. O’Reilly Media, Inc. [Online]. Available: <https://catalog.libraries.psu.edu/catalog/23192893>
- [16] “colab.google.” [Online]. Available: https://colab.google/?utm_source=chatgpt.com