

Segmentasi Berbasis Data Time Series Penjualan Produk Kopi Menggunakan Algoritma K-Means

Meri Anggaini¹, Herlawati Herlawati^{2*}, Rakhmat Purnomo³
^{1,2,3} Informatika/ Fakultas Ilmu Komputer/ Universitas Bhayangkara Jakarta Raya
*email: herlawati@ubharajaya.ac.id

DOI: <https://doi.org/10.31603/komtika.v9i2.15336>

Received: 14-11-2025, Revised: 25-11- 2025, Accepted: 27-11-2025

ABSTRACT

Coffee shops are businesses in the Food and Beverage (F&B) sector that contribute 7.15% to Indonesia's economy. The high demand for coffee has led to increasingly fierce competition. Kanae Coffee & Space in Bekasi faces challenges in maintaining customer loyalty and managing unpredictable demand. This study aims to apply the K-Means algorithm to cluster coffee products based on time series sales data, using the 6-step CRISP-DM methodology. The number of clusters was determined using the elbow method and confirmed with a silhouette coefficient of 0.5916 (good structure). The analysis resulted in five clusters with distinct characteristics: Cluster 0 (very low demand, stable trend, very high price), Cluster 1 (very high demand but sharply declining trend, very low price), Cluster 2 (moderately high demand, moderately stable trend, moderate price), Cluster 3 (moderate demand, slowly declining trend, moderately high price), and Cluster 4 (low demand, stable trend, moderately low price). These segmentation results are expected to serve as the basis for more effective marketing strategies and product management.

Keywords: Segmentation, Time Series, Coffee Products, K-Means, Coffee shop

ABSTRAK

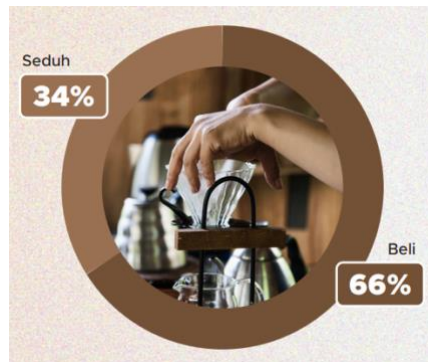
Coffee shop adalah usaha di bidang Food and Beverage (F&B) yang berkontribusi sebesar 7,15% terhadap perekonomian Indonesia. Tingginya minat konsumsi kopi memicu persaingan yang semakin ketat. Kanae Coffee & Space di Bekasi menghadapi tantangan dalam mempertahankan loyalitas pelanggan dan mengelola permintaan yang belum dapat diprediksi. Penelitian ini bertujuan menerapkan algoritma K-Means untuk mengelompokkan produk kopi berdasarkan data penjualan time series, dengan 6 tahapan CRISP-DM. Jumlah cluster ditentukan menggunakan metode elbow dan dikonfirmasi dengan silhouette coefficient sebesar 0,5916 (good structure). Hasil analisis menghasilkan lima cluster dengan karakteristik berbeda: Cluster 0 (permintaan sangat rendah, tren stabil, harga sangat mahal), Cluster 1 (permintaan sangat tinggi namun tren menurun tajam, harga sangat murah), Cluster 2 (permintaan cukup tinggi, tren cukup stabil, harga sedang), Cluster 3 (permintaan menengah, tren menurun perlahan, harga cukup mahal), dan Cluster 4 (permintaan rendah, tren stabil, harga cukup murah). Hasil segmentasi ini diharapkan menjadi dasar dalam strategi pemasaran dan pengelolaan produk yang lebih efektif.

Keywords: Segmentasi, Time Series, Produk Kopi, K-Means, Coffee shop

PENDAHULUAN

Industri Food and Beverage (F&B) adalah sektor bisnis yang bergerak di bidang makanan dan minuman. Berdasarkan www.id.crifasia.com, industri ini menyumbang 7,15% terhadap perekonomian Indonesia pada tahun 2024 dan tumbuh sebesar 4,53%, dengan peluang terus berkembang di tahun 2025. Salah satu bukti dari pesatnya pertumbuhan ini adalah menjamurnya kedai coffee shop yang mudah ditemukan. Coffee shop sendiri adalah usaha F&B yang menyajikan berbagai minuman kopi sebagai produk utama. Minum kopi di coffee shop

masih menjadi bagian dari tren gaya hidup masyarakat Indonesia. Menurut data dari goodstats.id, sebanyak 66% orang lebih memilih menikmati kopi di coffee shop karena cita rasanya yang khas berkat penggunaan mesin dan peralatan khusus. Selain itu, 40% responden mengaku dapat mengonsumsi dua gelas kopi dalam sehari, yang menunjukkan bahwa kopi telah menjadi bagian dari rutinitas masyarakat.



Sumber: <https://goodstats.id/article/survei-goodstats-kopi-jadi-bagian-dari-kehidupan-masyarakat-indonesia-D5iBT> (2024)

Gambar 1. Perbandingan Hasil Survei Antara Kopi Diseduh vs Kopi Beli di *Coffee shop*

Namun, di samping tingginya minat konsumsi kopi tersebut, tentu terdapat banyak persaingan pada bisnis coffee shop yang sejenis. Kondisi ini berdampak bagi para pemilik coffee shop yang menghadapi tantangan dalam mempertahankan loyalitas pelanggan. Oleh karena itu, penentuan strategi bisnis yang tepat diperlukan untuk mencegah dan meminimalkan potensi kerugian yang mungkin terjadi. Salah satu coffee shop yang mengalami situasi serupa adalah Kanae Coffee & Space. Frekuensi penjualan produk kopi pada Kanae Coffee & Space menunjukkan pola yang masih berfluktuasi dari waktu ke waktu. Selama ini, analisis data historis penjualan hanya dilakukan secara sederhana dan belum terdapat segmentasi produk untuk mengidentifikasi pola penjualan secara lebih mendalam. Kondisi ini membuat pengambilan keputusan menjadi kurang optimal, sehingga diperlukan pendekatan yang lebih sistematis melalui penerapan algoritma K-Means untuk melakukan segmentasi produk kopi berdasarkan data time series penjualan serta mengukur validitas cluster guna mendukung kualitas hasil yang terbentuk.

Sebagai dasar pertimbangan dalam memilih algoritma pengelompokkan untuk segmentasi produk, algoritma K-Means dipilih karena telah terbukti efektif dalam berbagai kasus serupa. Selain kemampuannya dalam mengelompokkan data secara efisien, K-Means juga mampu memberikan hasil klaster yang mudah diinterpretasikan oleh manajemen. K-Means merupakan salah satu metode klasterisasi yang membagi data ke dalam dua atau lebih kelompok berdasarkan kemiripan karakteristik antar data, sehingga sangat relevan untuk mengungkap pola penjualan produk kopi yang sebelumnya tidak terlihat melalui analisis manual [1]. Beberapa penelitian menunjukkan bahwa *K-Means* dapat diterapkan dalam mengelompokkan produk berdasarkan penjualan di toko sembako menjadi 3 kategori, dan dievaluasi menggunakan *Silhouette coefficient* dengan nilai sebesar 0,5793 [1]. Sementara itu, keunggulan algoritma *K-Means* juga terlihat dalam melakukan segmentasi pelanggan *mall* berdasarkan karakteristik tertentu. Dengan menggunakan metode *elbow*, diperoleh 4 *cluster* yang kemudian dievaluasi menggunakan *Silhouette coefficient* dengan nilai 0,553 [2]. Tak hanya itu, *K-Means* juga menunjukkan performa lebih baik dibandingkan *DB Scanner* dan

Hierarchical dalam melakukan analisis segmentasi pasar. Hasil didapatkan sebanyak 2 *cluster* dengan nilai *silhouette coefficient* tertinggi yaitu sebesar 0.716 [3].

Persaingan dalam industri kopi yang semakin ketat menuntut pelaku usaha untuk memahami pola permintaan produk secara lebih mendalam dan berbasis data. Namun, analisis penjualan yang dilakukan secara sederhana sering kali tidak mampu mengungkap perbedaan karakteristik tiap produk, terutama ketika data yang digunakan berupa *time series* yang memiliki dinamika tren dan fluktuasi dari waktu ke waktu. Hal ini menyebabkan *coffee shop* kesulitan dalam menentukan strategi pemasaran, pengendalian stok, serta evaluasi performa produk. Oleh karena itu, rumusan masalah penelitian ini adalah bagaimana mengidentifikasi segmentasi produk kopi secara lebih akurat menggunakan pendekatan klasterisasi yang dapat menggambarkan pola permintaan berdasarkan data *time series* penjualan.

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam melakukan segmentasi produk kopi berdasarkan perilaku penjualan *time series* sehingga dapat memberikan gambaran yang lebih jelas mengenai kelompok produk dengan karakteristik permintaan yang serupa. Penerapan metode ini dilakukan untuk menghasilkan *cluster* yang dapat menjadi dasar dalam evaluasi performa produk, penentuan strategi pengelolaan stok, serta perencanaan pemasaran yang lebih tepat sasaran. Selain itu, penelitian ini juga bertujuan menilai kualitas *cluster* yang terbentuk sehingga hasil segmentasi dapat dipercaya dan digunakan sebagai rekomendasi bagi pengambilan keputusan bisnis.

METODE

1. CRISP-DM

Cross-Industry Standard Process for Data Mining atau yang dikenal dengan *CRISP-DM* adalah standarisasi yang meliputi penggambaran alur kerja yang jelas dan terstruktur [4] serta membantu dalam melaksanakan proyek data mining yang dilakukan. Alur kerja *crisp-dm* (Gambar 2) terdiri dari 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [5] yaitu:

- a. *Business Understanding*. Tahap ini dilakukan untuk memastikan bahwa analisis data dan tujuan pemodelan dilakukan sesuai dengan tujuan bisnis yang ingin dicapai [6], yaitu mengelompokkan produk kopi untuk mengidentifikasi pola penjualan berdasarkan kemiripan permintaan pelanggan dari waktu ke waktu. Proses pengelompokan ini dilengkapi dengan evaluasi untuk memastikan kualitas segmentasi yang dihasilkan.
- b. *Data Understanding*. Tahap ini meliputi pengumpulan dan identifikasi sumber data, serta eksplorasi data [7]. Sumber data pada penelitian ini berasal dari data historis penjualan dari Kanae Coffee & Space selama 7 bulan terhitung dari bulan Juli 2024 sampai Januari 2025.
- c. *Data Preparation*. Setelah melakukan eksplorasi data, maka selanjutnya adalah memilih atribut yang relevan, melakukan transformasi data, meningkatkan kualitas data, serta melakukan normalisasi data [7].
- d. *Modeling*. Pada tahap ini digunakan teknik pemodelan dengan pendekatan *clustering* untuk melakukan segmentasi produk berdasarkan pola penjualan yaitu menggunakan algoritma *K-Means* karena mampu mengelompokkan data berdasarkan kemiripan karakteristik secara efektif [8].



Gambar 2. Tahapan Penelitian

- e. *Evaluation*. Tahap evaluasi difokuskan untuk menilai kualitas model *K-Means* yang dihasilkan, termasuk kesesuaian jumlah *cluster*, validitas segmentasi, serta memastikan seluruh tahapan sebelumnya telah dilakukan dengan benar dan mendukung tujuan analisis [9].
- f. *Deployment*. Tahap ini adalah tahap akhir yang bertujuan untuk memperoleh hasil dari proses pengujian, di mana hasil tersebut dapat memberikan wawasan dan informasi terkait pengelompokan data dalam pengambilan keputusan yang lebih tepat [7].

2. Metode Elbow

Metode *Elbow* adalah metode yang digunakan digunakan untuk memperoleh informasi guna menentukan *cluster* terbaik dengan menganalisis persentase perbandingan jumlah *cluster* yang memiliki bentuk titik-titik siku [10]. Proses ini mencakup menentukan jumlah *cluster* k dengan menganalisis nilai *Sum of Squared Errors* (*SSE*) pada setiap nilai k [11]. Rumus *Elbow* dituliskan sebagai berikut [10]:

$$SSE = \sum_{k=1}^K \sum_{i \in k} |X_i - C_k|^2 \quad (1)$$

Pada rumus (1), X_i adalah titik data dalam *cluster*, C_j adalah pusat *cluster* terdekat, k adalah jumlah *cluster*, dan n adalah jumlah total data.

3. K-Means Clustering

K-Means Clustering adalah algoritma yang termasuk ke dalam *unsupervised learning* (tidak mempunyai label) dan digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu [8]. Tujuannya adalah mengungkap pola tersembunyi dalam data, di mana data dalam satu kelompok memiliki kemiripan yang kuat satu sama lain [12]. Metode *clustering* ini dapat digunakan dalam analisa masalah bisnis, salah satunya untuk segmentasi pasar [13]. Proses pengelompokan dilakukan secara berulang hingga tidak ada data yang berpindah *cluster* [14], dengan menghitung jarak antara titik data dan pusat *cluster* menggunakan rumus *Euclidean Distance* [15]:

$$d_{(a,b)} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2)$$

Jarak antara dua objek, yaitu objek *a* dan objek *b*, dihitung menggunakan rumus *Euclidean Distance* di mana $d(a, b)$ menyatakan jarak antara kedua objek tersebut. Nilai *n* menunjukkan jumlah dimensi data, sementara a_i dan b_i masing-masing merepresentasikan koordinat dari objek *a* dan objek *b* pada dimensi ke-*i* dari *n* dimensi yang ada.

4. Silhouette coefficient

Silhouette coefficient adalah metrik yang mengukur kesamaan data dalam suatu *cluster* dan dihitung secara individu untuk setiap objek, di mana semakin mendekati nilai 1 menunjukkan kualitas pengelompokan yang baik, sedangkan nilai mendekati -1 menunjukkan pengelompokan yang kurang baik [10]. Dengan kata lain, *Silhouette coefficient* berfungsi untuk mengevaluasi kualitas serta kekuatan *cluster* yang dihasilkan. Adapun rumus *Silhouette coefficient* dirumuskan sebagai berikut [11]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

Setelah menghitung nilai *Silhouette coefficient*, tahap berikutnya adalah memahami standar untuk membaca hasil tersebut. Standar ini membantu melihat seberapa baik data yang dikelompokkan [16].

Tabel 1. Kriteria *Silhouette*

Score	Interpretation
0.71 – 1.00	Strong Structure
0.51 – 0.70	Good structure
0.26 – 0.50	Weak Structure
<=0.25	Bad Structure

5. Metode Pengumpulan Data

Dalam memperoleh informasi yang relevan guna mendukung analisis yang dilakukan dalam penelitian ini, peneliti melibatkan 3 cara, yaitu observasi dengan melakukan peninjauan langsung untuk mencatat fakta yang terjadi berikut dengan mengumpulkan data yang dibutuhkan, dilanjutkan dengan wawancara untuk mengetahui permasalahan yang terjadi, dan studi pustaka dengan mencari beberapa buku maupun penelitian sebelumnya dengan rentang waktu beberapa tahun terakhir yang berkaitan dengan topik penelitian yang dipilih, yaitu

algoritma *K-Means*. Adapun data yang digunakan untuk dianalisis lebih lanjut dapat dilihat pada Tabel 2.

Tabel 2. Data Penjualan Produk Kopi (Juli 2024 – Januari 2025)

<i>product</i>	<i>qty_</i> <i>Juli</i> <i>2024</i>	<i>qty_</i> <i>Agustus</i> <i>2024</i>	<i>qty_</i> <i>September</i> <i>2024</i>	<i>qty_</i> <i>Oktober</i> <i>2024</i>	<i>qty_</i> <i>November</i> <i>2024</i>	<i>qty_</i> <i>Desember</i> <i>2024</i>	<i>qty_</i> <i>Januari</i> <i>2025</i>	<i>harga_</i> <i>per_</i> <i>produk</i>
<i>Americano</i>	133	176	138	159	161	174	190	25000
<i>Beans</i>	1	2	1	1	0	2	0	25000
<i>Lokal Hot</i>								
<i>Japanese</i>								
.....								
<i>Okinawa</i>	142	99	90	84	80	86	79	28000
<i>Latte</i>								
<i>Red</i>	51	59	42	38	32	45	55	28000
<i>Presso</i>								

HASIL DAN PEMBAHASAN

Pembahasan pada hasil penelitian ini meliputi penerapan algoritma *K-Means* dalam melakukan segmentasi produk berdasarkan permintaan dari waktu ke waktu menggunakan metode *CRISP-DM* yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

1. Business Understanding

Kanae Coffee & Space berkomitmen untuk menghadirkan produk kopi dengan cita rasa autentik serta menjangkau berbagai segmen pelanggan. Namun, dinamika permintaan yang fluktuatif terhadap produk kopi yang dijual menimbulkan tantangan dalam memahami kebutuhan pelanggan dari waktu ke waktu yang pada akhirnya dapat berdampak pada kepuasan pelanggan itu sendiri. Penelitian ini bertujuan untuk mengelompokkan produk kopi guna mengidentifikasi pola penjualan berdasarkan kemiripan permintaan pelanggan dari waktu ke waktu, dengan menggunakan data historis penjualan Kanae Coffee & Space pada 26 jenis produk kopi yang terjual.

2. Data Understanding

Sumber data yang digunakan dalam penelitian ini adalah data historis penjualan dari Kanae Coffee & Space selama 7 bulan terhitung dari bulan Juli 2024 sampai Januari 2025. Data penjualan tersebut kemudian diolah menjadi 26 data berdasarkan jenis produk kopi yang dijual. Adapun atribut data yang tersedia terdiri dari *product* (nama produk), bulan, *sold_qty* (jumlah unit yang terjual), dan *harga_per_produk*.

3. Data Preparation

Langkah awal tahap ini dimulai dengan memilih atribut yang relevan agar data mencerminkan informasi yang dibutuhkan dan berkontribusi signifikan terhadap pengelompokan. Atribut yang digunakan telah melalui seleksi dan tercantum pada Tabel 3.

Tabel 3. Atribut Yang Digunakan

No	Atribut	Keterangan
1.	<i>product</i>	Menunjukkan nama produk kopi yang terjual.
2.	<i>qty</i> Juli 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan Juli 2024.
3.	<i>qty</i> Agustus 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan Agustus 2024.
4.	<i>qty</i> September 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan September 2024.
5.	<i>qty</i> Oktober 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan Oktober 2024.
6.	<i>qty</i> November 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan November 2024.
7.	<i>qty</i> Desember 2024	Menunjukkan jumlah unit produk kopi yang terjual pada bulan Desember 2024.
8.	<i>qty</i> Januari 2025	Menunjukkan jumlah unit produk kopi yang terjual pada bulan Januari 2025.
9.	harga per produk	Menunjukkan harga tiap unit produk kopi.

Setelah memilih atribut yang digunakan kemudian dilanjutkan dengan melakukan transformasi data, mengecek dan menangani *missing value*, melakukan *feature engineering* serta melakukan normalisasi data menggunakan *StandardScaler*.

product	qty_Juli_2024	qty_Agustus_2024	qty_September_2024	qty_Oktober_2024	qty_November_2024	qty_Desember_2024	qty_Januari_2025	harga_per_produk	slope	mean_qty	std_qty
Americano	0.518781655	1.183112562	1.011312163	1.596958978	1.383069336	1.613799647	1.778455492	-0.767209496	0.86167	1.25931	0.58539
Beans Lokal Hot Japanese	-0.801054254	-0.778911349	-0.825655672	-0.870249552	-0.829207409	-0.817774484	-0.828538612	-0.767209496	0.59873	-0.82603	-0.88753
Beans Lokal Hot V60	-0.561084088	-0.519563361	-0.557485185	-0.573559919	-0.554390422	-0.591581542	-0.567839202	-0.767209496	0.40322	-0.56444	-0.74037
Beans Lokal Iced Japanese	-0.391105221	-0.31659537	-0.329540271	-0.136330391	-0.073460695	-0.464348012	-0.581560223	-0.450583355	0.03915	-0.33624	-0.01339
Beans Lokal Iced V60	-0.801054254	-0.76763535	-0.839064196	-0.839019064	-0.774244011	-0.817774484	-0.828538612	-0.450583355	0.61896	-0.8149	-0.83121
Black Yuzu	-0.471095276	-0.54211536	-0.369765844	-0.386176992	-0.38950023	-0.577444483	-0.238534683	-0.292270284	0.73357	-0.43457	-0.35709
Cappuccino	0.088835109	0.066788612	0.206800702	-0.027026384	0.118911196	-0.195743893	-0.156208554	-0.292270284	-0.68223	0.02184	0.28899
Coffee Basil	-0.431100249	-0.294043371	-0.423399942	-0.308100773	-0.458204476	-0.337114482	-0.375744899	-0.292270284	0.37625	-0.38076	-0.52151
Espresso	-0.691067928	-0.654875355	-0.597710758	-0.636020894	-0.719280614	-0.74708919	-0.732491461	-0.767209496	0.37625	-0.68874	-0.61874
Flat White	-0.721064199	-0.778911349	-0.798838623	-0.760942845	-0.719280614	-0.775363308	-0.609002266	-0.292270284	0.84144	-0.7444	-0.60008
Guest Beans Hot V60	-0.801054254	-0.801463348	-0.839064196	-0.885864796	-0.815466559	-0.775363308	-0.746212482	2.399051916	0.80099	-0.8149	-0.75895
Guest Beans Iced Japanese	-0.791055497	-0.801463348	-0.839064196	-0.807788577	-0.774244011	-0.817774484	-0.677607374	2.715678057	0.86167	-0.79449	-0.66382
Guest Beans Iced V60	-0.801054254	-0.801463348	-0.839064196	-0.885864796	-0.829207409	-0.831911543	-0.828538612	2.715678057	0.63919	-0.82603	-0.88753
Honey Star Latte	2.828494497	2.750476491	2.539883938	2.440182146	2.56478238	2.504434358	2.478227594	-0.292270284	-2.94751	2.62852	2.99201
Kanae Machiatto	1.598647399	1.577772544	2.070585586	1.76872666	1.877739913	1.58552553	1.915665708	0.024355857	-0.86426	1.77323	0.91783
Kopi Susu Kanae	2.828494497	2.750476491	2.539883938	2.440182146	2.56478238	2.504434358	2.478227594	-0.767209496	-2.94751	2.62852	2.99201
Kuro Latte	-0.17113257	-0.260215373	-0.182046504	-0.292485529	-0.197128339	-0.09678448	-0.12876651	-0.292270284	0.40996	-0.19152	-0.36337
Latte	0.658764252	0.472724594	0.66269053	0.784966297	0.421209882	0.765576113	0.447516397	-0.292270284	-0.59459	0.60254	0.42691
Majic	-0.471095276	-0.587219358	-0.704978953	-0.682866626	-0.678058066	-0.605718601	-0.677607374	-0.292270284	0.14028	-0.62566	-0.32013
Mocca	-0.6510729	-0.801463348	-0.839064196	-0.885864796	-0.829207409	-0.831911543	-0.828538612	-0.292270284	0.30209	-0.80933	-0.49836
Mocha Ice	-0.681069171	-0.666151354	-0.678161904	-0.792173333	-0.746762313	-0.704678013	-0.677607374	-0.292270284	0.51783	-0.70915	-0.73351
Mocha Ice Cream	0.068837596	0.472724594	0.394520043	0.550737639	0.517395827	0.694890818	0.324027202	0.974234281	0.53131	0.41516	-0.14485
Okinawa Latte	0.608770467	0.314860601	0.367702995	0.425815688	0.270060539	0.369738464	0.255422094	-0.292270284	-0.89123	0.38547	0.69563
Red Presso	-0.301116409	-0.136179378	-0.275906174	-0.292485529	-0.38950023	-0.209880951	-0.0738882424	-0.292270284	0.45041	-0.24162	-0.23468
Shibuya Cookie	0.218818949	-0.023419384	0.126349556	0.004204104	0.146392895	0.072860227	-0.142487532	-0.292270284	-0.50694	0.06451	0.2995
Tokyo Bana Latte	0.11883138	-0.057247382	0.019081362	0.051049836	-0.087201544	0.086997286	0.022164727	-0.292270284	-0.06872	0.02369	-0.02317

Gambar 3. Hasil Data Preparation

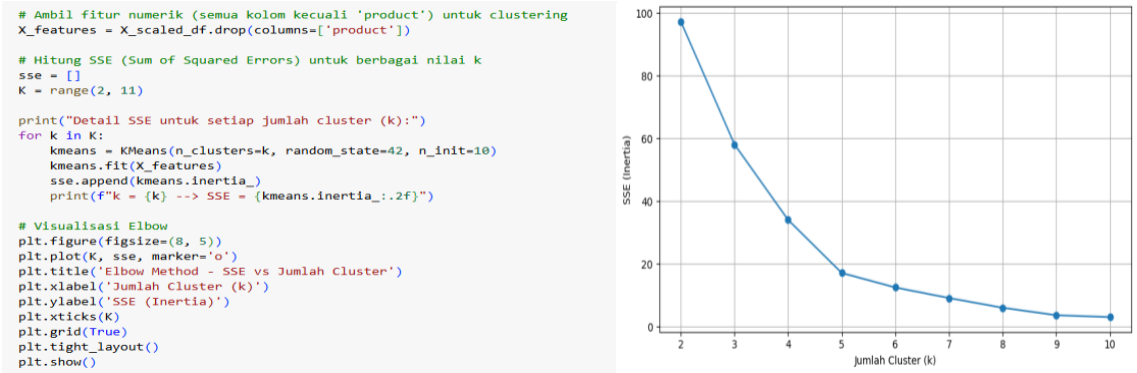
Gambar 3 Menunjukkan hasil dari *data preparation* yang mana memperlihatkan bahwa semua fitur telah dinormalisasi atau berada dalam rentang yang sebanding dengan menghasilkan 3 atribut baru dari hasil *feature engineering* yaitu *slope* (arah tren penjualan), *mean_qty* (rata-rata penjualan per produk) dan *std_qty* (tingkat variasi penjualan dari bulan ke bulan). Setelah proses tersebut dijalankan, data dinyatakan siap untuk diproses ke tahap selanjutnya.

4. Modeling

Pada tahap *modeling* ini, algoritma yang digunakan adalah algoritma *K-Means* untuk mengelompokkan produk kopi berdasarkan kemiripan pola penjualan dari waktu ke waktu. Data yang masuk ke dalam pengolahan adalah data yang telah direduksi sebanyak 26 data yang menunjukkan nama-nama produk kopi yang terjual. Proses pengolahan dan analisis menggunakan tools *Google Colab* dengan bahasa pemrograman *Python*.

a. Menentukan Jumlah *Cluster* Dengan Metode *Elbow*

Metode ini menghitung *SSE*, total jarak data ke pusat *cluster*, untuk berbagai jumlah *cluster* dari 2 ke atas. Hasilnya digambarkan pada Gambar 3 dengan sumbu *x* menunjukkan jumlah *cluster* dan sumbu *y* menunjukkan nilai *SSE*. Jumlah *cluster* dipilih berdasarkan titik di mana penurunan *SSE* melambat [17], seperti terlihat pada Gambar 4.



(a) (b)
Gambar 4. (a)Proses Metode *Elbow* (b)Hasil Metode *Elbow*

Gambar 4 memperlihatkan grafik *elbow* yang digunakan untuk menentukan jumlah *cluster* (*k*) sebelum melakukan pemodelan menggunakan algoritma *K-Means*. Nilai *SSE* tampak menurun tajam dari *k* = 2 hingga *k* = 5, lalu penurunannya melambat setelah *k* = 5. Rincian nilai *SSE* untuk setiap jumlah *cluster* disajikan pada Tabel 4.

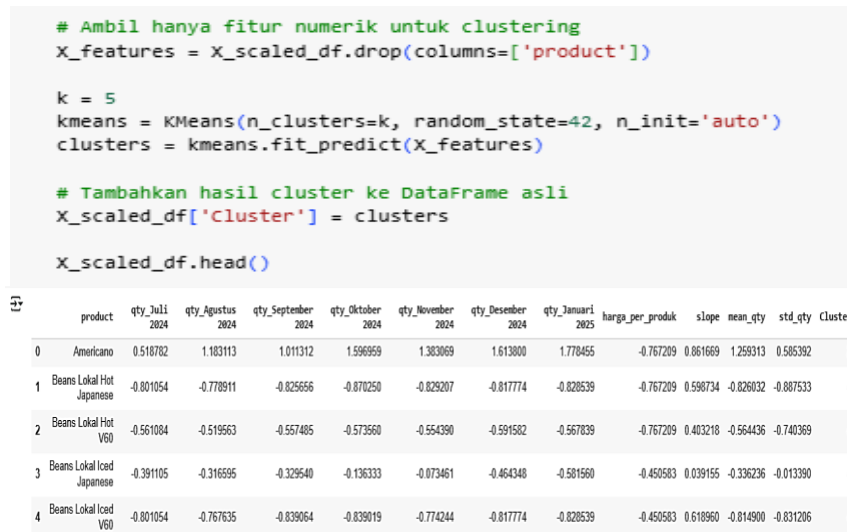
Tabel 4. *Elbow* Terhadap Penurunan Nilai *SSE*

Jumlah Cluster (<i>k</i>)	Nilai <i>SSE</i>
2	97,29
3	58,09
4	34,02
5	17,10
6	12,47
7	9,11
8	6,01
9	3,63
10	3,02

Berdasarkan Tabel 4, penurunan nilai *SSE* paling signifikan terjadi pada rentang *k* = 2 hingga *k* = 5, yaitu dari 97,29 menjadi 17,10. Setelah *k* = 5, penurunan nilai *SSE* mulai melambat, Oleh karena itu, jumlah *cluster* yang digunakan dalam penelitian ini ditetapkan sebanyak 5 *cluster*, karena di titik tersebut penurunan *SSE* mulai melambat.

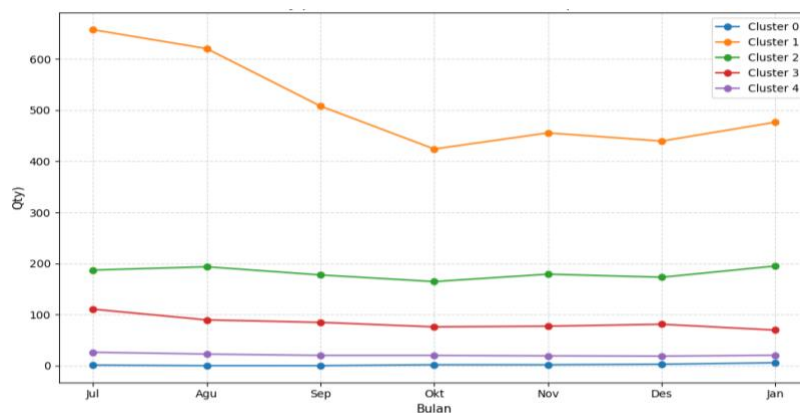
b. Penerapan Algoritma *K-Means*

Model *K-Means* diterapkan untuk membagi data ke dalam 5 kelompok atau *cluster* berdasarkan kemiripan nilai antar fitur. Proses pengelompokkan dilakukan pada atribut yang yang betipe numerik. Hasil pengelompokan ini disimpan dalam kolom baru bernama ‘*Cluster*’.



Gambar 5. Modeling K-Means

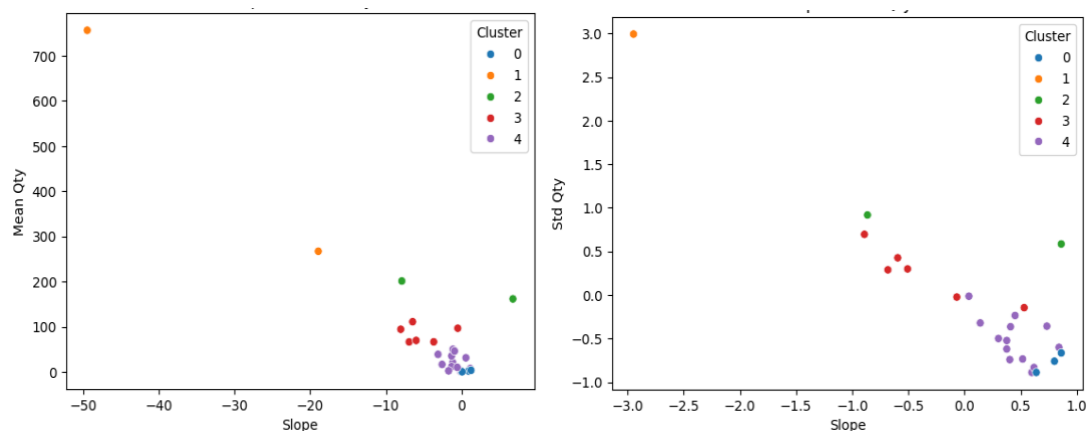
Gambar 6 menunjukkan pola tren penjualan rata-rata (Qty) dari Juli 2024 hingga Januari 2025 untuk masing-masing *cluster*, di mana setiap *cluster* menunjukkan pola penjualan yang berbeda. *Cluster 0* mencerminkan produk dengan volume penjualan sangat rendah dan tren permintaan stabil. *Cluster 1* berisi produk dengan volume penjualan tertinggi namun tren permintaannya menurun tajam. *Cluster 2* memiliki volume penjualan cukup tinggi dengan tren permintaan yang cukup stabil, menunjukkan permintaan yang konsisten. *Cluster 3* mencerminkan produk dengan volume penjualan menengah dan tren permintaan menurun perlahan. Sedangkan *Cluster 4* menunjukkan volume penjualan rendah namun stabil.



Gambar 6. Grafik Hasil Tren *clustering*

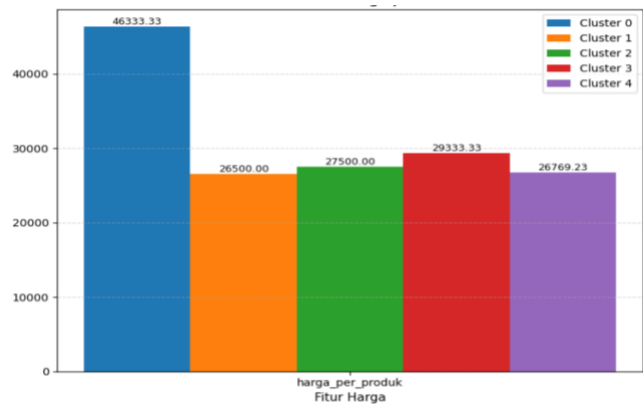
Selain itu, hasil analisis juga didukung berdasarkan hasil visualisasi pada Gambar 7 yang menunjukkan bahwa *cluster 0* merepresentasikan produk dengan penjualan sangat rendah, tren stabil dengan sedikit pertumbuhan (*slope* positif dan mendekati nol), *mean_qty* dan *std_qty* sangat rendah, menandakan minat produk yang sangat rendah. *Cluster 1* terdiri dari produk dengan volume penjualan tertinggi (*mean_qty* tinggi) namun mengalami tren penurunan tajam (*slope* sangat negatif) serta fluktuasi besar (*std_qty* tinggi), menunjukkan popularitas produk yang menurun drastis. *Cluster 2* mencakup produk dengan penjualan cukup tinggi, tren cukup stabil (*slope* mendekati nol), dan fluktuasi sedang (*std_qty* sedang), menggambarkan

permintaan yang cukup konsisten dan cocok untuk kebutuhan rutin. *Cluster 3* berisi produk dengan penjualan menengah, tren perlahan menurun (*slope* negatif), dan fluktuasi lebih rendah dari *cluster 2*, mengindikasikan produk yang mulai memasuki fase penurunan. Sedangkan *cluster 4* menunjukkan produk dengan penjualan rendah dan fluktuasi yang kecil (*slope* dan *std_qty* mendekati nol), yang menandakan permintaan pasar yang rendah dan konsisten kurang diminati dari waktu ke waktu.



Gambar 7. Grafik Hasil Tren *Clustering*

Kemudian dianalisis atribut rata-rata harga pada tiap *cluster*. Gambar 8 menunjukkan *cluster 0* memiliki rata-rata harga sangat tinggi, tergolong segmen produk berharga sangat tinggi. *Cluster 1* paling murah dan masuk kategori sangat murah. *Cluster 2* termasuk harga sedang. *Cluster 3* cukup tinggi meskipun masih jauh di bawah *cluster 0*. *Cluster 4* sedikit lebih rendah dari *cluster 2* dan dikategorikan cukup murah.



Gambar 8. Rata-Rata Harga Produk Tiap *Cluster*

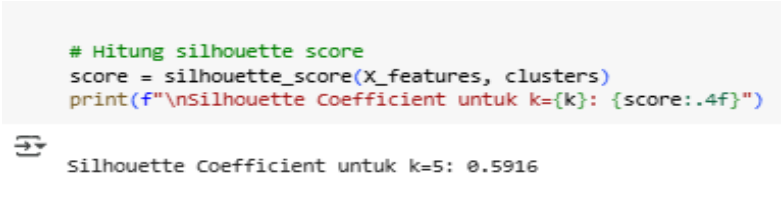
Selanjutnya pada Tabel 5 dapat dilihat hasil *cluster* yang menunjukkan produk-produk yang masuk ke dalam masing-masing *cluster* berdasarkan tingkat permintaan, tren penjualan, dan harga.

Tabel 5. Hasil *Clustering*

product	Cluster	Kategori (Permintaan -Tren - Harga)
Americano	2	Cukup Tinggi – Cukup Stabil – Sedang
Beans Lokal Hot Japanese	4	Rendah – Stabil – Cukup Murah
Beans Lokal Hot V60	4	Rendah – Stabil – Cukup Murah
Beans Lokal Iced Japanese	4	Rendah – Stabil – Cukup Murah
Beans Lokal Iced V60	4	Rendah – Stabil – Cukup Murah
Black Yuzu	4	Rendah – Stabil – Cukup Murah
Cappucino	3	Menengah –Menurun Perlahan – Cukup Mahal
Coffee Basil	4	Rendah – Stabil – Cukup Murah
Espresso	4	Rendah – Stabil – Cukup Murah
Flat White	4	Rendah – Stabil – Cukup Murah
Guest Beans Hot V60	0	Sangat Rendah – Stabil dan Tumbuh Sangat Ringan – Sangat Mahal
Guest Beans Iced Japanese	0	Sangat Rendah – Stabil dan Tumbuh Sangat Ringan – Sangat Mahal
Guest Beans Iced V60	0	Sangat Rendah –Stabil dan Tumbuh Sangat Ringan – Sangat Mahal
Honey Star Latte	1	Sangat Tinggi –Menurun Tajam – Sangat Murah
Kanae Machiatto	2	Cukup Tinggi – Cukup Stabil – Sedang
Kopi Susu Kanae	1	Sangat Tinggi –Menurun Tajam – Sangat Murah
Kuro Latte	4	Rendah – Stabil – Cukup Murah
Latte	3	Menengah – Menurun Perlahan – Cukup Mahal
Majic	4	Rendah – Stabil – Cukup Murah
Mocca	4	Rendah – Stabil – Cukup Murah
Mocha Ice	4	Rendah – Stabil – Cukup Murah
Mocha Ice Cream	3	Menengah –Menurun Perlahan – Cukup Mahal
Okinawa Latte	3	Menengah –Menurun Perlahan – Cukup Mahal
Red Presso	4	Rendah – Stabil – Cukup Murah
Shibuya Cookie	3	Menengah –Menurun Perlahan – Cukup Mahal
Tokyo Bana Latte	3	Menengah –Menurun Perlahan – Cukup Mahal

5. Evaluation

Untuk menilai kualitas hasil segmentasi produk yang telah dibuat, digunakan evaluasi menggunakan *Silhouette coefficient*. Evaluasi ini digunakan untuk mengukur seberapa baik suatu produk dikelompokkan ke dalam jumlah *cluster* yang terbentuk, seperti yang terlihat pada Gambar 9.



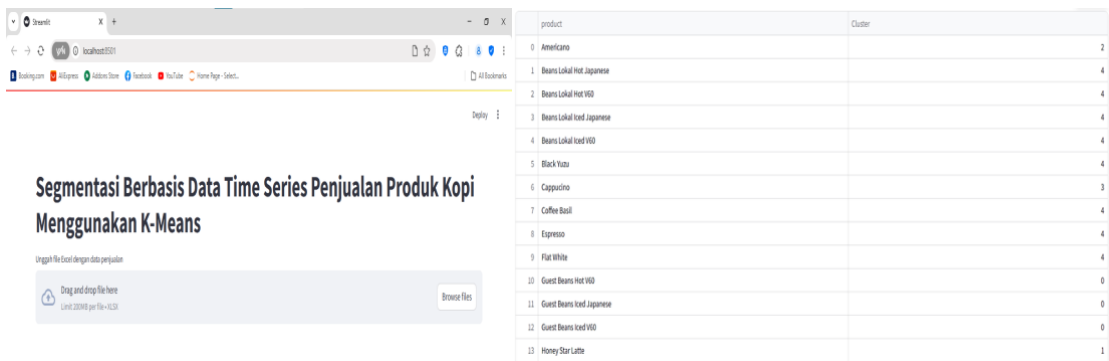
Gambar 9. *Silhouette coefficient*

Berdasarkan Gambar 9, hasil dari perhitungan *Silhouette coefficient* didapatkan sebesar 0.5916 dengan jumlah *cluster* sebanyak lima ($k = 5$). Berdasarkan kriteria penilaian, nilai tersebut termasuk dalam kategori "*Good structure*" [16]. Hal tersebut menandakan bahwa

struktur segmentasi yang terbentuk sudah cukup baik dan pengelompokan produk dapat dikatakan cukup jelas.

6. Deployment

Setelah evaluasi kualitas *cluster*, model dikembangkan ke dalam *Graphical User Interface (GUI)* sederhana menggunakan *Streamlit*. *GUI* ini menampilkan hasil pengelompokan produk, dimana pengguna dapat mengunggah data penjualan, dan sistem otomatis mengelompokkan produk berdasarkan pola penjualan. Hasil *cluster* dapat diunduh dalam format *excel*.



Gambar 10. Visualiasi Menu *GUI*

KESIMPULAN

Penelitian ini mengelompokkan 26 produk kopi berdasarkan tren penjualan (Juli 2024– Januari 2025), harga, dan tiga atribut *feature engineering* (*mean_qty*, *std_qty*, *slope*) menggunakan algoritma *K-Means*. Jumlah *cluster* ditentukan dengan metode *elbow* pada titik perlambatan penurunan *SSE*, yaitu lima *cluster* ($k = 5$). Evaluasi akhir dengan *Silhouette coefficient* menghasilkan nilai 0,5916, menunjukkan struktur pengelompokan “good structure”. *Cluster* 0: penjualan sangat rendah, tren stabil dan tumbuh ringan, harga sangat mahal. *Cluster* 1: penjualan sangat tinggi, tren menurun tajam, harga sangat murah. *Cluster* 2: penjualan cukup tinggi, tren cukup stabil, harga sedang. *Cluster* 3: penjualan menengah, tren menurun perlahan, harga cukup mahal. *Cluster* 4: penjualan rendah, tren stabil, harga cukup murah. Segmentasi ini bermanfaat untuk menjaga kualitas produk (*cluster* 0), promosi atau varian rasa baru (*cluster* 1), perluasan jangkauan (*cluster* 2), inovasi rasa atau kemasan (*cluster* 3), dan strategi promosi atau *bundling* (*cluster* 4). Hasil akhirnya diterapkan dalam *GUI* sederhana berbasis *Streamlit*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada para *reviewer* atas waktu, perhatian, dan masukan konstruktif yang diberikan selama proses peninjauan. Saran dan kritik yang disampaikan telah membantu meningkatkan kualitas serta kejelasan penelitian ini. Penulis sangat menghargai dedikasi para *reviewer* dalam memberikan evaluasi yang objektif dan mendalam, sehingga penelitian ini dapat berkembang menjadi karya yang lebih baik.”

DAFTAR PUSTAKA

- [1] A. Yahya and R. Kurniawan, "Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. January, pp. 350–358, 2025.
- [2] N. Rohman and A. Wibowo, "Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall," *SINTECH (Science and Information Technology) Journal*, vol. 7, no. 1, pp. 49–58, 2024.
- [3] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, DB Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 191, 2023.
- [4] D. Tri Cahaya, D. Puspita, and R. Syahri, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Potensi Padi Di Kota Pagar Alam," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2187–2193, 2024.
- [5] G. Urva, Desyanti, I. Albanna, M. S. Sungkar, I. M. A. O. Gunawan, I. Adhicandra, S. Ramadhan, R. L. Rahardian, Herlawati, R. T. Handayanto, A. A. G. Ariana, Hartatik, P. D. Atika, and S. Junaidi, *Penerapan Data Mining di Berbagai Bidang : Konsep, Metode, dan Studi Kasus*. Jambi, Indonesia: PT. Sonpedia Publishing Indonesia, 2023.
- [6] M. M. Muttaqin, Wahyu Wijaya Widiyanto, A. W. Green Ferry Mandias, Stenly Richard Pungus, S. A. H. Wiranti Kusuma Hapsari, E. F. B. Aslam Fatkhudin, Pasnur, and N. S. Mochammad Anshori, Suryani, *Pengenalan Data Mining*, no. July. Yayasan Kita Menulis, 2023.
- [7] Z. Purwanti and Sugiyono, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 3065–3079, 2024.
- [8] R. P. Widyastuti Andriyani, S. A. Hendrawan, A. I. Irvani, A. Sujarwo, Y. N. Asri, A. H. S. Jones, N. R. Feta, S. H. Ike Yuni Wulandari, H. Y. Suhendi, D. H. Ajif Yunizar Pratama Yusuf, and H. Lubis, *Data Sebagai Fondasi Kecerdasan Buatan*. Cv. Tohar Media, 2024.
- [9] D. Astuti, A. R. Iskandar, and A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, vol. 1, no. 2, pp. 60–72, 2019.
- [10] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator IPM Dengan K-Medoids," *Gaussian*, vol. 13, pp. 13–24, 2024.
- [11] S. Z. Rindiwana, A. N. Rachman, and V. Purwayoga, "Optimasi Jumlah Cluster untuk Analisis Penjualan Barang Kosmetik Menggunakan K-Medoids," *Jurnal Slistem dan Teknologi Informasi*, vol. 13, no. 1, pp. 148–165, 2025.
- [12] D. J. Kiwasi, F. Reba, S. A. Mandowen, A. M. Sroyer, H. E. Manurung, M. E. Koibur, E. H. Pujiarini, M. R. Edy, H. Yuliana, M. Yusuf, and R. M. Dinata, *Data Mining (Memahami Pola di Balik Angka)*. Purbalingga: Cv.Eureka Media Aksara, 2024.
- [13] S. A. Perdana, S. F. Florentin, and A. Santoso, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift," *Sebatik*, vol. 26, no.

- 2, pp. 446–457, 2022.
- [14] R. Purnomo, Wowon Priatna, and Tri Dharma Putra, “Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning,” *Journal of Informatic and Information Security*, vol. 2, no. 1, pp. 77–88, 2021.
- [15] R. R. Muhima, M. Kurniawan, S. R. Wardhana, A. Yudhana, Sunardi, W. M. Rahmawati, and G. E. Yuliasuti, *Kupas Tuntas Algoritma Clustering Konsep, Perhitungan Manual dan Program*. ANDI, 2021.
- [16] M. K. Anam, I. Y. Pasa, K. D. K. Wardhani, L. Efrizoni, and M. B. Firdaus, “K-Means Clustering to Identity Twitter Build Operate Transfer (BOT) on Influential Accounts,” *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 2, pp. 143–154, 2023.
- [17] A. M. Nur, H. Bahtiar., and M. A. Jannah, “Implementasi Algoritma K-Means Clustering Dalam Mengelompokkan Kepatuhan Wajib Pajak Bumi Dan Bangunan Dengan Optimasi Elbow,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 181–192, 2025.

