

Segmentasi Pola Transaksi Coffee Shop menggunakan Algoritma K-Means untuk Analisis Perilaku Pembelian

Resya Hidayatunnisa¹, Apriade Voutama²

Sistem Informasi, Ilmu Komputer, Universitas Singaperbangsa Karawang
¹resyahidayatunnisa7@gmail.com, ²apriade.voutama@staff.unsika.ac.id

Abstract

This study aims to segment customer transaction patterns in a coffee shop to understand purchasing behavior based on transaction data. The data used consisted of more than 3,500 transaction records obtained from the Kaggle platform, including transaction value, purchase time, and transaction day attributes. The method applied in this research was a data mining approach using the Cross Industry Standard Process for Data Mining (CRISP-DM) framework, with the K-Means algorithm as the clustering technique. The research stages included business understanding, data understanding, data preparation, modeling, and evaluation. In the modeling stage, data normalization was performed, and the optimal number of clusters was determined using the Elbow method, resulting in an inertia value of 5839.77 at $k = 3$. The results showed that the optimal number of clusters was three groups, representing low, medium, and high transaction characteristics at specific times. Each cluster exhibited different transaction patterns, which can be used as a basis for developing more targeted marketing strategies. This study concludes that the K-Means algorithm is effective in grouping transaction patterns and provides valuable insights into understanding customer purchasing behavior in coffee shops.

Keywords: transaction segmentation, k-means, data mining, customer behavior, coffee shop

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi pola transaksi pelanggan pada coffee shop guna memahami perilaku pembelian berdasarkan data transaksi. Data yang digunakan merupakan data penjualan coffee shop yang diperoleh dari platform Kaggle dengan jumlah lebih dari 3.500 transaksi, yang mencakup atribut nilai transaksi, waktu pembelian, dan hari transaksi. Metode yang digunakan dalam penelitian ini adalah pendekatan data mining dengan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM), serta algoritma K-Means sebagai teknik clustering. Tahapan penelitian meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi. Pada tahap pemodelan dilakukan normalisasi data serta penentuan jumlah cluster optimal menggunakan metode Elbow, dengan nilai inertia sebesar 5839,77 pada $k = 3$. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga kelompok yang merepresentasikan karakteristik transaksi dengan nilai rendah, sedang, dan tinggi pada waktu tertentu. Setiap cluster menunjukkan pola transaksi yang berbeda sehingga dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih terarah. Penelitian ini menyimpulkan bahwa algoritma K-Means mampu mengelompokkan pola transaksi secara efektif serta memberikan wawasan yang bermanfaat dalam memahami perilaku pembelian pelanggan pada coffee shop.

Kata kunci: segmentasi transaksi, k-means, data mining, perilaku pelanggan, coffee shop

© 2026 Author
Creative Commons Attribution 4.0 International License



1. Pendahuluan

Perkembangan industri coffee shop dalam beberapa tahun terakhir telah menyebabkan meningkatnya persaingan bisnis secara signifikan. Dalam situasi tersebut, pengambilan keputusan yang berbasis data menjadi aspek krusial untuk menunjang efektivitas strategi pemasaran maupun operasional. Data transaksi yang tersimpan pada sistem *point of sales* (POS) tidak sekadar berperan sebagai dokumentasi administratif, tetapi juga mencerminkan pola perilaku pembelian konsumen yang dapat dianalisis guna memperoleh wawasan strategis [1],[2]. Sejalan dengan kemajuan teknologi data mining, penggunaan algoritma clustering seperti K-Means semakin luas diterapkan dalam berbagai penelitian terkini untuk menemukan pola tersembunyi di dalam data [3],[4].

Nilai transaksi dan waktu pembelian merupakan indikator penting yang mampu merepresentasikan tingkat intensitas serta kecenderungan perilaku konsumsi pelanggan. Namun, dalam praktiknya, data transaksi pada banyak usaha coffee shop umumnya hanya digunakan untuk keperluan rekapitulasi penjualan semata tanpa dilakukan analisis pola secara terstruktur. Kondisi ini menyebabkan potensi informasi yang terkandung di dalam data belum dimanfaatkan secara maksimal untuk menunjang pengambilan keputusan yang berbasis analisis [1],[5].

Penelitian ini berfokus pada clustering untuk mengelompokkan data transaksi berdasarkan tingkat kemiripan karakteristik numerik yang dimiliki [2],[6]. Salah satu metode clustering yang umum digunakan dalam proses segmentasi adalah algoritma K-Means, yang bekerja dengan meminimalkan variasi dalam setiap kluster melalui perhitungan jarak terhadap pusat kluster (*centroid*) [1],[7]. Metode ini telah terbukti efektif dalam berbagai penelitian terkini, khususnya dalam analisis data transaksi dan segmentasi berbasis perilaku [8],[9],[10].

Sejumlah penelitian terdahulu telah mengimplementasikan algoritma K-Means untuk segmentasi pelanggan pada sektor ritel maupun layanan digital [2],[7]. Namun, mayoritas studi tersebut lebih menitikberatkan pada segmentasi individu pelanggan berdasarkan atribut demografis atau riwayat pembelian yang telah diagregasi. Pendekatan yang menggunakan data transaksi nyata secara langsung, dengan memperhatikan kombinasi variabel waktu dan nilai transaksi secara rinci, masih tergolong jarang diterapkan, khususnya pada industri coffee shop berskala kecil.

Berdasarkan hal tersebut, penelitian ini menawarkan kebaruan (*novelty*) melalui penerapan segmentasi pola transaksi dengan menggabungkan variabel waktu (*temporal*) dan nilai transaksi (*transactional*) secara detail guna mengidentifikasi karakteristik perilaku pembelian yang terbentuk secara alami. Penelitian ini bertujuan untuk mengelompokkan pola transaksi pada coffee shop berdasarkan nilai

pembelian dan waktu transaksi menggunakan algoritma K-Means, sehingga menghasilkan segmentasi yang dapat dimanfaatkan sebagai dasar dalam merancang strategi pemasaran dan operasional yang lebih tepat sasaran.

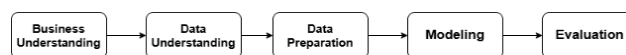
2. Metode Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan utama yang diawali dengan pengumpulan serta pemahaman terhadap dataset transaksi. Tahap berikutnya mencakup proses seleksi dan pra-pemrosesan data untuk memastikan kualitas dan kesiapan data sebelum dianalisis. Selanjutnya, diterapkan metode clustering menggunakan algoritma K-Means, kemudian hasil segmentasi dievaluasi guna menafsirkan pola transaksi yang terbentuk. Seluruh proses analisis dilakukan dengan menggunakan bahasa pemrograman *Python*.

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining untuk melakukan segmentasi pola transaksi. Kerangka kerja yang diterapkan adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai pedoman sistematis dalam menjalankan setiap tahapan analisis data, mulai dari pemahaman bisnis hingga evaluasi model.

Pada tahap *modeling*, penelitian ini menerapkan algoritma K-Means untuk melakukan proses pengelompokan data berdasarkan kemiripan karakteristik numerik yang dimiliki [11],[12].



Gambar 1. Alur Penelitian Menggunakan CRISP-DM

2.2. Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini berupa dataset transaksi coffee shop yang diperoleh dari platform Kaggle dengan nama file *Coffe_sales.csv*. Dataset tersebut terdiri atas lebih dari 3000 data transaksi.

Unit analisis pada penelitian ini adalah setiap transaksi penjualan yang tercatat. Variabel yang dimanfaatkan dalam proses segmentasi meliputi: *money* (nilai transaksi), *hour_of_day* (jam terjadinya transaksi), *Weekdaysort* (urutan hari berlangsungnya transaksi). Tabel 1 adalah contoh data transaksi.

Tabel 1. Tabel Contoh Data Transaksi

Transaction_id	Money	Hour_of_day	Weekdaysort
1	38.7	10	5
2	38.7	12	5
3	38.7	12	5
4	28.9	13	5
5	39.7	13	5

2.3. Data Preparation

Tahapan ini meliputi pembersihan data, pemilihan atribut yang relevan, serta proses normalisasi guna memastikan setiap variabel memiliki skala yang setara dalam proses clustering, sehingga hasil pengelompokan dapat lebih optimal [8].

Normalisasi dilakukan dengan metode standardisasi (*Z-score*), karena algoritma K-Means mengandalkan perhitungan jarak Euclidean yang peka terhadap perbedaan skala data. Rumus yang digunakan (1):

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

z = nilai hasil standardisasi (nilai yang sudah dinormalisasi)

x = nilai data asli dari suatu variabel

μ (μ) = nilai rata-rata (mean) dari seluruh data pada variabel tersebut

σ = standar deviasi

2.4. Modeling Menggunakan K-Means

Proses clustering dilakukan dengan menggunakan algoritma K-Means [13]. Algoritma ini mengelompokkan data ke dalam sejumlah k cluster berdasarkan tingkat kedekatan terhadap pusat cluster (centroid).

Jarak antar data dihitung menggunakan rumus Euclidean (2):

$$d(x_i, c_j) = \sqrt{\sum (x_{ik} - c_{jk})^2} \quad (2)$$

Keterangan simbol:

$d(x_i, c_j)$ = jarak antara data ke- i dan centroid ke- j

x_i = data ke- i

c_j = centroid ke- j

x_{ik} = nilai variabel ke- k pada data ke- i

c_{jk} = nilai variabel ke- k pada centroid ke- j

Σ (σ) = penjumlahan untuk seluruh variabel

k = jumlah variabel yang digunakan dalam clustering

Proses ini dilakukan secara iteratif, di mana data akan terus dikelompokkan ulang dan *centroid* diperbarui hingga posisi centroid tidak lagi mengalami perubahan (stabil).

2.5. Penentuan Jumlah Cluster

Jumlah cluster yang paling optimal ditentukan dengan menggunakan metode *Elbow Method*, yaitu dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa variasi jumlah cluster (k). Nilai WCSS tersebut kemudian divisualisasikan dalam bentuk grafik, dan titik yang membentuk pola siku (*elbow*) dijadikan sebagai acuan dalam menentukan jumlah cluster yang paling sesuai.

2.6. Evaluasi

Evaluasi dilakukan dengan menganalisis nilai inerti sebagai indikator tingkat homogenitas data dalam setiap cluster, di mana semakin kecil nilainya

menunjukkan bahwa anggota cluster semakin dekat dengan *centroid-nya* [14].

3. Hasil dan Pembahasan

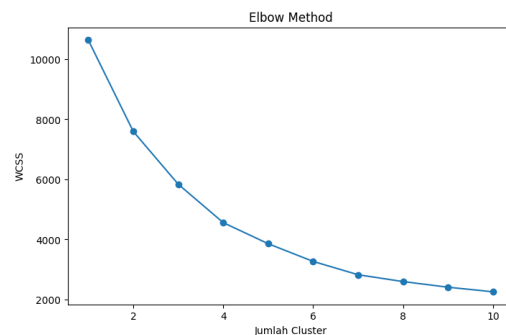
Hasil clustering yang diperoleh pada penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa algoritma K-Means efektif dalam melakukan segmentasi pelanggan berdasarkan pola transaksi.

Pada bagian ini disajikan hasil pengolahan data transaksi coffee shop dengan menggunakan algoritma K-Means Clustering. Hasil penelitian ditampilkan dalam bentuk tabel dan grafik guna mempermudah pemahaman terhadap pola transaksi yang terbentuk. Pembahasan dilakukan untuk menafsirkan karakteristik dari masing-masing cluster yang dihasilkan.

3.1. Hasil Pengujian Jumlah Cluster

Penentuan jumlah cluster optimal dilakukan dengan menggunakan metode *Elbow* melalui perhitungan nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa variasi jumlah cluster.

Dalam penelitian ini, pengujian jumlah cluster dilakukan pada rentang nilai $k = 1$ hingga $k = 10$. Hasil perhitungan WCSS selanjutnya divisualisasikan dalam bentuk grafik untuk mendukung identifikasi jumlah cluster yang paling optimal.



Gambar 1. Grafik Elbow Method

Berdasarkan grafik pada gambar 1, terlihat bahwa penurunan nilai WCSS terjadi secara signifikan hingga $k = 3$, kemudian setelah itu laju penurunan grafik mulai melandai. Kondisi ini menunjukkan bahwa jumlah cluster optimal yang digunakan dalam penelitian ini adalah sebanyak 3 cluster.

3.2. Hasil Clustering K-Means

Proses clustering dilakukan dengan menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga. Variabel yang digunakan dalam proses clustering meliputi nilai transaksi (*money*), waktu transaksi (*hour_of_day*), dan hari transaksi (*Weekdaysort*).

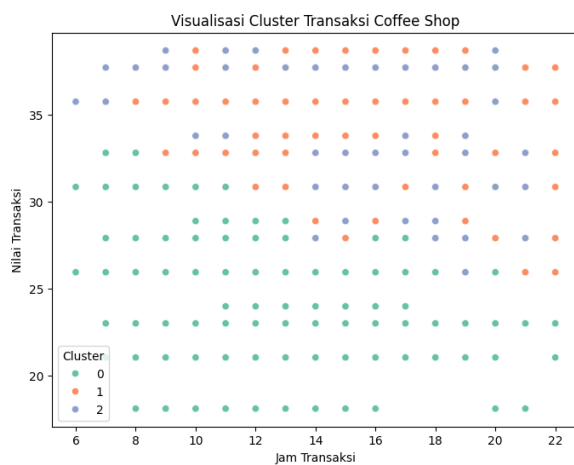
Tabel 2. Contoh Hasil Clustering

Money	Hour_of day	Weekdaysort	Cluster
38.7	10	5	1
38.7	12	5	1
38.7	12	5	1
38.7	13	5	0
38.7	13	5	1

Pada tabel 2, setiap data transaksi berhasil dikelompokkan ke dalam cluster berdasarkan tingkat kemiripan karakteristik masing-masing data.

3.3. Visualisasi Cluster

Untuk mempermudah pemahaman terhadap hasil clustering, dilakukan visualisasi menggunakan grafik scatter.



Gambar 2. Visualisasi Hasil Clustering

Berdasarkan Gambar 2, terlihat bahwa data terbagi ke dalam tiga kelompok cluster dengan pola distribusi yang cukup jelas berdasarkan nilai transaksi (*money*) dan waktu transaksi (*hour_of day*). Masing-masing cluster menunjukkan karakteristik yang berbeda satu sama lain.

3.4 Analisis Karakteristik Cluster

Untuk mengetahui karakteristik masing-masing cluster, dilakukan perhitungan nilai rata-rata dari setiap variabel pada tiap cluster. Hasil perhitungan tersebut ditampilkan pada Tabel 3.

Tabel 3. Rata-rata Karakteristik Cluster

Cluster	Money	Hour_of day	Weekdaysort
0	26.11	11.80	3.97
1	34.51	15.46	5.53
2	34.07	15.18	1.88

Berdasarkan hasil rata-rata setiap cluster, terlihat bahwa masing-masing kelompok memiliki karakteristik pola transaksi yang berbeda. Cluster 0 memiliki nilai transaksi terendah dengan waktu pembelian cenderung terjadi pada siang hari dan

berada di pertengahan minggu, yang menunjukkan pola pembelian rutin dengan nilai rendah.

Cluster 1 memiliki nilai transaksi tertinggi dengan waktu pembelian pada sore hari dan mendekati akhir pekan. Hal ini menunjukkan bahwa cluster ini merupakan kelompok pelanggan dengan kontribusi transaksi yang tinggi.

Sementara itu, cluster 2 juga memiliki nilai transaksi yang tinggi dengan waktu pembelian pada sore hari, namun terjadi di awal minggu. Hal ini menunjukkan adanya perbedaan pola waktu transaksi dibandingkan cluster 1 meskipun memiliki nilai transaksi yang hampir sama.

Hasil segmentasi yang diperoleh menunjukkan bahwa algoritma K-Means mampu mengelompokkan pola transaksi pelanggan secara efektif berdasarkan nilai transaksi dan waktu pembelian. Temuan ini selaras dengan penelitian terdahulu yang menyatakan bahwa metode K-Means dapat dimanfaatkan untuk mengidentifikasi karakteristik pelanggan berdasarkan perilaku transaksinya [15].

Informasi hasil segmentasi ini dapat dimanfaatkan oleh pelaku usaha coffee shop sebagai dasar dalam merancang strategi pemasaran yang lebih tepat sasaran, misalnya melalui pemberian promosi khusus pada waktu tertentu maupun penawaran produk premium kepada segmen pelanggan dengan nilai transaksi tinggi.

3.5 Implikasi Strategi Berdasarkan Cluster

Berdasarkan hasil segmentasi yang diperoleh, terdapat beberapa implikasi strategis yang dapat diterapkan oleh pelaku usaha *coffee shop*, antara lain: (1). Pemberian promosi pada jam tertentu untuk meningkatkan transaksi pada *cluster* dengan nilai pembelian rendah; (2). Penargetan promosi pada akhir pekan untuk mempertahankan pelanggan dengan nilai transaksi tinggi; (3). Penerapan strategi *upselling* dan *cross-selling* pada *cluster* dengan potensi transaksi menengah hingga tinggi.

Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa segmentasi berbasis *clustering* mampu membantu dalam penyusunan strategi pemasaran yang lebih tepat sasaran [2],[8]. Namun, penelitian ini memberikan kontribusi tambahan dengan mengintegrasikan variabel waktu dan nilai transaksi secara simultan, sehingga menghasilkan segmentasi yang lebih kontekstual, khususnya pada industri *coffee shop* skala kecil.

4. Kesimpulan

Penelitian ini membuktikan bahwa penggunaan algoritma K-Means dalam segmentasi pola transaksi coffee shop mampu menghasilkan pengelompokan data yang optimal dengan tiga cluster, sebagaimana ditentukan melalui metode Elbow dengan nilai inerti sebesar 5839,77. Secara akademis, penelitian ini berkontribusi dalam pengembangan analisis

segmentasi berbasis data mining dengan menggabungkan variabel nilai transaksi dan waktu pembelian (temporal) secara bersamaan dalam proses clustering. Pendekatan ini memungkinkan identifikasi pola transaksi yang lebih kontekstual dibandingkan metode segmentasi konvensional yang umumnya hanya memanfaatkan data agregat atau atribut pelanggan. Lebih lanjut, hasil penelitian menunjukkan bahwa kombinasi variabel transaksi dan waktu mampu menghasilkan karakteristik cluster yang berbeda secara signifikan, sehingga memperkaya metode segmentasi dalam analisis perilaku konsumen. Untuk penelitian berikutnya, disarankan menambahkan variabel lain seperti jenis produk dan frekuensi pembelian, serta melakukan perbandingan dengan metode clustering lain guna meningkatkan kualitas dan akurasi hasil segmentasi.

Ucapan Terimakasih

Terima kasih kepada semua pihak yang telah memberikan dukungan dalam penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada penyedia dataset yang digunakan melalui platform Kaggle, serta kepada dosen pengampu mata kuliah yang telah memberikan arahan dan masukan selama proses penyusunan tugas.

Daftar Rujukan

- [1] J. Han, Jiawei; Kamber, Micheline; Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington: Morgan Kaufmann, 2011.
- [2] N. A. M. Hadi, A. A.; Kamaruddin, M. H.; Zin, "Customer segmentation using K-means clustering technique," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, p. International Journal of Advanced Computer Science, 2018.
- [3] M. Ahmed, R. Seraj, S. Mohammed, and S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [4] A. Fahad *et al.*, "IEEE TRANSACTIONS ON A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 3, pp. 267–279, 2014, doi: 10.1109/TETC.2014.2330519.
- [5] F. Dolnicar, Sara; Grün, Bettina; Leisch, *No TitleMarket Segmentation Analysis: Understanding It, Doing It, and Making It Useful*. Singapore: Springer, 2018.
- [6] W. A. Wedel, Michel; Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, 3rd ed. New York: Springer, 2017.
- [7] E. Sarstedt, Marko; Mooi, *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*. Springer: Springer, 2019.
- [8] R. S. Cordova, "CUSTOMER SEGMENTATION IN THE ONLINE RETAIL INDUSTRY USING BIG DATA ANALYTICS," vol. 102, no. 22, pp. 8312–8324, 2024.
- [9] K. R. Kashwan and C. M. Velu, "Customer Segmentation Using Clustering and Data Mining Techniques," vol. 5, no. 6, pp. 1–6, 2013, doi: 10.7763/IJCTE.2013.V5.811.
- [10] C. Shahapure, K. R., & Nicholas, "Cluster Quality Analysis Using Silhouette Score. Proceedings of the IEEE International Conference on Data Mining Workshops," 2020, doi: <https://doi.org/10.1109/ICDMW51313.2020.00096>.
- [11] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
- [12] S. Arthur, David; Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, 2007.
- [13] C. C. Aggarwal, "Data mining: The textbook," *Knowledge-Based Syst.*, vol. 177, pp. 1–10, 2019.
- [14] I. Arbelaitz, O.; Gurrutxaga, I.; Muguerza, J.; Pérez, J. M.; Perona, "An extensive comparative study of cluster validity indices," *Indices, An extensive Comp. study Clust. validity*, vol. 253, pp. 1–14, 2013.
- [15] W. Kumar, V.; Reinartz and J. P. C. Science, "Customer relationship management: Concept, strategy, and tools," *Procedia Comput. Sci.*, vol. 132, pp. 1552–1559, 2018, doi: 10.1016/j.procs.2017.11.116.