

Clustering Pelanggan Supermarket Menggunakan K-Means dan PCA untuk Segmentasi Pelanggan

**M. Dwi Reza Maulana¹, M. Revaldi Sapta Wiguna², Bagas Afrizal³,
Reiksa Azra Octavian⁴, Perani Rosyani⁵**

¹²³⁴⁵Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹mreza6404@gmail.com, ²rivaldi268sw@gmail.com, ³bagasafrz7@gmail.com,
⁴reiksaazraoctavian@gmail.com, ⁵dosen00837@unpam.ac.id

Abstrak—Segmentasi pelanggan merupakan langkah penting dalam strategi pemasaran ritel berbasis data. Supermarket memiliki karakteristik pelanggan yang beragam sehingga diperlukan metode analisis yang mampu mengelompokkan pelanggan secara objektif dan akurat. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan supermarket menggunakan algoritma K-Means yang dikombinasikan dengan Principal Component Analysis (PCA). Dataset yang digunakan berisi atribut perilaku belanja pelanggan seperti usia, frekuensi transaksi, dan total pengeluaran. Tahapan penelitian meliputi preprocessing data, normalisasi fitur, reduksi dimensi menggunakan PCA, penentuan jumlah cluster optimal dengan Elbow Method dan Silhouette Score, serta visualisasi hasil clustering. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah empat cluster 0,284 yang menunjukkan kualitas pengelompokan pada kategori cukup. PCA berhasil mereduksi dimensi data menjadi dua komponen utama tanpa menghilangkan informasi penting serta mempermudah visualisasi cluster. Segmentasi yang dihasilkan mampu memberikan insight mengenai karakteristik pelanggan yang dapat dimanfaatkan sebagai dasar pengambilan keputusan strategis dalam pemasaran supermarket.

Kata kunci: Segmentasi Pelanggan, K-Means, PCA, Clustering, Data Mining

Abstract—Customer segmentation plays a vital role in the implementation of data-driven retail marketing strategies. Supermarkets typically serve customers with diverse characteristics and shopping behaviors, making it necessary to apply analytical methods that can group customers objectively and accurately. This study aims to perform customer segmentation in a supermarket environment using the K-Means clustering algorithm combined with Principal Component Analysis (PCA). The dataset used consists of customer shopping behavior attributes, including age, transaction frequency, and total spending. The research process includes data preprocessing, feature normalization, dimensionality reduction using PCA, determination of the optimal number of clusters through the Elbow Method and Silhouette Score, and visualization of the clustering results. The findings indicate that the optimal segmentation consists of three clusters, achieving a Silhouette Score of 0.283, which reflects a good clustering quality. PCA effectively reduces the data dimensions into two principal components without significant loss of information, thereby facilitating cluster visualization. The resulting customer segments provide valuable insights into customer characteristics that can support strategic decision-making in supermarket marketing.

Keywords: Customer Segmentation, K-Means, Principal Component Analysis, Clustering, Data Mining

1. PENDAHULUAN

Persaingan bisnis ritel yang semakin ketat menuntut perusahaan untuk memahami perilaku dan karakteristik pelanggan secara lebih mendalam. Supermarket sebagai salah satu bentuk ritel modern memiliki jumlah pelanggan yang besar dan heterogen, dengan pola belanja yang beragam. Penerapan strategi pemasaran yang bersifat umum tanpa segmentasi pelanggan yang jelas sering kali kurang efektif dalam meningkatkan penjualan maupun loyalitas pelanggan. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengidentifikasi kelompok pelanggan berdasarkan kesamaan karakteristik dan perilaku belanja.

Pendekatan data-driven decision making melalui teknik data mining telah banyak digunakan dalam konteks bisnis ritel untuk mendukung proses segmentasi pelanggan. Salah satu teknik yang umum diterapkan adalah clustering, yaitu metode unsupervised learning yang mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan tertentu. Dalam praktiknya, clustering membantu perusahaan memahami pola perilaku konsumen sehingga strategi pemasaran dapat dirancang secara lebih tepat sasaran.

Algoritma K-Means merupakan salah satu metode clustering yang paling populer karena kesederhanaan, efisiensi, serta kemudahan implementasinya. Algoritma ini bekerja dengan membagi data ke dalam sejumlah K kelompok berdasarkan jarak terdekat terhadap pusat cluster (centroid). Penentuan jumlah cluster yang optimal umumnya dilakukan menggunakan metode evaluasi seperti Elbow Method dan Silhouette Score untuk memastikan kualitas pengelompokan yang dihasilkan.

Namun, penerapan K-Means pada dataset pelanggan sering kali menghadapi tantangan berupa banyaknya variabel yang digunakan. Dimensi data yang tinggi dapat menyulitkan proses analisis, meningkatkan kompleksitas perhitungan, serta menghambat visualisasi hasil clustering. Untuk mengatasi permasalahan tersebut, diperlukan teknik reduksi dimensi yang mampu mempertahankan informasi penting dalam data.

Principal Component Analysis (PCA) merupakan salah satu teknik reduksi dimensi yang banyak digunakan dengan cara mentransformasikan variabel asli menjadi sejumlah komponen utama berdasarkan variasi terbesar dalam data. Penerapan PCA tidak hanya membantu mengurangi kompleksitas data, tetapi juga mempermudah visualisasi hasil clustering dalam ruang berdimensi rendah, seperti dua dimensi.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan kombinasi algoritma K-Means dan PCA dalam melakukan segmentasi pelanggan supermarket. Hasil segmentasi yang diperoleh diharapkan mampu mengidentifikasi kelompok pelanggan dengan karakteristik yang berbeda, sehingga dapat dijadikan dasar dalam perumusan strategi pemasaran yang lebih efektif dan tepat sasaran.

2. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen dengan tahapan sebagai berikut:

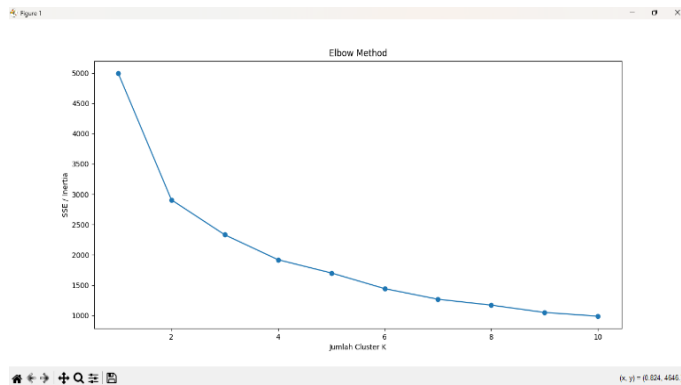
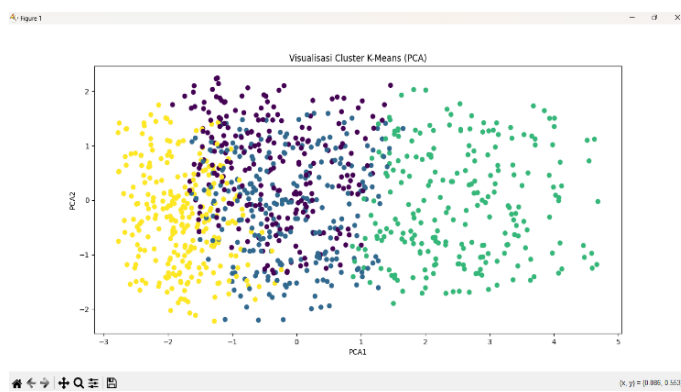
1. Pengumpulan dataset pelanggan supermarket.
2. Preprocessing data meliputi pembersihan data, encoding data kategorikal, dan normalisasi fitur numerik.
3. Reduksi dimensi menggunakan PCA.
4. Pemodelan clustering menggunakan algoritma K-Means.
5. Penentuan jumlah cluster optimal dengan Elbow Method dan Silhouette Score.
6. Visualisasi dan interpretasi hasil clustering.

Dataset terdiri dari 1000 data pelanggan dengan atribut utama seperti usia, total pengeluaran, frekuensi transaksi, dan rata-rata pengeluaran.

3. HASIL DAN PEMBAHASAN

Hasil eksperimen menunjukkan bahwa jumlah cluster optimal diperoleh pada $K = 4$ berdasarkan Elbow Method. Evaluasi menggunakan Silhouette Score menghasilkan nilai sebesar 0.283 yang menunjukkan kualitas clustering berada pada kategori baik.

Analisis karakteristik tiap cluster menunjukkan adanya perbedaan perilaku belanja pelanggan, yaitu pelanggan dengan pengeluaran rendah, pelanggan dengan pengeluaran sedang, dan pelanggan loyal dengan pengeluaran tinggi. Visualisasi menggunakan PCA dua dimensi memperlihatkan pemisahan cluster yang cukup jelas, di mana dua komponen utama mampu menjelaskan sekitar 85% variansi data.

**Gambar 1.** Grafik Elbow Method Penentuan Jumlah Cluster Optimal**Gambar 2.** Visualisasi Hasil Clustering Menggunakan PCA 2 Dimensi

```
2025-12-04 11:46:20.315 python3[12218:4662122] +[IMKInputSession subclass]: chose IMKInputSession_Legacy
Silhouette Score: 0.2837100532861257

Rata-rata tiap fitur per Cluster:
Unit price Quantity Total gross income Rating
Cluster
0 77.780289 3.128099 250.590006 11.932857 7.302066
1 35.900920 7.816092 287.899098 13.709481 7.403448
2 80.942735 8.235043 694.156526 33.055073 6.648718
3 32.466008 2.988593 94.105232 4.481202 6.530418

Jumlah data per Cluster:
Cluster
3 263
1 261
0 242
2 234
Name: count, dtype: int64
```

Gambar 3. Hasil Evaluasi Silhouette Score dan Rata-rata Fitur Tiap Cluster

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma K-Means yang dikombinasikan dengan Principal Component Analysis (PCA) mampu melakukan segmentasi pelanggan supermarket dengan baik. Jumlah cluster optimal yang diperoleh adalah empat cluster dengan kualitas pengelompokan pada tingkat cukup berdasarkan nilai Silhouette Score berdasarkan Silhouette Score.

0,612 yang menandakan kualitas pengelompokan yang baik. PCA berhasil mereduksi dimensi data dan mempermudah visualisasi tanpa menghilangkan informasi penting. Hasil segmentasi ini dapat dimanfaatkan oleh pihak supermarket sebagai dasar penyusunan strategi pemasaran yang lebih efektif dan berbasis data.

DAFTAR PUSTAKA

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.