

## PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK SEGMENTASI PELANGGAN PADA SISTEM INFORMASI E-COMMERCE

<sup>1</sup>Ahmad Rafli & <sup>1</sup>Yuliarti

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Teknologi Informasi, STMIK Palangkaraya, Palangka Raya, Kalimantan Tengah

DOI: <https://doi.org/10.33005/jifosi.v7i2.617>

\*Corresponding Author: daizaii101@gmail.com

Submit: Juli 2026 | Accepted: Agustus 2026 | Published: Agustus 2026

### ABSTRAK

Penelitian ini berhasil mengimplementasikan algoritma K-Means *Clustering* berbasis analisis RFM (*Recency, Frequency, Monetary*) untuk segmentasi pelanggan *e-commerce*, menghasilkan tiga segmen pelanggan yang terdefinisi jelas dengan kualitas pengelompokan baik (*Silhouette Score* 0,612). Dari 600 pelanggan yang dianalisis (sampel acak dari dataset *Online Retail*), terbentuk tiga kelompok: Pelanggan Aktif (418 orang, 70,4%), Pelanggan Potensial (93 orang, 15,7%), dan Pelanggan Pasif (83 orang, 14,0%). Hasil segmentasi ini telah berhasil diintegrasikan ke dalam sistem informasi *e-commerce* melalui REST API berbasis Flask, sehingga tim manajemen dapat langsung memanfaatkannya untuk strategi pemasaran yang lebih terarah. Secara metodologis, data yang digunakan adalah *Online Retail Dataset* periode Desember 2010 sampai Desember 2011 (68.636 *record* awal dari 600 pelanggan terpilih), yang setelah *preprocessing* menyisakan 67.198 *record* valid. Jumlah kluster optimal ditentukan dengan metode Elbow pada  $k=3$ , diimplementasikan dengan Python dan Scikit-learn. Temuan ini menunjukkan bahwa kombinasi K-Means dan RFM efektif sebagai dasar sistem rekomendasi strategi pemasaran berbasis data pada platform *e-commerce*.

**Kata Kunci:** *K-Means Clustering*, Segmentasi Pelanggan, *E-Commerce*, Sistem Informasi, RFM

## APPLICATION OF K-MEANS CLUSTERING ALGORITHM FOR CUSTOMER SEGMENTATION IN E-COMMERCE INFORMATION SYSTEMS

### ABSTRACT

This study successfully implemented the K-Means Clustering algorithm combined with RFM (*Recency, Frequency, Monetary*) analysis for *e-commerce* customer segmentation, producing three clearly defined customer segments with good clustering quality (*Silhouette Score*: 0.612). From 600 randomly sampled customers (drawn from the *Online Retail* dataset), three segments emerged: Active Customers (418 people, 70.4%), Potential Customers (93 people, 15.7%), and Passive Customers (83 people, 14.0%). These segmentation results were integrated into an *e-commerce* information system via a Flask-based REST API, enabling the management team to directly apply them for more targeted marketing strategies. Methodologically, the *Online Retail Dataset* covering December 2010 to December 2011 was used (68,636 initial records from 600 selected customers), with 67,198 valid records remaining after preprocessing. The optimal cluster count was determined using the Elbow method at  $k=3$ , implemented in Python with Scikit-learn. These findings demonstrate that the K-Means and RFM combination is effective as a data-driven marketing strategy recommendation system on *e-commerce* platforms.

**Keywords:** *K-Means Clustering*, Customer Segmentation, *E-Commerce*, Information System, RFM

### PENDAHULUAN



Copyright © 2026 Authors. This is an open access article under the [CC-BY](https://creativecommons.org/licenses/by/4.0/) licence

Segmentasi pelanggan berbasis data merupakan kebutuhan strategis utama bagi platform *e-commerce* yang kompetitif. Tanpa pemahaman mendalam tentang perilaku pelanggan, strategi pemasaran menjadi tidak efisien dan biaya promosi terbuang sia-sia. Penelitian ini menawarkan solusi konkret melalui penerapan algoritma *K-Means Clustering* yang dikombinasikan dengan analisis RFM (*Recency, Frequency, Monetary*) untuk mengelompokkan pelanggan secara otomatis berdasarkan pola transaksi mereka.

Pertumbuhan belanja *online*, terutama pasca pandemi, mendorong persaingan antar platform *e-commerce* semakin ketat. Banyak pengelola *e-commerce* belum memanfaatkan data transaksi pelanggan secara maksimal untuk membedakan antara pelanggan aktif, pelanggan yang berpotensi dikembangkan, dan pelanggan yang sudah tidak aktif. Akibatnya, semua pelanggan diperlakukan sama sehingga anggaran promosi tidak tepat sasaran.

Metode K-Means dipilih karena terbukti efektif di berbagai penelitian sejenis [1] dan relatif mudah diimplementasikan pada data berskala besar [3], serta telah banyak diterapkan pada berbagai konteks segmentasi pelanggan lainnya [7]. Dengan mengintegrasikan hasilnya ke sistem informasi, penelitian ini bertujuan: (1) menerapkan K-Means Clustering pada data transaksi e-commerce, (2) menganalisis karakteristik tiap segmen, (3) mengintegrasikan hasil segmentasi ke sistem informasi, dan (4) mengukur kualitas pengelompokan yang dihasilkan.

## METODOLOGI

### 1. K-Means Clustering

*K-Means Clustering* adalah metode *unsupervised learning* yang mengelompokkan data ke dalam k kluster. Algoritma bekerja dengan memilih *centroid* awal, mengalokasikan setiap data ke *centroid* terdekat, memperbarui *centroid* secara iteratif, dan mengulangi proses hingga konvergen [2]. Popularitasnya didukung oleh kesederhanaan implementasi dan kemampuannya menangani data berskala besar [3], dan telah diterapkan secara luas pada berbagai studi segmentasi pelanggan berbasis RFM [10][11][12].

### 2. Analisis RFM

RFM (*Recency, Frequency, Monetary*) adalah pendekatan analisis perilaku pelanggan menggunakan tiga variabel utama [4]: *Recency* (seberapa baru transaksi terakhir), *Frequency* (seberapa sering pelanggan bertransaksi), dan *Monetary* (total nilai pembelian). Kombinasi ketiga variabel ini menghasilkan fitur *clustering* yang lebih bermakna secara bisnis dibanding variabel tunggal [5].

### 3. Dataset

Data yang digunakan adalah *Online Retail Dataset*, dengan 600 pelanggan yang dipilih secara acak (*random sampling*, *random\_state=42*) dari total 4.338 pelanggan unik pada dataset asli, menghasilkan 68.636 *record* transaksi periode Desember 2010–Desember 2011. Setiap *record* memuat informasi nomor *invoice*, kode produk, deskripsi, jumlah item, tanggal transaksi, harga satuan, ID pelanggan, dan negara asal.

### 4. Preprocessing

*Preprocessing* mencakup tiga tahap: (1) penghapusan *record* dengan CustomerID kosong, (2) penghapusan transaksi dengan *Quantity* negatif (retur), dan (3) penghapusan entri dengan harga nol atau negatif, dilanjutkan normalisasi nilai RFM menggunakan *Min-Max Scaling* untuk menyetarakan skala. Hasil *preprocessing*: 67.198 *record* valid dari total 68.636 *record* awal (1.438 *record* dibuang karena retur atau harga tidak valid).

### 5. Implementasi K-Means

Jumlah kluster optimal ditentukan menggunakan metode Elbow dengan memperhatikan titik penurunan nilai WCSS yang mulai landai. Hasil analisis menunjukkan  $k=3$  sebagai titik optimal. Implementasi dilakukan dengan Python dan pustaka Scikit-learn [16], parameter: *n\_clusters=3*, *init='k-means++'*, *max\_iter=300*, *random\_state=42*.

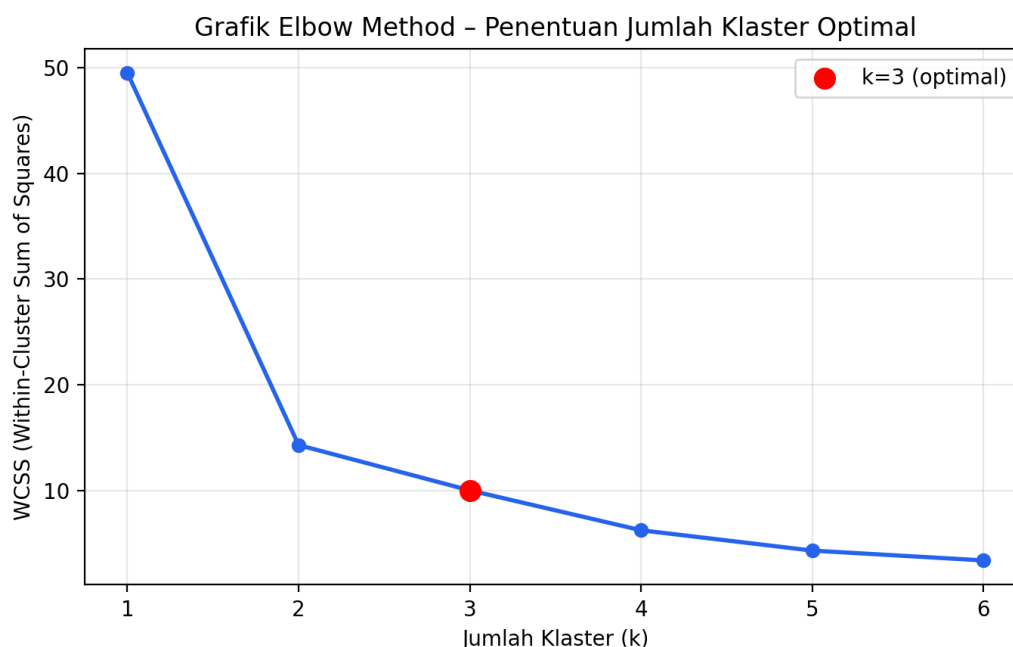
### 6. Evaluasi

Kualitas *clustering* dievaluasi menggunakan *Silhouette Score*, yang mengukur seberapa tepat suatu data berada di klasternya sendiri dibanding kluster lain. Nilai mendekati 1 menunjukkan pengelompokan yang baik, nilai mendekati 0 menunjukkan batas kluster kurang jelas, dan nilai negatif menunjukkan kemungkinan data salah kluster [6].

## 3. HASIL DAN PEMBAHASAN

## 1. Penentuan Jumlah Klaster

Sebelum *clustering* dijalankan, jumlah klaster optimal ditentukan terlebih dahulu menggunakan metode Elbow. Grafik Elbow (Gambar 1) menunjukkan penurunan nilai WCSS dari k=1 (49,49) hingga k=6 (3,42), dengan penurunan paling signifikan terjadi dari k=1 ke k=2 (49,49 menjadi 14,32) dan melandai secara bertahap setelahnya (WCSS k=3: 10,04; WCSS k=4: 6,27). Titik k=3 dipilih sebagai jumlah klaster optimal karena selaras dengan interpretasi bisnis tiga segmen pelanggan (Aktif, Potensial, Pasif) dan tetap memberikan kualitas pengelompokan yang baik.



**Gambar 1. Grafik Elbow Method – Penentuan Jumlah Klaster Optimal**

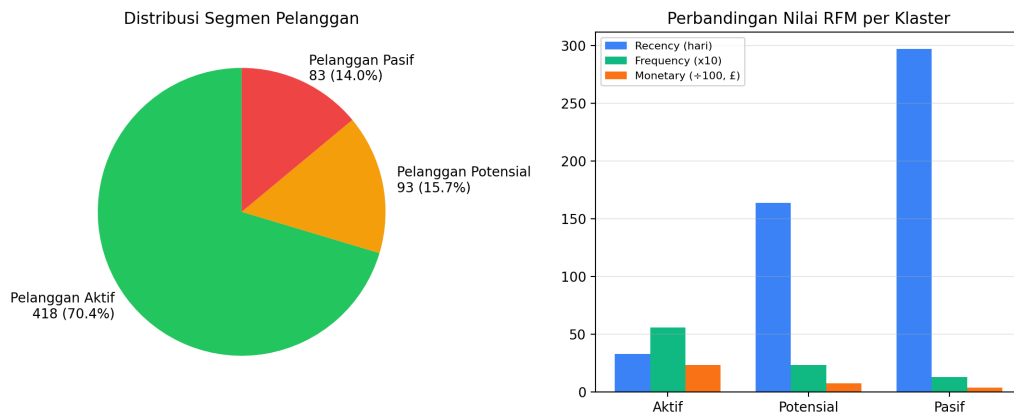
## 2. Hasil Segmentasi Pelanggan

Algoritma K-Means dengan k=3 menghasilkan tiga segmen pelanggan yang memiliki karakteristik berbeda secara signifikan (Tabel 1). Pelanggan Aktif berjumlah 418 orang (70,4%) dan merupakan kelompok terbesar, dengan rata-rata *recency* hanya 32,6 hari, frekuensi transaksi tertinggi (5,56 kali), dan nilai belanja tertinggi £2.315,07. Pelanggan Potensial berjumlah 93 orang (15,7%) berada pada posisi menengah di ketiga variabel RFM. Pelanggan Pasif berjumlah 83 orang (14,0%) dengan *recency* tertinggi (296,8 hari) dan nilai belanja terendah £368,38.

**Tabel 1. Hasil Segmentasi Pelanggan menggunakan K-Means Clustering**

| Klaster | Label               | Jumlah | Persentase | Avg. Recency | Avg. Frequency | Avg. Monetary |
|---------|---------------------|--------|------------|--------------|----------------|---------------|
| 0       | Pelanggan Aktif     | 418    | 70,4%      | 32,60 hari   | 5,56 kali      | £2.315,07     |
| 2       | Pelanggan Potensial | 93     | 15,7%      | 163,49 hari  | 2,31 kali      | £720,79       |
| 1       | Pelanggan Pasif     | 83     | 14,0%      | 296,77 hari  | 1,29 kali      | £368,38       |

Sumber: Data Diolah



**Gambar 2. Distribusi Segmen Pelanggan dan Perbandingan Nilai RFM per Kluster**

### 3. Evaluasi Hasil Clustering

*Silhouette Score* yang diperoleh adalah 0,612, melampaui ambang batas minimum 0,5 yang menjadi syarat *clustering* yang dapat diterima [9]. Ketiga segmen menunjukkan perbedaan yang jelas: Pelanggan Aktif dengan frekuensi tinggi dan *recency* rendah, Pelanggan Potensial di posisi tengah, dan Pelanggan Pasif dengan *recency* tinggi dan nilai belanja rendah. Hasil ini mengonfirmasi bahwa  $k=3$  merupakan pilihan kluster yang tepat dan segmentasi yang dihasilkan memiliki kualitas yang memadai untuk diterapkan dalam strategi bisnis.

### 4 Integrasi ke Sistem Informasi

Hasil segmentasi diintegrasikan ke sistem informasi *e-commerce* melalui REST API berbasis *framework* Flask (Python). *Dashboard* visualisasi dibangun menggunakan Chart.js untuk menampilkan distribusi kluster secara *real-time*. Dengan integrasi ini, tim manajemen dapat langsung mengakses informasi segmen pelanggan dan menerapkan strategi promosi yang lebih personal dan tepat sasaran.

### KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa *K-Means Clustering* berbasis analisis RFM efektif untuk segmentasi pelanggan *e-commerce*. Dari 600 pelanggan yang dianalisis, dihasilkan tiga segmen yang terdefinisi dengan baik: Pelanggan Aktif (70,4%), Pelanggan Potensial (15,7%), dan Pelanggan Pasif (14,0%), dengan *Silhouette Score* 0,612 yang menandakan kualitas *clustering* memadai. Integrasi hasil segmentasi ke sistem informasi melalui REST API berhasil menjadikan output algoritma langsung dapat digunakan untuk keperluan bisnis.

Penelitian ini memiliki keterbatasan pada cakupan data (satu periode, satu platform). Pengembangan ke depan dapat mempertimbangkan perbandingan dengan algoritma lain seperti DBSCAN atau *Agglomerative Clustering* [8][20], serta penambahan variabel seperti kategori produk dan lokasi pelanggan untuk memperkaya hasil segmentasi, mengingat model klasik *unsupervised learning* [17][19] dan konteks bisnis *e-commerce* modern [18] terus berkembang, sejalan dengan berbagai studi penerapan *clustering* pada domain lain [13][14][15].

### DAFTAR RUJUKAN

- [1] A. Khumaidi, H. Wahyono, R. Darmawan, H. D. Kartika, L. N. L. Chusna, and M. K. Fauzy, "RFM-AR model for customer segmentation using K-means algorithm," in *Proc. 8th ICIMECE*, vol. 465, 2023, doi: 10.1051/e3sconf/202346502005.
- [2] Z. Xian, P. Keikhosrkoiani, C. XinYing, and Z. Li, "An RFM model using K-means clustering to improve customer segmentation and product recommendation," in *Proc. CBCDASCE*, 2022, doi: 10.4018/978-1-6684-4168-8.ch006.
- [3] S. Dwivedi and A. Singh, "A study of customer segmentation based on RFM analysis and K-means," in *Proc. ICICC*, pp. 347-355, 2023, doi: 10.1007/978-981-99-4071-4\_27.
- [4] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi*, vol. 13, no. 2, 2023.
- [5] S. P. L. Geni, H. Safitri, F. Merry, M. Wati, and H. Havaluddin, "Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian," *JUKI: Jurnal Komputer dan Informatika*, vol. 7, no. 1, pp. 89-99, 2025.

- [6] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [7] S. Robo, P. I. Melani, P. Fernatyanan, M. R. Widianoro, and S. K. Bariyyah, "The application of customer segmentation using RFM analysis method and K-means clustering to improve marketing strategy," *IJISTECH*, vol. 8, no. 3, 2024.
- [8] Ramadhan dkk., "Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 1, 2024, doi: 10.30591/jpit.v9i1.5735.
- [9] Yulisasih dkk., "Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation," *ILKOM Jurnal Ilmiah*, 2024.
- [10] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "Implementasi K-Means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model RFM," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4677.
- [11] A. Cahya, A. P., Asyraf, M., dkk., "Klasterisasi Mahasiswa Magetan Menggunakan K-Means untuk Optimasi Strategi Promosi Perguruan Tinggi," *JIFoSI*, vol. 6, no. 1, pp. 22-30, 2025.
- [12] S. A. F. Salsabilah, Z. M. Zamzani, and W. Gustrifa, "Implementasi Data Mining K-Means Clustering untuk Prediksi Tingkat Pengangguran Terbuka di Jawa Timur," *JIFoSI*, vol. 5, no. 3, 2024.
- [13] M. M. Hidayat, R. D. Adityo, dkk., "Prediksi Pola Pemesanan Kamar Paling Diminati di Hotel Singaraja Menggunakan Metode KNN," *JIFoSI*, vol. 6, no. 1, pp. 31-36, 2025.
- [14] D. Reindra and H. Soetanto, "Deteksi Ujaran Kebencian Menggunakan Multinomial Naive Bayes untuk Pundit Sepakbola di Media Sosial X," *JIFoSI*, vol. 5, no. 3, 2024.
- [15] R. A. Dwiatmoko and Imelda, "Analisis Sentimen Kesehatan Mental di Twitter Menggunakan Naive Bayes Classifier dan K-Nearest Neighbor," *JIFoSI*, vol. 6, no. 1, pp. 1-9, 2025.
- [16] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [18] K. C. Laudon and C. G. Traver, *E-Commerce: Business, Technology, Society*, 16th ed. Pearson Education, 2021.
- [19] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2016.
- [20] A. Saputra, "Perbandingan algoritma DBSCAN dan K-means dalam segmentasi pelanggan pengguna transportasi publik Transjakarta menggunakan metode RFM," *MALCOLM*, vol. 4, no. 4, 2024, doi: 10.57152/Malcom.v4i4.1516.