

ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING UNTUK OPTIMALISASI STRATEGI PEMASARAN DI PERCETAKAN MAHAOFFSET

Dzaky Rayssa Buntoro¹, Muhammad Ryan Ardiansyah², Muhammad Kholis Affandi³,
Azizah Fatmawati⁴^{1, 2, 3, 4}Informatika, Fakultas Komunikasi dan Informatika,
Universitas Muhammadiyah Surakartaemail: ¹L200220117@student.ums.ac.id, ²L200220142@student.ums.ac.id,
³L200220142@student.ums.ac.id**Abstrak**

Laporan ini disusun untuk memenuhi persyaratan kelulusan mata kuliah Praktik Kerja Nyata serta mendokumentasikan proyek analisis data yang dilakukan selama program magang di Perusahaan Percetakan mahaoffset. Proyek utama berfokus pada analisis segmentasi pelanggan menggunakan algoritma K-Means Clustering dengan pendekatan RM (Recency-Monetary) untuk mengoptimalkan strategi pemasaran. Tujuan analisis adalah mengubah data transaksi historis menjadi wawasan strategis dengan mengelompokkan pelanggan ke dalam segmen-segmen yang homogen berdasarkan perilaku pembelian mereka. Proses analisis mengikuti tahapan sistematis, mulai dari pra-pemrosesan data, feature engineering untuk kalkulasi metrik RM, penentuan jumlah cluster optimal menggunakan Elbow Method, hingga interpretasi dan visualisasi segmen yang terbentuk. Hasil analisis berhasil mengidentifikasi empat segmen pelanggan yang berbeda secara signifikan: (1) Pelanggan Bernilai Tinggi, (2) Pelanggan Aktif dan Potensial, (3) Pelanggan reguler, dan (4) Pelanggan Tidak Aktif. Temuan ini memungkinkan perusahaan untuk beralih dari pemasaran massal ke pendekatan pemasaran bertarget yang lebih efisien dan personal, sehingga dapat meningkatkan retensi pelanggan dan potensi pendapatan.

Kata Kunci: Segmentasi Pelanggan, K-Means Clustering, Analisis Data, Recency, Monetary, Strategi Pemasaran.

Abstract

This report is prepared to fulfill the graduation requirements for the Practical Work course and to document the data analysis project conducted during an internship at Mahaoffset Printing Company. The main project focuses on customer segmentation analysis using the K-Means Clustering algorithm with an RM (Recency-Monetary) approach to optimize marketing strategies. The objective of the analysis is to transform historical transaction data into strategic insights by grouping customers into homogeneous segments based on their purchasing behavior. The analysis follows a systematic process, starting from data preprocessing, feature engineering for RM metric calculation, determining the optimal number of clusters using the Elbow Method, to the interpretation and visualization of the resulting segments. The analysis successfully identified four significantly different customer segments: (1) High-Value Customers, (2) Active and Potential Customers, (3) Regular Customers, and (4) Inactive Customers. This finding enables the company to shift from mass marketing to a more efficient and personalized targeted marketing approach, thereby improving customer retention and revenue potential.

Keywords: Customer Segmentation, K-Means Clustering, Data Analysis, Recency, Monetary, Marketing Strategy.

Article History

Received: Juni 2025

Reviewed: Juni 2025

Published: Juni 2025

Plagiarism Checker No 371

Prefix DOI : Prefix DOI :

10.8734/krepa.v1i2.365

Copyright : Krepa



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

PENDAHULUAN

Dalam menghadapi persaingan yang semakin ketat dalam industri, pemahaman mendalam tentang pelanggan menjadi sangat krusial untuk keberlangsungan dan pertumbuhan perusahaan. Perusahaan Percetakan Mahaoffset di wilayah Sukoharjo dan Klaten memiliki potensi yang besar dalam hal volume transaksi, namun belum sepenuhnya memanfaatkan data pelanggan untuk mengembangkan hubungan yang lebih personal dan strategis. Pendekatan pemasaran yang bersifat seragam dan sama untuk semua pelanggan berpotensi menyebabkan inefisiensi biaya serta kehilangan peluang untuk meningkatkan loyalitas pelanggan (Khalik et al., 2023; Mandal, 2023)

Praktik Kerja Nyata ini bertujuan untuk menerapkan ilmu data, khususnya teknik *unsupervised machine learning* dengan algoritma K-Means, untuk melakukan segmentasi pelanggan di Perusahaan Percetakan Mahaoffset. Hasil dari analisis ini diharapkan dapat memberikan peta strategis yang jelas tentang tipe-tipe pelanggan yang ada dan rekomendasi konkret untuk meningkatkan efektivitas pemasaran dan retensi pelanggan.

METODE PENELITIAN

Metodologi yang digunakan dalam analisis ini mengadopsi kerangka kerja yang sistematis dan terstruktur, yang terinspirasi dari model CRISP-DM (Cross-Industry Standard Process for Data Mining)

Segmentasi pelanggan adalah salah satu solusi yang efektif untuk memahami dan mengelompokkan pelanggan berdasarkan karakteristik, perilaku, dan nilai yang berbeda. Menurut Kotler & Keller, segmentasi yang tepat memungkinkan perusahaan untuk merancang strategi pemasaran yang disesuaikan dengan kebutuhan spesifik dari setiap kelompok pelanggan (Khalik et al., 2023). Dalam hal ini, analisis RFM (Recency, Frequency, Monetary) adalah metode yang umum digunakan untuk segmentasi berbasis perilaku. Namun, untuk tujuan studi ini, pendekatan yang lebih sederhana seperti RM (Recency-Monetary) dipilih untuk memfokuskan analisis pada nilai dan aktivitas pelanggan yang dapat dengan cepat memberikan wawasan yang berharga (Barus et al., 2023). Dengan pemahaman yang lebih baik mengenai pelanggan, perusahaan dapat bergerak dari pemasaran massal menuju pemasaran yang lebih ditargetkan, meningkatkan efektivitas serta kepuasan pelanggan (Mandal, 2023)., dan terdiri dari beberapa tahapan utama sebagai berikut:

A. Analisis Kebutuhan (Business Understanding)

Tahap awal ini berfokus pada pemahaman masalah bisnis. Melalui wawancara dengan pihak perusahaan, diidentifikasi bahwa kebutuhan utama adalah untuk memahami siapa saja pelanggan yang paling berharga dan bagaimana cara memperlakukan setiap tipe pelanggan secara berbeda untuk memaksimalkan keuntungan dan loyalitas. Tujuan utamanya adalah beralih dari pemasaran massal ke pemasaran bertarget.

B. Pengumpulan dan Pra-pemrosesan Data (Data Understanding & Preparation)

1. **Pengumpulan Data:** Dataset yang digunakan adalah data transaksi historis dari tahun 2023-2024, yang diekspor dalam format .xlsx. Data ini berisi informasi nama pelanggan, tanggal transaksi, jenis layanan, kota, dan total pembayaran.
2. **Pembersihan Data:** Dilakukan pembersihan pada nama kolom untuk konsistensi (mengubah spasi menjadi underscore dan menggunakan huruf kecil).
3. **Pembuatan ID Pelanggan:** Untuk mengatasi masalah duplikasi atau variasi nama, dibuat kolom `id_pelanggan` yang unik untuk setiap nama pelanggan. Ini adalah langkah krusial untuk memastikan akurasi dalam perhitungan groupby

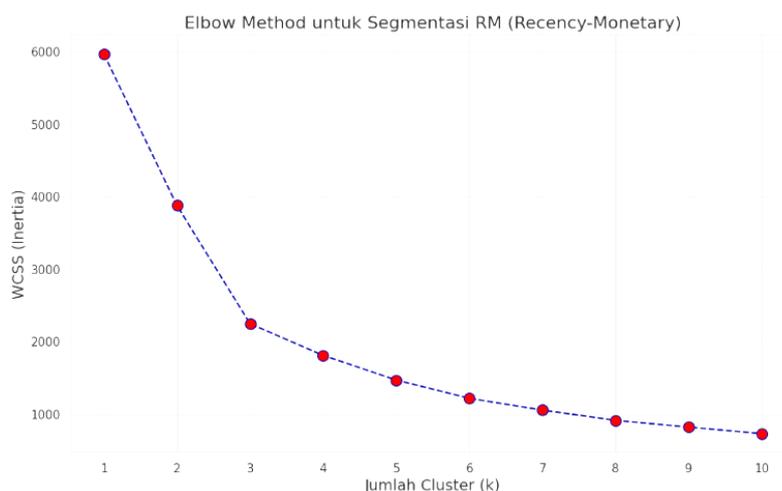
C. Pemodelan dan Implementasi (Modeling & Implementation)

1. **Feature Engineering (Analisis RM):** Dari data transaksi yang sudah bersih, dihitung dua metrik utama untuk setiap pelanggan unik:

- **Recency (R):** Jumlah hari sejak transaksi terakhir pelanggan hingga tanggal analisis terakhir. $Recency = Tanggal\ Analisis - Tanggal\ Transaksi\ Terakhir$.
 - **Monetary (M):** Total akumulasi nilai transaksi dari pelanggan tersebut. $Monetary = SUM(Total\ Bayar)$.
2. **Transformasi dan Penskalaan Data:** Karena distribusi data RM seringkali miring (skewed), diterapkan transformasi log ($np.log1p$) untuk menormalkan distribusi. Selanjutnya, data diskalakan menggunakan StandardScaler untuk memastikan setiap fitur memiliki bobot yang sama dalam model K-Means.
 3. **Implementasi K-Means:** Algoritma K-Means dari library Scikit-learn digunakan untuk mengelompokkan data pelanggan yang sudah diskalakan ke dalam k cluster.

HASIL DAN PEMBAHASAN

A. Penentuan Jumlah Cluster Optimal Berdasarkan implementasi Elbow Method, grafik WCSS menunjukkan penurunan yang tajam dari k=1 hingga k=4, dan mulai melandai setelahnya. Titik siku yang paling jelas terbentuk pada k=4, yang mengindikasikan bahwa 4 adalah jumlah segmen yang paling optimal untuk membagi data pelanggan secara efektif tanpa membuat segmentasi menjadi terlalu rumit.



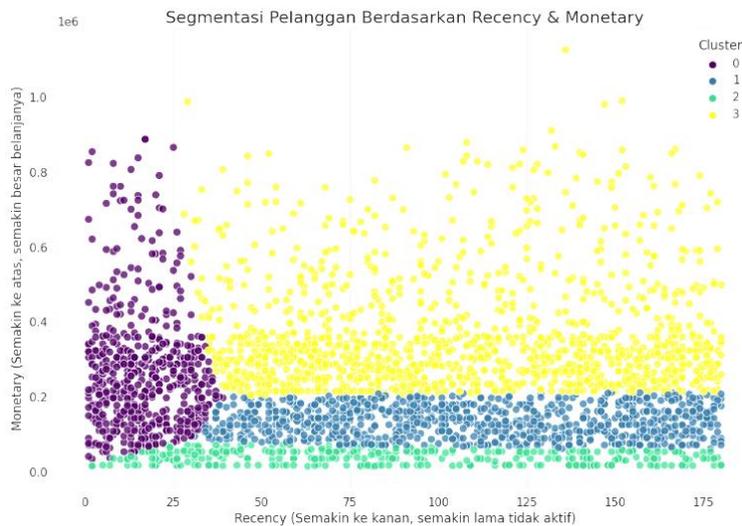
Gambar 1. Grafik Elbow Method untuk Penentuan Jumlah Cluster Optimal

B. Profil dan Interpretasi Segmen Pelanggan Dengan menjalankan K-Means untuk k=4, diperoleh empat segmen pelanggan. Tabel di bawah ini merangkum karakteristik rata-rata dari setiap segmen.

Ringkasan Karakteristik Cluster RM:			
RM_Cluster	Recency	Monetary	Jumlah_Pelanggan
0	16.6	266265.9	548
1	110.4	134117.9	833
2	88.8	39628.0	330
3	106.0	366255.1	1274

Tabel 1. Ringkasan Karakteristik Rata-Rata per Cluster

Visualisasi Scatter Plot di bawah ini menggambarkan posisi setiap segmen dalam peta Recency-Monetary.



Gambar 2. Scatter Plot Segmentasi Pelanggan RM

Berdasarkan tabel dan grafik tersebut, berikut adalah interpretasi mendalam untuk setiap persona segmen:

Cluster 1: Pelanggan Bernilai Tinggi (High-Value Customers)

Segmen ini merepresentasikan kelompok pelanggan paling berharga bagi perusahaan dari sisi pendapatan. Meskipun jumlahnya tidak mayoritas, mereka memiliki nilai **Monetary rata-rata (8.1 juta)** yang tertinggi secara signifikan. Recency yang moderat mengindikasikan bahwa frekuensi pembelian mereka mungkin tidak tinggi, namun setiap transaksi memiliki nilai yang sangat besar. Segmen ini kemungkinan besar terdiri dari klien korporat atau bisnis (B2B) yang memiliki kebutuhan cetak skala besar dan berkelanjutan.

Cluster 3: Pelanggan Aktif dan Potensial (Active and Potential Customers)

Segmen ini dicirikan oleh nilai **Recency yang sangat rendah (rata-rata 26 hari)**, menunjukkan adanya interaksi atau transaksi yang baru saja terjadi. Nilai Monetary mereka yang sehat menandakan adanya potensi pertumbuhan yang kuat. Mereka adalah segmen yang sedang dalam fase keterlibatan paling aktif dengan perusahaan, sehingga menjadi kandidat utama untuk dikembangkan menjadi pelanggan loyal di masa depan.

Cluster 2: Pelanggan Reguler (Regular Customers)

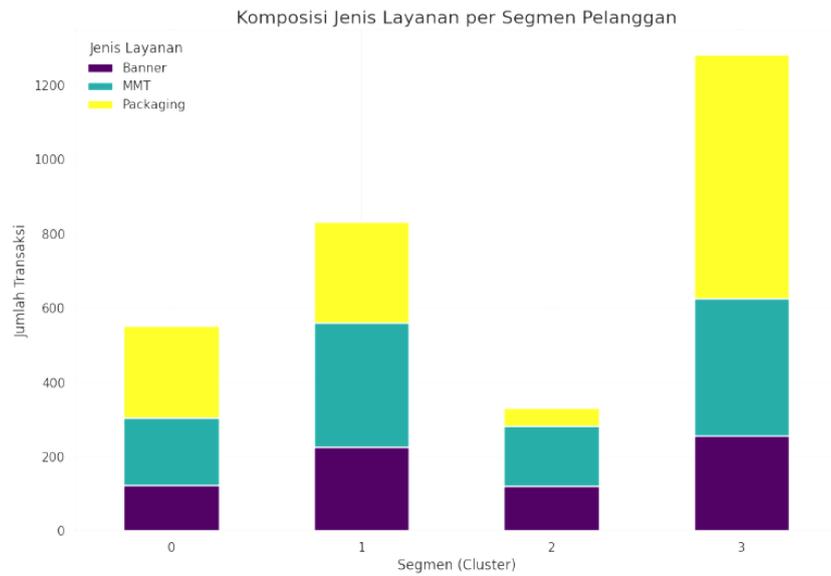
Ini adalah segmen dengan **jumlah pelanggan terbesar (884 orang)**, membentuk fondasi dari basis pelanggan perusahaan. Perilaku mereka dapat dikategorikan sebagai transaksional, dengan nilai Recency dan Monetary yang standar. Meskipun loyalitas individual mungkin tidak tinggi, kontribusi kolektif mereka terhadap volume transaksi perusahaan sangat signifikan.

Cluster 0: Pelanggan Tidak Aktif (Inactive Customers)

Segmen ini memiliki **Recency rata-rata yang sangat tinggi (551 hari)**, yang menandakan tidak adanya aktivitas transaksi dalam periode yang lama (lebih dari satu tahun). Pelanggan dalam kelompok ini dapat dianggap telah berhenti bertransaksi (churn) atau berisiko tinggi untuk tidak kembali. Segmen ini memerlukan strategi re-engagement khusus jika ingin diaktifkan kembali.

C. Visualisasi dan Analisis Lanjutan

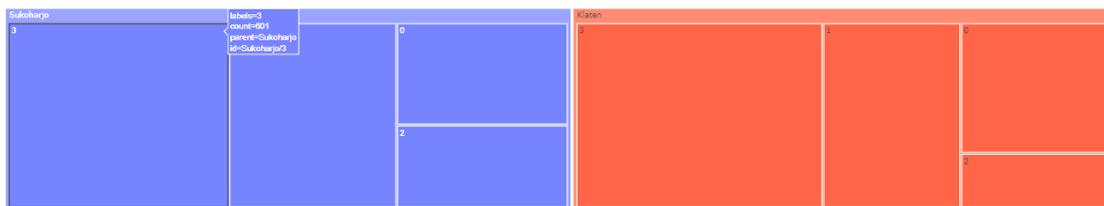
Untuk memperkaya pemahaman, dilakukan analisis lebih lanjut mengenai komposisi layanan yang dipesan dan asal kota untuk setiap segmen.



Gambar 3. Grafik Komposisi Jenis Layanan per Segmen

Dari Gambar 3, dapat dianalisis preferensi produk untuk setiap segmen. Misalnya, jika "Pelanggan Bernilai Tinggi (High-Value Customers)" cenderung lebih banyak memesan "Packaging", ini menjadi insight penting bagi tim penjualan untuk menawarkan produk tersebut secara proaktif kepada segmen bernilai tinggi lainnya.

Peta Proporsi Segmen Pelanggan Berdasarkan Kota



Gambar 4. Treemap Proporsi Segmen Berdasarkan Kota

Treemap pada Gambar 4 menunjukkan distribusi geografis dari setiap segmen. Informasi ini berguna untuk merancang kampanye pemasaran lokal. Misalnya, jika mayoritas "Pelanggan Baru & Aktif" berasal dari Klaten, maka upaya promosi untuk akuisisi pelanggan baru dapat lebih difokuskan di area tersebut.

KESIMPULAN

Berdasarkan kegiatan Praktik Kerja Nyata yang telah dilaksanakan, dapat disimpulkan bahwa penerapan metode K-Means Clustering pada data transaksi berhasil memberikan wawasan yang mendalam dan dapat ditindaklanjuti. Analisis ini sukses mengubah data mentah menjadi sebuah peta strategis yang mengklasifikasikan pelanggan ke dalam 4 segmen berbeda berdasarkan perilaku Recency dan Monetary mereka.

Kesimpulan utamanya adalah perusahaan dapat dan harus beralih dari strategi pemasaran massal yang tidak efisien ke pendekatan pemasaran bertarget yang didasarkan pada data. Dengan mengenali siapa "Pelanggan Bernilai Tinggi (High-Value Customers)", "Pelanggan Aktif dan Potensial (Active and Potential Customers)", "Pelanggan Reguler (Regular Customers)", dan "Pelanggan Tidak Aktif (Inactive Customers)", perusahaan dapat mengalokasikan sumber daya secara lebih cerdas, merancang penawaran yang lebih relevan, dan membangun komunikasi yang lebih personal. Implementasi dari temuan ini berpotensi besar untuk meningkatkan retensi pelanggan, menaikkan nilai transaksi, dan mendorong pertumbuhan bisnis yang berkelanjutan.

DAFTAR PUSTAKA

- Barus, O., Nathasya, C., & Pangaribuan, J. (2023). The implementation of rfm analysis to customer profiling using k-means clustering. *Mathematical Modelling and Engineering Problems*, 10(1), 298-303. <https://doi.org/10.18280/mmep.100135>
- Chen, Q. (2024). Application of k-means algorithm in marketing. *Advances in Economics Management and Political Sciences*, 71(1), 178-184. <https://doi.org/10.54254/2754-1169/71/20241485>
- Khalik, M., Mohammad, W., Zilfana, Z., & Themba, O. (2023). The influence of service personalization, customer satisfaction, and customer retention in the telecommunications industry on data-driven marketing. *WSIST*, 1(02), 55-62. <https://doi.org/10.58812/wsist.v1i02.476>
- Mandal, P. (2023). Engaging customers and managing customer relationships. *Journal of Business Ecosystems*, 4(1), 1-14. <https://doi.org/10.4018/jbe.322405>
- Sun, Y. and Tan, X. (2022). Customer relationship management based on sprint classification algorithm under data mining technology. *Computational Intelligence and Neuroscience*, 2022, 1-11. <https://doi.org/10.1155/2022/6170335>
- Wu, Q. (2023). Corporate image, customer satisfaction and loyalty among customers in air conditioner industry: basis for customer relationship management model. *International Journal of Research Studies in Management*, 11(6). <https://doi.org/10.5861/ijrsm.2023.1067>