

Sistem Segmentasi Pelanggan Berbasis Web Menggunakan Algoritma K-Means Clustering

Muhhammad Rafli Anwar^{1*}, Eko Purwanto², Rudi Susanto³

¹Program Studi Teknik Informatika,
Universitas Duta Bangsa Surakarta
^{1*}202020123@mhs.udb.ac.id

²Program Studi Sistem Informasi,
Universitas Duta Bangsa Surakarta
^{2*}eko_purwanto@udb.ac.id

³Program Studi Teknik Informatika,
Universitas Duta Bangsa Surakarta
^{3*}rudi_susanto@udb.ac.id

Abstrak—Memahami karakteristik konsumen melalui segmentasi pelanggan menjadi strategi krusial bagi dunia bisnis untuk menyusun kebijakan pemasaran yang lebih terarah. Penelitian ini bertujuan merancang sistem berbasis web yang menerapkan algoritma K-Means Clustering dalam proses segmentasi pelanggan secara otomatis. Metode penelitian yang digunakan adalah Research and Development dengan pendekatan kuantitatif. React JS dipilih sebagai teknologi frontend dan Flask sebagai backend API yang memanfaatkan library scikit-learn untuk kebutuhan clustering. Data yang dipakai merupakan data simulasi pelanggan dengan atribut total belanja dan frekuensi transaksi. Evaluasi kualitas clustering dilakukan menggunakan Elbow Method dan Silhouette Score. Hasil pengujian memperlihatkan bahwa sistem mampu mengelompokkan pelanggan ke dalam tiga cluster dengan Silhouette Score 0,6792 yang menandakan kualitas clustering tergolong baik. Kata kunci—Segmentasi Pelanggan; K-Means Clustering; React JS; Flask; Sistem Berbasis Web

Abstract—Understanding consumer characteristics through customer segmentation has become a vital strategy for businesses in designing more targeted marketing policies. This research aims to develop a web-based system that applies the K-Means Clustering algorithm for automated customer segmentation. The system employs React JS as the frontend framework and Flask as the backend API, utilizing the scikit-learn library for clustering tasks. The dataset consists of simulated customer data with attributes of total spending and transaction frequency. Clustering quality was evaluated using the Elbow Method and Silhouette Score. Experimental results demonstrated that the system successfully grouped customers into three clusters with a Silhouette Score of 0.6792, indicating good clustering quality.

Keywords—Customer Segmentation; K-Means Clustering; React JS; Flask; Web-Based System

V. PENDAHULUAN

Dalam dunia bisnis, pemahaman terhadap profil konsumen melalui segmentasi pelanggan menjadi kunci utama dalam menyusun strategi pemasaran yang efektif. [1] menyatakan bahwa segmentasi pelanggan memungkinkan perusahaan membagi pasar ke dalam kelompok-kelompok dengan karakteristik homogen, sehingga setiap segmen dapat didekati dengan strategi bisnis yang sesuai. Sayangnya, masih banyak perusahaan khususnya Usaha Kecil dan Menengah (UKM) yang melakukan segmentasi secara manual memanfaatkan rekam sederhana, atau bahkan belum melakukannya sama sekali. Akibatnya, penargetan pasar dan alokasi sumber daya pemasaran menjadi tidak efektif [2].

Kompleksitas permasalahan ini semakin bertambah seiring pertumbuhan volume data pelanggan yang dihasilkan dari transaksi harian. Informasi bernilai tinggi seperti kebiasaan belanja, frekuensi transaksi, dan preferensi produk kerap terkubur dalam kumpulan data berukuran besar dan sulit diekstraksi tanpa bantuan teknik data mining [3]. Clustering hadir sebagai salah satu solusi yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik antar data tanpa memerlukan label sebelumnya.

K-Means menjadi algoritma clustering yang banyak diadopsi karena implementasinya yang sederhana, efisiensi komputasi yang tinggi, serta kapasitasnya dalam memproses data bervolume besar [4]; [5]. Beberapa studi sebelumnya telah mengaplikasikan K-Means untuk segmentasi pelanggan, misalnya [2] yang berhasil menetapkan pelanggan ritel dan [6] yang memanfaatkannya untuk pengambilan keputusan bisnis. Meskipun demikian, pendekatan yang digunakan dalam penelitian-penelitian tersebut masih bersifat offline sehingga membutuhkan instalasi perangkat lunak tambahan dan keahlian teknis tertentu.

Perkembangan teknologi web masa kini membuka peluang untuk menjawab tantangan tersebut. Kerangka kerja seperti React JS di sisi

frontend dan Flask di sisi backend memungkinkan pengembangan sistem berbasis web yang modular, responsif, dan mudah dijangkau. Berdasarkan hal tersebut, penelitian ini bertujuan mengembangkan sistem berbasis web yang menerapkan algoritma K-Means untuk segmentasi pelanggan secara otomatis, dilengkapi visualisasi interaktif serta metrik evaluasi seperti Silhouette Score dan Elbow Method, guna mempermudah dan mempercepat proses analisis data pelanggan secara lebih informatif.

VI. METODOLOGI PENELITIAN

Penelitian ini mengadopsi metode Research and Development [7] dengan pendekatan kuantitatif sederhana. Fokus utama penelitian adalah membangun sistem berbasis web yang menerapkan algoritma K-Means untuk proses segmentasi pelanggan. Data yang dipakai merupakan data simulasi (dummy) yang mencakup informasi total belanja dan jumlah transaksi, disusun sedemikian rupa layaknya kondisi data pelanggan sesungguhnya.

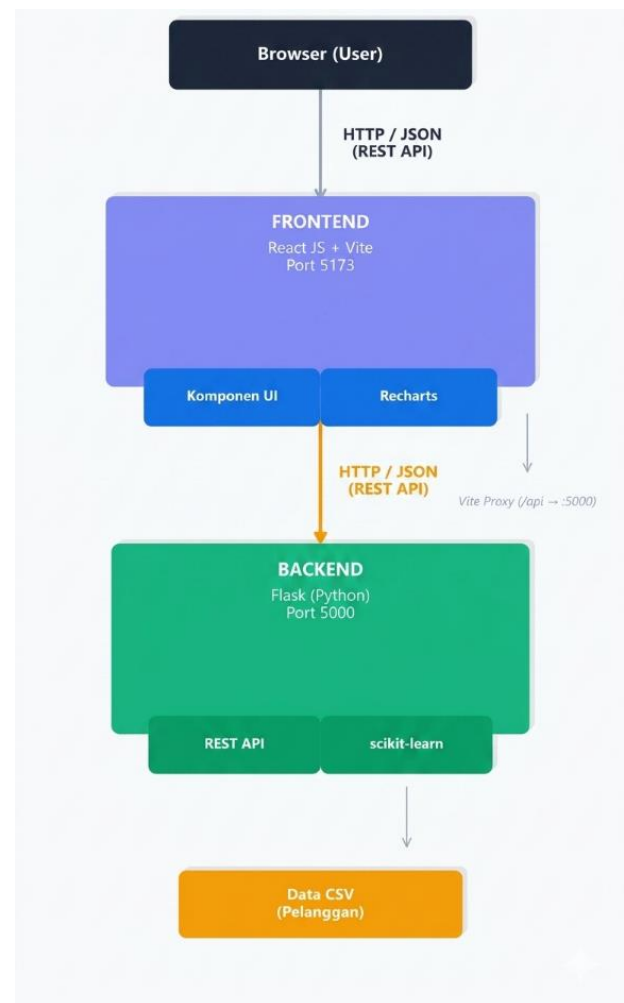
Pengembangan sistem dilakukan dengan model Waterfall [8] melalui lima tahapan: (1) analisis kebutuhan yang meliputi kebutuhan fungsional dan non-fungsional; (2) perancangan sistem mencakup arsitektur sistem, flowchart algoritma K-Means, serta antarmuka pengguna; (3) implementasi menggunakan React JS sebagai frontend dan Flask (Python) sebagai backend dengan library scikit-learn; (4) pengujian sistem melalui black-box testing disertai evaluasi kualitas clustering memakai Silhouette Score dan Elbow Method; serta (5) evaluasi akhir untuk perbaikan apabila ditemukan kendala.

Tahapan penelitian mencakup identifikasi masalah, studi literatur, pengumpulan data, perancangan sistem, implementasi, pengujian, dan analisis hasil. Arsitektur sistem terdiri dari frontend (React JS) dan backend (Flask) yang saling berkomunikasi melalui REST API. Algoritma K-Means beroperasi melalui langkah-langkah berikut: (1)

penetapan jumlah cluster K, (2) inialisasi centroid awal secara acak, (3) perhitungan jarak setiap data ke centroid dengan Euclidean Distance, (4) pengelompokan data ke centroid dengan jarak terdekat, (5) pembaruan centroid berdasarkan rata-rata data dalam tiap cluster, dan (6) pengulangan langkah (3) hingga (5) hingga tercapai konvergensi.[4][5]



Gambar 30. Alur Penelitian



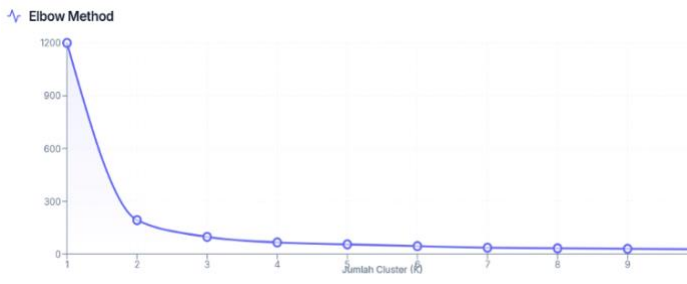
Gambar 2. Arsitektur Sistem

VII. HASIL DAN PEMBAHASAN

Sistem telah berhasil dibangun menggunakan React JS di sisi frontend dan Flask sebagai backend API. Pengguna dapat mengunggah file CSV, memilih kolom fitur yang akan dianalisis, menentukan jumlah cluster K, lalu menjalankan proses clustering. Hasil akhir ditampilkan dalam sebuah dashboard yang memuat statistik utama, tabel karakteristik cluster, scatter plot interaktif, dan grafik Elbow Method.

K. Elbow Method

Penentuan jumlah cluster optimal dilakukan dengan Elbow Method melalui perhitungan inertia untuk rentang K=1 hingga K=10 [9]. Hasil perhitungan memperlihatkan titik siku (elbow) terletak pada K=2 atau K=3, dengan inertia yang turun secara signifikan dari 1200 (K=1) menjadi 193,11 (K=2) dan 97,92 (K=3), kemudian penurunan cenderung melambat pada nilai K berikutnya. Temuan ini mengindikasikan bahwa jumlah cluster optimal untuk dataset yang digunakan adalah K=3.



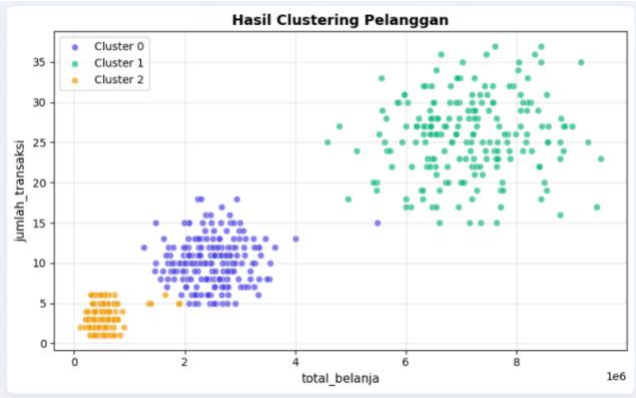
Gambar 3. Grafik Elbow Method

L. Hasil Clustering

Proses clustering dengan K=3 menghasilkan tiga kelompok pelanggan dengan profil yang berbeda. Cluster 0 (Menengah) memiliki rata-rata belanja Rp2.493.289 per pelanggan dengan 10 kali transaksi. Cluster 1 (Tinggi) mencatat rata-rata belanja Rp7.136.513 dengan 26 transaksi, sementara Cluster 2 (Rendah) hanya memiliki rata-rata belanja Rp521.773 dengan 3 transaksi. Karakteristik detail dari masing-masing cluster disajikan pada Tabel 1.

Tabel 12. Karakteristik Cluster

Cluster	Total Belanja (mean)	Jumlah Transaksi (mean)	Jumlah Pelanggan
Cluster 0 (Menengah)	Rp2.493.289	10	196
Cluster 1 (Tinggi)	Rp7.136.513	26	199
Cluster 2 (Rendah)	Rp521.773	3	205



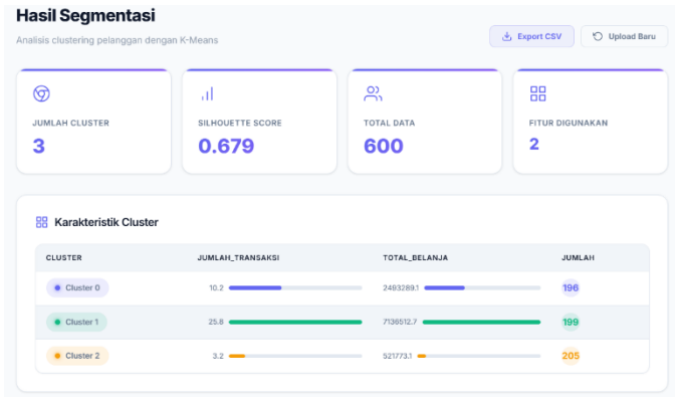
Gambar 4. Scatter Plot Hasil Clustering

M. Evaluasi Silhouette Score

Evaluasi kualitas clustering menggunakan Silhouette Score dilakukan pada berbagai variasi nilai K. Hasil perbandingan memperlihatkan bahwa K=2 memperoleh Silhouette Score 0,7410, K=3 sebesar 0,6792, K=4 sebesar 0,5850, dan K=5 sebesar 0,5194. Dengan nilai 0,6792 pada K=3, kualitas clustering berada pada kategori baik menurut kriteria [10], yang mengonfirmasi bahwa algoritma K-Means mampu memisahkan data pelanggan ke dalam kelompok-kelompok yang kohesif.

Tabel 2. Perbandingan Silhouette Score

Jumlah Cluster (K)	Silhouette Score
2	0,7410
3	0,6792
4	0,5850
5	0,5194



Gambar 5. Dashboard Hasil Segmentasi

N. Analisis Hasil Segmentasi

Ketiga cluster hasil segmentasi memperlihatkan profil yang saling berbeda. Temuan ini sejalan dengan riset [11] pada transaksi ritel online dan [12] dengan pendekatan RFM, yang juga menghasilkan tiga kelompok pelanggan dengan pola serupa. Cluster 2 (Rendah) merepresentasikan pelanggan pasif dengan nilai belanja dan frekuensi transaksi rendah sehingga memerlukan strategi aktivasi khusus. Cluster 0 (Menengah) adalah pelanggan regular yang potensial untuk ditingkatkan melalui program promosi dan loyalitas. Sementara Cluster 1 (Tinggi) merupakan pelanggan premium dengan nilai tertinggi yang perlu dipertahankan melalui layanan eksklusif. Pola segmentasi tiga kelompok ini juga dikonfirmasi oleh studi [13] dan [2] dalam konteks ritel. Informasi segmentasi ini dapat menjadi landasan bagi perusahaan dalam mengambil keputusan strategis dan menyusun strategi pemasaran yang lebih terarah.

Validasi fungsionalitas sistem dilakukan melalui pengujian black-box pada seluruh fitur. Hasil pengujian mengonfirmasi bahwa semua fitur berjalan sesuai yang diharapkan, mencakup unggah CSV (valid maupun invalid), pemilihan fitur, Elbow Method, clustering, scatter plot, dan ekspor data.

VIII. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means ke dalam sebuah sistem berbasis web untuk segmentasi pelanggan secara otomatis menggunakan React JS dan Flask. Sistem yang dikembangkan mampu menerima masukan data CSV, memproses clustering dengan algoritma K-Means, serta menyajikan hasil secara interaktif melalui dashboard yang terdiri dari statistik utama, tabel karakteristik cluster, scatter plot, dan grafik Elbow Method.

Sistem mampu memetakan pelanggan ke dalam tiga cluster dengan profil yang berbeda: Cluster 0 (Menengah) dengan rata-rata belanja Rp2.493.289 dan 10 transaksi, Cluster 1 (Tinggi) dengan rata-rata Rp7.136.513 dan 26 transaksi, serta Cluster 2 (Rendah) dengan rata-rata Rp521.773 dan 3 transaksi. Silhouette Score sebesar 0,6792 menegaskan bahwa kualitas clustering yang dihasilkan berada pada kategori baik.

Fitur Elbow Method dan Silhouette Score yang tersedia pada sistem memudahkan pengguna dalam menentukan jumlah cluster terbaik dan mengevaluasi kualitas hasil segmentasi. Pengembangan ke depannya

dapat menambahkan algoritma clustering lain seperti DBSCAN, mengintegrasikan data riil perusahaan, serta melengkapi sistem dengan fitur autentikasi pengguna.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Duta Bangsa Surakarta atas dukungan dan fasilitas yang telah diberikan selama proses penelitian ini. Penulis juga menyampaikan rasa terima kasih kepada kedua orang tua tercinta atas doa dan dukungan yang tiada henti, serta kepada segenap akademisi yang telah memberikan bimbingan, masukan, dan arahan yang berharga selama pengerjaan penelitian ini.

REFERENSI

- ADDIN Mendeley Bibliography CSL_BIBLIOGRAPHY 1. Pressman RS. *Software Engineering: A Practitioner's Approach*. 8th ed. McGraw-Hill; 2015.
2. Sugiyono. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta; 2013.
 3. MacQueen J. Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. 1967. p. 281–97.
 4. Prasetyo E. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Andi Offset; 2014.
 5. Kodinariya TM, Makwana PR. Review on Determining Number of Cluster in K-Means Clustering. *Int J Adv Res Comput Sci Manag Stud*. 2013;1(6):90–5.
 6. Larose DT. *Discovering Knowledge in Data: An Introduction to Data Mining*. 2nd ed. Wiley; 2014.
 7. Sarkar M, Puja AR, Chowdhury FR. Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis. *J Bus Manag Stud*. 2024;6(2):54–60.
 8. Rajput L, Singh SN. Customer Segmentation of E-Commerce Data Using K-Means Clustering Algorithm. In: *Proceedings of the 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. 2023. p. 658–64.
 9. Awaliyah DA, Prasetyo B, Muzaynah R, Lestari AD. Optimizing Customer Segmentation in Online Retail Transactions Through the Implementation of the K-Means Clustering Algorithm. *Sci J Informatics*. 2024;11:539.
 10. Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data. *Inf Sci (Ny)*. 2023;622:178–210.
 11. Gomes FC, Meisen T. A Review on Customer Segmentation Methods for Personalized Services. In: *Procedia Computer Science*. 2023. p. 582–9.
 12. Shi H, Xu M. A Data-Driven Customer Segmentation Method Based on K-Means and Improved RFM. In: *Procedia Computer Science*. 2023. p. 590–7.
 13. Arnesta AR, Fahmi A, Nurhadi N. Implementasi Algoritma K-Means untuk Segmentasi Pelanggan Berdasarkan Data Transaksi. *J Teknol Inf*. 2025;9(1):45–56.