Jurnal Minfo Polgan (JMP)
Volume 15, Nomor 3, September-Desember 2026
DOI : https://doi.org/10.33395/jmp.v15i3.16877
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Segmentasi Pelanggan Berbasis *Rfm* dan *K-Means* pada Sistem Analitik Penjualan

[1]Rayhan Arma Novriansyah, [2]Yohanni Syahra
[1,2*]Program Studi Teknologi Informasi dan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
*Korespondensi: yohannisyahra@umsu.ac.id

Submit : 26 Sept 2026 | **Diterima** : 01 Nov 2026 | **Terbit** : 08 Nov 2026

ABSTRACT

The growth of daily transaction data volumes in the retail sector often ends up as passive digital archives due to resource limitations in manual analysis. This condition makes it difficult for management to map consumer loyalty objectively, causing marketing strategies to miss the target. This study aims to design a web-based analytical information system using Python to automate customer segmentation. The approach used is the Knowledge Discovery in Database (KDD) cycle, integrating RFM (Recency, Frequency, Monetary) model feature extraction with the K-Means Clustering Unsupervised Machine Learning algorithm. From a total of 2,590 rows of raw transaction data, the system successfully pre-processed and aggregated them into 60 unique customer entities ready for clustering. Testing using the Elbow Method validated that the optimal number of partitions for this dataset is three clusters. Based on K-Means computation using Euclidean Distance, the system reached absolute convergence at the eighth iteration. This study successfully partitioned consumers into three segments: High Priority or Loyal Cluster, Medium Priority or Regular Cluster, and Low Priority or Passive Cluster. This analytical system is proven effective in visualizing segmentation autonomously, utilized as business intelligence to design targeted promotions and minimize customer churn risk. For future research, it is recommended to integrate directly into the point-of-sale system database in real-time and add customer demographic variables to make the analysis more comprehensive.

***Keywords**: Data Mining, K-Means Clustering, Model RFM, Python, Segmentasi Pelanggan.*

ABSTRAK

Pertumbuhan volume data transaksi harian pada sektor retail seringkali hanya berujung menjadi arsip digital pasif akibat keterbatasan sumber daya dalam menganalisisnya secara manual. Kondisi ini menyulitkan pihak manajemen dalam memetakan loyalitas konsumen secara objektif, sehingga strategi pemasaran sering salah sasaran. Penelitian ini bertujuan untuk merancang bangun sistem informasi analitik berbasis web menggunakan bahasa pemrograman Python untuk mengotomatisasi segmentasi pelanggan. Pendekatan yang digunakan adalah siklus Knowledge Discovery in Database (KDD) yang mengintegrasikan ekstraksi fitur model *RFM* (Recency, Frequency, Monetary) dengan algoritma Unsupervised Machine Learning yaitu *K-Means* Clustering. Dari total 2.590 baris data log transaksi mentah, sistem berhasil melakukan pra-pemrosesan dan agregasi menjadi 60 entitas pelanggan unik yang siap diklasterisasi. Pengujian menggunakan Elbow Method memvalidasi bahwa jumlah partisi paling optimal untuk dataset ini adalah tiga klaster. Berdasarkan hasil komputasi *K-Means* menggunakan pengukuran Euclidean Distance, sistem mencapai titik konvergen mutlak pada iterasi kedelapan. Hasil penelitian ini sukses mempartisi konsumen ke dalam tiga segmen, yaitu Klaster Prioritas Tinggi atau Loyal, Klaster Prioritas Sedang atau Reguler, dan Klaster Prioritas Rendah atau Pasif. Sistem analitik ini terbukti efektif dalam memvisualisasikan segmentasi secara otonom, yang dapat dimanfaatkan oleh manajemen sebagai intelijen bisnis untuk merancang kampanye promosi tertarget dan meminimalisir risiko kehilangan pelanggan. Untuk penelitian selanjutnya, disarankan melakukan integrasi langsung ke basis data sistem kasir secara waktu nyata serta menambahkan variabel demografis pelanggan agar analisis lebih komprehensif.

Kata Kunci: Data Mining, *K-Means* Clustering, Model *RFM*, Python, Segmentasi Pelanggan.

@2026 The Author Published By Politeknik Ganesha Medan
This is an open access article under the CC BY licence (http://creativecommons.org/licenses/by/4.0/). which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properlycited.

PENDAHULUAN

Pertumbuhan *volume* data transaksi pada sektor perdagangan *retail* saat ini menghasilkan tumpukan informasi historis yang sangat besar. Namun, tantangan utama yang dihadapi oleh banyak pelaku usaha adalah keterbatasan sistem dan sumber daya dalam mengolah data tersebut untuk memahami karakteristik pelanggan. Tanpa pengolahan analitik yang tepat, ribuan catatan transaksi harian hanya akan berujung menjadi arsip *digital* pasif. Kondisi tersebut selaras dengan permasalahan nyata pada rekam jejak sistem penjualan *retail*, di mana ditemukan tumpukan log transaksi yang sangat masif mencapai 2.590 baris data transaksi murni dalam kurun waktu observasi hingga bulan Juli 2025. Pihak manajemen mengalami kesulitan untuk memetakan loyalitas pelanggannya secara objektif akibat keterbatasan dalam menganalisis ribuan baris data secara manual. Akibatnya, strategi pemasaran tidak memiliki target yang spesifik karena sulit membedakan secara pasti mana pelanggan yang berkontribusi tinggi, pelanggan menengah, dan pelanggan pasif yang berisiko berpindah ke kompetitor.

Untuk memecahkan permasalahan tersebut, diperlukan penerapan disiplin ilmu Data *Mining* tingkat lanjut menggunakan teknologi *Machine Learning*. Algoritma yang terbukti andal dalam mempartisi data berskala besar tanpa label awal adalah *K-Means Clustering*, yang bekerja secara mandiri dengan mengelompokkan data konsumen ke dalam beberapa klaster berdasarkan jarak kemiripan terdekat *Euclidean Distance*. Penelitian terdahulu oleh Syahra, Fadlil, dan Yuliansyah (2025) menegaskan bahwa integrasi algoritma *K-Means* dengan model ekstraksi *Recency*, *Frequency*, *Monetary* (*RFM*) mampu menghasilkan pemetaan pelanggan yang jauh lebih akurat pada industri *ritail*. Chandra dan Fenriana (2024) juga membuktikan bahwa pendekatan *RFM* menjadi indikator absolut dari tingkat loyalitas, di mana aplikasi analitik terbukti mempercepat pembuatan keputusan pemasaran. Selain itu, kajian oleh Shirole, Salokhe, dan Jadhav (2021) memvalidasi bahwa sistem pengelompokan yang menggunakan segmentasi perilaku masa lalu sangat efektif untuk meningkatkan retensi konsumen secara pesat.

Oleh karena komputer tidak dapat langsung memproses baris *log* transaksi mentah, penelitian ini mengintegrasikan pemodelan analitik perilaku *RFM* untuk mengekstraksi nilai pelanggan secara kuantitatif berdasarkan dimensi waktu transaksi terakhir, intensitas kedatangan, dan total pengeluaran uang. Komputasi iteratif yang rumit dari algoritma tersebut diimplementasikan ke dalam sebuah rancangan sistem informasi web menggunakan Python agar dapat divisualisasikan menjadi dasbor analitik yang mudah dipahami. Tujuan penelitian ini berfokus pada penerapan tahapan Data Mining untuk mereduksi log massal, mengimplementasikan algoritma *K-Means Clustering* untuk identifikasi pola tersembunyi, menggunakan Elbow Methode untuk menentukan hasil klaster yang akan di peroleh, serta menghasilkan arsitektur sistem informasi visual untuk mendukung kebijakan manajerial promosi perusahaan.

TINJAUAN PUSTAKA

Penelitian mengenai segmentasi pelanggan telah banyak dikembangkan dengan memanfaatkan kerangka kerja berbasis penambangan data. Syahra, Fadlil, dan Yuliansyah (2025) melakukan penelitian mengenai implementasi metode *K-Means* Clustering yang digabungkan dengan pemodelan *RFM* guna mendukung sistem Customer Relationship Management (CRM) pada sektor retail. Hasil penelitian mereka menunjukkan bahwa segmentasi pelanggan berdasarkan rentang waktu, frekuensi kedatangan, dan total pembelanjaan terbukti dapat membedakan pelanggan bernilai tinggi dari kelompok lainnya secara signifikan. Algoritma berbasis jarak Euclidean ini memberikan objektivitas matematis yang mencegah bias interpretasi manusia dalam menentukan strategi pemasaran.

Studi relevan berikutnya dilakukan oleh Chandra dan Fenriana (2024) yang merancang aplikasi segmentasi pelanggan otonom menggunakan K-Means. Penelitian tersebut menyoroti urgensi ketersediaan sistem antarmuka bagi para pengambil keputusan di suatu perusahaan, di mana proses ekstraksi matriks *RFM* yang terkomputerisasi mampu memangkas waktu pengolahan data transaksi penjualan berskala masif dengan sangat efisien. Pendekatan

terpusat pada kemudahan penggunaan sistem ini menjamin manajer yang tidak memiliki latar belakang pemrograman tetap mampu mengeksekusi analisis klasterisasi dengan baik.

Selanjutnya, penelitian oleh Shirole, Salokhe, dan Jadhav (2021) mengkaji keandalan algoritma clustering yang berjalan tanpa label pengawasan atau Unsupervised Machine Learning dalam memecah pasar konsumen menjadi beberapa partisi homogen. Penemuan mereka memvalidasi bahwa kelompok konsumen yang diklasterisasi melalui model *RFM* menunjukkan tingkat respons yang jauh lebih tinggi terhadap kampanye promosi dibandingkan jika perusahaan menggunakan segmentasi demografis konvensional. Evaluasi hasil kelompok juga diperkuat dengan metode pengukuran jarak kuadrat untuk memastikan sebaran partisi benar-benar berada pada titik yang paling rapat.

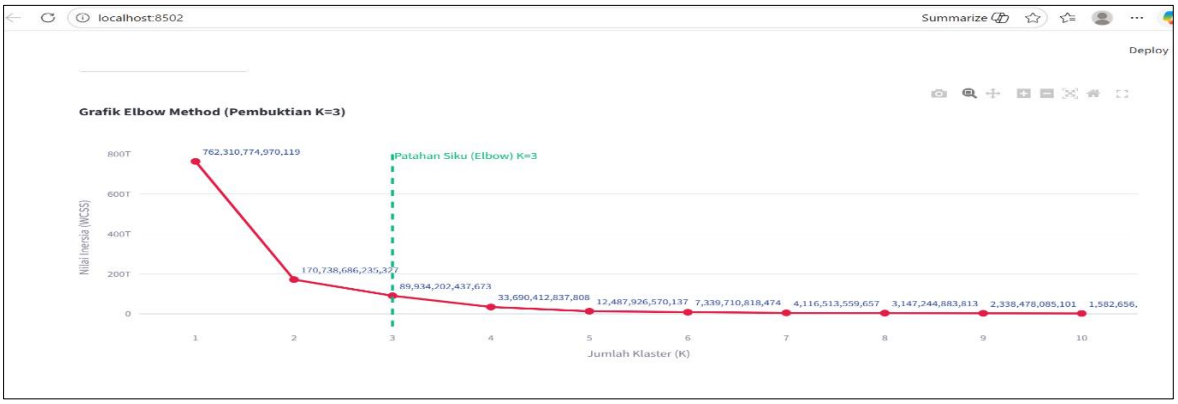
METODE PENELITIAN

Penelitian ini menerapkan tahapan analitik dari kerangka kerja *Knowledge Discovery in Databases* (KDD) yang merupakan proses metodologis dan sistematis untuk menyaring data mentah menjadi pengetahuan fungsional. Tahapan awal dimulai dengan proses seleksi data secara otomatis di mana sistem membaca seluruh judul kolom pada berkas database log transaksi yang diunggah, lalu menyaring dan mengisolasi variabel fitur operasional seperti tanggal transaksi, entitas pembeli, dan total uang yang dibelanjakan. Setelah itu, data memasuki fase pembersihan atau Data Cleaning untuk mensterilisasi anomali teks, menyeragamkan kapitalisasi entitas pelanggan, menghilangkan seluruh karakter alfabetis pada nominal uang, dan mengonversinya ke dalam tipe data bilangan desimal berpresisi tinggi.

Selanjutnya, data mentah ditransformasikan menggunakan metode agregasi model *RFM*. Data mentah diperoleh dari Rifqil Group Barokah yang merupakan sebuah Perusahaan Retail, periode data yang diperoleh dari Januari sampai Desember 2025. Nilai Recency dihitung dari selisih hari mutlak antara tanggal batas akhir observasi dengan tanggal transaksi terakhir konsumen, nilai Frequency dihitung dengan mengekstraksi seluruh kuantitas unit kunjungan, dan nilai Monetary didapatkan dari penjumlahan akumulasi total aktivitas finansial tiap konsumen. Matriks matang ini lalu dieksekusi menggunakan algoritma *K-Means* Clustering. Evaluasi untuk menentukan jumlah klaster target yang optimal dilakukan menggunakan Elbow Method, yang mengukur penurunan nilai Sum of Squared Errors (SSE) hingga membentuk patahan siku grafis. Komputasi penentuan jarak klaster antara objek data konsumen dengan titik sentroidnya diproses secara spasial dan iteratif menggunakan perhitungan Euclidean Distance menggunakan bahasa pemrograman Python.

HASIL DAN PEMBAHASAN

Tahapan pra-pemrosesan berhasil mengagregasi total 2.590 baris log transaksi mentah menjadi matriks *RFM* tunggal, menghasilkan 60 baris entitas pelanggan unik yang sepenuhnya siap untuk diklasterisasi. Setelah matriks *RFM* terbentuk, pengujian dilanjutkan dengan memvalidasi jumlah klaster yang optimal pada aplikasi web menggunakan Elbow Method. Kanvas grafik garis interaktif memproyeksikan penurunan inersia jarak, dan titik patahan sudut siku yang tajam terjadi tepat pada konfigurasi tiga klaster atau K=3. Konfigurasi ini merepresentasikan segmentasi tiga kelas, yakni Klaster Tinggi, Klaster Sedang, dan Klaster Rendah.



Gambar 1. Grafik Elbow *K = 3*

Sistem analitik berbasis web ini dibangun bukan sekadar sebagai purwarupa untuk satu kasus penggunaan saja, melainkan dirancang secara mutlak sebagai perangkat lunak operasional yang fungsional bagi berbagai pengguna umum yang ingin mengolah data transaksi mereka. Antarmuka pengguna secara khusus mengusung tema desain ocean blue serenity dengan konsep minimalis untuk memberikan pengalaman visual yang profesional serta memastikan kejelasan navigasi. Seluruh modul komputasi mulai dari pemilihan sentroid awal untuk titik M1, M2, dan M3 diproses dengan kalkulasi perulangan menggunakan geometri Euclidean Distance. Pergeseran rata-rata klaster terus dieksekusi oleh mesin algoritma secara otonom hingga menemukan stabilitas. Berdasarkan log pengujian sistem, kondisi konvergen mutlak tanpa adanya objek data yang bergeser keanggotaan berhasil dicapai pada siklus iterasi kedelapan, di mana nilai Sum of Squared Errors (SSE) terkunci stabil di angka 122.419.554.167.615.

Hasil klasifikasi akhir sukses memetakan 60 entitas tersebut ke dalam tiga kelompok perilaku homogen. Klaster pertama merupakan kelompok Prioritas Tinggi atau Pelanggan Loyal yang memiliki kontribusi finansial paling masif, frekuensi berbelanja yang repetitif, serta jeda waktu kedatangan terakhir yang sangat singkat. Klaster kedua merupakan kelompok Prioritas Sedang atau Pelanggan Reguler yang menunjukkan rutinitas dan pengeluaran menengah, sehingga ideal untuk didorong menjadi konsumen VIP melalui pemberian diskon. Sedangkan klaster ketiga merupakan kelompok Prioritas Rendah atau Pelanggan Pasif yang memiliki jeda hari transaksi terlama dengan nominal sangat kecil. Informasi analitik ini divisualisasikan oleh sistem dalam bentuk proyeksi grafis 3D Scatter Plot interaktif guna memudahkan manajemen merancang program pemulihan atau win-back campaign bagi pelanggan pasif.

Tabel Keputusan Segmentasi per Klaster

Prioritas Tinggi / Loyal (5 Pelanggan)

No.	Konsumen	Recency (R)	Frequency (F)	Monetary (M)	Status Pelanggan
1	C 013	189	553	11856500	Prioritas Tinggi (Loyal)
2	C 026	30	1153	10696000	Prioritas Tinggi (Loyal)
3	C 031	2	194	8467000	Prioritas Tinggi (Loyal)
4	C 032	471	178	11074000	Prioritas Tinggi (Loyal)
5	C 048	20	379	18809000	Prioritas Tinggi (Loyal)

Gambar 2. Hasil *Cluster* 1

Prioritas Sedang / Reguler (12 Pelanggan)

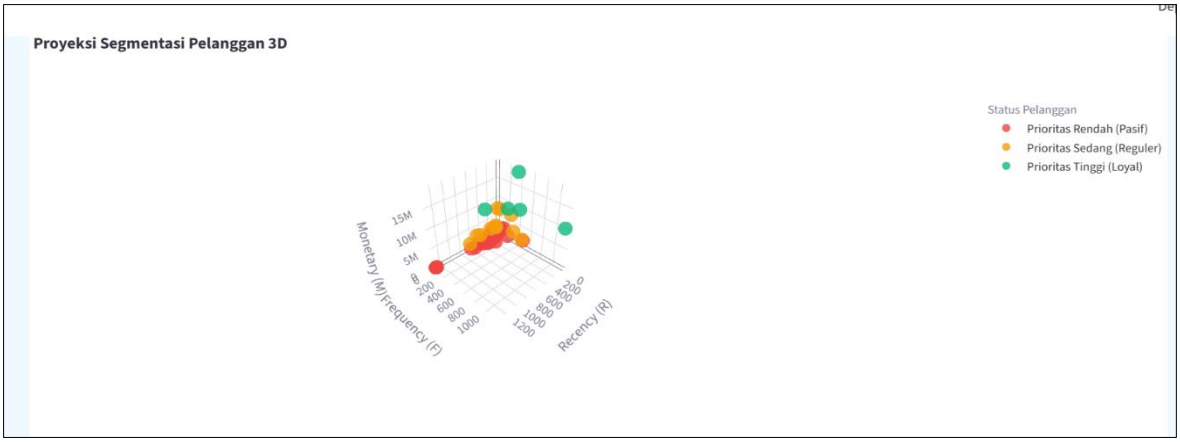
No.	Konsumen	Recency (R)	Frequency (F)	Monetary (M)	Status Pelanggan
1	C 007	2	259	7264000	Prioritas Sedang (Reguler)
2	C 008	17	313	3068000	Prioritas Sedang (Reguler)
3	C 015	113	57	2977000	Prioritas Sedang (Reguler)
4	C 017	735	124	3253000	Prioritas Sedang (Reguler)
5	C 020	250	72	3634000	Prioritas Sedang (Reguler)
6	C 021	106	42	2793000	Prioritas Sedang (Reguler)
7	C 022	506	107	3910000	Prioritas Sedang (Reguler)
8	C 023	20	62	7587000	Prioritas Sedang (Reguler)
9	C 027	526	54	3525000	Prioritas Sedang (Reguler)
10	C 041	63	4	2725000	Prioritas Sedang (Reguler)

Gambar 3. Hasil *Cluster* 2

Prioritas Rendah / Pasif (43 Pelanggan)

No.	Konsumen	Recency (R)	Frequency (F)	Monetary (M)	Status Pelanggan
1	C 001	1296	9	63000	Prioritas Rendah (Pasif)
2	C 002	1303	5	50000	Prioritas Rendah (Pasif)
3	C 003	199	12	155000	Prioritas Rendah (Pasif)
4	C 004	331	90	2035000	Prioritas Rendah (Pasif)
5	C 005	1284	7	70000	Prioritas Rendah (Pasif)
6	C 006	304	41	875000	Prioritas Rendah (Pasif)
7	C 009	333	68	859000	Prioritas Rendah (Pasif)
8	C 010	610	98	1037000	Prioritas Rendah (Pasif)
9	C 011	174	95	1165000	Prioritas Rendah (Pasif)
10	C 012	184	63	681500	Prioritas Rendah (Pasif)

Gambar 4. Hasil Cluster 3



Gambar 5. Proyeksi Segmentasi Pelanggan 3D

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem analitik penjualan berbasis web dengan mengintegrasikan model *RFM* (Recency, Frequency, Monetary) dan algoritma *K-Means* Clustering. Melalui tahapan Knowledge Discovery in Databases (KDD), sebanyak 2.590 baris data transaksi berhasil diproses dan diagregasi menjadi 60 entitas pelanggan unik yang digunakan sebagai data analisis. Penentuan jumlah klaster menggunakan Elbow Method menghasilkan tiga klaster optimal (K=3), sedangkan proses komputasi *K-Means* menggunakan Euclidean Distance mencapai kondisi konvergen pada iterasi kedelapan. Hasil pengelompokan menunjukkan tiga karakteristik pelanggan, yaitu Klaster Prioritas Tinggi atau Loyal dengan frekuensi dan nilai transaksi yang relatif besar, Klaster Prioritas Sedang atau Reguler dengan pola kunjungan dan pengeluaran menengah, serta Klaster Prioritas Rendah atau Pasif yang memiliki interval waktu transaksi terakhir lebih panjang dan nilai transaksi relatif kecil. Hasil segmentasi tersebut kemudian divisualisasikan melalui antarmuka web dan grafik 3D Scatter Plot untuk membantu manajemen memahami karakteristik pelanggan berdasarkan perilaku transaksinya. Dengan demikian, sistem yang dikembangkan dapat digunakan sebagai alat bantu analisis dalam mengidentifikasi kelompok pelanggan dan menyediakan informasi pendukung bagi penyusunan strategi pemasaran yang lebih terarah. Namun, penelitian ini masih terbatas pada analisis data historis berdasarkan variabel *RFM* dan belum melakukan pengukuran churn aktual maupun pengujian langsung terhadap efektivitas strategi pemasaran. Oleh karena itu, hasil segmentasi, khususnya pada Klaster Prioritas Rendah, hanya menunjukkan indikasi pelanggan dengan aktivitas transaksi yang relatif rendah, bukan bukti bahwa pelanggan tersebut telah berpindah atau akan berpindah ke kompetitor. Penelitian selanjutnya disarankan untuk mengintegrasikan sistem dengan basis data kasir secara real-time, menambahkan variabel demografis pelanggan, serta melakukan pengujian terhadap efektivitas strategi pemasaran dan tingkat retensi pelanggan.

REFERENSI

Chandra, N., & Fenriana, I. (2024). Perancangan Aplikasi Segmentasi Pelanggan Menggunakan *K-Means* Dengan Model *RFM*. Akselerator: Jurnal Sains Terapan dan Teknologi, 5(1), 46-59.

Cui, M. (2020). On the Elbow Method. Clausius Scientific Press, 1(1), 5-8.

Hidayat, R., & Saputra, A. (2023). Implementasi *K-Means* Clustering untuk Segmentasi Pelanggan E-Commerce. Jurnal Teknik Informatika, 4(2), 112-120.

Madyatmadja, E. D., Kusumawati, L., Jamil, S. P., & Kusumawardhana, W. (2021). Infotech: Journal of Technology Information. Raden Ario Damar, 7(1), 55-62.

Marten, T., & Hubbard, J. (2025). Seeing Structure: Measuring UML Diagram Density with Vision and Clustering Algorithms. Journal of Software Engineering, 12(3), 115-125.

Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. Journal of Information Technology and Computer Science (INTECOMS), 1(1), 72-77.

Rahman, M. A., & Ali, M. (2022). Customer segmentation using machine learning techniques. Journal of Big Data, 9(1), 1-15.

Setiawan, A., & Larasati, B. (2024). Analisis Perilaku Konsumen Menggunakan Model *RFM* dan Algoritma K-Means. Jurnal Sistem Informasi Bisnis, 14(1), 45-55.

Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer Segmentation Using *RFM* Model and *K-Means* Clustering. International Journal of Scientific Research in Science and Technology, 591-597.

Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from Data Mining and Machine Learning. Social Science Research, 110, 102817.

Sipayung, M., Maharani, H., & Paskhadira, B. A. (2017). Designing Customer Target Recommendation System Using *K-Means* Clustering Method. IJITEE (International Journal of Information Technology and Electrical Engineering), 1(1).

Syahra, Y. (2018). Penerapan Data Mining Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma *K-Means* Clustering. Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer), 17(2), 228.

Syahra, Y., Fadlil, A., & Yuliansyah, H. (2025). Customer Segmentation Using *RFM* and *K-Means* Clustering to Support CRM in Retail Industry. Sinkron, 9(3), 1120-1131.

Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Unsupervised Learning. American Journal of Orthodontics and Dentofacial Orthopedics, 163(6), 877-882.

Wijaya, K., & Santoso, H. (2022). Optimasi *K-Means* dengan Metode Elbow dalam Pengelompokan Data Penjualan. Jurnal Ilmu Komputer, 10(2), 78-85.