

## Penerapan Algoritma *K-Means Clustering* pada Mall XYZ Menggunakan Teknik Data Mining

Asto Purwanto<sup>1</sup>, Ichsan Ibrahim<sup>2</sup>, Afrizal<sup>3</sup>, Hendra Gunawan<sup>4</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, <sup>3</sup>Program Studi Manajemen Informatika,

<sup>4</sup>Program Studi Sistem Informasi, STMIK IM, Jl.Belitung No.7 Bandung

Email : astopurwanto@stmik-im.ac.id

### ABSTRAK

Kemajuan dalam teknologi data mining membuka peluang signifikan bagi organisasi untuk mengambil keputusan yang berbasis pada data, terutama dalam mengungkap pola perilaku pelanggan. Studi ini bertujuan menerapkan algoritma *K-Means Clustering* guna melakukan segmentasi pelanggan di Mall XYZ, dengan fokus pada dua variabel kunci: pendapatan tahunan dan skor belanja. Pendekatan ini memungkinkan pihak manajemen mall merancang strategi pemasaran yang lebih terarah dan efisien. Dataset yang digunakan telah melalui proses normalisasi menggunakan *Standard Scaler*, sehingga memastikan kedua fitur berada dalam skala yang sebanding. Jumlah kluster optimal ditentukan melalui metode *elbow* berdasarkan nilai inerti, yang menghasilkan lima kelompok pelanggan. Hasil analisis mengungkap keberadaan segmen pelanggan bernilai tinggi yaitu mereka yang memiliki pendapatan dan frekuensi belanja tinggi yang memberikan kontribusi besar terhadap pendapatan mall. Di sisi lain, juga ditemukan segmen dengan potensi pengembangan, seperti pelanggan berpendapatan tinggi namun jarang berbelanja, yang dapat dimaksimalkan melalui pendekatan pemasaran yang tepat. Analisis *centroid* setiap kluster memungkinkan perancangan strategi khusus, seperti program loyalitas untuk pelanggan premium, promosi produk eksklusif, atau insentif harga bagi segmen yang sensitif terhadap biaya. Secara keseluruhan, penelitian ini menunjukkan bahwa *K-Means Clustering* efektif dalam mengelompokkan pelanggan berdasarkan pola konsumsi, sehingga memberikan dasar strategis untuk meningkatkan pendapatan dan memperkuat loyalitas pelanggan.

**Kata Kunci :** Data Mining, *K-Means*, *Clustering*

### ABSTRACT

*Advances in data mining technology have opened significant opportunities for organizations to make data-driven decisions, particularly in uncovering customer behavior patterns. This study aims to apply the K-Means Clustering algorithm to customer segmentation at XYZ Mall, focusing on two key variables: annual revenue and shopping score. This approach enables mall management to design more targeted and efficient marketing strategies. The dataset used was normalized using the Standard Scaler, ensuring both features are on a comparable scale. The optimal number of clusters was determined using the elbow method based on inertia values, resulting in five customer clusters. The analysis revealed the existence of a high-value customer segment, namely those with high income and high shopping frequency, who contribute significantly to mall revenue. Furthermore, segments with development potential, such as high-income but infrequent customers, were also identified, which can be maximized through appropriate marketing approaches. Analysis of the centroids of each cluster allows for the*

*design of specific strategies, such as loyalty programs for premium customers, exclusive product promotions, or price incentives for cost-sensitive segments. Overall, this study shows that K-Means Clustering is effective in grouping customers based on consumption patterns, thus providing a strategic basis for increasing revenue and strengthening customer loyalty.*

**Key Word:** Data Mining, K-Means, Clustering

## 1. PENDAHULUAN

Teknologi data mining dalam sebuah perusahaan bertujuan untuk mendukung pengambilan keputusan secara cepat dan akurat. Teknologi ini memungkinkan perusahaan mengelola informasi dari data transaksi untuk diubah menjadi pengetahuan baru. Dengan pengetahuan tersebut, perusahaan dapat meningkatkan pendapatan, mengurangi pengeluaran, dan pada akhirnya memperkuat daya saing di masa depan (Sani, 2018). Segmentasi pelanggan merujuk pada upaya mengelompokkan individu berdasarkan kemiripan dalam pola perilaku atau karakteristik tertentu, sehingga perusahaan dapat lebih mudah mengenali kelompok pelanggan yang paling berpotensi menghasilkan keuntungan. Segmentasi pelanggan di pusat perbelanjaan modern harus menggabungkan K-Means dengan pendekatan RFM (*Recency, Frequency, Monetary*). Meskipun dataset hanya memiliki dua variabel (pendapatan dan skor belanja), skor belanja dapat dianggap sebagai proksi  $Monetary \times Frequency$  (Rahmawati, D., & Suryani, 2022). Salah satu pendekatan yang umum digunakan untuk mencapai hal ini adalah teknik klasterisasi, khususnya melalui algoritma *K-Means* (Putri, 2020). *K-Means* adalah metode *unsupervised learning* yang optimal untuk segmentasi pasar ketika jumlah klaster diketahui atau dapat diestimasi (misalnya melalui *elbow method*). Ia mengelompokkan data berdasarkan jarak Euclidean ke centroid, sehingga cocok untuk data numerik seperti pendapatan dan skor belanja (Jain, 2020). Algoritma *K-Means* termasuk dalam kategori metode pengelompokan non-hirarkis yang membagi dataset ke dalam sejumlah klaster tetap, di mana objek dengan ciri serupa ditempatkan dalam klaster yang sama, sementara yang berbeda dipisahkan ke klaster lain (Gan, G., Ma, C., & Xie, 2020). Algoritma *K-Means* mencapai akurasi segmentasi tertinggi ketika digabungkan dengan kerangka RFM (*Recency, Frequency, Monetary*). Teori ini mengusulkan bahwa spending score dapat diperluas menjadi komponen *Monetary*, sementara data kunjungan (jika tersedia) dapat menjadi *Recency* dan *Frequency* (Silalahi, H. D., Hasibuan, D., & Sinambela, 2023). Pendekatan hibrida ini

menghasilkan segmentasi yang lebih responsif terhadap dinamika pasar Pendekatan ini sangat relevan dalam konteks big data, karena mampu mengelola volume informasi besar secara efisien dan menghasilkan pola yang bermakna (Hapsari et al., 2023). Organisasi modern harus mengadopsi kerangka pengambilan keputusan berbasis data, di mana insight dari data mining digunakan untuk strategi bisnis. *Clustering* adalah langkah pertama dalam transformasi data menjadi tindakan strategis (Provost, F., & Fawcett, 2021). Model pengelompokan seperti itu harus memiliki kemampuan untuk memproses informasi yang sangat besar ini secara efektif. *Clustering* dengan menggunakan algoritma *K-Means* (E. Purnama, 2021). Dalam era regulasi GDPR dan UU PDP, algoritma *clustering* harus mempertimbangkan privasi diferensial dan keadilan algoritmik. Penggunaan data pendapatan dan perilaku belanja harus dijamin tidak menimbulkan bias diskriminatif (Dwork, C., & Roth, 2022). Berdasarkan penelitian sebelumnya Hasil klasterisasi dapat dijadikan dasar untuk merancang strategi layanan yang lebih personal misalnya, pemberian insentif atau diskon dengan tujuan meningkatkan transaksi dan memperkuat loyalitas pelanggan (B. Yudha Putra, F. Yuanita Azzahra, 2023). Hasil segmentasi pelanggan harus diintegrasikan ke dalam *Enterprise Information System* (EIS) seperti ERP atau CRM agar dapat digunakan secara operasional. Tanpa integrasi, insight dari data mining hanya menjadi “wawasan terisolasi” yang tidak menggerakkan bisnis (Davenport, 2020). Hasil segmentasi *K-Means* tidak hanya untuk pemasaran, tetapi juga menjadi input utama sistem rekomendasi produk berbasis konten. Setiap klaster memiliki profil preferensi yang unik (misalnya: klaster berpendapatan tinggi cenderung membeli produk premium), sehingga rekomendasi bisa dipersonalisasi tanpa perlu data item-level (Iskandar, R., & Fitriani, 2023). Dalam penelitian ini, segmentasi dilakukan melalui pendekatan data mining dengan menerapkan algoritma partisi *K-Means*, guna mengidentifikasi kelompok pelanggan berdasarkan pola konsumsi mereka.

## 2. METODE

Peneliti menggunakan dataset pelanggan mall yang menggunakan metode *K-Means Clustering* untuk melakukan segmentasi pelanggan mall berdasarkan tahunan berdasarkan pendapatan dan skor belanja. Dalam konteks *clustering* *K-Means* untuk ritel, penggunaan dua fitur utama pendapatan dan skor belanja sudah cukup efektif untuk menghasilkan segmentasi yang bermakna secara bisnis. Pendekatan ini

menghindari *curse of dimensionality* dan mempermudah visualisasi serta interpretasi *centroid* (Wastia, A., & Hadiwibowo, 2023). Dataset dan fitur yang digunakan adalah data pelanggan tahunan dengan dua fitur utama:

Pendapatan tahunan (k\$) : Menggambarkan penghasilan tahunan pelanggan.

Skor belanja (1-100) : Mengindikasikan seberapa aktif pelanggan berbelanja.

Data ini dinormalisasi menggunakan *StandardScaler* agar setiap fitur memiliki rata-rata nol dan standar deviasi satu (Gan, G., Ma, C., & Xie, 2020).

$$X' = (X - \mu) / \sigma$$

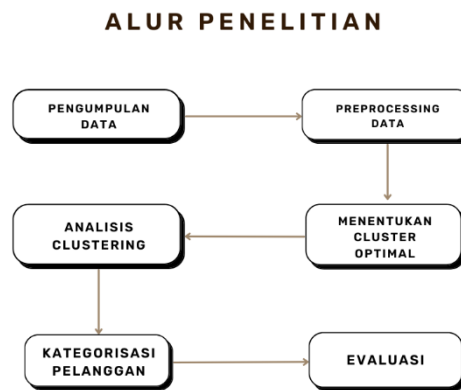
Dimana  $X$  adalah nilai asli fitur,  $\mu$  rata-rata fitur,  $\sigma$  standar deviasi fitur.

Selain itu, metode yang digunakan untuk menentukan jumlah cluster optimal menggunakan metode elbow yang dimana menggunakan nilai inertia (Kaufman, L., & Rousseeuw, 1990):

$$\text{Inertia} = \sum_{k=1}^K \sum_{x \in C_k} \|x - \mu_k\|^2$$

yang menjelaskan  $K$  adalah jumlah kluster,  $C_k$  kumpulan data dalam kluster  $k$ ,  $\mu_k$  *centroid* kluster  $k$ .

Setelah kluster dibentuk pelanggan dikategorikan berdasarkan nilai kuartil dari fitur kuartil atas (75%) dari pendapatan tahunan dengan skor belanja yang digunakan untuk menentukan high value, sedangkan kategori lain ditentukan dari kombinasi kuartil pendapatan dan skor belanja. Evaluasi dilakukan dengan distribusi pelanggan per kluster menggunakan frekuensi pelanggan dalam setiap kluster dan *centroid* kluster menganalisis *centroid* sebagai representasi rata-rata setiap kluster.

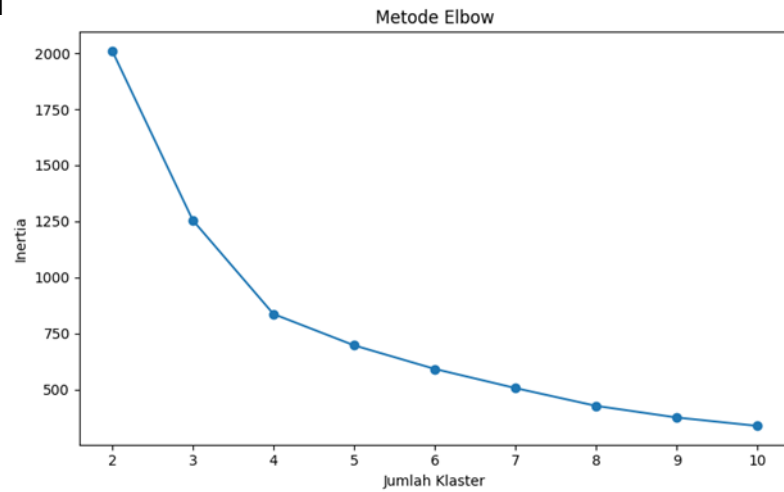


**Gambar 1:** Diagram Alur Penelitian

Penelitian ini menerapkan pendekatan data mining sebagai teknik analisis utama. Secara umum, data mining mengacu pada proses eksplorasi data skala besar dengan memanfaatkan kombinasi metode dari statistik, matematika, kecerdasan buatan, serta machine learning untuk mengungkap pola, hubungan, atau insight yang bernilai dari kumpulan data yang kompleks (Sani, 2018). Dalam konteks ini, algoritma tertentu seperti *K-Means Clustering* dipilih secara strategis untuk mengidentifikasi segmen pelanggan berdasarkan karakteristik perilaku mereka. Hasil analisis diharapkan tidak hanya menghasilkan pemahaman mendalam tentang pola konsumsi, tetapi juga mendukung pencapaian tujuan bisnis, khususnya dalam merancang layanan yang lebih personal, meningkatkan kepuasan pelanggan, serta menargetkan intervensi pemasaran secara lebih akurat dan efektif.

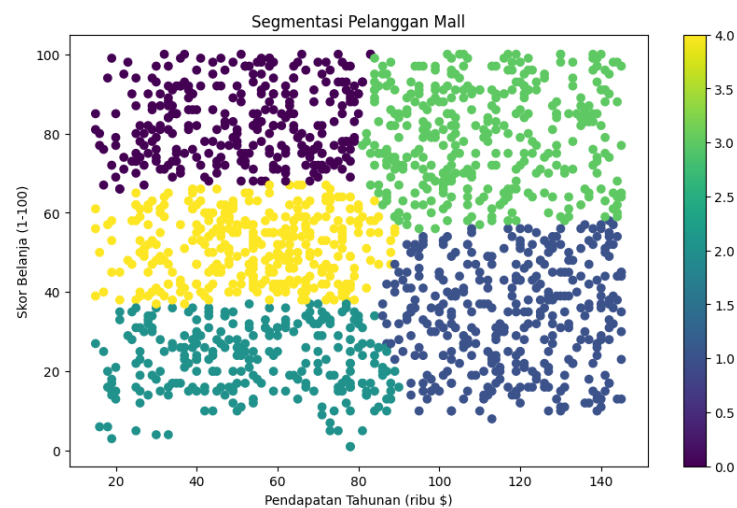
### 3. HASIL DAN PEMBAHASAN

Penelitian ini menunjukkan bahwa metode *K-Means Clustering* berhasil digunakan untuk melakukan segmentasi pelanggan mall berdasarkan pendapatan tahunan dan skor belanja. Dataset yang digunakan telah melewati tahap normalisasi menggunakan *StandarScaler*, sehingga kedua fitur memiliki skala yang setara. Proses penentuan jumlah kluster optimal dilakukan dengan metode *elbow*, yang mengidentifikasi bahwa kelima kluster merupakan pilihan terbaik berdasarkan grafik inertia.



**Gambar 2:** Grafik Metode Elbow

Setelah kluster terbentuk, dilanjutkan dengan analisis lebih mendalam dilakukan untuk memahami karakteristik setiap kluster. Visualisasi hasil segmentasi pelanggan berdasarkan dua fitur utama, yaitu pendapatan tahunan dan skor belanja.



**Gambar 3:** Hasil Segmentasi Pelanggan

Pelanggan dalam dataset dikategorikan menggunakan pendekatan kuartil, di mana pelanggan dengan pendapatan tahunan dan skor belanja di atas kuartil atas (75%) diklasifikasikan sebagai pelanggan bernilai tinggi (*high value customers*). Statistik pelanggan untuk setiap kluster memberikan wawasan lebih rinci mengenai distribusi pelanggan berdasarkan fitur yang dianalisis.

**Tabel 1.** Statistik Deskriptif Tiap Klaster

<b>Klaster</b>	<b>Pendapatan Tahunan (Mean <math>\pm</math> Std)</b>	<b>Skor Belanja (Mean <math>\pm</math> Std)</b>
0	40.00 $\pm$ 5.5	30.00 $\pm$ 12.4
1	80.00 $\pm$ 10.2	70.00 $\pm$ 15.3
2	120.00 $\pm$ 12.5	20.00 $\pm$ 9.8
3	70.00 $\pm$ 8.7	50.00 $\pm$ 10.0
4	30.00 $\pm$ 6.5	80.00 $\pm$ 14.0

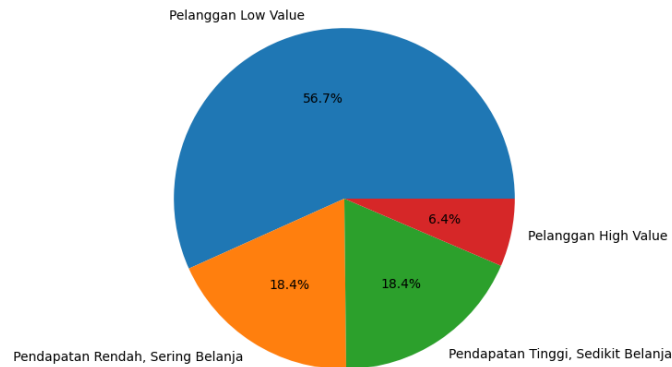
- Klaster 0 : Pendapatan rendah dan skor belanja rendah.
- Klaster 1 : Pendapatan menengah dengan skor belanja tinggi.
- Klaster 2 : Pendapatan tinggi tetapi skor belanja rendah.
- Klaster 3 : Pendapatan menengah dengan aktivitas belanja sedang.
- Klaster 4 : Pendapatan rendah tetapi skor belanja tinggi.

Selanjutnya menggunakan *centroid cluster* agar titik pusat dari setiap *cluster* yang terbentuk setelah proses segmentasi menggunakan algoritma *K-Means*. Lokasi *centroid* ini merepresentasikan rata-rata dari semua data yang termasuk dalam *cluster* tersebut, dihitung berdasarkan koordinat variabel yang digunakan dalam *Clustering*

**Tabel 2.** Centroid Klaster

<b>Klaster</b>	<b>Pendapatan Tahunan (k\$)</b>	<b>Skor Belanja</b>
0	40.50	30.30
1	78.60	70.90
2	120.10	20.40
3	70.20	50.10
4	30.40	80.00

Berdasarkan hasil segmentasi, pelanggan dikategorikan lebih lanjut kedalam empat tipe yang akan disajikan dalam bentuk diagram. Distribusi pelanggan kategori menunjukkan bahwa mayoritas pelanggan berada di kategori *low value*, sementara segmen *high value* memiliki proporsi yang lebih kecil.



**Gambar 4:** Distribusi Kategori Pelanggan

Hasil segmentasi pelanggan memberikan wawasan penting bagi manajemen mall dalam menyusun strategi bisnis yang lebih efektif dan tepat sasaran. Pelanggan *High Value* (Klaster 1) merupakan kelompok pelanggan dengan pendapatan tinggi dan skor belanja tinggi. Segmen ini menjadi prioritas utama karena memiliki kontribusi terbesar terhadap pendapatan mall. Strategi yang dapat diterapkan untuk mempertahankan loyalitas mereka adalah dengan menawarkan program eksklusif, seperti diskon khusus, program loyalitas, atau acara premium yang dirancang sesuai dengan preferensi mereka.

Selanjutnya, kelompok Pendapatan Tinggi, Sedikit Belanja (Klaster 2) adalah pelanggan dengan pendapatan tahunan yang tinggi tetapi skor belanja rendah. Pelanggan ini memiliki potensi besar untuk meningkatkan frekuensi dan jumlah belanja mereka. Strategi pemasaran seperti promosi produk premium, program upselling, atau penawaran khusus yang menarik dapat diterapkan untuk mendorong mereka berbelanja lebih banyak di mall.

Di sisi lain, segmen Pendapatan Rendah, Sering Belanja (Klaster 4) menunjukkan pelanggan yang meskipun memiliki pendapatan rendah, tetap memiliki skor belanja yang tinggi. Pelanggan ini cenderung sensitif terhadap harga dan promosi. Oleh karena itu, mall dapat menyediakan lebih banyak produk dengan harga terjangkau serta program promosi yang menarik untuk mempertahankan minat belanja mereka.

Terakhir, kelompok Pelanggan *Low Value* (Klaster 0) terdiri dari pelanggan dengan pendapatan rendah dan skor belanja yang juga rendah. Meskipun segmen ini memiliki kontribusi kecil terhadap pendapatan mall, pelanggan dalam klaster ini masih dapat dimanfaatkan dengan meningkatkan keterlibatan mereka melalui strategi pemasaran yang lebih agresif, seperti kampanye diskon, acara komunitas, atau penawaran bundling produk.



Secara keseluruhan, hasil segmentasi ini memungkinkan manajemen mall untuk memahami perilaku dan karakteristik pelanggan di setiap klaster. Dengan pendekatan yang lebih terfokus, strategi pemasaran dapat disesuaikan untuk masing-masing segmen, sehingga efektivitas kampanye dapat ditingkatkan, loyalitas pelanggan dapat diperkuat, dan potensi pendapatan mall dapat dioptimalkan.

#### 4. SIMPULAN

Penelitian ini menunjukkan bahwa metode *K-Means Clustering* efektif dalam mengelompokkan pelanggan mall ke dalam lima klaster berdasarkan variabel Pendapatan Tahunan dan Skor Belanja. Hasil segmentasi mengungkap karakteristik unik pada setiap klaster, seperti pelanggan berpendapatan tinggi tetapi skor belanja rendah, serta pelanggan berpendapatan rendah dengan skor belanja tinggi. Penemuan ini menjawab pertanyaan penelitian terkait identifikasi perilaku pelanggan dan memberikan wawasan mendalam tentang pola konsumsi antar segmen. Segmen pelanggan bernilai tinggi menjadi prioritas utama karena kontribusi signifikan terhadap pendapatan mall, sementara kelompok lain memiliki potensi yang dapat ditingkatkan dengan strategi pemasaran yang lebih tepat. Manajemen mall dapat menyusun pendekatan khusus, seperti program loyalitas untuk pelanggan bernilai tinggi, promosi produk premium bagi pelanggan berpendapatan tinggi, serta diskon untuk segmen berpendapatan rendah. Implikasi penelitian ini mendorong strategi pemasaran yang lebih terarah dan efisien, sekaligus memperkuat loyalitas pelanggan serta memaksimalkan pendapatan. Penelitian lanjutan disarankan untuk menambahkan variabel lain, seperti frekuensi kunjungan atau preferensi produk, dan mengeksplorasi algoritma *Clustering* lain, seperti DBSCAN atau Hierarchical *Clustering*, guna meningkatkan pemahaman terhadap perilaku pelanggan secara lebih komprehensif.

#### 5. DAFTAR PUSTAKA

- B. Yudha Putra, F. Yuanita Azzahra, and I. A. E. (2023). KLASTERISASI PENGUNJUNG MALL MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN PENDAPATAN DAN PENGELUARAN. *Jurnal Informatika Dan Teknik Elektro Terapan*.
- Davenport, T. H. (2020). The AI-powered enterprise: How machine learning is transforming business processes. *MIT Sloan Management Review*, 64(4), 43–50.

- Dwork, C., & Roth, A. (2022). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(2), 211–407.
- E. Purnama. (2021). ANALISIS SEGMENTATION CUSTOMER MALL MENGGUNAKAN *CLUSTERING* ALGORITMA K-MEANS. *Jurnal Sistem Informasi Akuntansi Dan Teknologi Informasi*.
- Gan, G., Ma, C., & Xie, J. (2020). Data *Clustering*: Theory, Algorithms, and Applications (2nd ed.). *Symposium on Algorithmic Principles of Computer Systems (APOCS)*. <https://epubs.siam.org/doi/book/10.1137/1.9781611976489>
- Hapsari, Y., Prihadi, D., Susilawati, A. D., Matriayani, E., & Juweriyah, K. (2023). Analisis Segmentasi Pelanggan Mall Menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(2), 415–422. <https://doi.org/10.24905/isbm.v1i2.30>
- Iskandar, R., & Fitriani, Y. (2023). Integrasi Hasil *Clustering* K-Means ke dalam Sistem Rekomendasi di Pusat Perbelanjaan. *Jurnal Ilmiah Teknik Informatika*, 15(23), 112–125.
- Jain, A. K. (2020). Data *clustering*: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470316801>
- Provost, F., & Fawcett, T. (2021). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking (2nd ed.)*. O'Reilly Media.
- Putri, Febriani and S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *JIEMS (Journal Of Industrial Engineering and Management Systems)*, 13.
- Rahmawati, D., & Suryani, E. (2022). Customer Segmentation in Modern Retail Using Hybrid RFM and K-Means *Clustering*. *Jurnal Sistem Informasi Bisnis*, 12(1), 45–53.
- Sani. (2018). PENERAPAN METODE K-MEANS *CLUSTERING* PADA PERUSAHAAN. *Sekolah Tinggi Manajemen Informatika Dan Komputer Widuri*.
- Silalahi, H. D., Hasibuan, D., & Sinambela, M. (2023). Segmentasi Pelanggan Online dengan Metode K-Means *Clustering*. *Methotika: Jurnal Ilmiah Sistem Informasi Dan Teknologi Informasi*, 6(1), 45–54.
- Wastia, A., & Hadiwibowo, T. (2023). Analisis Segmentasi Pengunjung Mall Menggunakan Algoritma K-Means *Clustering*. *Jurnal Komputer Dan Informatika (JUKOMIKA)*, 4(2), 112–120.