

Segmentasi Pelanggan Menggunakan K-Means Clustering Pada Data Transaksi Online Retail

Robby Satria Darma¹, Lukman Sunardi^{1,*}, Asep Toyib²

¹Fakultas Ilmu Teknik, Program Studi Informatika, Universitas Bina Insan, Lubuk Linggau, Indonesia

²Fakultas Ilmu Teknik, Program Studi Sistem Informasi, Universitas Bina Insan, Lubuk Linggau, Indonesia

Email: ¹robisatria523@gmail.com, ^{2,*}lukmanmmci@gmail.com, ³asepfighter@gmail.com

Email Penulis Korespondensi: lukmanmmci@gmail.com

Submitted: 07/08/2025; Accepted: 30/11/2025; Published: 30/11/2025

Abstrak—Industri telekomunikasi menghadapi tantangan dalam memahami karakteristik pelanggan akibat besarnya volume data dan keberagaman perilaku penggunaan layanan. Segmentasi pelanggan menjadi pendekatan strategis untuk mendukung penyusunan strategi pemasaran yang lebih tepat sasaran dan efektif. Penelitian ini bertujuan menerapkan algoritma *K-Means clustering* untuk melakukan segmentasi pelanggan berbasis data pada perusahaan telekomunikasi. Dataset pelanggan diproses melalui tahapan prapemrosesan yang meliputi penanganan nilai hilang, pengkodean variabel kategorikal menggunakan *One-Hot Encoding*, serta penskalaan fitur menggunakan *StandardScaler*. Penentuan jumlah cluster optimal dilakukan menggunakan metode *Elbow*. Hasil penelitian menunjukkan terbentuknya tiga segmen pelanggan dengan karakteristik berbeda berdasarkan atribut *tenure*, *monthly charges*, *total charges*, dan tingkat *churn*. Visualisasi menggunakan *Principal Component Analysis (PCA)* memperlihatkan pemisahan antar cluster secara jelas. Temuan menarik menunjukkan bahwa segmen dengan nilai pelanggan tertinggi memiliki tingkat *churn* yang paling tinggi, sehingga membutuhkan strategi retensi yang lebih spesifik. Kontribusi penelitian ini adalah memberikan pendekatan segmentasi pelanggan berbasis *unsupervised learning* yang dapat dimanfaatkan perusahaan untuk merancang strategi pemasaran yang lebih efektif, meningkatkan retensi pelanggan, serta mendukung pengambilan keputusan berbasis data.

Kata Kunci: Segmentasi pelanggan; K-Means; Clustering; Strategi pemasaran; Telekomunikasi

Abstract—The telecommunications industry faces challenges in understanding customer characteristics due to large data volumes and diverse service usage behaviors. Customer segmentation becomes a strategic approach to support more targeted and effective marketing strategies. This study aims to apply the *K-Means clustering* algorithm to perform data-driven customer segmentation in a telecommunications company. The customer dataset undergoes preprocessing stages including missing value handling, categorical variable encoding using *One-Hot Encoding*, and feature scaling with *StandardScaler*. The optimal number of clusters is determined using the *Elbow Method*. The results show the formation of three customer segments with distinct characteristics based on *tenure*, *monthly charges*, *total charges*, and *churn rate*. Visualization using *Principal Component Analysis (PCA)* clearly illustrates the separation among clusters. An interesting finding reveals that the segment with the highest customer value also has the highest *churn* rate, indicating the need for more specific retention strategies. The contribution of this study lies in providing an *unsupervised learning*-based customer segmentation approach that can assist companies in designing more effective marketing strategies, improving customer retention, and supporting data-driven decision making.

Keywords: Customer segmentation; K-Means; Clustering; Marketing strategy; Telecommunications

1. PENDAHULUAN

Perkembangan teknologi digital mendorong industri telekomunikasi mengalami pertumbuhan yang sangat pesat, ditandai dengan meningkatnya jumlah pelanggan serta keberagaman pola penggunaan layanan. Kondisi ini menuntut perusahaan telekomunikasi untuk mampu memahami karakteristik pelanggan secara lebih mendalam agar strategi pemasaran yang diterapkan menjadi lebih efektif dan tepat sasaran. Strategi pemasaran konvensional yang tidak berbasis data sering kali gagal dalam mengakomodasi kebutuhan spesifik pelanggan, sehingga diperlukan pendekatan analisis data pelanggan secara sistematis [1], [2]. Segmentasi pelanggan merupakan salah satu pendekatan yang dapat digunakan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik tertentu. Dengan segmentasi yang baik, perusahaan dapat mengidentifikasi pelanggan bernilai tinggi, pelanggan berisiko *churn*, serta merancang strategi promosi yang sesuai dengan karakteristik masing-masing segmen [3], [4], [5]. Namun, proses segmentasi secara manual menjadi sangat sulit dilakukan ketika perusahaan dihadapkan pada data pelanggan dalam jumlah besar dan atribut yang kompleks.

Pemanfaatan teknik *data mining* dan *machine learning* khususnya pendekatan *unsupervised learning* menjadi solusi yang relevan dalam melakukan segmentasi pelanggan berbasis data [6], [7]. Salah satu algoritma clustering yang paling banyak digunakan adalah K-Means karena kemampuannya dalam mengelompokkan data tanpa label serta efisiensi komputasinya [8], [9]. Beberapa penelitian terdahulu telah berhasil menerapkan algoritma K-Means dalam segmentasi pelanggan di berbagai bidang industri [10], [11], [12]. Penelitian pada industri e-sports menunjukkan bahwa kombinasi K-Means dan analisis RFM mampu meningkatkan strategi retensi pelanggan.

Permasalahan utama yang dihadapi perusahaan telekomunikasi adalah kesulitan dalam mengidentifikasi pola dan karakteristik pelanggan akibat besarnya volume data dan keberagaman atribut pelanggan. Kondisi ini menyebabkan strategi pemasaran yang diterapkan kurang efektif karena tidak didasarkan pada pemahaman mendalam mengenai segmen pelanggan. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu

mengelompokkan pelanggan secara otomatis berdasarkan karakteristik yang dimiliki. Penelitian ini menawarkan pendekatan segmentasi pelanggan menggunakan algoritma K-Means yang dipadukan dengan tahapan prapemrosesan data, penentuan jumlah cluster optimal menggunakan metode *Elbow*, serta visualisasi hasil segmentasi menggunakan *Principal Component Analysis (PCA)*.

Studi lain pada klinik kecantikan menunjukkan bahwa segmentasi berbasis K-Means dapat mengidentifikasi pelanggan yang paling *profitable*. Penelitian pada penjualan online juga membuktikan bahwa K-Means efektif dalam memahami pola perilaku pelanggan untuk meningkatkan efisiensi promosi. Selain itu, segmentasi pelanggan pada industri produk konsumsi menggunakan K-Means mampu mengelompokkan pelanggan berdasarkan preferensi pembelian. Penelitian lain menunjukkan bahwa K-Means memberikan hasil *clustering* yang stabil serta mudah diinterpretasikan dalam konteks pemasaran berbasis data [13], [14], [15].

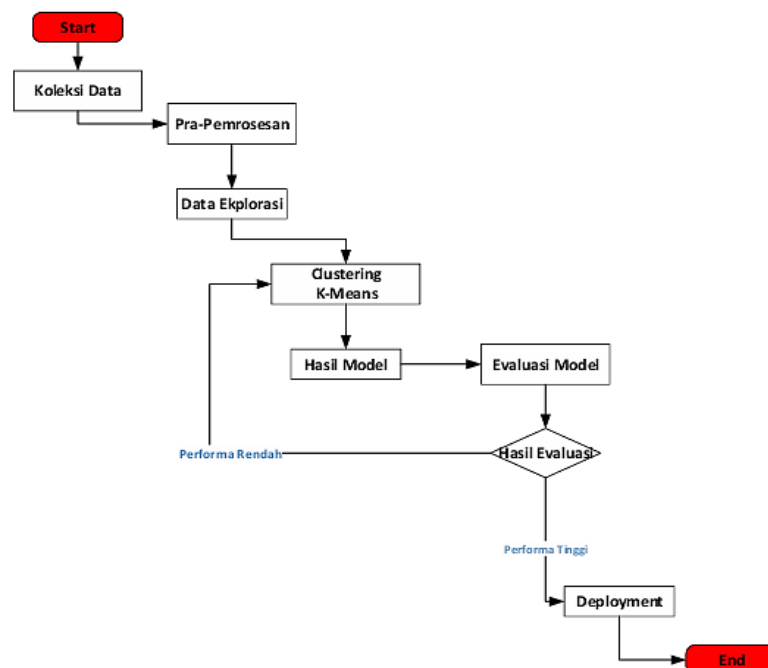
Meskipun berbagai penelitian tersebut menunjukkan keberhasilan penggunaan K-Means, sebagian besar penelitian masih berfokus pada atribut RFM (*Recency, Frequency, Monetary*) atau atribut demografis sebagai dasar pengelompokan. Penelitian-penelitian tersebut belum banyak mengeksplorasi pemanfaatan atribut perilaku pelanggan seperti *tenure, monthly charges, total charges*, serta keterkaitannya dengan tingkat *churn* dalam industri telekomunikasi. Selain itu, sebagian penelitian hanya berhenti pada proses pengelompokan tanpa mengaitkan hasil segmentasi dengan rekomendasi strategi pemasaran yang aplikatif. Berdasarkan kajian tersebut, terdapat *research gap* berupa belum optimalnya pemanfaatan algoritma K-Means dalam segmentasi pelanggan telekomunikasi yang mengintegrasikan atribut perilaku pelanggan dengan analisis tingkat *churn* sebagai dasar penyusunan strategi pemasaran. Selain itu, visualisasi hasil segmentasi menggunakan teknik reduksi dimensi seperti *Principal Component Analysis (PCA)* juga masih jarang digunakan untuk memperjelas pemisahan antar segmen secara visual [16], [13].

Tujuan penelitian ini adalah menerapkan K-Means dalam segmentasi pelanggan telekomunikasi, menganalisis karakteristik setiap segmen yang terbentuk, serta mengaitkannya dengan tingkat *churn* pelanggan sebagai dasar penyusunan strategi pemasaran yang lebih tepat sasaran. Dengan demikian, penelitian ini tidak hanya berkontribusi dalam penerapan *unsupervised learning* pada segmentasi pelanggan, tetapi juga memberikan implikasi praktis bagi perusahaan telekomunikasi dalam meningkatkan efektivitas strategi pemasaran berbasis analisis data pelanggan. Kontribusi penelitian ini meliputi: Mengembangkan segmentasi pelanggan berbasis atribut perilaku layanan, Mengintegrasikan analisis *churn* dalam proses clustering, Menyediakan interpretasi hasil segmentasi untuk strategi pemasaran berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dirancang untuk melakukan segmentasi pelanggan telekomunikasi menggunakan pendekatan *unsupervised learning* melalui algoritma K-Means. Tahapan penelitian disusun secara sistematis agar proses pengolahan data hingga analisis hasil segmentasi dapat berjalan terstruktur dan menghasilkan segmentasi yang representatif. Adapun tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Segmentasi Pelanggan Menggunakan K-Means

Berdasarkan Gambar 1 Tahapan penelitian meliputi pengumpulan dataset pelanggan, proses preprocessing data, penentuan jumlah cluster menggunakan metode Elbow, penerapan algoritma K-Means, visualisasi menggunakan PCA, serta analisis karakteristik cluster dan interpretasi strategi pemasaran.

2.2 Preprocessing Data

Tahapan preprocessing seperti normalisasi dan encoding sangat penting dalam meningkatkan kualitas data sebelum dilakukan proses clustering[17]. Data pelanggan yang diperoleh sering kali mengandung nilai kosong (*missing value*) dan variabel kategorikal yang tidak dapat langsung diproses oleh algoritma K-Means.

Penanganan *missing value* dilakukan dengan menghapus atau mengisi nilai kosong berdasarkan distribusi data. Variabel kategorikal seperti jenis layanan dikonversi menggunakan teknik *One-Hot Encoding*[18]. Selanjutnya, dilakukan normalisasi data menggunakan *StandardScaler* agar setiap atribut memiliki skala yang sebanding dan tidak mendominasi proses perhitungan jarak pada K-Means.

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

2.3 Penentuan Jumlah Cluster dengan Elbow Method

Penentuan jumlah cluster optimal dilakukan menggunakan metode *Elbow*[9]. Metode ini bekerja dengan menghitung nilai WCSS pada beberapa percobaan jumlah cluster. Titik siku (*elbow point*) pada grafik menunjukkan jumlah cluster yang paling optimal, di mana penurunan nilai WCSS mulai melambat[18].

Secara matematis, WCSS dirumuskan sebagai berikut:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - c_k)^2 \quad (2)$$

k adalah jumlah cluster, x_i adalah data dalam cluster, μ_k adalah centroid cluster ke- i .

2.4 Penerapan Algoritma K-Means

Algoritma K-Means merupakan metode clustering berbasis partisi yang bekerja dengan mengelompokkan data berdasarkan kedekatan terhadap centroid[8], [19]. Proses clustering dilakukan secara iteratif dengan memperbarui posisi centroid hingga mencapai kondisi konvergen. Pendekatan ini bertujuan untuk meminimalkan variasi dalam satu cluster dan memaksimalkan perbedaan antar cluster. Adapun rumus yang digunakan dapat dilihat sebagai berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (3)$$

Langkah selanjutnya mengulangi proses hingga konvergen, rumus *Euclidean Distance* yang digunakan adalah:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

2.5 Visualisasi Hasil Clustering Menggunakan PCA

Untuk memudahkan interpretasi hasil clustering, digunakan teknik reduksi dimensi *Principal Component Analysis (PCA)*[20]. PCA mengubah data berdimensi tinggi menjadi dua dimensi utama tanpa menghilangkan informasi penting, sehingga pemisahan cluster dapat divisualisasikan dengan jelas.

$$Z = XW \quad (5)$$

2.6 Analisis Karakteristik Setiap Cluster

Setelah cluster terbentuk, dilakukan analisis nilai rata-rata setiap atribut dalam cluster untuk mengetahui karakteristik masing-masing segmen pelanggan. Atribut yang dianalisis meliputi: *Tenure*, *Monthly Charges*, *Total Charges*, *Churn Rate*. Hasil analisis ini digunakan sebagai dasar dalam menyusun rekomendasi strategi pemasaran yang sesuai untuk setiap segmen.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Prapemrosesan Data

Tahap awal implementasi adalah memuat dataset pelanggan dan melakukan pembersihan data. Proses ini mencakup penanganan *missing value*, transformasi variabel kategorikal menggunakan *One-Hot Encoding*, serta normalisasi fitur dengan *StandardScaler*. Tujuan normalisasi adalah memastikan setiap fitur berada pada skala yang sebanding sehingga tidak mendominasi perhitungan jarak pada K-Means. Gambar ilustrasi tahapan prapemrosesan dapat dilihat pada Gambar 2 berikut.

```
# Separate features for clustering (excluding customerID and Churn)
features = df_encoded.drop(columns=['customerID', 'churn'])

# Scale the features
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
features_scaled = scaler.fit_transform(features)

# Display the first few rows of the scaled features
print("Scaled Features:")
display(pd.DataFrame(features_scaled, columns=features.columns, head=5))
```

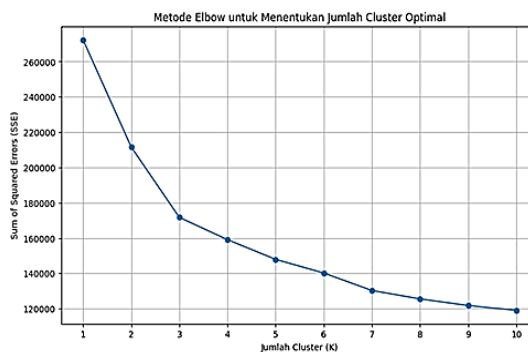
	tenure	monthlyCharges	totalCharges	gender_male	seniorCitizen_1	seniorCitizen_false	seniorCitizen_true	Partner_No	Partner_true	Partner_yes	...	streamingService_free	streamingService
0	-1.287404	-1.172126	-1.001342	-1.014064	-2.70560	0.996613	-0.322466	-0.511871	1.565001	-0.496093	...	-0.761178	-0.421
1	0.056031	-0.270752	-0.181297	0.960237	-2.70560	0.996613	-0.322466	-0.511871	-0.628957	-0.496093	...	-0.761178	-0.421
2	-1.248633	-0.373862	-0.966814	0.960237	-2.70560	0.996613	-0.322466	-0.511871	-0.628957	-0.496093	...	-0.761178	-0.421
3	0.506510	-0.758020	-0.202794	0.960237	-2.70560	0.996613	-0.322466	-0.511871	-0.628957	-0.496093	...	-0.761178	-0.421
4	-1.248633	0.190380	-0.947632	-1.014064	-2.70560	0.996613	-0.322466	-0.511871	-0.628957	-0.496093	...	-0.761178	-0.421

Gambar 2. Ilustrasi Tahapan Prapemrosesan (Encoding Dan Scaling) Yang Diterapkan Pada Dataset

Sebagaimana ditunjukkan pada Gambar 2, proses preprocessing menghasilkan data yang telah dinormalisasi dan siap digunakan dalam proses clustering.

3.2 Penentuan Jumlah Cluster dengan Elbow Method

Penentuan jumlah cluster optimal dilakukan menggunakan *Elbow Method* dengan menghitung nilai *Within Cluster Sum of Squares (WCSS)* pada beberapa kandidat jumlah cluster. Grafik menunjukkan titik siku pada k=3, yang mengindikasikan jumlah cluster paling representatif. Penentuan Cluster Optimal dapat dilihat pada Gambar 3.



Gambar 3. Penentuan Cluster Optimal

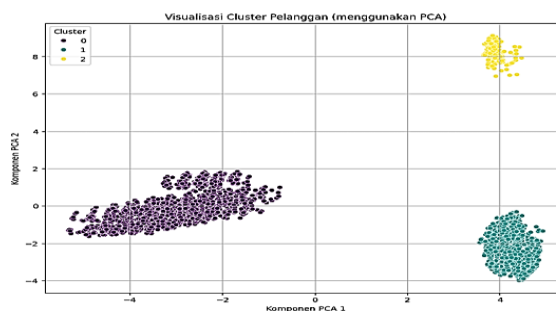
Temuan ini konsisten dengan visual yang Anda tampilkan pada proses penentuan jumlah cluster di skripsi (penentuan *cluster optimal*)

3.3 Penerapan K-Means dan Pembentukan Cluster

Dengan k=3, algoritma K-Means dijalankan hingga centroid konvergen. Setiap pelanggan kemudian memiliki label cluster yang merepresentasikan kemiripan karakteristik perilakunya. Hasil *clustering* menunjukkan distribusi pelanggan yang relatif seimbang pada ketiga cluster. Ini menandakan bahwa atribut yang digunakan (*tenure, monthly charges, total charges, churn*) cukup representatif dalam membedakan karakteristik pelanggan. Hasil ini sejalan dengan penelitian sebelumnya [9], [11].

3.4 Visualisasi Pemisahan Cluster Menggunakan PCA

Untuk memudahkan interpretasi, hasil *clustering* divisualisasikan dengan *Principal Component Analysis (PCA)* ke dalam dua dimensi. Gambar visualisasi dapat dilihat pada Gambar 4.



Gambar 4. Visualisasi Pca Yang Memperlihatkan Pemisahan Tiga Cluster

Visualisasi hasil clustering pada Gambar 4 menunjukkan pemisahan yang cukup jelas antar cluster tanpa adanya tumpang tindih signifikan.

3.5 Analisis Karakteristik Tiap Cluster

Hasil analisis karakteristik setiap cluster ditunjukkan pada Tabel 1, yang memperlihatkan perbedaan nilai atribut antar segmen pelanggan.

Tabel 1. Analisis Tiap Cluster

Atribut	Cluster 1	Cluster 2	Cluster 3
Tenure	Rendah	Tinggi	Sedang
Monthly Charges	Tinggi	Sedang	Rendah
Total Charges	Rendah	Tinggi	Sedang
Churn Rate	Tinggi	Rendah	Sedang

Berdasarkan Tabel 1, Cluster 1 merupakan pelanggan baru dengan tingkat churn tinggi yang ditandai dengan nilai tenure rendah dan biaya bulanan tinggi, sehingga cenderung sensitif terhadap harga. Cluster 2 menunjukkan pelanggan loyal dengan nilai total pembayaran tinggi dan tingkat churn rendah, sehingga merupakan segmen yang paling menguntungkan. Sementara itu, Cluster 3 merupakan pelanggan dengan karakteristik menengah yang relatif stabil namun masih memiliki potensi untuk ditingkatkan.

3.6 Implikasi Terhadap Strategi Pemasaran

Hasil segmentasi dapat diimplementasikan dalam strategi pemasaran yang berbeda untuk setiap cluster. Cluster 1 memerlukan strategi retensi awal seperti pemberian diskon dan edukasi layanan. Cluster 2 dapat difokuskan pada program loyalitas dan upselling. Sedangkan Cluster 3 dapat ditargetkan dengan strategi promosi untuk meningkatkan penggunaan layanan. Pendekatan ini juga didukung oleh penelitian terkait retensi pelanggan berbasis segmentasi [17].

3.7 Perbandingan dengan Penelitian Terdahulu

Dibandingkan dengan penelitian sebelumnya yang umumnya menggunakan atribut RFM atau demografi, penelitian ini menggunakan atribut perilaku layanan pelanggan serta mengintegrasikan analisis churn dan visualisasi PCA. Pendekatan ini memberikan hasil yang lebih aplikatif dalam mendukung pengambilan keputusan pemasaran berbasis data [9], [11].

4. KESIMPULAN

Penerapan algoritma *K-Means clustering* pada data pelanggan telekomunikasi dalam penelitian ini berhasil mengelompokkan pelanggan ke dalam tiga segmen yang memiliki karakteristik perilaku berbeda berdasarkan atribut *tenure*, *monthly charges*, *total charges*, dan tingkat *churn*. Melalui tahapan *preprocessing* yang mencakup penanganan *missing value*, *One-Hot Encoding*, dan normalisasi menggunakan *StandardScaler*, data dapat dipersiapkan dengan baik sehingga proses *clustering* menghasilkan pemisahan segmen yang representatif. Penentuan jumlah cluster menggunakan metode *Elbow* menunjukkan bahwa tiga cluster merupakan jumlah yang paling optimal untuk merepresentasikan variasi karakteristik pelanggan tanpa menyebabkan segmentasi yang berlebihan. Visualisasi menggunakan *Principal Component Analysis (PCA)* turut memperjelas pemisahan antar segmen secara visual, sehingga memudahkan interpretasi hasil *clustering*. Hasil analisis menunjukkan temuan penting bahwa segmen pelanggan dengan nilai penggunaan layanan tertinggi justru memiliki tingkat *churn* paling tinggi. Temuan ini memberikan wawasan strategis bahwa pelanggan bernilai tinggi tidak selalu identik dengan loyalitas tinggi, sehingga memerlukan pendekatan retensi yang lebih spesifik. Sementara itu, segmen pelanggan dengan biaya layanan rendah menunjukkan tingkat *churn* yang paling kecil dan cenderung stabil, sedangkan segmen menengah memiliki potensi untuk ditingkatkan melalui strategi pemasaran yang tepat. Kontribusi utama penelitian ini terletak pada pemanfaatan pendekatan *unsupervised learning* untuk menghasilkan segmentasi pelanggan yang tidak hanya informatif secara analitis, tetapi juga aplikatif dalam penyusunan strategi pemasaran berbasis data. Hasil segmentasi dapat langsung diterjemahkan menjadi program retensi, *upselling*, *cross-selling*, serta program loyalitas yang disesuaikan dengan karakteristik tiap segmen pelanggan. Dengan demikian, perusahaan telekomunikasi dapat meningkatkan efektivitas pemasaran, mengurangi tingkat *churn*, dan mendukung pengambilan keputusan yang lebih akurat. Penelitian ini masih memiliki keterbatasan pada penggunaan atribut perilaku pelanggan yang terbatas pada beberapa variabel utama. Penelitian selanjutnya dapat menambahkan atribut lain seperti data demografis, pola penggunaan layanan yang lebih rinci, atau mengombinasikan metode *clustering* lain untuk mendapatkan segmentasi yang lebih kaya dan komprehensif.

REFERENCES

- [1] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. Dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023.
- [2] K. K. Tsipitsis and A. Chorianopoulos, *Data Mining Techniques in CRM: Inside Customer Segmentation*. John Wiley & Sons, 2011.
- [3] F. Leisch, S. Dolnicar, and B. Grün, *Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful*. 2018. Accessed: Jan. 30, 2026. [Online]. Available: <https://library.oapen.org/handle/20.500.12657/51281>
- [4] A. Pramuditya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Ilmu Data*, vol. 2, no. 10, 2022, Accessed: Feb. 04, 2026. [Online]. Available: <http://www.ilmudata.org/index.php/ilmudata/article/view/235>
- [5] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf. JSI*, vol. 8, no. 1, pp. 958–965, 2016.
- [6] "(PDF) Customer Segmentation Using K- Means Clustering Algorithm." Accessed: Jan. 16, 2026. [Online]. Available: https://www.researchgate.net/publication/355587534_Customer_Segmentation_Using_K-_Means_Clustering_Algorithm
- [7] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," in *Emerging Technology in Modelling and Graphics*, vol. 937, J. K. Mandal and D. Bhattacharya, Eds., in *Advances in Intelligent Systems and Computing*, vol. 937. , Singapore: Springer Singapore, 2020, pp. 69–83. doi: 10.1007/978-981-13-7403-6_9.
- [8] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022. Accessed: Jan. 15, 2026. [Online]. Available: https://books.google.com/books?hl=id&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=%5B1%5D+J.+Han,+M.+Kamber,+and+J.+Pei,+Data+Mining:+Concepts+and+Techniques,+3rd+ed.+Morgan+Kaufmann,+2011.+doi:10.1016/C2009-0-61819-5&ots=_N8GRHjpr3&sig=a7jQCeig1APDYuP22XHdVqTA5qA
- [9] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, sydney, Australia: IEEE, Oct. 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.
- [10] B. Walek, "Fuzzy-Expert System for Customer Behavior Prediction," in *Artificial Intelligence and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference*, 2018, p. 122. Accessed: Jan. 16, 2026. [Online]. Available: https://books.google.com/books?hl=id&lr=&id=s5RdDwAAQBAJ&oi=fnd&pg=PA122&dq=Customer+behavior+analysis,+%E2%80%9D+Expert+Systems+&ots=NT9TYDWRbu&sig=u5xzbzp8gTXcxsbaGIBpyz0iJ_Y
- [11] H. Abbasimehr and M. Shabani, "A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers," *Kybernetes*, vol. 50, no. 2, pp. 221–242, 2021.
- [12] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," in *Emerging Technology in Modelling and Graphics*, vol. 937, J. K. Mandal and D. Bhattacharya, Eds., in *Advances in Intelligent Systems and Computing*, vol. 937. , Singapore: Springer Singapore, 2020, pp. 69–83. doi: 10.1007/978-981-13-7403-6_9.
- [13] K. Z. Wijaya, A. Djunaidi, and F. Mahananto, "Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri," *J. Tek. ITS*, vol. 10, no. 2, pp. A300–A237, 2021.
- [14] A. D. Savitri, F. A. Bachtiar, and N. Y. Setyawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus: Belle Crown Malang)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 2, no. 9, pp. 2957–2966, 2018.
- [15] I. Ariati, R. N. Norsa, L. Akhsan, and J. Heikal, "Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan Uht Milk Greenfield," *Cerdika J. Ilm. Indones.*, vol. 3, no. 7, pp. 729–743, 2023.
- [16] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4. 5, SVM, dan naive Bayes," *J. Teknol. Dan Sist. Komput.*, vol. 8, no. 2, pp. 89–93, 2020.
- [17] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. in *Intelligent Systems Reference Library*, vol. 72. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-10247-4.
- [18] N. Yudistira, "Peran big data dan deep learning untuk menyelesaikan permasalahan secara komprehensif," *Expert*, vol. 11, no. 2, pp. 78–89, 2021.
- [19] D. Arthur and S. Vassilvtskii, "k-means++: The advantages of careful seeding," Stanford, 2006. Accessed: Jan. 15, 2026. [Online]. Available: <http://ilpubs.stanford.edu:8090/778/>
- [20] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, 2016.