Jurnal Komputer Antartika
Volume 4 Nomor 2 2026 Halaman 93-98
ISSN (Online) 2988-5272
DOI: https://doi.org/10.70052/jka.v4i2.1375

Segmentasi Perilaku Pengguna Kartu Kredit untuk Prediksi Default dengan K-Means Clustering

Faiz Daffa Kurnia[1], Supriatin[2*], Dwi Nurani[3], Alfie Nur Rahmi[4]

[1,2,3,4] Universitas Amikom Yogyakarta
Jl Ringroad Utara, Condongcatur, Sleman, Yogyakarta, Indonesia

e-mail korespondensi: supriatin@amikom.ac.id

Submit: 22-05-2026 | Revisi: 10-06-2026 | Terima: 20-06-2026 | Terbit online: 04-07-2026

Abstrak - Peningkatan penggunaan kartu kredit di berbagai lapisan masyarakat diikuti oleh meningkatnya risiko *credit default* yang berpotensi menimbulkan kerugian finansial bagi institusi keuangan. Penelitian ini bertujuan untuk melakukan segmentasi perilaku pengguna kartu kredit serta menganalisis potensi default pada setiap segmen menggunakan metode K-Means Clustering. Data yang digunakan berasal dari *Credit Card User Behavior and Default Dataset* dari Kaggle, yang mencakup 10.000 nasabah dengan 24 atribut. Metodologi penelitian meliputi praproses data, seleksi fitur berbasis korelasi, penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Analysis, serta visualisasi dengan Principal Component Analysis (PCA). Hasil penelitian menunjukkan bahwa jumlah cluster optimal diperoleh pada K=2 dengan nilai Silhouette Score sebesar 0,250, lebih tinggi dibandingkan K=3 (0,195) dan K=4 (0,178). Segmentasi yang dihasilkan mampu mengidentifikasi perbedaan karakteristik perilaku nasabah yang berkaitan dengan tingkat risiko default. Kontribusi penelitian ini terletak pada integrasi metode unsupervised learning untuk segmentasi perilaku dengan analisis risiko pada tiap cluster, serta pengembangan aplikasi berbasis web menggunakan Streamlit untuk mendukung prediksi segmen nasabah secara interaktif.

Kata Kunci : K-Means Clustering, Segmentasi Pengguna, Kartu Kredit, Credit Default, Data Mining

Abstract *- The increasing use of credit cards across various segments of society has led to a higher risk of credit default, potentially causing financial losses for financial institutions. This study aims to segment credit card user behavior and analyze default risk within each segment using the K-Means Clustering method. The dataset used is the Credit Card User Behavior and Default Dataset from Kaggle, consisting of 10,000 customers with 24 attributes. The research methodology includes data preprocessing, correlation-based feature selection, determination of the optimal number of clusters using the Elbow Method and Silhouette Analysis, and visualization using Principal Component Analysis (PCA). The results indicate that the optimal number of clusters is K=2, achieving the highest Silhouette Score of 0.250, compared to K=3 (0.195) and K=4 (0.178). The resulting segmentation identifies distinct behavioral characteristics associated with different levels of default risk. The main contribution of this study lies in integrating unsupervised learning for behavioral segmentation with risk analysis in each cluster, as well as implementing the model in a web-based application using Streamlit to support interactive decision-making.*

Keywords : K-Means Clustering, User Segmentation, Credit Card, Credit Default, Data Mining

1. Pendahuluan

Perkembangan teknologi digital telah mendorong peningkatan penggunaan kartu kredit sebagai alat transaksi yang praktis dan fleksibel. Namun, peningkatan tersebut juga diikuti oleh risiko *credit default* yang dapat menimbulkan kerugian finansial bagi lembaga keuangan. Risiko ini sering kali dipengaruhi oleh perbedaan karakteristik dan perilaku nasabah, seperti pola pembayaran, tingkat pengeluaran, serta pemanfaatan batas kredit. Keterbatasan dalam memahami pola perilaku tersebut menyebabkan lembaga keuangan kesulitan dalam mengidentifikasi nasabah dengan potensi gagal bayar secara dini.

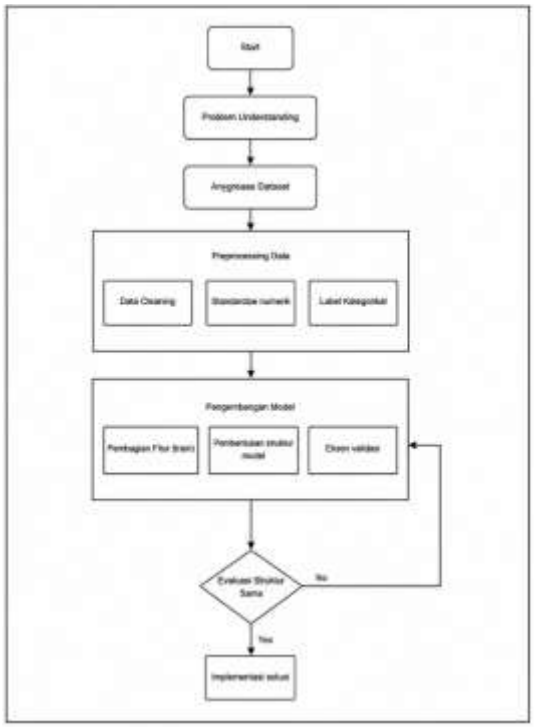
Pendekatan berbasis *data mining*, khususnya teknik *clustering*, banyak digunakan untuk mengelompokkan nasabah berdasarkan kemiripan karakteristik tanpa memerlukan label awal[1]. Beberapa penelitian menunjukkan bahwa algoritma K-Means Clustering efektif dalam melakukan segmentasi pelanggan kartu kredit dibandingkan metode lain seperti DBSCAN dan Hierarchical Clustering [2][3][4], meskipun hasilnya bergantung pada karakteristik data yang digunakan. Di sisi lain, penelitian terkait prediksi default umumnya menggunakan metode klasifikasi seperti Random Forest dan XGBoost yang berfokus pada akurasi prediksi, namun kurang mengeksplorasi hubungan antara segmentasi perilaku dan risiko default[1][3][5].

Meskipun berbagai penelitian telah menerapkan K-Means Clustering untuk segmentasi pelanggan kartu kredit dan penelitian lain berfokus pada prediksi credit default menggunakan algoritma klasifikasi machine learning, masih terdapat keterbatasan dalam menghubungkan hasil segmentasi perilaku dengan analisis risiko default secara langsung. Sebagian besar penelitian hanya menghasilkan kelompok pelanggan berdasarkan karakteristik penggunaan kartu kredit tanpa mengevaluasi tingkat risiko gagal bayar pada setiap kelompok. Kebaruan penelitian ini terletak pada integrasi metode K-Means Clustering dengan analisis distribusi risiko default pada masing-masing cluster sehingga tidak hanya menghasilkan segmentasi perilaku nasabah, tetapi juga memberikan interpretasi profil risiko yang dapat dimanfaatkan dalam pengambilan keputusan kredit. Selain itu, model yang dihasilkan diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit untuk mendukung prediksi segmen pelanggan secara interaktif dan mudah digunakan.

Berdasarkan hal tersebut, penelitian ini mengusulkan pendekatan yang mengintegrasikan segmentasi perilaku pengguna kartu kredit menggunakan K-Means Clustering dengan analisis risiko default pada setiap cluster [6]. Pendekatan ini tidak hanya menghasilkan pengelompokan data, tetapi juga memberikan interpretasi risiko yang lebih aplikatif bagi pengambilan keputusan. Selain itu, model yang dihasilkan diimplementasikan dalam aplikasi berbasis web untuk mendukung prediksi segmen nasabah secara interaktif [7]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan pendekatan *data-driven* untuk pengelolaan risiko kredit [8].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan *data mining* dengan metode *unsupervised learning* untuk melakukan segmentasi perilaku pengguna kartu kredit dan menganalisis potensi risiko *default* pada setiap kelompok. Alur penelitian mengacu pada tahapan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang meliputi pemahaman masalah, pengolahan data, pemodelan, evaluasi, dan implementasi.



Gambar 1. Alur Penelitian

Alur penelitian yang dilakukan pada penelitian ini terdiri atas enam tahapan utama, yaitu pemahaman masalah, pengumpulan data, preprocessing data, pengembangan model, validasi dan interpretasi hasil, serta deployment sistem sebagaimana ditunjukkan pada Gambar 1. Tahapan penelitian diawali dengan pemahaman masalah untuk mengidentifikasi segmen pengguna kartu kredit berdasarkan perilaku pembayaran dan tingkat risiko default. Selanjutnya, dilakukan pengumpulan data menggunakan *Credit Card User Behavior and Default Dataset* yang diperoleh dari platform Kaggle dan diolah melalui lingkungan Google Colab. Data yang diperoleh kemudian melalui tahap *preprocessing* yang meliputi penanganan nilai hilang (*missing values*), transformasi variabel kategorikal menggunakan *One-Hot Encoding*, serta standarisasi fitur numerik menggunakan *StandardScaler*. Setelah data siap digunakan, dilakukan pengembangan model dengan mengimplementasikan algoritma K-Means Clustering menggunakan pustaka *scikit-learn*. Pada tahap ini juga dilakukan seleksi fitur berdasarkan analisis korelasi serta penentuan jumlah cluster optimal menggunakan *Elbow Method*. Model yang dihasilkan kemudian dievaluasi menggunakan *Silhouette Score* untuk mengukur kualitas clustering dan divisualisasikan menggunakan *Principal Component Analysis* (PCA) guna mempermudah interpretasi hasil segmentasi. Tahap terakhir adalah

deployment, yaitu mengimplementasikan model yang telah dikembangkan ke dalam aplikasi berbasis web menggunakan framework Streamlit sehingga dapat digunakan secara interaktif sebagai alat bantu dalam proses pengambilan keputusan.

2.1. Pengumpulan dan Persiapan Data

Dataset yang digunakan dalam penelitian ini adalah *Credit Card User Behavior and Default Dataset* yang diperoleh dari platform Kaggle. Dataset tersebut terdiri dari 10.000 data nasabah dengan 24 atribut yang mencakup informasi demografis, kondisi keuangan, perilaku transaksi, serta status gagal bayar (*default*). Variabel yang digunakan meliputi atribut numerik dan kategorikal yang relevan untuk menggambarkan perilaku penggunaan kartu kredit.

Tahap persiapan data diawali dengan proses *preprocessing* untuk memastikan data siap digunakan pada algoritma K-Means Clustering. Proses ini meliputi pemeriksaan dan penanganan nilai hilang (*missing values*), transformasi variabel kategorikal menggunakan teknik *One-Hot Encoding*, serta normalisasi data numerik menggunakan *StandardScaler* untuk menyamakan skala antar fitur sehingga tidak terjadi dominasi atribut tertentu dalam proses clustering [9]. Selain itu, dilakukan *Exploratory Data Analysis* (EDA) untuk memahami distribusi data dan hubungan antar variabel yang digunakan dalam penelitian.

Setelah proses *preprocessing*, dilakukan seleksi fitur untuk meningkatkan kualitas clustering dengan memilih atribut yang paling relevan [9], [10]. Seleksi fitur dilakukan berdasarkan analisis korelasi antar variabel dan kontribusinya terhadap variasi data. Hasil seleksi menunjukkan bahwa atribut Age, Annual Income, Total Spend Last Year, Credit Score, dan Customer Lifetime Value (CLV) merupakan fitur yang paling representatif untuk digunakan dalam proses segmentasi.

2.2. Pengembangan Model Clustering

Proses segmentasi dilakukan menggunakan algoritma K-Means Clustering dengan bantuan library *scikit-learn*. Algoritma ini bekerja dengan mengelompokkan data berdasarkan kedekatan jarak terhadap centroid sehingga variasi dalam setiap cluster (*intra-cluster variance*) dapat diminimalkan [11]. Sebelum proses clustering dilakukan, jumlah cluster optimal ditentukan menggunakan *Elbow Method* berdasarkan nilai *Within-Cluster Sum of Squares* (WCSS) [12] serta *Silhouette Analysis* untuk mengevaluasi kualitas pemisahan antar cluster [13]. Pengujian dilakukan pada beberapa nilai cluster, yaitu K=2, K=3, dan K=4.

Evaluasi model dilakukan menggunakan *Silhouette Score*, di mana nilai yang lebih tinggi menunjukkan kualitas pemisahan cluster yang lebih baik [13]. Selain itu, hasil clustering divisualisasikan menggunakan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data dan memproyeksikan hasil segmentasi ke dalam ruang dua dimensi sehingga pola distribusi cluster dapat diamati dengan lebih jelas [14]. PCA bekerja dengan mentransformasikan sejumlah variabel yang saling berkorelasi menjadi sejumlah komponen utama (*principal components*) yang bersifat saling bebas (*orthogonal*) dan mampu mempertahankan variasi informasi terbesar dalam data [14]. Pada penelitian ini, data yang telah dinormalisasi digunakan untuk membentuk matriks kovarians guna mengidentifikasi hubungan antar fitur. Selanjutnya dilakukan dekomposisi eigen untuk memperoleh *eigenvalue* dan *eigenvector*, di mana *eigenvector* dengan nilai *eigenvalue* terbesar dipilih sebagai komponen utama yang paling representatif terhadap variasi data. Dua komponen utama pertama, yaitu PC1 dan PC2, digunakan sebagai sumbu visualisasi untuk menggambarkan distribusi cluster hasil K-Means. Hasil visualisasi menunjukkan bahwa konfigurasi K=2 menghasilkan pemisahan cluster yang lebih baik dibandingkan konfigurasi lainnya, sejalan dengan hasil evaluasi *Silhouette Score*. Interpretasi hasil kemudian dilakukan dengan menganalisis rata-rata nilai fitur pada setiap cluster dan menghubungkannya dengan variabel *default* untuk mengidentifikasi kelompok nasabah yang memiliki tingkat risiko gagal bayar lebih tinggi.

2.3. Implementasi Sistem

Model clustering yang telah diperoleh pada tahap pengembangan kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Streamlit. Implementasi ini bertujuan untuk mempermudah pengguna dalam melakukan segmentasi nasabah secara interaktif tanpa perlu menjalankan proses pemodelan secara langsung. Model K-Means yang telah dilatih beserta objek normalisasi (*scaler*) disimpan dalam format *.pkl* sehingga dapat digunakan kembali pada saat proses prediksi.

Aplikasi dikembangkan menggunakan bahasa pemrograman Python dan dijalankan melalui platform Visual Studio Code sebelum dideploy menggunakan Streamlit Community Cloud. Pengguna dapat memasukkan data nasabah berupa usia (Age), pendapatan tahunan (Annual Income), total pengeluaran tahun sebelumnya (Total Spend Last Year), skor kredit (Credit Score), dan Customer Lifetime Value (CLV). Data masukan tersebut kemudian diproses menggunakan *scaler* yang sama dengan tahap pelatihan sebelum dilakukan prediksi cluster menggunakan model K-Means yang dipilih.

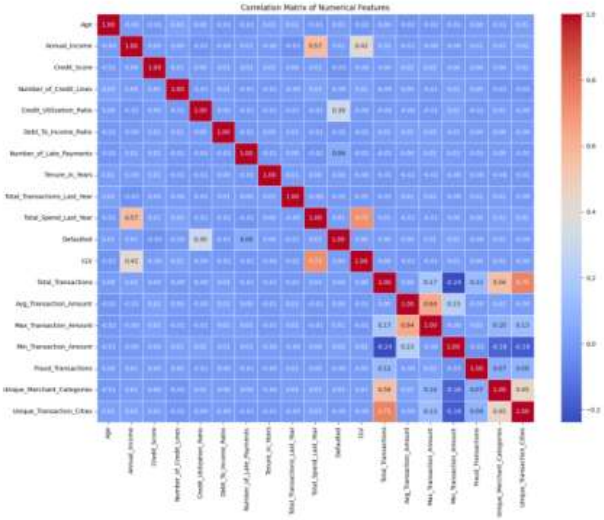
Hasil yang ditampilkan oleh sistem berupa informasi segmen nasabah, karakteristik utama cluster, serta estimasi tingkat risiko default berdasarkan profil cluster yang terbentuk. Dengan adanya implementasi ini, hasil penelitian tidak hanya bersifat teoritis tetapi juga dapat dimanfaatkan sebagai alat bantu pengambilan keputusan berbasis data dalam pengelolaan risiko kredit.

3. Hasil dan Pembahasan

3.1 Pemrosesan data

Langkah pertama yang dilakukan adalah pemeriksaan nilai hilang menggunakan isnull().sum(), hasilnya

menunjukkan tidak adanya nilai hilang. Variabel kategorikal dikonversi menggunakan one-hot encoding dengan drop_first=True untuk menghindari multikolinearitas. Kemudian terakhir fitur numerik distandarisasi menggunakan StandardScaler untuk menghilangkan perbedaan skala antar variabel. Tahap preprocessing ini penting dilakukan agar data siap untuk dikembangkan dengan lebih efisien dalam pembuatan model. Pada gambar 2 Matriks korelasi menunjukkan mayoritas korelasi antar fitur tergolong lemah hingga sedang.



Gambar 2. Visualisasi Matriks Korelasi

3.2 Hasil Segmentasi dengan K-Means Clustering

Proses clustering dilakukan menggunakan algoritma K-Means pada data yang telah melalui tahap *preprocessing* dan seleksi fitur. Berdasarkan hasil pengujian dengan Elbow Method, penurunan nilai *inertia* mulai melandai pada rentang K=2 hingga K=4. Untuk memperkuat pemilihan jumlah cluster, dilakukan evaluasi menggunakan *Silhouette Score*.

Hasil evaluasi menunjukkan bahwa nilai *Silhouette Score* tertinggi diperoleh pada K=2 sebesar 0,250, diikuti oleh K=3 sebesar 0,195 dan K=4 sebesar 0,178. Meskipun nilai yang diperoleh relatif rendah, hal ini mencerminkan karakteristik data perilaku keuangan yang cenderung memiliki distribusi kompleks dan *overlapping* antar kelompok [2][10]. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa data kartu kredit sering kali tidak membentuk cluster yang terpisah secara sempurna.

Tabel 1. Hasil Evaluasi *silhouette score*

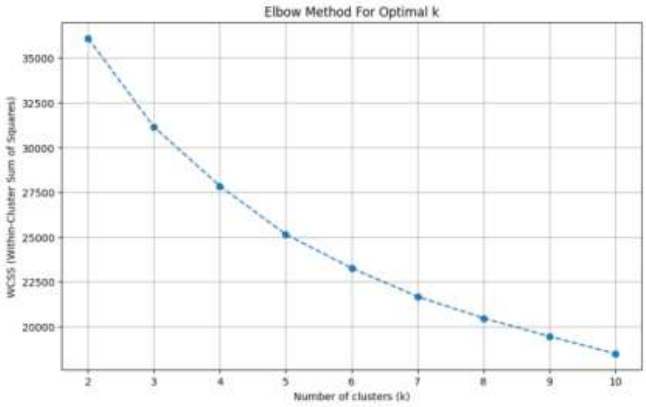
Jumlah Cluster (K)	Silhouette Score
K = 2	0,250
K = 3	0,195
K = 4	0,178
K = 5	0,178
K = 6	0,175
K = 7	0,173
K = 8	0,174
K = 9	0,165

Visualisasi menggunakan *Principal Component Analysis* (PCA) pada Tabel 1 menunjukkan bahwa pada K=2 terdapat pemisahan yang cukup jelas antar cluster, sedangkan pada K=3 dan K=4 terlihat peningkatan segmentasi namun disertai dengan tumpang tindih antar kelompok. Oleh karena itu, konfigurasi K=2 dipilih sebagai model dengan kualitas pemisahan terbaik, sementara K=3 dan K=4 digunakan sebagai alternatif untuk analisis segmentasi yang lebih granular.

3.3 Interpretasi Karakteristik Cluster

Setiap cluster yang dihasilkan menunjukkan karakteristik perilaku nasabah yang berbeda. Cluster dengan nilai *Annual Income* dan *Credit Score* yang tinggi serta tingkat pengeluaran yang stabil cenderung memiliki risiko *default* yang rendah. Sebaliknya, cluster dengan skor kredit rendah, rasio pengeluaran tinggi, dan riwayat pembayaran yang kurang baik menunjukkan kecenderungan risiko gagal bayar yang lebih tinggi.

Analisis rata-rata nilai variabel *default* masing-masing cluster pada Gambar 3 menunjukkan adanya perbedaan tingkat risiko yang signifikan antar kelompok. Hal ini menunjukkan bahwa pendekatan segmentasi berbasis clustering mampu mengidentifikasi pola risiko yang tidak terlihat secara langsung pada data mentah.



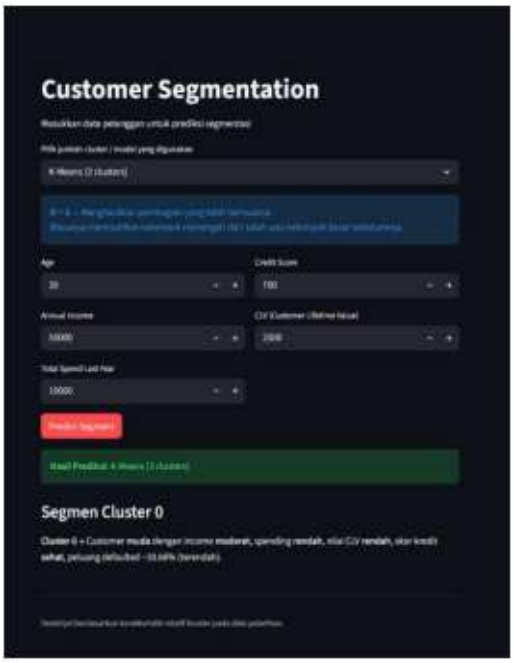
Gambar 3. Visualisasi grafik Elbow Plot

3.4 Implikasi terhadap Analisis Risiko Default

Hasil segmentasi menunjukkan bahwa pendekatan clustering tidak hanya berfungsi untuk mengelompokkan data, tetapi juga dapat digunakan sebagai dasar dalam analisis risiko kredit [15]. Dengan menghubungkan hasil cluster dengan variabel *default*, diperoleh pemetaan profil risiko nasabah yang lebih terstruktur. Pendekatan ini memberikan alternatif terhadap metode klasifikasi tradisional yang umumnya berfokus pada prediksi label, dengan menawarkan pemahaman yang lebih mendalam terhadap perilaku nasabah. Dengan demikian, lembaga keuangan dapat memanfaatkan hasil segmentasi ini untuk mengembangkan strategi mitigasi risiko yang lebih tepat sasaran, seperti penyesuaian limit kredit atau kebijakan penagihan.

3.5 Implementasi Sistem

Implementasi dalam sistem ini berupa sebuah aplikasi web sederhana berbasis *Streamlit*. Sistem ini dikembangkan menggunakan Bahasa pemrograman python di platfrom visual studio code. Ketiga model clustering yang sebelumnya telah diperoleh beserta scaler nya dari google colab di export secara local dengan ekstensi .pkl. Kemudian setelah kode sudah selesai dikembangkan, sistem dapat langsung di deploy menggunakan Streamlit Community Cloud. Setelah proses deployment selesai, sistem dapat digunakan untuk melakukan prediksi segmentasi nasabah.



Gambar 4. Tampilan system

Aplikasi yang ditampilkan pada Gambar 4 memungkinkan pengguna memasukkan data pelanggan secara manual (usia, pendapatan tahunan, total pengeluaran tahun lalu, skor kredit, dan CLV). Setelah memilih salah satu model (K=2, K=3, atau K=4) dan menekan tombol "Predict Segment", sistem menghasilkan prediksi segmen nasabah beserta deskripsi karakteristik cluster, termasuk tingkat pendapatan, pola pengeluaran, nilai CLV, skor kredit, dan estimasi peluang default pada cluster tersebut.

4. Kesimpulan

Penelitian ini berhasil menerapkan metode K-Means Clustering untuk melakukan segmentasi perilaku pengguna kartu kredit berdasarkan karakteristik demografis, kondisi finansial, dan aktivitas transaksi. Hasil penelitian menunjukkan bahwa jumlah cluster optimal diperoleh pada K=2 dengan nilai Silhouette Score sebesar 0,250. Segmentasi yang dihasilkan mampu mengelompokkan nasabah ke dalam kelompok yang memiliki karakteristik dan kecenderungan risiko default yang berbeda, sehingga memberikan gambaran yang lebih jelas mengenai profil risiko masing-masing segmen pelanggan. Temuan penelitian menunjukkan bahwa pendekatan clustering tidak hanya bermanfaat untuk memahami pola perilaku pengguna kartu kredit, tetapi juga dapat digunakan sebagai alat bantu dalam identifikasi kelompok nasabah yang berpotensi memiliki risiko gagal bayar lebih tinggi. Selain itu, implementasi model ke dalam aplikasi berbasis web menggunakan Streamlit memungkinkan proses segmentasi dilakukan secara interaktif dan mendukung pengambilan keputusan berbasis data. Penelitian ini masih memiliki keterbatasan pada penggunaan satu metode clustering dan evaluasi yang berfokus pada Silhouette Score. Oleh karena itu, penelitian selanjutnya dapat mengkaji metode clustering lain seperti DBSCAN, Agglomerative Clustering, atau Gaussian Mixture Model, serta mengombinasikan hasil segmentasi dengan algoritma klasifikasi seperti Random Forest, XGBoost, atau Neural Network untuk meningkatkan kemampuan prediksi risiko default dan menghasilkan model yang lebih akurat.

Referensi

[1] A. Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation," *J. City Dev. Vol. 3, 2021, Pages 12-30*, vol. 3, no. 1, pp. 12–30, 2021, doi: 10.12691/jcd-3-1-3.

[2] H. Ramadhan, M. R. Abdan Kamaludin, M. A. Nasrullah, and D. Rolliawati, "Comparison of Hierarchical, K-Means and DBSCAN Clustering Methods for Credit Card Customer Segmentation Analysis Based on Expenditure Level," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 246–251, 2023, doi: 10.30871/jaic.v7i2.5790.

[3] F. D. S. Alhamdani, A. A. Dianti, and Y. Azhar, "Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode K-Means Clustering," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 70–77, 2021, doi: 10.14421/jiska.2021.6.2.70-77.

[4] A. H. Zikri and R. D. Risanty, "Segmentasi Pemegang Kartu Kredit melalui Klasterisasi K-Means Menggunakan Bahasa Pemrograman Python," *J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 15, no. 3, pp. 490–503, 2025, [Online]. Available: https://jurnal.umj.ac.id/index.php/just-it/index

[5] I. Muslim, K. Karo, A. Yusmanto, R. Setiawan, U. Telkom, and U. Surya, "Segmentation of Credit Card Customers Based on Their Credit Card Usage Behavior using The K-Means Algorithm," vol. 2, no. 2, pp. 55–64, 2021, [Online]. Available: file:///D:/%23TRIDHARMA/%23PENELITIAN/Genap 2025-2026/Publikasi Faiz/referensi/4_Segmentation_of_Credit_Card_Customers_Based_on_The.pdf

[6] R. Bhandary and B. K. Ghosh, "Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods," *J. Risk Financ. Manag.*, vol. 18, no. 1, 2025, doi: 10.3390/jrfm18010023.

[7] F. Wahab, I. Khan, and S. Sabada, "Credit card default prediction using ML and DL techniques," *Internet Things Cyber-Physical Syst.*, vol. 4, no. June, pp. 293–306, 2024, doi: 10.1016/j.iotcps.2024.09.001.

[8] E. Costa e Silva, I. C. Lopes, A. Correia, and S. Faria, "A logistic regression model for consumer default risk," *J. Appl. Stat.*, vol. 47, no. 13–15, pp. 2879–2894, 2020, doi: 10.1080/02664763.2020.1759030.

[9] W. Cai, F. Yang, B. Yao, C. Li, and G. Sun, *An adaptive k-means clustering algorithm based on grid and domain centroid weights for digital twins in the context of digital transformation*, vol. 12, no. 1. Springer International Publishing, 2025. doi: 10.1186/s40537-025-01180-z.

[10] A. A. Wani, "Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions," *PeerJ Comput. Sci.*, vol. 10, pp. 1–45, 2024, doi: 10.7717/PEERJ-CS.2286.

[11] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vol. 07-09-Janu, pp. 1027–1035, 2007.

[12] Purnima Bholowalia and Arvind Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 975–8887, 2014.

[13] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[14] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

[15] F. S. Wewengkang and A. Wibowo, "Analysis Of Online Loan Regional Clustering in Indonesia in 2024 Based On Outstanding And Default Rate (TWP90) Using K-Means Clustering," vol. 8, no. 1, pp. 2237–2246, 2026.