

# PENGELOMPOKAN NASABAH BERDASARKAN TINGKAT RISIKO KREDIT MENGUNAKAN ALGORITMA K-MEANS

Moh Sutrisno<sup>1</sup>, Denni Pratama<sup>2</sup>.

Program Studi Komputerisasi Akuntansi<sup>1,2</sup>

STMIK IKMI Cirebon  
<https://ikmi.ac.id/page/18/?lang=de>  
[mohsutrisno450@gmail.com](mailto:mohsutrisno450@gmail.com)

Published ; 30 Maret 2026

**Abstract**—Banking plays a strategic role in maintaining economic stability through credit distribution to the public and business sectors. However, manual credit risk assessment processes are prone to bias, delays in decision-making, and increased Non-Performing Loan (NPL) risk. Therefore, a data-driven approach is required to objectively and efficiently classify customer credit risk. This study applies the K-Means Clustering method to group customers based on several parameters, including credit limit, customer income, payment history, and length of relationship with the bank. The research stages consist of data collection, data cleaning, normalization, implementation of the K-Means algorithm, and evaluation of clustering results based on centroid distance. The system is developed using a visual data mining application without coding to facilitate analysis. The results indicate that the K-Means method successfully forms several customer clusters with different credit risk levels, namely low, medium, and high risk. These results can assist banks in more accurately identifying high-risk customers and serve as a basis for determining appropriate credit policies and risk mitigation strategies to minimize credit default.

**Keywords:** clustering, credit risk, customer, K-Means, data mining.

**Abstrak**— Perbankan memiliki peran strategis dalam menjaga stabilitas ekonomi melalui penyaluran kredit kepada masyarakat dan pelaku usaha. Namun, proses penilaian risiko kredit yang masih dilakukan secara manual berpotensi menimbulkan bias, keterlambatan pengambilan keputusan, serta meningkatkan risiko kredit macet (Non-Performing Loan/NPL). Oleh karena itu, diperlukan pendekatan berbasis data untuk membantu pengelompokan risiko kredit nasabah secara objektif dan efisien. Penelitian ini menerapkan metode K-Means Clustering untuk mengelompokkan nasabah berdasarkan parameter nilai kredit (plafon), pendapatan, riwayat pembayaran, dan lama hubungan dengan bank. Tahapan penelitian meliputi pengumpulan data, pembersihan data, normalisasi, penerapan algoritma K-Means, serta evaluasi hasil clustering berdasarkan jarak centroid. Sistem dikembangkan menggunakan aplikasi data mining berbasis visual tanpa coding. Hasil penelitian menunjukkan bahwa metode K-Means mampu membentuk beberapa cluster nasabah dengan tingkat risiko berbeda, yaitu risiko rendah, menengah, dan tinggi. Pengelompokan ini dapat membantu pihak bank dalam mengidentifikasi nasabah berisiko tinggi secara lebih akurat serta menjadi dasar dalam penyusunan kebijakan kredit dan strategi mitigasi risiko guna meminimalkan terjadinya kredit macet.

**Kata kunci:** clustering, risiko kredit, nasabah, K-Means, data mining.

## INTRODUCTION

Perkembangan industri perbankan dan lembaga keuangan mengalami pertumbuhan yang signifikan seiring dengan meningkatnya kebutuhan masyarakat terhadap layanan pembiayaan dan kredit. Kredit menjadi salah satu produk utama yang berperan penting dalam mendukung aktivitas ekonomi, baik bagi sektor usaha maupun individu. Namun, peningkatan penyaluran kredit juga diiringi dengan meningkatnya risiko kredit, yaitu risiko

terjadinya kegagalan nasabah dalam memenuhi kewajiban pembayaran pinjaman sesuai dengan perjanjian yang telah disepakati [1]

Risiko kredit merupakan salah satu risiko utama yang dapat memengaruhi stabilitas dan kinerja lembaga keuangan[2]. Tingginya tingkat risiko kredit dapat menyebabkan peningkatan kredit bermasalah atau *non-performing loan* (NPL), yang pada akhirnya berdampak pada penurunan profitabilitas dan kepercayaan terhadap institusi keuangan[3] [4] [5]Oleh karena itu, pengelolaan

risiko kredit yang efektif menjadi faktor krusial dalam menjaga keberlanjutan operasional lembaga keuangan.

Kemajuan teknologi informasi turut berperan penting dalam mendukung pengelolaan risiko kredit. Penerapan sistem informasi perbankan memungkinkan lembaga keuangan untuk mengumpulkan, menyimpan, dan mengelola data nasabah dalam jumlah besar secara lebih terstruktur dan akurat. Data historis nasabah, seperti data pembayaran, pendapatan, dan karakteristik kredit, dapat dimanfaatkan sebagai dasar analisis untuk mengidentifikasi tingkat risiko kredit secara lebih objektif dan berbasis data.[6]

Meskipun demikian, pengelompokan tingkat risiko kredit nasabah secara konvensional masih menghadapi berbagai kendala. Proses analisis yang dilakukan secara manual cenderung memakan waktu, bersifat subjektif, dan kurang optimal dalam menangani data dengan jumlah besar serta karakteristik yang kompleks. Hal ini mendorong perlunya pendekatan yang lebih sistematis dan otomatis dalam mengelompokkan nasabah berdasarkan tingkat risiko kredit.[7], [8], [9]

Data mining menjadi salah satu pendekatan yang banyak digunakan untuk mengolah dan menganalisis data dalam jumlah besar guna menemukan pola-pola tersembunyi yang berguna dalam pengambilan keputusan[2]. Salah satu teknik data mining yang sering diterapkan dalam analisis risiko kredit adalah clustering. Teknik clustering bertujuan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan tertentu tanpa memerlukan label kelas sebelumnya [10][11].

Metode K-Means merupakan salah satu algoritma clustering yang paling populer karena memiliki konsep yang sederhana, mudah diimplementasikan, dan efisien dalam mengelompokkan data berskala besar [12][13]. Metode ini bekerja dengan membagi data ke dalam sejumlah cluster berdasarkan jarak terdekat antara data dan pusat cluster (centroid). Penerapan K-Means dalam pengelompokan nasabah memungkinkan lembaga keuangan untuk mengidentifikasi kelompok nasabah dengan tingkat risiko rendah, sedang, dan tinggi secara lebih sistematis.

Beberapa penelitian sebelumnya menunjukkan bahwa penggunaan metode K-Means dalam analisis risiko kredit mampu memberikan hasil pengelompokan yang cukup baik dan membantu lembaga keuangan dalam menyusun strategi mitigasi risiko serta pengambilan keputusan kredit [14]. Dengan mengetahui karakteristik masing-masing kelompok risiko, pihak bank dapat menyesuaikan kebijakan pemberian kredit,

penentuan suku bunga, serta strategi pengawasan kredit secara lebih efektif [15].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan pengelompokan nasabah berdasarkan tingkat risiko kredit menggunakan metode K-Means. Hasil pengelompokan diharapkan dapat memberikan informasi yang lebih jelas mengenai klasifikasi risiko nasabah sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan kredit serta mendukung penerapan manajemen risiko kredit yang lebih efektif dan berbasis data.

## MATERIALS AND METHODS

### Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data mining, khususnya teknik clustering. Pendekatan kuantitatif dipilih karena mampu mengolah data numerik secara objektif dan sistematis untuk mengidentifikasi pola dan karakteristik risiko kredit nasabah berdasarkan variabel-variabel yang relevan [16]. Teknik clustering digunakan karena penelitian ini tidak bertujuan melakukan klasifikasi berdasarkan label tertentu, melainkan mengelompokkan nasabah ke dalam beberapa kelompok risiko berdasarkan tingkat kemiripan karakteristik data [17], [18], [19].

Metode K-Means clustering diterapkan dalam penelitian ini karena memiliki kemampuan yang baik dalam mengelompokkan data berukuran besar, bersifat sederhana, serta efisien dari sisi komputasi [20]. Metode ini banyak digunakan dalam analisis risiko kredit karena mampu menghasilkan kelompok nasabah dengan karakteristik risiko yang relatif homogen, sehingga memudahkan proses interpretasi dan pengambilan keputusan [21], [22].

### Sumber dan Jenis Data

Dataset yang digunakan dalam penelitian ini terdiri dari 10.000 data nasabah kredit yang masing-masing merepresentasikan satu individu nasabah. Jumlah data yang relatif besar ini dipilih untuk memastikan bahwa proses clustering dapat menghasilkan kelompok nasabah yang representatif serta mencerminkan variasi karakteristik risiko kredit secara lebih akurat. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle, yang berisi dataset terkait data kredit nasabah. Dataset tersebut bersifat publik dan sering digunakan dalam penelitian akademik untuk

analisis risiko kredit dan perilaku pembayaran nasabah [23], [24].

Jenis data yang digunakan adalah data numerik dan kategorikal yang merepresentasikan karakteristik demografis, kondisi keuangan, serta riwayat pembayaran kredit nasabah [25]. Penggunaan data sekunder dipilih karena data telah terstruktur dengan baik, memiliki jumlah observasi yang memadai, serta relevan untuk penerapan metode clustering dalam pengelompokan tingkat risiko kredit [26].

#### Penjelasan Dataset Risiko Kredit

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle dengan *format Comma Separated Values* (.csv) yang berjumlah 10.000 data nasabah kredit dan selanjutnya diproses menggunakan perangkat lunak RapidMiner/AI Studio. Dataset terdiri dari sejumlah atribut yang merepresentasikan informasi kredit nasabah, seperti limit kredit, jenis kelamin, tingkat pendidikan, status pernikahan, usia, serta riwayat pembayaran dan jumlah tagihan kredit.

Secara umum, dataset mencakup atribut-atribut utama yang berpengaruh terhadap tingkat risiko kredit, antara lain:

- Limit kredit nasabah,
- Riwayat keterlambatan pembayaran,
- Jumlah tagihan kredit,
- Jumlah pembayaran yang dilakukan,
- Karakteristik demografis nasabah.

Atribut-atribut tersebut dipilih karena memiliki keterkaitan langsung dengan kemampuan dan perilaku pembayaran kredit nasabah, sehingga relevan digunakan sebagai dasar pengelompokan tingkat risiko kredit. Sebelum dilakukan proses clustering, data diseleksi dan dipersiapkan agar hanya atribut yang relevan yang digunakan dalam analisis untuk meningkatkan kualitas hasil pengelompokan.

#### Teknik Pengolahan dan Analisis Data

Pengolahan dan analisis data dilakukan menggunakan metode K-Means clustering dengan bantuan perangkat lunak RapidMiner/AI Studio. Tahapan analisis data dalam penelitian ini meliputi beberapa langkah sebagai berikut:

##### 1. Pengimporan Data

Dataset kredit nasabah diimpor ke dalam RapidMiner/AI Studio menggunakan operator Read CSV. Pada tahap ini dilakukan pemeriksaan struktur data, tipe atribut, serta

kelengkapan data untuk memastikan data siap digunakan dalam proses analisis.

##### 2. Praproses Data

Tahap praproses data meliputi penanganan nilai kosong (missing values), normalisasi data numerik, serta transformasi atribut kategorikal menjadi bentuk numerik. Normalisasi dilakukan untuk memastikan setiap atribut memiliki skala yang sebanding sehingga tidak mendominasi proses perhitungan jarak dalam algoritma K-Means.

##### 3. Penentuan Jumlah Cluster

Jumlah cluster (k) ditentukan berdasarkan pertimbangan karakteristik data dan tujuan penelitian, yaitu pengelompokan nasabah ke dalam tingkat risiko rendah, sedang, dan tinggi. Penentuan jumlah cluster ini juga didukung oleh evaluasi hasil clustering menggunakan metode validasi internal, seperti *Davies-Bouldin Index* (DBI).

##### 4. Proses Clustering Menggunakan K-Means

Proses clustering dilakukan menggunakan operator K-Means pada RapidMiner/AI Studio. Algoritma K-Means mengelompokkan data berdasarkan jarak terdekat antara data dan pusat cluster (centroid) hingga mencapai kondisi konvergen, di mana perubahan anggota cluster sudah minimal.

##### 5. Analisis Hasil Clustering

Hasil clustering dianalisis secara deskriptif dengan mengamati karakteristik masing-masing cluster yang terbentuk. Analisis difokuskan pada interpretasi tingkat risiko kredit setiap cluster berdasarkan nilai rata-rata atribut keuangan dan riwayat pembayaran nasabah.

#### Diagram Alur Tahapan Penelitian

Untuk memperjelas alur penelitian, digunakan diagram alur yang menggambarkan tahapan penelitian mulai dari pengumpulan data, praproses data, proses clustering menggunakan metode K-Means, hingga analisis hasil pengelompokan. Diagram alur ini disajikan pada bagian akhir metode penelitian untuk memberikan gambaran visual yang sistematis dan memperkuat pemahaman terhadap proses pengelompokan

nasabah berdasarkan tingkat risiko kredit yang diterapkan dalam penelitian ini.



Gambar 1 Diagram Alur Tahapan Perancangan

**RESULTS AND DISCUSSION**

Bagian ini membahas hasil penelitian serta pembahasan yang diperoleh dari penerapan algoritma K-Means dalam pengelompokan nasabah berdasarkan tingkat risiko kredit. Penyajian hasil disusun secara sistematis dan selaras dengan tujuan penelitian, yaitu mengelompokkan nasabah ke dalam beberapa tingkat risiko kredit berdasarkan karakteristik data historis serta menginterpretasikan hasil pengelompokan sebagai dasar pendukung pengambilan keputusan kredit.

**Hasil Pembangunan Model K-MEANS**

Proses pengelompokan nasabah dilakukan menggunakan algoritma K-Means dengan memanfaatkan dataset kredit nasabah yang diperoleh dari platform Kaggle. Dataset ini memuat berbagai atribut yang merepresentasikan karakteristik kredit nasabah, seperti batas kredit, riwayat pembayaran, serta informasi tagihan dan pembayaran. Sebelum proses clustering dilakukan, data melalui tahapan prapemrosesan yang meliputi seleksi atribut, penanganan data kosong, serta

normalisasi data untuk memastikan seluruh atribut berada pada skala yang sebanding.

Penelitian ini menetapkan jumlah cluster sebanyak tiga ( $k = 3$ ), yang ditentukan untuk merepresentasikan tingkat risiko kredit nasabah, yaitu risiko rendah, sedang, dan tinggi. Proses clustering dilakukan menggunakan perangkat lunak RapidMiner dengan operator K-Means hingga diperoleh hasil pengelompokan yang konvergen dan stabil.

Alur tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 1, yang menggambarkan proses mulai dari input dataset, prapemrosesan data, penerapan algoritma K-Means, hingga analisis hasil clustering.



Gambar 2 Hasil Pembentukan Kluster

**Distribusi Jumlah Nasabah pada Setiap Cluster**

Hasil pengelompokan nasabah menggunakan algoritma K-Means menghasilkan tiga cluster dengan jumlah anggota yang berbeda pada setiap cluster. Distribusi jumlah nasabah pada masing-masing cluster ditampilkan pada Gambar 3

Cluster	Jumlah Nasabah	Karakteristik Utama	Tingkat Risiko Kredit
Cluster 0	348	Nasabah dengan batas kredit menengah, riwayat pembayaran relatif stabil, namun masih terdapat keterlambatan pembayaran pada beberapa periode	Risiko Kredit <b>Sedang</b>
Cluster 1	1.193	Nasabah dengan batas kredit rendah, frekuensi keterlambatan pembayaran tinggi, serta tingkat tunggakan yang relatif besar	Risiko Kredit <b>Tinggi</b>
Cluster 2	5.327	Nasabah dengan batas kredit tinggi, riwayat pembayaran baik, dan tingkat tunggakan yang rendah	Risiko Kredit <b>Rendah</b>

Gambar 3 Distribusi jumlah nasabah pada setiap cluster

Berdasarkan hasil clustering, diketahui bahwa cluster\_2 merupakan cluster dengan jumlah nasabah terbanyak, yaitu sebanyak 5.327 nasabah. Selanjutnya,

cluster\_0 memiliki jumlah nasabah sebesar 3.480 nasabah, sedangkan cluster\_1 merupakan cluster dengan jumlah nasabah paling sedikit, yaitu sebanyak 1.193 nasabah.

Perbedaan jumlah anggota pada setiap cluster menunjukkan adanya variasi karakteristik kredit nasabah yang membentuk kelompok risiko yang berbeda. Dominasi jumlah nasabah pada cluster\_2 mengindikasikan bahwa sebagian besar nasabah memiliki karakteristik kredit yang relatif baik dan stabil dibandingkan dengan cluster lainnya.

### **Analisis Karakteristik dan Tingkat Risiko Kredit Nasabah**

Meskipun penelitian ini tidak menyajikan grafik perbandingan rata-rata atribut secara visual, interpretasi karakteristik masing-masing cluster dilakukan berdasarkan pola umum atribut kredit yang membentuk cluster tersebut. Hasil analisis menunjukkan bahwa cluster\_2 didominasi oleh nasabah dengan batas kredit yang relatif tinggi, riwayat pembayaran yang baik, serta tingkat tunggakan yang rendah, sehingga cluster ini dikategorikan sebagai nasabah dengan risiko kredit rendah.

Selanjutnya, cluster\_0 memiliki karakteristik nasabah dengan batas kredit menengah dan pola pembayaran yang relatif stabil, meskipun masih ditemukan keterlambatan pembayaran pada beberapa periode. Oleh karena itu, cluster\_0 diklasifikasikan sebagai nasabah dengan tingkat risiko kredit sedang.

Sementara itu, cluster\_1 didominasi oleh nasabah dengan batas kredit rendah, frekuensi keterlambatan pembayaran yang tinggi, serta tingkat tunggakan yang lebih besar dibandingkan cluster lainnya. Karakteristik tersebut menunjukkan bahwa nasabah dalam cluster\_1 memiliki potensi risiko gagal bayar yang lebih tinggi, sehingga dikategorikan sebagai nasabah dengan risiko kredit tinggi.

#### **Pembahasan**

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan nasabah ke dalam cluster yang memiliki karakteristik risiko kredit yang berbeda secara jelas. Perbedaan jumlah nasabah pada setiap cluster memberikan gambaran mengenai proporsi tingkat risiko kredit dalam keseluruhan data nasabah.

Dominasi cluster dengan risiko rendah menunjukkan bahwa sebagian besar nasabah memiliki perilaku kredit yang baik, namun keberadaan cluster risiko sedang dan tinggi tetap menjadi perhatian penting bagi pihak perbankan. Informasi hasil clustering ini dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pengelolaan risiko kredit, seperti penyesuaian kebijakan pemberian kredit, pengawasan nasabah

berisiko tinggi, serta perancangan program mitigasi risiko yang lebih tepat sasaran.

Keunggulan penggunaan algoritma K-Means dalam penelitian ini terletak pada kemampuannya dalam mengelompokkan data berukuran besar secara efisien tanpa memerlukan data berlabel. Namun demikian, metode ini memiliki keterbatasan karena hasil clustering sangat dipengaruhi oleh pemilihan jumlah cluster dan tidak mempertimbangkan hubungan sebab-akibat antar atribut kredit. Oleh karena itu, pengembangan penelitian selanjutnya dapat dilakukan dengan membandingkan metode clustering lain atau menggabungkannya dengan teknik klasifikasi untuk meningkatkan akurasi analisis risiko kredit.

### **CONCLUSION**

Penelitian ini bertujuan untuk mengelompokkan nasabah berdasarkan tingkat risiko kredit menggunakan algoritma K-Means berbasis data historis nasabah. Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Means berhasil diterapkan untuk membentuk kelompok nasabah dengan tingkat risiko kredit yang berbeda, yaitu risiko rendah, menengah, dan tinggi.

Proses pengelompokan dilakukan menggunakan perangkat lunak RapidMiner/AI Studio melalui tahapan pengimporan data, pra-proses data, normalisasi, penerapan algoritma K-Means, serta analisis hasil clustering. Hasil pengelompokan menunjukkan bahwa setiap cluster memiliki karakteristik risiko kredit yang berbeda dan dapat diinterpretasikan secara jelas berdasarkan atribut kredit yang digunakan.

Hasil penelitian ini menunjukkan bahwa algoritma K-Means cukup representatif dan layak digunakan sebagai alat bantu dalam analisis risiko kredit. Pengelompokan nasabah berdasarkan tingkat risiko kredit dapat membantu pihak perbankan dalam mengidentifikasi nasabah berisiko tinggi secara lebih objektif serta mendukung penyusunan kebijakan kredit dan strategi mitigasi risiko guna meminimalkan terjadinya kredit bermasalah.

Meskipun demikian, penelitian ini masih memiliki keterbatasan karena hanya memanfaatkan data internal kredit nasabah. Oleh karena itu, penelitian selanjutnya disarankan untuk menambahkan variabel eksternal, membandingkan algoritma *clustering* lain, serta menggabungkan metode clustering dengan teknik klasifikasi untuk meningkatkan akurasi analisis risiko kredit.

### **REFERENCE**

- [1] A. Alamsyah, A. A. Hafidh, and A. D. Mulya, "Innovative Credit Risk Assessment: Leveraging Social Media Data for Inclusive Credit Scoring in Indonesia's Fintech Sector," *Journal of Risk and Financial Management*, vol. 18, no. 2, p. 74, 2025, doi: 10.3390/jrfm18020074.
- [2] M. Bui and B. Q. Ta, "An Interpretable Decision Tree Ensemble Model for Imbalanced Credit Scoring Datasets," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 6, pp. 10853–10864, 2023, doi: 10.3233/jifs-230825.
- [3] M. M. A. Sharif, H. M. Alamgir, and Y. KiKwan, "Determinants of Non-Performing Loans(NPLs):," *Diujbe*, vol. 15, no. 02, pp. 50–63, 2022, doi: 10.36481/diujbe.v015i2.wk2akg32.
- [4] R. Karim, Md. M. Roshid, F. B. Shamme, and Md. M. Hasan, "Non-Performing Loans and Bank Profitability," *International Journal of Finance & Banking Studies (2147-4486)*, vol. 11, no. 4, pp. 95–102, 2023, doi: 10.20525/ijfbs.v11i4.2314.
- [5] S. K. Oli, "Deprived Sector Lending and Non-Performing Loans in Nepal," *Applied Economics and Finance*, vol. 8, no. 4, p. 1, 2021, doi: 10.11114/ae.v8i4.5261.
- [6] N. Eligüzel, S. Aydoğan, and İ. M. Eligüzel, "Clustering Application and Evaluation of the Countries' Word Risk and Climate Risk Indices," *International Journal of Applied Mathematics Electronics and Computers*, vol. 11, no. 1, pp. 13–19, Mar. 2023, doi: 10.18100/ijamec.1217399.
- [7] B. Siregar and Y. Yosia, "Implementation of K-means Clustering Algorithm for the Indonesian Stock Exchange," *JURNAL SISFOTEK GLOBAL*, vol. 14, no. 1, p. 49, Mar. 2024, doi: 10.38101/sisfotek.v14i1.10860.
- [8] T. Martuti, E. T. Putri, and R. Gusmana, "Application of K-Means Clustering for Student Class Division System," *J. Of Big Data Analytic and Artificial Intelligence*, vol. 6, no. 2, pp. 17–24, 2023, doi: 10.71302/jbidai.v6i2.35.
- [9] B. Siregar and Y. Yosia, "Implementation of K-means Clustering Algorithm for the Indonesian Stock Exchange," *JURNAL SISFOTEK GLOBAL*, vol. 14, no. 1, p. 49, Mar. 2024, doi: 10.38101/sisfotek.v14i1.10860.
- [10] E. S. Dalmaijer, C. L. Nord, and D. E. Astle, "Statistical Power for Cluster Analysis," *BMC Bioinformatics*, vol. 23, no. 1, 2022, doi: 10.1186/s12859-022-04675-1.
- [11] N. Eligüzel, S. Aydoğan, and İ. M. Eligüzel, "Clustering Application and Evaluation of the Countries' Word Risk and Climate Risk Indices," *International Journal of Applied Mathematics Electronics and Computers*, vol. 11, no. 1, pp. 13–19, 2023, doi: 10.18100/ijamec.1217399.
- [12] R. Prediction *et al.*, "Jurnal Sains Informatika Terapan ( JSIT ) RISK PREDICTION OF CORONARY HEART DISEASE USING A DECISION," no. 2021, pp. 206–210, 2025.
- [13] T. Z. Fate, D. S. E. D. S. Ezra, and I. A. Samuel, "Comparative Analysis of Clustering Techniques for Customer Segmentation: Evaluating K-Means, Hierarchical, and DBSCAN Models Alongside RFM Frameworks to Enhance Marketing Strategies Through Behavioral, Demographic, and Transactional Insights," *Ijaem*, vol. 7, no. 4, pp. 34–43, 2025, doi: 10.35629/5252-07043443.
- [14] G. Yao, X. Hu, T. Zhou, and Y. Zhang, "Enterprise Credit Risk Prediction Using Supply Chain Information: A Decision Tree Ensemble Model Based on the Differential Sampling Rate, Synthetic Minority Oversampling Technique and <sc>AdaBoost</Sc>," *Expert Syst.*, vol. 39, no. 6, 2022, doi: 10.1111/exsy.12953.
- [15] R. Muhammad, "Analysis of Credit Risk, Intellectual Capital and Financial Performance of Listed Deposit Money Banks in Nigeria," *Asian Journal of Economics and Business*, vol. 3, no. 2, pp. 263–289, 2022, doi: 10.47509/ajeb.2022.v03i02.05.
- [16] R. K. Martin *et al.*, "Unsupervised Machine Learning of the Combined Danish and Norwegian Knee Ligament Registers: Identification of 5 Distinct Patient Groups With Differing ACL Revision Rates," *Am. J. Sports Med.*, vol. 52, no. 4, pp. 881–891, 2024, doi: 10.1177/03635465231225215.
- [17] E. A. Amira, M. Musiega, and B. O. Alala, "Influence of Credit Risk Management on Financial Performance of Commercial Banks in Kenya," *Cradle of Knowledge African Journal of Educational and Social Science Research (The)*, vol. 11, no. 3, pp. 138–145, 2023, doi: 10.4314/ajessr.v11i3.5.
- [18] L. Zhu, W. Liu, R. Zhang, and B. Dong, "Credit Risk Evaluation of Supply Chain Finance Based on K-Means-SVM Model," *Advances in Engineering Technology Research*, vol. 2, no. 1, p. 221, Sep. 2022, doi: 10.56028/aetr.2.1.221.
- [19] J. Wang, W. Rong, Z. Zhang, and D. Mei, "Credit Debt Default Risk Assessment Based

- on the XGBoost Algorithm: An Empirical Study From China,” *Wirel. Commun. Mob. Comput.*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/8005493.
- [20] Y. Zhao, “Empowering Sustainable Finance: The Convergence of AI, Blockchain, and Big Data Analytics,” *Advances in Economics Management and Political Sciences*, vol. 85, no. 1, pp. 267–273, 2024, doi: 10.54254/2754-1169/85/20240925.
- [21] I. W. Pelealu and F. G. Worang, “Analysis the Effect of Loan Loss Provision on Bank Profitability Analisa Dampak Cadangan Kerugian Penurunan Nilai Terhadap Profitabilitas Bank,” *Jurnal EMBA*, vol. 6, no. 4, pp. 3278–3287, 2018.
- [22] P. Octaviandy and O. Pribadi, “Implementasi Metode Artificial Neural Network Untuk Analisis Resiko Kredit Pada Sistem Peer to Peer Lending,” *Bulletin of Computer Science Research*, vol. 3, no. 1, pp. 116–122, 2022, doi: 10.47065/bulletincsr.v3i1.208.
- [23] S. N. Kalid, K. C. Khor, K. H. Ng, and G. K. Tong, “Detecting Frauds and Payment Defaults on Credit Card Data Inherited with Imbalanced Class Distribution and Overlapping Class Problems: A Systematic Review,” *IEEE Access*, vol. 12, no. December 2023, pp. 23636–23652, 2024, doi: 10.1109/ACCESS.2024.3362831.
- [24] T. Zhang, X. Qian, Y. Zhou, G. Xu, and M. Wu, “Robust Clustering and Anomaly Detection of User Electricity Consumption Behavior Based on Correntropy,” *Iet Generation Transmission & Distribution*, vol. 19, no. 1, 2025, doi: 10.1049/gtd.2.70027.
- [25] M. Li *et al.*, “Multivariate Analysis for Data Mining to Characterize Poultry House Environment in Winter,” *Poult. Sci.*, vol. 103, no. 5, p. 103633, 2024, doi: 10.1016/j.psj.2024.103633.
- [26] S. C. Hicks, R. Liu, Y. Ni, E. Purdom, and D. Risso, “Mbkmeans: Fast Clustering for Single Cell Data Using Mini-Batch K-Means,” *PLoS Comput. Biol.*, vol. 17, no. 1, p. e1008625, 2021, doi: 10.1371/journal.pcbi.1008625.