

PENERAPAN ALGORITMA X – MEANS UNTUK SEGMENTASI NASABAH BERDASARKAN DATA KEUANGAN

Wisnu Adimas¹, Denni Pratama².

Program Studi Komputerisasi Akuntansi^{1,2}

STMIK IKMI Cirebon

<https://ikmi.ac.id/page/18/?lang=de>
wisnuadimas1211@gmail.com

Published : 30 Maret 2026

Abstract— The utilization of large-scale data in the financial sector plays an important role in supporting effective decision-making. Customer data stored in information systems must be processed optimally to generate valuable information. However, customer data management is often not supported by a structured segmentation process based on transaction behavior and payment delay risk, making it difficult for institutions to understand customer characteristics and determine appropriate strategies. This problem occurs due to the absence of a data-driven customer segmentation method. Therefore, this study applies the X-Means clustering algorithm to automatically group customers based on similarity in their characteristics. The research stages include customer data collection and data preprocessing, consisting of data cleaning, handling missing values, and data normalization, followed by the clustering process using the X-Means algorithm. The objective of this study is to perform customer segmentation based on transaction frequency and payment delay risk scores and to identify the characteristics of each resulting cluster. The results indicate that the X-Means algorithm successfully forms several clusters with different transaction patterns and risk levels, which can be used as a basis for risk management and business strategy development.

Keywords: clustering, customer segmentation, X-Means algorithm.

Abstrak— Pemanfaatan data dalam jumlah besar pada sektor keuangan berperan penting dalam mendukung pengambilan keputusan yang efektif. Data nasabah yang tersimpan dalam sistem informasi perlu diolah secara optimal agar menghasilkan informasi yang bernilai. Namun, pengelolaan data nasabah sering kali belum didukung oleh proses segmentasi yang terstruktur berdasarkan perilaku transaksi dan risiko keterlambatan pembayaran, sehingga menyulitkan institusi dalam memahami karakteristik nasabah dan menyusun strategi yang tepat. Permasalahan tersebut disebabkan oleh belum diterapkannya metode segmentasi nasabah yang berbasis data. Oleh karena itu, penelitian ini menggunakan algoritma X-Means untuk melakukan pengelompokan nasabah secara otomatis berdasarkan tingkat kemiripan karakteristik. Tahapan penelitian meliputi pengumpulan data nasabah, pra-pemrosesan data berupa pembersihan data, penanganan nilai hilang, serta normalisasi data, kemudian dilanjutkan dengan proses clustering menggunakan algoritma X-Means. Tujuan penelitian ini adalah melakukan segmentasi nasabah berdasarkan frekuensi transaksi dan skor risiko keterlambatan, serta mengidentifikasi karakteristik setiap cluster. Hasil penelitian menunjukkan bahwa algoritma X-Means mampu membentuk beberapa cluster dengan pola transaksi dan tingkat risiko yang berbeda sebagai dasar pengelolaan risiko dan strategi bisnis.

Kata kunci: clustering, segmentasi nasabah, algoritma X-Means

INTRODUCTION

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam industri keuangan, khususnya dalam pengelolaan dan

pemanfaatan data nasabah. Lembaga keuangan semakin mengandalkan pendekatan berbasis data untuk memahami perilaku, preferensi, serta karakteristik keuangan nasabah. Salah satu penerapan penting dari analisis data dalam konteks

ini adalah segmentasi nasabah, yang memungkinkan lembaga keuangan mengelompokkan nasabah dengan karakteristik yang serupa guna mendukung proses pengambilan keputusan strategis.

Segmentasi nasabah memiliki peranan penting dalam meningkatkan efektivitas strategi pemasaran, pengembangan produk, serta personalisasi layanan. Dengan mengidentifikasi kelompok nasabah yang homogen, lembaga keuangan dapat merancang produk dan layanan yang lebih sesuai dengan kebutuhan nasabah, sekaligus meningkatkan kepuasan dan loyalitas pelanggan. Namun, meningkatnya volume dan kompleksitas data keuangan menimbulkan tantangan bagi metode segmentasi konvensional yang umumnya bersifat manual, berbasis asumsi, dan kurang efektif dalam menangani data berskala besar.

Teknik *data mining* menawarkan solusi yang efektif untuk menggali pola dan informasi penting dari kumpulan data yang besar dan kompleks. Salah satu teknik *data mining* yang banyak digunakan untuk segmentasi nasabah adalah *clustering*, karena mampu mengelompokkan data berdasarkan tingkat kemiripan tanpa memerlukan label awal. Melalui pendekatan *clustering*, struktur tersembunyi dalam data nasabah dapat diidentifikasi sehingga memberikan wawasan yang lebih mendalam mengenai karakteristik masing-masing segmen.

Algoritma *K-Means* merupakan metode *clustering* yang paling umum digunakan karena kesederhanaan dan efisiensi komputasinya. Namun demikian, algoritma ini memiliki keterbatasan utama, yaitu perlunya menentukan jumlah cluster di awal proses analisis. Penentuan jumlah cluster yang tidak tepat dapat menghasilkan segmentasi yang kurang optimal dan mengurangi akurasi hasil analisis.

Untuk mengatasi keterbatasan tersebut, algoritma *X-Means* dikembangkan sebagai perluasan dari algoritma *K-Means*. Algoritma *X-Means* memiliki kemampuan untuk menentukan jumlah cluster secara otomatis dengan mengevaluasi berbagai model pengelompokan berdasarkan kriteria tertentu. Kemampuan ini menjadikan *X-Means* lebih fleksibel dan sesuai untuk segmentasi nasabah berbasis data keuangan, terutama ketika jumlah cluster optimal belum diketahui sebelumnya.

Penelitian ini berfokus pada penerapan algoritma *X-Means* untuk melakukan segmentasi nasabah berdasarkan data keuangan. Melalui penerapan metode ini, penelitian ini bertujuan untuk mengidentifikasi kelompok nasabah yang memiliki karakteristik keuangan yang berbeda. Hasil segmentasi yang diperoleh diharapkan dapat memberikan informasi yang bernilai bagi lembaga

keuangan dalam mendukung pengambilan keputusan berbasis data, khususnya dalam perumusan strategi pemasaran yang lebih tepat sasaran serta peningkatan pengelolaan hubungan dengan nasabah.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik *data mining* untuk melakukan segmentasi nasabah berdasarkan data keuangan. Algoritma *X-Means* digunakan sebagai metode *clustering* karena kemampuannya dalam menentukan jumlah cluster secara otomatis, sehingga sesuai untuk data keuangan yang bersifat kompleks dan beragam.

Data yang digunakan terlebih dahulu melalui tahap pra-pemrosesan yang meliputi pembersihan, transformasi, normalisasi, dan seleksi fitur guna memastikan kualitas data. Proses pengelompokan dilakukan menggunakan algoritma *X-Means* dengan ukuran jarak *Euclidean* untuk memperoleh struktur cluster yang optimal.

Kualitas hasil *clustering* dievaluasi menggunakan *Davies-Bouldin Index* dan *Silhouette Score*. Hasil pengelompokan kemudian dianalisis untuk mengidentifikasi karakteristik setiap segmen nasabah sebagai dasar pendukung pengambilan keputusan berbasis data di industri keuangan.

Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan dataset nasabah keuangan berjumlah 2.500 baris data yang bersifat sekunder dan digunakan untuk keperluan penelitian akademis. Dataset ini merepresentasikan perilaku keuangan nasabah melalui lima atribut utama, yaitu total pendapatan, total pengeluaran, frekuensi transaksi, keterlambatan pembayaran, dan skor risiko keuangan. Seluruh atribut tersebut dirancang untuk mencerminkan karakteristik finansial nasabah secara komprehensif dan relevan dalam proses segmentasi.

Sebagian besar atribut dalam dataset berbentuk numerik, sebagaimana lazim digunakan dalam analisis data keuangan. Namun, atribut total pendapatan dan total pengeluaran pada awalnya memiliki format teks akibat penggunaan pemisah ribuan, sehingga diperlukan proses pembersihan dan transformasi data sebelum dilakukan analisis lebih lanjut. Transformasi ini bertujuan untuk memastikan seluruh atribut berada dalam format numerik yang konsisten dan dapat diproses secara optimal oleh algoritma *clustering*.

Kelima atribut yang digunakan memiliki peran penting dalam menggambarkan perilaku keuangan nasabah. Dengan mengelompokkan nasabah berdasarkan atribut tersebut, proses segmentasi

diharapkan mampu menghasilkan kelompok nasabah yang memiliki karakteristik keuangan yang serupa. Dataset ini menjadi fondasi utama dalam pembangunan model segmentasi nasabah, sehingga kualitas dan keterstrukturannya sangat menentukan keakuratan hasil analisis serta relevansinya dalam mendukung pengambilan keputusan di sektor keuangan.

Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan metode dokumentasi dengan memanfaatkan dataset keuangan nasabah yang telah tersedia. Dataset tersebut digunakan sebagai sumber data utama karena telah memenuhi kebutuhan penelitian dan siap untuk diproses dalam analisis *data mining*. Pemanfaatan data sekunder ini memungkinkan peneliti untuk lebih fokus pada proses analisis dan interpretasi tanpa memerlukan pengumpulan data primer yang memakan waktu dan sumber daya tambahan.

Proses pengumpulan data diawali dengan pengunduhan dataset dari sumber yang terpercaya, kemudian dilanjutkan dengan analisis struktur dataset untuk memahami format data, tipe atribut, jumlah baris, serta rentang nilai setiap variabel. Tahap ini bertujuan untuk memastikan bahwa data dapat diproses secara optimal pada tahap analisis selanjutnya. Selain itu, dilakukan penilaian kualitas data yang mencakup pemeriksaan nilai kosong, keberadaan *outlier*, serta kesesuaian tipe data numerik guna menjaga integritas dan keandalan data yang digunakan.

Seluruh proses pengumpulan dan pengolahan data dilakukan dengan bantuan aplikasi RapidMiner Studio, yang menyediakan berbagai fasilitas analisis dan visualisasi data. Pendekatan pengumpulan data yang sistematis, mulai dari pemilihan sumber data hingga evaluasi kualitas data, menjadi fondasi penting dalam menghasilkan hasil analisis yang akurat, relevan, dan dapat diimplementasikan dalam konteks pengambilan keputusan di industri keuangan.

Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas dan kesiapan data sebelum diterapkan proses *clustering*. Tahapan ini menjadi krusial karena algoritma *X-Means* menggunakan perhitungan jarak berbasis *Euclidean*, sehingga perbedaan skala dan format data dapat memengaruhi hasil pengelompokan. Oleh karena itu, proses pra-pemrosesan dilakukan secara sistematis untuk meminimalkan potensi bias dan kesalahan analisis.

Langkah awal pra-pemrosesan adalah pembersihan format data numerik. Dua atribut, yaitu *total pendapatan* dan *total pengeluaran*, pada awalnya masih menggunakan format teks dengan pemisah ribuan berupa tanda koma. Format

tersebut tidak dapat diproses secara numerik oleh algoritma *clustering*, sehingga dilakukan penghapusan karakter non-numerik dan konversi data ke dalam tipe numerik agar dapat digunakan dalam proses perhitungan

Selanjutnya dilakukan pemeriksaan *missing value* untuk memastikan tidak terdapat nilai kosong yang dapat mengganggu proses analisis. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki *missing value*, sehingga tidak diperlukan proses imputasi atau penghapusan data pada tahap ini. Selain itu, dilakukan pemeriksaan *outlier* untuk mengidentifikasi adanya nilai ekstrem yang berpotensi memengaruhi posisi centroid cluster. Pemeriksaan *outlier* dilakukan melalui analisis visual dan perhitungan statistik. Apabila ditemukan *outlier* yang bersifat ekstrem, penanganan dapat dilakukan melalui penghapusan atau transformasi data, namun pada penelitian ini tidak ditemukan *outlier* yang signifikan sehingga data tetap dipertahankan

Tahap berikutnya adalah normalisasi data untuk menyamakan skala antar atribut. Normalisasi diperlukan karena perbedaan skala antar fitur dapat menyebabkan atribut tertentu memiliki pengaruh yang lebih dominan dalam perhitungan jarak. Metode normalisasi yang digunakan adalah *Z-Score Standardization*, yang mengubah setiap atribut sehingga memiliki nilai rata-rata nol dan deviasi standar satu. Dengan normalisasi ini, setiap fitur dapat berkontribusi secara proporsional dalam proses *clustering*.

Pada tahap seleksi fitur, seluruh atribut yang tersedia dalam dataset digunakan karena masing-masing memiliki relevansi langsung terhadap perilaku keuangan nasabah. Atribut yang digunakan meliputi pendapatan, pengeluaran, frekuensi transaksi, keterlambatan pembayaran, dan skor risiko. Penggunaan seluruh fitur ini diharapkan mampu menghasilkan segmentasi nasabah yang lebih representatif dan mencerminkan kondisi keuangan secara menyeluruh

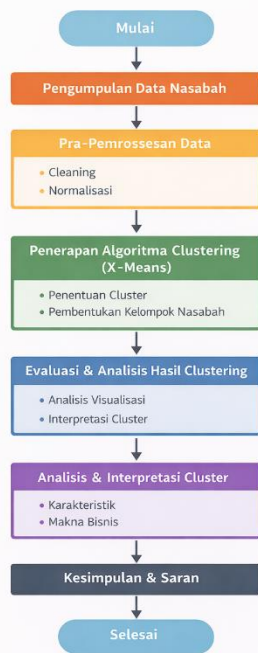
Diagram Alur Penelitian

Diagram alur penelitian digunakan untuk menggambarkan tahapan penelitian secara keseluruhan dalam penerapan algoritma *X-Means* untuk segmentasi nasabah berbasis data keuangan. Alur penelitian dimulai dari pengumpulan data nasabah yang relevan, kemudian dilanjutkan dengan tahap pra-pemrosesan data yang mencakup pembersihan dan normalisasi untuk memastikan kualitas data sebelum dilakukan proses *clustering*.

Tahap berikutnya adalah penerapan algoritma *X-Means* untuk mengelompokkan data nasabah secara otomatis dengan menentukan jumlah cluster yang optimal. Hasil *clustering* kemudian dievaluasi dan dianalisis untuk memahami pola serta karakteristik setiap cluster. Analisis lanjutan dilakukan untuk

menginterpretasikan hasil pengelompokan dari sudut pandang bisnis sebagai dasar pengambilan keputusan. Tahap akhir penelitian adalah penyusunan kesimpulan dan saran berdasarkan hasil analisis yang diperoleh.

Algoritma *X-Means* dipilih karena kemampuannya dalam menentukan jumlah cluster secara otomatis. Parameter utama yang digunakan meliputi jumlah cluster minimum dan maksimum, metode pengukuran jarak berbasis *Euclidean*, serta kriteria evaluasi cluster yang disesuaikan dengan karakteristik data numerik yang telah dinormalisasi.



Gambar 1 Diagram Alur Penelitian

RESULTS AND DISCUSSION

Bagian ini menyajikan hasil penerapan algoritma *X-Means* dalam melakukan segmentasi nasabah berdasarkan data keuangan. Hasil yang diperoleh merupakan implementasi langsung dari metodologi penelitian yang telah dijelaskan pada bagian sebelumnya dan disusun untuk menjawab tujuan penelitian yang telah ditetapkan.

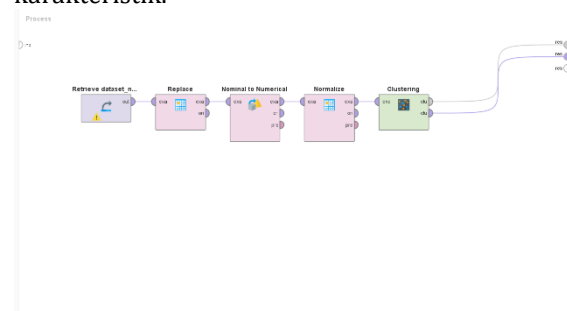
Proses Pengolahan data dan *clustering* menggunakan *X-Means*

Alur proses *clustering* dimulai dengan pengambilan dataset nasabah keuangan melalui operator *Retrieve*. Tahap ini bertujuan untuk memuat data ke dalam lingkungan analisis sebelum dilakukan pengolahan lebih lanjut.

Selanjutnya, dilakukan tahap *Replace* untuk menangani data yang tidak konsisten atau bernilai kosong guna memastikan kualitas data yang

digunakan. Data yang masih bersifat kategorikal kemudian dikonversi ke dalam bentuk numerik menggunakan operator *Nominal to Numerical* agar dapat diproses oleh algoritma *clustering*. Setelah proses konversi, data dinormalisasi menggunakan operator *Normalize* untuk menyamakan skala antar atribut sehingga setiap fitur memiliki kontribusi yang seimbang dalam perhitungan jarak.

Tahap akhir dari proses ini adalah penerapan algoritma *X-Means* untuk melakukan segmentasi nasabah berdasarkan karakteristik keuangan yang dimiliki. Algoritma *X-Means* secara otomatis menentukan jumlah cluster yang optimal dan mengelompokkan data nasabah ke dalam beberapa segmen yang memiliki tingkat kemiripan karakteristik.



Gambar 2 Alur Proses *Clustering* Menggunakan *X-Means*

Hasil Clustering Nasabah

Hasil segmentasi data nasabah menggunakan algoritma *X-Means* ditunjukkan pada Gambar 3.

Row No.	Id	cluster	total_peng_1	total_peng_2	total_peng_3	total_peng_4	total_peng_5	total_peng_6	total_peng_7	total_peng_8
1	1	cluster_2	49.980	48.679	-0.000	48.019	-0.900	-0.000	-0.800	-0.000
2	2	cluster_2	-0.000	49.998	-0.000	48.019	-0.900	-0.000	-0.800	-0.000
3	3	cluster_2	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
4	4	cluster_1	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
5	5	cluster_4	-0.000	48.679	-0.000	48.019	49.999	-0.000	-0.800	-0.000
6	6	cluster_3	-0.000	48.679	-0.000	48.019	-0.900	49.999	-0.000	-0.000
7	7	cluster_2	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
8	8	cluster_2	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
9	9	cluster_3	-0.000	48.679	-0.000	48.019	0.900	-0.000	-0.800	-0.000
10	10	cluster_3	-0.000	48.679	-0.000	48.019	0.900	-0.000	-0.800	-0.000
11	11	cluster_1	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
12	12	cluster_2	0.000	49.999	0.000	0.000	0.000	0.000	0.000	0.000
13	13	cluster_2	-0.000	48.679	-0.000	48.019	-0.900	-0.000	-0.800	-0.000
14	14	cluster_2	-0.000	48.679	-0.000	48.019	-0.900	-0.000	-0.800	-0.000
15	15	cluster_2	-0.000	49.999	-0.000	48.000	0.000	-0.000	-0.800	-0.000
16	16	cluster_2	-0.000	49.999	-0.000	48.000	0.000	-0.000	-0.800	-0.000
17	17	cluster_2	-0.000	48.679	-0.000	48.019	0.900	-0.000	-0.800	-0.000

Gambar 3 Hasil Segmentasi *Data Nasabah* Menggunakan Algoritma *X-Means*

clustering menghasilkan beberapa kelompok nasabah yang dibentuk berdasarkan kemiripan karakteristik atribut keuangan. Setiap nasabah dikelompokkan ke dalam satu cluster tertentu, sehingga hasil segmentasi ini dapat digunakan sebagai dasar untuk analisis lebih lanjut terkait pola dan karakteristik perilaku keuangan pada masing-masing kelompok.

Distribusi jumlah nasabah pada setiap cluster disajikan pada Tabel 1.

Cluster	Jumlah Nasabah
Cluster A	420

Cluster B	410
Cluster C	430
Cluster D	405
Cluster E	415
Cluster F	420
Total Nasabah	2500

Hasil tersebut menunjukkan bahwa jumlah anggota pada setiap cluster relatif seimbang, dengan kisaran antara 405 hingga 430 nasabah per cluster dari total 2.500 data. Distribusi yang proporsional ini mengindikasikan bahwa algoritma *X-Means* mampu membagi data secara efektif tanpa adanya dominasi yang signifikan pada salah satu cluster tertentu.

Interpretasi Karakteristik Cluster

Interpretasi karakteristik cluster dilakukan untuk memahami perbedaan pola perilaku nasabah pada setiap kelompok yang terbentuk. Analisis ini difokuskan pada atribut utama yang digunakan dalam proses *clustering*, khususnya frekuensi transaksi dan skor risiko keterlambatan pembayaran.

Cluster A merepresentasikan nasabah dengan frekuensi transaksi relatif tinggi dan skor risiko keterlambatan pada tingkat menengah. Kelompok ini menunjukkan tingkat aktivitas transaksi yang cukup tinggi, namun masih memiliki potensi risiko keterlambatan pembayaran yang perlu mendapat perhatian. Cluster B terdiri dari nasabah dengan frekuensi transaksi tinggi dan skor risiko keterlambatan yang rendah, sehingga dapat dikategorikan sebagai kelompok nasabah aktif dengan perilaku pembayaran yang baik.

Nasabah pada Cluster C memiliki frekuensi transaksi dan skor risiko keterlambatan pada tingkat sedang, yang mencerminkan perilaku keuangan yang relatif stabil tanpa kecenderungan ekstrem. Cluster D menunjukkan karakteristik nasabah dengan frekuensi transaksi rendah hingga sedang serta skor risiko keterlambatan yang rendah, menggambarkan kelompok nasabah yang kurang aktif namun memiliki tingkat kepatuhan pembayaran yang baik.

Cluster E didominasi oleh nasabah dengan frekuensi transaksi rendah dan skor risiko keterlambatan yang cukup tinggi. Karakteristik ini mengindikasikan adanya potensi risiko keterlambatan pembayaran pada kelompok nasabah yang kurang aktif bertransaksi. Sementara itu, Cluster F menunjukkan nasabah dengan frekuensi transaksi rendah dan skor risiko keterlambatan yang tinggi, sehingga kelompok ini dapat dikategorikan sebagai segmen nasabah dengan tingkat risiko yang paling perlu mendapatkan perhatian dalam pengelolaan dan pengambilan keputusan.

Hasil pengelompokan diperoleh dari keluaran proses *clustering* menggunakan algoritma *X-Means* pada aplikasi Altair AI Studio. Output berupa *ExampleSet (Clustering)* yang memuat data nasabah beserta label cluster pada setiap baris data. Keberadaan label cluster tersebut menunjukkan bahwa seluruh data nasabah telah berhasil dikelompokkan berdasarkan kemiripan karakteristiknya.

Ringkasan hasil pengelompokan disajikan pada Tabel 2

Cluster	Jumlah Nasabah	Frekuensi Transaksi	Skor Risiko	Karakteristik Umum
Cluster A	420	Tinggi	Sedang	Nasabah aktif dengan risiko menengah
Cluster B	410	Sedang	Rendah	Nasabah cukup aktif dan stabil
Cluster C	430	Rendah	Tinggi	Nasabah berisiko tinggi
Cluster D	405	Tinggi	Rendah	Nasabah aktif dan aman
Cluster E	415	Sedang	Sedang	Nasabah rata-rata
Cluster F	420	Rendah	Sedang	Nasabah pasif

yang memperlihatkan distribusi jumlah anggota pada setiap cluster beserta karakteristik umumnya berdasarkan frekuensi transaksi dan skor risiko. Hasil *clustering* menunjukkan bahwa algoritma *X-Means* mampu mengelompokkan data nasabah ke dalam beberapa segmen dengan karakteristik perilaku keuangan yang berbeda dan relatif seimbang dari sisi jumlah anggota.

CONCLUSION

Penelitian ini menerapkan algoritma *X-Means* untuk melakukan segmentasi nasabah berdasarkan data keuangan. Hasil penelitian menunjukkan bahwa algoritma *X-Means* mampu mengelompokkan nasabah ke dalam beberapa cluster dengan karakteristik yang berbeda berdasarkan frekuensi transaksi dan skor risiko. Proses penentuan jumlah cluster secara otomatis menjadi keunggulan utama metode ini dalam menangani data keuangan yang bersifat numerik dan beragam.

Hasil segmentasi menunjukkan adanya perbedaan tingkat aktivitas transaksi dan risiko keuangan antar kelompok nasabah, mulai dari nasabah aktif dengan risiko rendah hingga nasabah dengan aktivitas transaksi rendah dan tingkat risiko yang lebih tinggi. Temuan ini membuktikan bahwa algoritma *X-Means* efektif dalam mengidentifikasi pola perilaku keuangan nasabah secara lebih terstruktur dibandingkan metode *clustering* konvensional.

Secara praktis, hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan manajerial, khususnya dalam pengelolaan risiko dan penyusunan strategi segmentasi nasabah. Segmentasi yang dihasilkan

dapat membantu institusi keuangan dalam merancang layanan dan strategi yang lebih tepat sasaran sesuai dengan karakteristik masing-masing kelompok nasabah. Penelitian selanjutnya disarankan untuk menggunakan data yang lebih beragam serta menambahkan variabel lain agar hasil segmentasi yang diperoleh menjadi lebih komprehensif.

REFERENCE

- [1] Adeniran, I. A., Abhulimen, A. O., Obiki-Osafiele, A. N., Osundare, O. S., Agu, E. E., & Efunniyi, C. P. (2024). Data-Driven Approaches to Improve Customer Experience in Banking: Techniques and Outcomes. *International Journal of Management & Entrepreneurship Research*, 6(8), 2797–2818. <https://doi.org/10.51594/ijmer.v6i8.1467>
- [2] Adhitama, R., Burhanuddin, A., & Febriani, A. (2022). Penerapan X Means Clustering Pada UMKM Kab Banyumas Yang Mendukung Mega Shifting Consumer Behavior Akibat Covid-19. *Journal of Informatics Information System Software Engineering and Applications (Inista)*, 4(1), 71–80. <https://doi.org/10.20895/inista.v4i1.429>
- [3] Wahab, N. D., Nasib, S. K., Nurwan, N., Wungguli, D., & Yahya, N. I. (2025). Pengelompokan Data Stunting Di Indonesia Menggunakan Metode X-Means Dan Agglomerative Hierarchical Clustering. *Research in the Mathematical and Natural Sciences*, 4(1), 52–64. <https://doi.org/10.55657/rmns.v4i1.201>
- [4] Wahidin, A. J., & Sensuse, D. I. (2021a). Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *Jurnal Ict Information Communication & Technology*, 20(2), 298–302. <https://doi.org/10.36054/jict-ikmi.v20i2.387>
- [5] Wahidin, A. J., & Sensuse, D. I. (2021b). Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *Jurnal Ict Information Communication & Technology*, 20(2), 298–302. <https://doi.org/10.36054/jict-ikmi.v20i2.387>
- [6] Wulandari, V., Syarif, Y., Alfian, Z., Althof, M. A., & Mufidah, M. (2024). Comparison of Density-Based Spatial Clustering of Applications With Noise (DBSCAN), K-Means and X-Means Algorithms on Shopping Trends Data. *Ijatis*, 1(1), 1–8. <https://doi.org/10.57152/ijatis.v1i1.1135>
- [7] Oti, E. U., Olusola, M. O., Eze, F., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>
- [8] Adniana, R., Solihudin, D., & Narasati, R. (2024). Optimasi Analisis Data Kepuasan Pelanggan Cv Mega Baja Bintaro Dengan Penerapan Algoritma X-Means Clustering. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 445–453. <https://doi.org/10.36040/jati.v8i1.8722>
- [9] Arifianto, F., Hasudungan, J., Muzaky, A., & Achsan, H. T. Y. (2024). Segmentasi Pelanggan Berdasarkan Recency, Frequency, Dan Monetary Dengan K-Means Clustering: Studi Kasus Toko Pakaian Almost Famous. *Jurnal Teknologi Informatika Dan Komputer*, 10(1), 122–135. <https://doi.org/10.37012/jtik.v10i1.2096>
- [10] Djun, S. F., Gunadi, I. G. A., & Sariyasa, S. (2024). Analisis Segmentasi Pelanggan Pada Bisnis Dengan Menggunakan Metode K-Means Clustering Pada Model Data RFM. *Jtim Jurnal Teknologi Informasi Dan Multimedia*, 5(4), 354–364. <https://doi.org/10.35746/jtim.v5i4.434>
- [11] Kishendrian, H., Hanum, N., Prianto, C., & Rahayu, W. I. (2023). Penerapan Metode Clustering Dalam Segmentasi Pelanggan Perusahaan Logistik. *Sintech (Science and Information Technology) Journal*, 6(3), 137–146. <https://doi.org/10.31598/sintechjournal.v6i3.1432>
- [12] Royanti, N. I., Indrayant, I., & Amalia, N. (2023). Penerapan Metode K-Means Untuk Segmentasi Keluhan Pelanggan (Studi Kasus: Hanatex). *Ic-Tech*, 18(2), 1–7. <https://doi.org/10.47775/ictech.v18i2.282>
- [13] Siagian, R., Sirait, P. S. P., & Halima, A. (2021). E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model. *Journal of Informatics and Telecommunication Engineering*, 5(1), 21–30. <https://doi.org/10.31289/jite.v5i1.5182>
- [14] Hartama, D., & Oktaviani, S. (2025). Optimization of K-Means and K-Medoids Clustering Using Dbi Silhouette Elbow on Student Data. *Jurteks (Jurnal Teknologi Dan Sistem Informasi)*, 11(2), 289–296.

- <https://doi.org/10.33330/jurteks.v11i2.3531>
- [15] Situmorang, A., Arifin, A., Rusilpan, I., & Juliane, C. (2022). Analisa Dan Penerapan Metode Algoritma K-Means Clustering Untuk Mengidentifikasi Rekomendasi Kategori Baru Pada List Movie IMDb. *Jurnal Media Informatika Budidarma*, 6(4), 2171. <https://doi.org/10.30865/mib.v6i4.4729>
- [16] Murpratiwi, S. I., Indrawan, I. G. A., & Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 18(2), 152. <https://doi.org/10.23887/jptk-undiksha.v18i2.37426>
- [17] Mutawalli, L., Fadli, S., & Supardianto, S. (2023). Komparasi Metode Perhitungan Jarak K-Means Paling Baik Terhadap Pembentukan Pola Kunjungan Wisatawan Mancanegara. *Journal of Information System Research (Josh)*, 5(1), 159–166. <https://doi.org/10.47065/josh.v5i1.4377>
- [18] Shang, R., Ara, B., Zada, I., Nazir, S., Ullah, Z., & Khan, S. (2021). Analysis of Simple K-Mean and Parallel K-Mean Clustering for Software Products and Organizational Performance Using Education Sector Dataset. *Scientific Programming*, 2021, 1–20. <https://doi.org/10.1155/2021/9988318>
- [19] Ashtiani, M. N., & Raahemi, B. (2022). Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review. *Ieee Access*, 10, 72504–72525. <https://doi.org/10.1109/access.2021.3096799>
- [20] He, W., Hung, J.-L., & Liu, L. (2022). Impact of Big Data Analytics On banking: A Case Study. *Journal of Enterprise Information Management*. <https://doi.org/10.1108/jeim-05-2020-0176>
- [21] Li, L. (2023). The Digital Transformation of the Banking Industry in the Era of Big Data. *Advances in Economics Management and Political Sciences*, 39(1), 24–30. <https://doi.org/10.54254/2754-1169/39/20231929>
- [22] Saxena, A., Singh, A., & Govindaraj, M. (2023). Analyzing Customer Churn in Banking: A Data Mining Framework. *Multidisciplinary Science Journal*, 5, 2023ss0310. <https://doi.org/10.31893/multiscience.2023ss0310>
- [23] Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection. *Academic Journal of Science and Technology*, 10(1), 33–39. <https://doi.org/10.54097/74414c90>
- [24] Wang, A., & Yu, H. (2022). The Construction and Empirical Analysis of the Company's Financial Early Warning Model Based on Data Mining Algorithms. *Journal of Mathematics*, 2022(1). <https://doi.org/10.1155/2022/3808895>
- [25] Yuan, D., Gazi, Md. A. I., Harymawan, I., Dhar, B. K., & Hossain, A. I. (2022). Profitability Determining Factors of Banking Sector: Panel Data Analysis of Commercial Banks in South Asian Countries. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1000412>