

Pemanfaatan *Data Mining* Untuk Segmentasi Pelanggan Kartu Kredit Dengan Metode *K-Means*

¹Intan Pramesta Nurhayati, ²Helmayana, ³Adistiani, ⁴Kezia Maruenci, ⁵Yuriana Sari Harahap, ⁶Maulana Fanyuri.

¹Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

²Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

³Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

⁴Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

⁵Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

⁶Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan

¹pramesta.tn@gmail.com, ²helmayana14@gmail.com, ³adistiani102@gmail.com, ⁴maruencikezia3@gmail.com,
⁵yurianasari0@gmail.com, ⁶dosen02359@unpam.ac.id

Abstract

This study aims to cluster credit card users based on demographic information and card usage behavior using K-Means clustering algorithms. The BankChurners.xlsx dataset, which contains over 10,000 customer data, was analyzed using RapidMiner software. The analysis process includes data preprocessing steps, including normalization, attribute selection, and categorical data encoding. The K-Means algorithm is then used to group customers into two clusters. The results of this clustering show the existence of two main segments with different characteristics, where the majority of customers fall into one larger group. Cluster quality assessment using the Davies-Bouldin index shows satisfactory separation results. This result can serve as a basis for strategic decision-making, particularly in designing marketing plans and developing services that are more precise and suited to the characteristics of each customer segment

Keywords: Data Mining, K-Means, RapidMiner, Customer Segmentation, Credit Card

Abstrak

Studi ini memiliki tujuan untuk mengelompokkan pengguna kartu kredit berdasarkan informasi demografis dan tingkah laku penggunaan kartu melalui pemanfaatan algoritma *K-Means clustering*. Dataset *BankChurners.xlsx* yang berisi lebih dari 10.000 data pelanggan dianalisis menggunakan perangkat lunak *RapidMiner*. Proses analisis meliputi langkah-langkah pra-pemrosesan data, termasuk normalisasi, pemilihan atribut, dan pengkodean data kategorikal. Algoritma *K-Means* kemudian digunakan untuk mengelompokkan pelanggan menjadi dua kluster. Hasil dari pengelompokan ini menunjukkan adanya dua segmen utama dengan ciri-ciri yang berbeda, di mana mayoritas pelanggan masuk dalam satu kelompok yang lebih besar. Penilaian kualitas kluster menggunakan *indeks Davies-Bouldin* menunjukkan hasil pemisahan yang memuaskan. Hasil ini dapat dijadikan landasan dalam pengambilan keputusan strategis, khususnya dalam merancang rencana pemasaran dan mengembangkan layanan yang lebih tepat dan sesuai dengan karakteristik masing-masing segmen pelanggan.

Kata Kunci: Data Mining, K-Means, RapidMiner, Segmentasi Pelanggan, Kartu Kredit

A. PENDAHULUAN

Di era digitalisasi dengan kemajuan dalam sistem pembayaran elektronik telah menyebabkan peningkatan penggunaan kartu kredit di masyarakat. Studi yang ada menunjukkan bahwa metode pembayaran non-tunai, seperti kartu debit, kartu kredit, dan uang digital,

berkontribusi secara positif terhadap pertumbuhan ekonomi nasional, baik dalam jangka pendek maupun jangka panjang. Meningkatnya jumlah pengguna kartu kredit menimbulkan tantangan untuk mengenali perilaku dan preferensi pengguna secara lebih mendalam, sehingga metode analisis yang mampu mengelola data dalam jumlah besar dan kompleks sangat dibutuhkan.

Data yang dikumpulkan oleh perusahaan penyedia kartu kredit sangat bervariasi dan dalam jumlah yang besar, mulai dari informasi demografis, data perilaku, serta status keaktifan pelanggan. Informasi ini sangat penting untuk mengevaluasi perilaku pelanggan dan menemukan segmen pasar yang berbeda. Dengan segmentasi yang akurat, perusahaan bisa menyesuaikan cara pemasaran dan layanan mereka agar lebih efisien dan lebih personal.

Salah satu cara yang efektif untuk mengelompokkan pelanggan berdasarkan ciri dan perilaku mereka adalah dengan menggunakan metode *K-Means* dalam data mining. *K-Means* mampu mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan kemiripan atribut, seperti pelanggan yang aktif dengan banyak transaksi, pelanggan setia yang memiliki limit tinggi, atau pelanggan yang berisiko berhenti menggunakan layanan. Penelitian terdahulu menunjukkan bahwa *K-Means* dapat menghasilkan segmentasi yang berguna dan membantu lembaga keuangan merencanakan produk serta strategi pemasaran yang lebih terfokus. Selain itu, penerapan metode *K-Means* dalam segmentasi pelanggan kartu kredit telah terbukti cepat dan efektif dalam mengidentifikasi berbagai profil pelanggan.

Studi ini memanfaatkan data pelanggan kartu kredit yang mencakup atribut-atribut demografi dan perilaku aktual dari dataset "BankChurners. xlsx". Dengan menerapkan metode *K-Means* clustering, diharapkan dapat diperoleh segmentasi pelanggan yang relevan yang bermanfaat bagi pengambilan keputusan strategis perusahaan kartu kredit, khususnya dalam mengembangkan layanan dan pemasaran yang lebih terarah.

B. PELAKSAAAN DAN METODE

Data yang digunakan dalam penelitian ini diambil dari situs Kaggle, dengan judul dataset "BankChurners. csv". Dataset ini memuat rincian demografis serta perilaku riil pelanggan kartu kredit, dan telah banyak dimanfaatkan dalam penelitian mengenai segmentasi pelanggan di bidang pengolahan data.

Dataset "BankChurners.csv" terdiri dari 10.127 entri data pelanggan dengan 20 atribut yang merepresentasikan data demografis, perilaku penggunaan kartu kredit, serta informasi keaktifan pelanggan. Beberapa atribut penting dalam dataset ini meliputi:

1. CLIENTNUM: Nomor identifikasi unik untuk setiap pelanggan
2. Attrition_Flag: Status pelanggan (Existing Customer atau Attrited Customer)
3. Customer_Age: Usia pelanggan
4. Gender: Jenis kelamin pelanggan (M/F)
5. Dependent_count: Jumlah tanggungan
6. Education_Level: Tingkat pendidikan pelanggan
7. Marital_Status: Status pernikahan
8. Income_Category: Kategori pendapatan

9. Card_Category: Kategori kartu kredit (Blue, Silver, Gold, Platinum)
10. Months_on_book: Durasi hubungan dengan bank (dalam bulan)
11. Total_Relationship_Count: Jumlah produk yang dimiliki pelanggan
12. Months_Inactive_12_mon: Jumlah bulan tidak aktif dalam 12 bulan terakhir
13. Contacts_Count_12_mon: Jumlah kontak dengan bank dalam 12 bulan terakhir
14. Credit_Limit: Batas kredit pada kartu kredit
15. Total_Revolving_Bal: Total saldo bergulir
16. Avg_Open_To_Buy: Rata-rata kredit yang tersedia untuk digunakan
17. Total_Amt_Chng_Q4_Q1: Perubahan jumlah transaksi dari Q1 ke Q4
18. Total_Trans_Amt: Jumlah total transaksi (dalam 12 bulan terakhir)
19. Total_Trans_Ct: Jumlah total transaksi (dalam 12 bulan terakhir)
20. Total_Ct_Chng_Q4_Q1: Perubahan jumlah transaksi dari Q1 ke Q4
21. Avg_Utilization_Ratio: Rasio pemanfaatan kartu kredit rata-rata

Dataset ini juga mencakup dua kolom tambahan yang berisi nilai klasifikasi *Naive Bayes* yang tidak akan digunakan dalam analisis clustering ini.

C. HASIL DAN PEMBAHASAN

Data yang digunakan adalah dataset "*BankChurners.xlsx*" dari Kaggle, berisi 10.127 entri pelanggan dan 20 atribut yang mencakup informasi demografis dan penggunaan kartu kredit. Atribut seperti usia, jenis kelamin, pendapatan, kategori kartu, jumlah transaksi, dan saldo digunakan untuk klusterisasi.

[illegible]

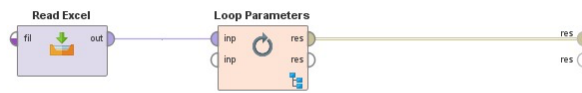
2	Graduate	Married	Less than seven	25	6	2	4	20388	1991	8427	0.703	20094
---	----------	---------	-----------------	----	---	---	---	-------	------	------	-------	-------

Gambar 1 Data Set Bank Churners

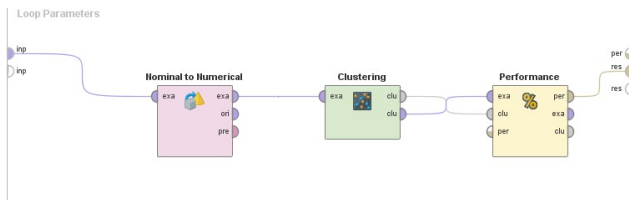
Metode yang digunakan adalah algoritma *K-Means* clustering dengan tahapan sebagai berikut:

1. Menentukan jumlah klaster ($K=2$).
2. Menyusun dan normalisasi data
3. Menjalankan algoritma *K-Means* di *RapidMiner*.
4. Mengevaluasi hasil klaster menggunakan Davies-Bouldin Index dan Silhouette Coefficient.

Langkah preprocessing termasuk pembersihan data, pemilihan atribut relevan, transformasi data kategorikal, dan normalisasi nilai numerik. Seluruh proses dilakukan menggunakan *RapidMiner Studio* yang menyediakan antarmuka visual untuk pengolahan data.



Gambar 2 Diagram Alur Preprocessing, Clustering, dan Evaluasi Kinerja pada RapidMiner

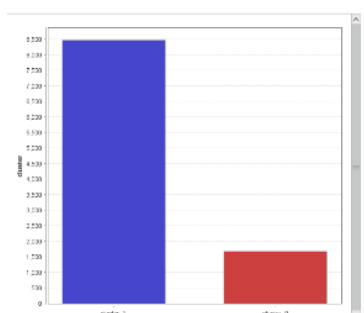


Gambar 3 Diagram alur proses yang terdiri dari tahap konversi data nominal ke numerik, proses clustering, dan evaluasi kinerja clustering menggunakan RapidMiner.

Visualisasi hasil clustering menunjukkan dua kluster utama:

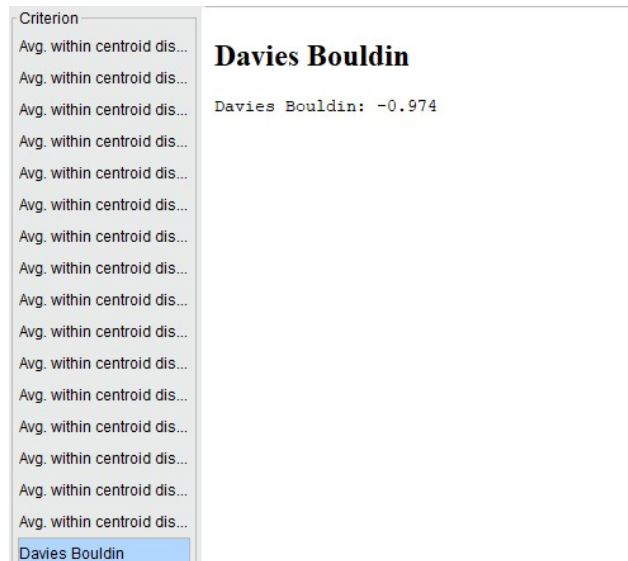
Table 1 Distribusi Jumlah Pelanggan Per Kluster

Kluster	Jumlah Pelanggan
0	1.668
1	8.459



Gambar 4 Visualisasi Distribusi Kluster Pelanggan dalam Bentuk Diagram Batang dari RapidMiner

Evaluasi kluster menggunakan Davies-Bouldin Index menghasilkan nilai -0,974, yang menunjukkan pemisahan kluster cukup baik (lihat Gambar 2). Klasterisasi ini membantu mengidentifikasi kelompok pelanggan pasif dan aktif. Kelebihan *K-Means*:



Gambar 5 Evaluasi Davies-Bouldin Index Hasil Clustering RapidMiner

Kutipan dan Acuan

Kajian oleh Alhamdani, Dianti, dan Azhar (2021) empat metode clustering: *K-Means*, Agglomerative Clustering, GMM, dan DBSCAN untuk pemisahan pelanggan kartu kredit berdasarkan perilaku penggunaan mereka. Dataset 9000 data pengguna aktif mempunyai 18 fitur untuk dipakai penelitian ini. Hasil analisis dilakukan menggunakan nilai silhouette sebagai indikator kualitas cluster menunjukkan bahwa metode *K-Means* menghasilkan tiga cluster utama dengan nilai silhouette tertinggi 0,207014. Kesimpulannya menunjukkan bahwa *K-Means* dapat mengelompokkan pelanggan berdasarkan pola penggunaan kartu kredit sudah sangat tepat. ini dapat dipakai untuk mendukung strategi pemasaran yang lebih tepat sasaran.

Penelitian menunjukkan bahwa algoritma *K-Means* memiliki performa clustering yang lebih optimal dibandingkan K-Medoids, dilihat dari nilai Davies Bouldin Index yang lebih rendah (Qomariyah & Siregar, 2022).

Model RFM membantu bisnis pengelompokan pelanggan berdasarkan umur, frekuensi, dan total uang yang mereka keluarkan. Hasilnya, strategi pemasaran dapat disesuaikan untuk meningkatkan loyalitas pelanggan (Sajidah, Herdiana, & Solihudin, 2023). *K-Means* dapat menjadi solusi metode dalam mengatasi permasalahan pelanggan mall karena *K-Means* adalah algoritma data mining untuk mengelompokkan atau mengklasifikasikan 'N' objek berdasarkan atribut atau fitur mereka menjadi 'K' kelompok (Cahyono, 2023).

Penelitian ini mengidentifikasi lima kluster pengguna Amazon yang memiliki karakteristik berbeda, dengan pendekatan *K-Means* Clustering untuk memahami kebiasaan dan preferensi belanja mereka (Rimakka & Aras, 2023). Penelitian ini mengelompokkan pengguna Cash Management System Bank BRI berdasarkan volume dan frekuensi transaksi dengan metode

K-Means, menggunakan pendekatan kerangka kerja CRISP-DM

untuk menghasilkan segmentasi pelanggan yang lebih efektif (Hesananda & Apriliga, 2024). Kelompok pengguna yang Berulang kali melakukan transaksi besar dan selalu membayar tagihannya penuh menunjukkan salah satu klaster yang terbentuk. Kemungkinan besar bahwa kelompok ini akan menjadi sasaran untuk program loyalitas seperti cashback atau penawaran eksklusif dari bank (Devi, 2025).

Algoritma *K-Means* membantu perusahaan untuk membentuk strategi pemasaran yang lebih idealis dengan menyesuaikan pola pembelian berdasarkan harga, jenis paket, dan durasi (Naufal, Martanto, Dikananda, & Rohman, 2025).

D. PENUTUP

Simpulan

Dengan menggunakan data BankChurners.xlsx, algoritma *K-Means* berhasil dipakai untuk membagi pelanggan kartu kredit. dilihat dari survei demografis dan perilaku penggunaan kartu, pelanggan dapat dibagi menjadi dua bagian utama, dilihat dari hasil pengelompokan. Kelompok yang dominan mempunyai mayoritas pelanggan. Kelompok lain mempunyai bagian yang berbeda. Dengan memakai proses pembagian ini, bisnis dapat memperoleh pemahaman yang lebih baik tentang profil pelanggan mereka dan membuat rencana pemasaran yang lebih sesuai dan berhasil. Kualitas pengelompokan yang sangat baik dengan pemisahan kelompok yang jelas ditunjukkan dalam evaluasi yang dilakukan dengan indeks Davies-Bouldin.

Saran

Untuk meningkatkan efektivitas penerapan data mining dalam segmentasi pengguna, perusahaan penyedia kartu kredit disarankan untuk memanfaatkan hasil klasterisasi ini dalam merumuskan strategi pemasaran yang lebih fokus dan sesuai dengan karakteristik masing-masing bagian pengguna. Selain itu, penelitian berikutnya sebaiknya mempertimbangkan penggunaan metode pengelompokan lain, seperti DBSCAN atau Hierarchical Clustering, untuk mendapatkan perbandingan terkait efektivitas dan akurasi segmentation yang dihasilkan. Penambahan data seperti riwayat transaksi yang lebih detail atau data dari sumber lain juga dapat memperkaya analisis serta meningkatkan ketepatan model. Untuk menjaga agar segmentasi tetap relevan dengan perubahan perilaku pengguna, pembaruan dan evaluasi model secara rutin sangat dianjurkan, sehingga keputusan strategis yang diambil oleh perusahaan selalu didasarkan pada informasi yang terbaru dan tepat.

E. DAFTAR PUSTAKA

- Al Isra Denk Rimakka1, R. A. (2023). USER SEGMENTATION BASED ON PURCHASING HABITS AND PREFERENCES ON THE AMAZON PLATFORM USING *K-MEANS* CLUSTERING. *Jurnal Teknologi Informasi dan Komunikasi* , 103-109.
- Armadyah Amborowati , . E. (2014). REVIEW PEMANFAATAN TEKNIK DATA MINING DALAM SEGMENTASI KONSUMEN . 66-73.
- Arwansyah, S. (2020). Clusterisasi Varietas Benih Tanaman Padi Menggunakan Algoritma *K-Means*. *PROSIDINGSEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, 1–10.
- Aulia Dewi Savitri1, F. A. (2018). Segmentasi Pelanggan Menggunakan Metode *K-Means* Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang) . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* , 2957-2966.
- CAHYONO, G. D. (2023). IMPLEMENTASI ALGORITMA *K-MEANS* PADA KLASIFIKASI PENGELUARAN BELANJA PELANGGAN MALL . 1-127.
- CHANDRA, K. (2013). CLUSTERING NASABAH BANK MENGGUNAKAN *K-MEANS* BERDASARKAN NILAI RFM (RECENCY, FREQUENCY, MONETARY) . *Media Informatika*, 39-55.
- Eddiwan Dwiguna, A. B. (2024). PENERAPAN DATA MINING UNTUK MENENTUKAN PENERIMA BANTUAN BLT MENGGUNAKAN METODE CLUSTERING *K-MEANS* PADA DESA PAMULIHAN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 1382-1388.
- Eko Saputra1, H. O. (n.d.). SEGMENTASI DAERAH RENTAN STUNTING DI INDONESIA MENGGUNAKAN METODE *K-MEANS*. 1-11.
- Fahmi Naufal1*, M. A. (2025). ANALISIS SEGMENTASI PELANGGAN VOUCHER WIFI DENGAN METODE *K-MEANS* . *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, 81-90 .
- Fatimah Defina Setiti Alhamdani (1), A. A. (2021). Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode *K-Means* Clustering. *JISKA*, 70 – 77 .
- Imas Maskanah1, A. P. (2020). Segmentasi Pelanggan Toko Purnama dengan Algoritma *K-Means* dan Model RFM untuk Perancangan Strategi Pemasaran. *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, 218-228.

- Irfan Nafis Sjamsuddin^{1*}, D. A. (2025). Klasterisasi Pasien Rawat Inap BPJS pada RS Islam Assyifa Sukabumi menggunakan Metode *K-Means*. *JURNAL KRIDATAMA SAINS DAN TEKNOLOGI* , 355- 368.
- Kurniati, E. A. (2015). VISUALISASI SEGMENTASI FITUR ATM MENGGUNAKAN ANALISIS RFM DAN ALGORITMA *K-MEANS* UNTUK MENDUKUNG EVALUASI FITUR ATM DI PT. BANK BRISYARIAH .
- Miftahul Fajar, N. R. (2024). ANALISIS POLA PENJUALAN OBAT DI APOTEK AN-NAAFI MENGGUNAKAN METODE *K-MEANS* CLUSTERING. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 486-492.
- Muhammad Syam Al ghifari 1, M. 2. (2024). PENGELOMPOKAN TRANSAKSI PENJUALAN AKSESORIS HP DAN PULSA DENGAN METODE *K-MEANS* UNTUK MENINGKATKAN STRATEGI PEMASARAN DI TOKO BAGUS CELLULER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2838-2849.
- Nurul Muna Fithryani^{1*}, M. A. (2025). ALGORITMA *K-MEANS* UNTUK MENINGKATKAN SEGMENTASI POLA KEKERASAN . *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* , 997-1003.
- Qomariyah (1), M. U. (2022). Comparative Study of *K-Means* Clustering Algorithm and K-Medoids Clustering in Student Data Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)* , 91 – 99 .
- Rafflesia Mukhlis Juliyanto¹, F. P. (2024). PENERAPAN ALGORITMA CLUSTERING UNTUK SEGMENTASI . 117-127.
- Rangga Bayu Prasetyo, Y. A. (2023). IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA *K-MEANS* CLUSTERING PENYAKIT PASIEN RAWAT JALAN PADA KLINIK DR. ATIRAH DESA SIOYONG, SULTENG . *JATI (Jurnal Mahasiswa Teknik Informatika)* , 2114-2151.
- Rizki Hesanda¹, P. A. (2024). Customer Segmentation of Cash Management System Using KMeans Clustering. *Journal of Applied and Research Computer Science and Information Systems* , 191-202.
- Sajidah, R. H. (2023). SEGMENTASI PELANGGAN SALON NUII BEAUTY GLOW MENGGUNAKAN *K-MEANS* CLUSTERING . *JATI (Jurnal Mahasiswa Teknik Informatika)*, 558-566.
- Sari Indah, F. G. (2023). PENERAPAN DATA MINING MENGGUNAKAN METODE *K-MEANS* UNTUK PENENTUAN REWARD PELANGGAN (Studi Kasus: UD. Penyubur Tani) . *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi* , 201-207.
- Sri Adiningsi, W. S. (2023). KLASIFIKASI BENTUK TELUR AYAM BERBASIS CITRA DIGITAL MENGGUNAKAN METODE *K-MEANS* CLUSTERING. *JATI (Jurnal Mahasiswa Teknik Informatika)* , 1507-1512.
- Suharmanto¹, W. S. (2023). Penerapan Data Mining Menggunakan Algoritma *K-Means* Untuk Clustering Perokok Usia Lebih dari 15 Tahun . *Bulletin of Information Technology (BIT)* , 501 - 507 .
- Yolanda¹, I. R. (2024). Identifikasi Cerdas Apel Fuji dan Apel Hijau: Pendekatan *K-Means* Clustering untuk Segmentasi Buah. *Journal of Education Research* , 4232-4239.