

PENGELOMPOKAN TRANSAKSI KARTU DEBIT PERBANKAN MENGUNAKAN ALGORITMA K-MEANS

Iwan Irawan¹, Reza Rahman², Arief Wibowo³

^{1,2,3}Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi,
Universitas Budi Luhur

Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan 12260

email: *¹2211602152@student.budiluhur.ac.id, ²2211602111@student.budiluhur.ac.id,
³arief.wibowo@budiluhur.ac.id

Abstract

One of bank customers' most widely used non-cash payment methods is making payments to merchants using debit cards. The data generated from these transactions can be utilized effectively by banks. This study analyzes customer spending habits through debit card transactions, employing a data mining technique called K-means clustering. By identifying patterns in customer transactions, the research aims to assist business units in developing targeted product strategies. The analysis determined that four clusters were optimal, resulting in a tightly grouped dataset with an average distance of 5.764 from the respective cluster centers. Grouping nominal transactions based on the date and time of the transaction can provide valuable insights for bank management when considering customer fund allocation.

Keywords: Banking, Clustering, Data Mining, Debit Card Transactions, K-Means Algorithm

PENDAHULUAN

Jenis pembayaran transaksi *merchant*, *store* atau toko tanpa menggunakan uang tunai (*Cashless*), menggunakan kartu debit atau kartu kredit yang juga dikenal sebagai uang elektronik, semakin banyak digunakan saat ini. Metode pembayaran non-tunai dianggap lebih praktis dan aman karena mengurangi risiko kehilangan uang dan penipuan (Bilal, Rasyid, & Rasyid, 2022).

Kemajuan teknologi saat ini mampu menyediakan kenyamanan dalam sistem pembayaran (Bilal et al., 2022). Perkembangan teknologi membawa banyak perubahan salah satunya dalam alat pembayaran elektronik yang dapat mengubah cara masyarakat dalam bertransaksi. Alat pembayaran elektronik ini memberikan kemudahan dan manfaat kepada masyarakat dalam menjalankan aktivitas, terutama dalam memenuhi kebutuhan sehari-hari (Achir Ainun Yaumul & Kusumaningrum Trias Madanika, 2021). Dalam konteks makroekonomi, semua pengeluaran dianggap sebagai ukuran tingkat konsumsi suatu negara, perilaku masyarakat ini menjadi faktor penting (Bilal et al., 2022; Pradana & Pratama, 2022). Uang elektronik memungkinkan transaksi non-tunai seperti *debit*, *kredit*, *E-Money*, dan *e-wallet* (Tanu Wijaya Abe & Subandi, 2024). Salah satu alat pembayaran elektronik yang dikeluarkan oleh bank adalah kartu debit, atau *debit card*, yang digunakan oleh nasabah bank untuk melakukan transaksi.

Jumlah transaksi bank yang dilakukan meningkat seiring dengan jumlah pelanggan yang menggunakan kartu debit (Budiansyah Hasibuan & Mahyuni, 2022). Kemungkinan besar data akan menjadi tidak produktif dan hanya akan disimpan di bank jika dibiarkan begitu saja. Pengolahan menggunakan teknik *data mining* yang efektif diperlukan untuk mengoptimalkan informasi dari data yang telah dikumpulkan. Analisis data historis untuk mengidentifikasi pola dalamnya adalah salah satu manfaat utama *data mining* (Budiansyah Hasibuan & Mahyuni, 2022; Purnama & Arianto, 2024). Persaingan yang begitu ketat membuat peningkatan dalam mutu layanan kepada pelanggan menjadi sangat krusial, mengingat bisnis perbankan pada dasarnya adalah bisnis pelayanan (Nyoman & Guna, 2023).

Ardhianto, Relawanto, & Wibowo (2021) melakukan penelitian dengan pendekatan data mining algoritma *K-Means* untuk mengelompokkan kantor cabang yang terbagi menjadi tiga, cabang besar, cabang menengah, cabang kecil. Pengelompokan cabang tersebut berdasarkan jumlah transaksi per bulan. Klasterisasi *K-means* terbukti dapat membantu pihak manajemen bank dalam proses pengelompokan cabang sehingga diketahui perencanaan dan pengelolaan cabang bank sesuai hasil pengelompokan.

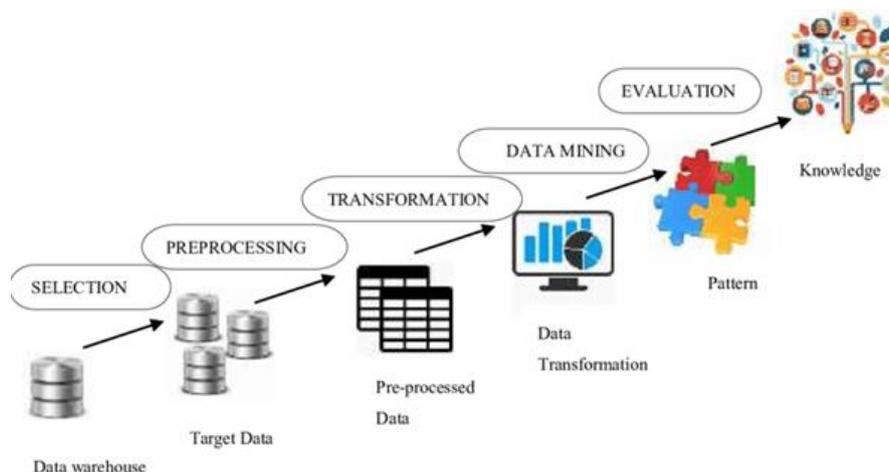
Wibowo & Sasongko (2022) melakukan penelitian yang berhubungan dengan data mining menggunakan metode *k-means* untuk pengelompokan *profitabilitas* suku bunga. Dari hasil *k-means clustering* didapatkan beberapa klaster dari yang suku bunga terbesar sampai terkecil serta bertujuan memberi pandangan terhadap investor yang akan melakukan investasi dengan acuan tarif suku bunga yang stabil.

Bank Capital adalah salah satu bank di Indonesia yang menggunakan *Data Warehouse* untuk menyimpan semua informasi yang terjadi di dalamnya. Dengan menggunakan *Data Warehouse*, bank tersebut dapat menggunakannya sebagai sumber informasi untuk keperluan *Data Mining*. *Data mining* menganalisis kebutuhan pelanggan bank dan mengamati pola perilaku pelanggan seperti Mempertahankan pelanggan (menjaga pelanggan) dan menemukan penipuan (menemukan penipuan) adalah dua tujuan utama data mining dalam industri perbankan. Studi ini menggunakan *algoritma clustering K-Means* untuk membagi pelanggan berdasarkan frekuensi, nominal, dan jenis transaksi. Studi tersebut membandingkan *K-Means* dengan sejumlah metode pengelompokan lainnya dalam penelitian yang dijalankan. Hasilnya, teknik *K-Means* menunjukkan kinerja yang lebih unggul dalam hal nilai *Davies-Bouldin Index (DBI)* jika dibandingkan dengan pendekatan menggunakan metode *K-Medoids*. Penelitian ini diharapkan dapat memberikan garis besar untuk pengembangan lebih lanjut dari penggunaan *Data Mining* di Bank Capital.

METODE PENELITIAN

1. Tahapan Penelitian

Sistematika penelitian ini memiliki struktur yang teratur berdasarkan metode *Knowledge Discovery in Databases (KDD)*. KDD merupakan suatu cara yang digunakan dalam mendapatkan pengetahuan baru dari data yang terkumpul. Proses ini menganalisis tabel-tabel basis data yang saling berhubungan untuk menemukan hubungan antara datanya (Sunata, Azrullah, & Rianto, 2020). Proses KDD terdiri dari beberapa langkah, dimulai dengan pemilihan data dan berakhir dengan penemuan pengetahuan. Pengetahuan yang ditemukan dari proses ini dapat digunakan untuk mendukung pengambilan keputusan dan strategi.



Gambar 1. Tahap Knowledge Discovery on Database (KDD)
(Rohman & Wibowo, 2024)

a. *Data Selection*

Proses pemilihan data atau *dataset* yang terkumpul adalah salah satu langkah penting dalam proses pemodelan data, karena dapat menentukan keberhasilan atau kegagalan pemodelan. Proses pemilihan data melibatkan pengolahan *dataset* dengan menggunakan algoritma yang telah ditentukan, dan hasilnya dapat disimpan dalam *file* yang terpisah dari database awal (Rojih & Heikal, 2023).

Data yang digunakan dalam penelitian ini berasal dari transaksi perbankan selama satu bulan dari Bank Capital yang terjadi dari tanggal 1 hingga 29 Desember 2023, dengan jumlah data sebanyak 2209. *Dataset* tersebut mencakup berbagai *variabel*, seperti *tanggal report rintis*, *tanggal transaksi T24*, *tanggal Settlement*, *kode ISS*, *kode ACQ*, dan lain sebagainya. Namun, untuk penerapan data mining, data yang dipilih hanya melibatkan jenis transaksi awal dan total transaksi. Berdasarkan masukan dari tim operasional *ATM Center* kantor pusat, tanggal transaksi dipilih sebagai variabel yang paling sesuai untuk mencerminkan tingkat keaktifan nasabah dalam menggunakan *Debit card*.

b. *Preprocessing/Cleaning*

Sebelum tahap *data mining* dimulai, proses pembersihan data dilakukan dengan tujuan untuk memproses data yang menjadi fokus *Knowledge Discovery in Databases (KDD)*, seperti menghapus duplikat data, menangani ketidak konsistenan data, dan memperbaiki kesalahan data. Proses pengayaan data juga penting untuk membuat informasi yang ada menjadi lebih sesuai dengan kebutuhan (Wardhana Muhammad Rizki, Triayudi Agung, & Hayati Nur, 2021).

c. *Transformation*

Data diubah melalui proses transformasi menjadi format atau struktur yang lebih sesuai dan bermanfaat untuk digunakan dalam proses *data mining*. Transformasi ini dapat mencakup berbagai langkah seperti normalisasi, pengkodean ulang, atau penambahan atribut baru. Ini dilakukan agar analisis *data mining* dapat dilakukan dengan lebih efektif dan temuan analisis dapat memberikan pemahaman yang lebih mendalam (Rojih & Heikal, 2023).

d. *Data Mining*

Data Mining merupakan alur mengidentifikasi struktur dan *trend* data yang dapat digunakan untuk membuat keputusan atau meningkatkan pemahaman. Untuk menemukan pola dan *trend* dalam data, data mining menggunakan berbagai metode dan algoritma. Tujuan dan tahapan proses *Knowledge Discovery in Databases (KDD)* biasanya menentukan teknik atau algoritma yang akan digunakan. Dalam penelitian ini, *Rapid Miner* digunakan sebagai alat untuk mengatur data dengan teknik *Clustering K-Means* (A'yuni, Nazir, Handayani, & Afrianty, 2023).

e. *Evaluation*

Tahapan ini mencakup penelitian terhadap apakah pola atau informasi yang ditemukan konsisten dengan pengetahuan yang sudah ada sebelumnya. Evaluasi model adalah langkah penting untuk memastikan bahwa model tersebut berfungsi dengan baik dan menghasilkan hasil yang akurat (Intan, Sari, Utami, Yudi, & Wijaya, 2024). Pengujian ini melibatkan variasi nilai *K* pada teknik *Clustering K-Means* dengan menggunakan beberapa titik *centroid*. Setelahnya, dilakukan pemeriksaan apakah titik *centroid* sudah sesuai dengan harapan, dan hal ini umumnya dilakukan dengan menggunakan teknik *Elbow-Method*. Tujuan dari pemeriksaan ini adalah untuk menentukan nilai *k* ideal yang akan memberikan hasil klusterisasi yang paling sesuai dengan data yang sedang diproses.

f. *Knowledge*

Pada tahap ini, pengguna menerima pola dan struktur informasi yang ditemukan. Tujuan proses ini adalah untuk memastikan semua orang memahami dan dapat menggunakan pengetahuan ini untuk membuat keputusan yang lebih baik. Dengan cara ini, Dimungkinkan untuk menyajikan data yang kompleks dengan cara yang lebih mudah dipahami dan mudah

dipahami oleh pengguna. Presentasi hasil dari proses Data Mining ini memainkan peran kunci dalam mendukung pengambilan keputusan yang informatif dan efektif.

2. Data Visualization

Proses penyajian data untuk memudahkan penerimaan dan pemahaman dilakukan melalui visualisasi. Pemahaman ini berkembang seiring kemampuan simbol-simbol visual untuk menyampaikan makna dari suatu informasi (Defina Fatimah et al., 2021). Dengan melakukan visualisasi, data diubah menjadi informasi yang dapat dimengerti secara universal. Pendekatan visual dalam penyajian data membantu mengatasi kompleksitas informasi dan memungkinkan *audiens* untuk lebih mudah menangkap, memproses, dan menginterpretasi informasi yang disajikan (Al Indiani et al., 2022; Suwirya, Candiasa, & Dantes, 2022).

3. Langkah Algoritma K-Means

Algoritma K-Means membagi data berdasarkan jumlah kelompok yang memiliki persamaan karakteristik satu sama lain. jumlah pengelompokan data awal yang diinginkan (nilai k) ditentukan oleh pengguna. Algoritma ini bekerja secara *iteratif*, dimulai dengan pemilihan titik awal secara acak untuk setiap kelompok (Ardhianto et al., 2021). Fungsi utama dari pengelompokan ini adalah untuk meminimalisir kesamaan suatu kelompok dan memaksimalkan fungsi kesamaan di dalam kelompok tersebut. Alur kerja algoritma K-Means dapat dijelaskan menggunakan persamaan berikut:

- tentukan nilai k yang diinginkan sebagai jumlah *cluster*,
- melakukan inisialisasi nilai awal *cluster* yang dilakukan secara acak.,
- melakukan perhitungan jarak pada salah satu data ke titik pusat *cluster* menggunakan persamaan 1 yang dinamakan *Euclidean Distance*.

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

dimana:

D : jarak antara objek dan *centroid*

n : jumlah variabel

x_i : (x_1, x_2, \dots) yaitu data ke- i

y_j : (y_1, y_2, \dots) *centroid* ke- j

- menghitung pusat *cluster* baru menggunakan persamaan 2.

$$C = \frac{\sum m}{n} \quad (2)$$

dimana:

C : *centroid* data

m : nilai jarak pada suatu cluster

n : banyaknya data pada suatu cluster

- Apabila dalam iterasi pada langkah-langkah tersebut tidak terjadi lagi perpindahan anggota (data) antar kelompok atau tidak ada perubahan pada posisi pusat *cluster*, maka *algoritma K-Means* dianggap telah mencapai *konvergensi* (Rohman & Wibowo, 2024). Pada titik ini, kelompok atau *cluster* yang terbentuk dianggap sudah stabil, dan tidak ada lagi perubahan yang signifikan dalam penempatan anggota ke dalam kelompok (Harahap, 2020).

4. Elbow Method

Elbow Method adalah metode untuk menentukan nilai k yang ideal atau efektif pada *algoritma K-Means*. Proses ini dilakukan seperti berikut:

- Menentukan suatu nilai $k > 0$, yang akan menjadi jumlah *cluster* yang dimuat dalam *dataset* oleh *algoritma K-Means*.

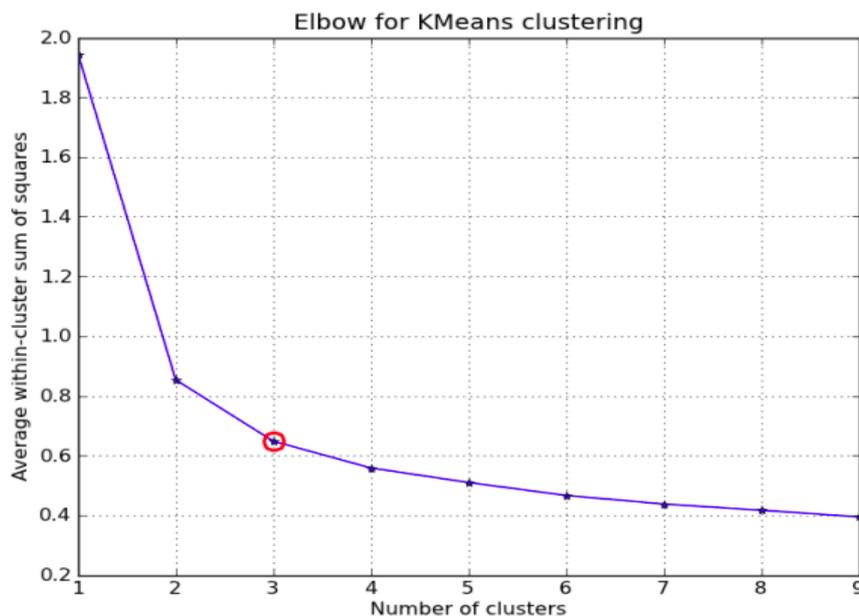
- b. Melakukan penjumlahan pada nilai k yang sering disebut sebagai $var(k)$. Ini mencerminkan sejauh apa data dalam suatu *cluster* berkumpul di sekitar pusat *clusternya*.
- c. Membuat kurva yang menggambarkan hubungan antara nilai k dan $var(k)$.
- d. Menentukan titik belok pada kurva, yang merupakan titik di mana penurunan varians menjadi kurang signifikan atau membentuk siku seperti "*elbow*". Titik ini mencerminkan jumlah *cluster* yang "tepat" atau optimal.
- e. Nilai (k) yang terdapat pada titik siku tersebut dianggap sebagai jumlah *cluster* optimal untuk analisis *K-Means*.
- f. Perhitungan *sum of within-cluster variance* (*SSE* atau *var*) pada data dengan nilai k dilakukan menggunakan persamaan 3.

$$SSE = \sum(Y_i - \tilde{Y}_i)^2 \quad (3)$$

dimana:

Y_i : adalah nilai yang terdapat pada setiap data *cluster*

\tilde{Y}_i : merupakan nilai titik tengah dari tiap *cluster*.



Gambar 2. Kurva *Elbow Method*

Pada Gambar 2 terlihat bahwa metode *Elbow* membantu menentukan jumlah cluster yang optimal dengan melihat di mana peningkatan jumlah cluster tidak memberikan penurunan varian yang signifikan, atau dengan kata lain, titik di mana penurunan varian menurun secara drastis.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang bersumber dari laporan transaksi nasabah menggunakan *Debit Card* Bank Capital. Pada tahap awal dilakukan *preprocessing* pada data sumber sehingga menghasilkan *dataset* yang siap diolah. Selain itu, dilakukan pemilihan *field-field* yang ada pada laporan transaksi *Debit Card* kemudian dilakukan *transformation*. Data yang digunakan adalah data transaksi *Debit Card* bulan Desember 2023 sebanyak 2.209 transaksi. Tabel 1 berisi 8 (delapan) atribut yang dipilih untuk tahap pemodelan *data mining*. Atribut-atribut yang dipilih adalah yang relevan dengan tujuan penelitian untuk memastikan hasil penelitian yang akurat dan bermanfaat.

Tabel 1. Daftar Atribut Terpilih untuk Proses Data Mining

No	Nama Atribut	Tipe Data	Range Data Setelah Transform
1	Tgl Trx T24	Date	1 = Tanggal 1 sd 7 (Minggu ke-1) 2 = Tanggal 8 sd 14 (Minggu ke-2) 3 = Tanggal 15 sd 21 (Minggu ke-3) 4 = Tanggal > 21 (Minggu ke-4)
2	Jam Trx	Char(8)	Jam 00 sd 23
3	Merchant Code	Char(4)	1 = 0001 - 1499 (Agricultural Services) 2 = 1500 - 2999 (Contracted Services) 3 = 3000 - 3299 (Airlines) 4 = 3300 - 3499 (Car Rental) 5 = 3500 - 3999 (Lodging) 6 = 4000 - 4799 (Transportation Services) 7 = 4800 - 4999 (Utility Services) 8 = 5000 - 5599 (Retail Outlet Services) 9 = 5600 - 5699 (Clothing Stores) 10 = 5700 - 7299 (Miscellaneous Stores) 11 = 7300 - 7999 (Business Services) 12 = 8000 - 8999 (Professional Services and Membership Organizations) 13 = 9000 - 9999 (Government Services)
4	Nominal Trx	Double	0 = 1 - 500.000 1 = 500.001 - 2.000.000 2 = 2.000.001 - 5.000.000 3 = 5.000.001 - 10.000.000 4 = 10.000.001 - 20.000.000 5 = 20.000.001 - 50.000.000
5	Interchange Fee	Double	0 = 1 - 1.000 1 = 1.001 - 2.000 2 = 2.001 - 5.000 3 = 5.001 - 10.000 4 = 10.001 - 50.000 5 = Lebih Besar 50.000

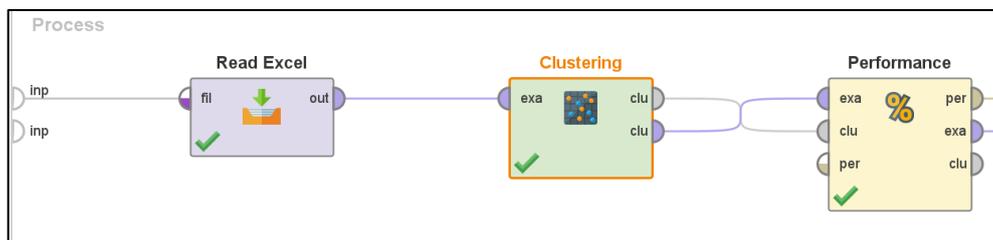
Tabel 1 menunjukkan atribut yang dipilih untuk proses pemodelan. Atribut bernama Tgl Trx T24 adalah tanggal transaksi *Debit Card* dibukukan pada *Core Banking System* T24, tanggal ini ditransformasikan menjadi periode minggu yaitu tanggal 1 sampai dengan 7 menjadi kode '1', tanggal 8 sampai dengan 14 menjadi kode '2', tanggal 15 sampai dengan 21 menjadi kode '3' dan tanggal diatas 21 menjadi kode '4'. Kemudian nama atribut 'Tgl Trx T24' diubah menjadi 'Minggu ke'. Jam Trx adalah jam transaksi dilakukan, data ini ditransformasi menjadi 2 angka (diambil jam saja, menit dan detik di abaikan). *Merchant Code* adalah *Merchant Category Codes* (MCC), kode ini dilakukan tranformasi sesuai kategori dari *Merchant*. MCC adalah empat digit angka yang digunakan penerbit kartu kredit (*Card Issuer*) untuk mengklasifikasikan pembelian individu menurut jenis pedagang tempat pembelian tersebut dilakukan. Kode ini digunakan untuk berbagai tujuan, termasuk perhitungan MDR (*Merchant Discount Rate*) dan penerbitan hadiah kartu kredit. Nominal Trx adalah nominal atau nilai transaksi, nilai ini ditransformasikan seperti pada tabel 1 diatas. Sebagai contoh, nilai transaksi Rp.1.500.000 akan ditransformasikan menjadi kode '2' dan nilai transaksi Rp.7.000.000 akan ditransformasikan menjadi kode '3'. Interchange Fee adalah biaya transaksi, nilai biaya ini akan ditransformasikan seperti Tabel 1. Sebagai contoh, biaya Rp.7.500 akan ditransformasikan menjadi kode '3'.

column index	attribute meta data information	column selected	type	attribute
0	Minggu ke	<input checked="" type="checkbox"/>	integer	attribute
1	Jam Trx	<input checked="" type="checkbox"/>	integer	attribute
2	Merchant Code	<input checked="" type="checkbox"/>	integer	attribute
3	Nominal Trx	<input checked="" type="checkbox"/>	integer	attribute
4	Interchange Fee	<input checked="" type="checkbox"/>	integer	attribute

Gambar 3. Struktur data pada *RapidMiner*

Data transaksi *Debit Card* yang di-*upload* ke dalam aplikasi *RapidMiner* adalah data yang telah diolah dan ditransformasikan dalam format file excel (*.xls). Pada gambar 3 terlihat hasil pembacaan data dari file excel ke dalam aplikasi *Rapid Miner* menggunakan *Operator Read Excel*. Pada atribut ‘Minggu ke’ type datanya adalah *Integer*, atribut ‘Jam Trx’ type datanya adalah *Integer*, ‘Merchant Code’ type datanya adalah *integer*, *Nominal Trx* type datanya adalah *integer* dan *Interchange Fee* type datanya adalah *integer*. Seluruh type data ini yaitu *Integer* sudah sesuai dan dapat digunakan untuk pemrosesan menggunakan aplikasi *Rapid Miner*.

Pada gambar 4 menunjukkan penggunaan *Operator Cluster Distance Performance* untuk menghitung nilai *Average within centroid distance* pada *Rapid Miner*. Nilai ini digunakan untuk menentukan nilai *k* yang paling optimal pada algoritma *K-Means*.



Gambar 4. Algoritma *K-Means Clustering* dan *Operator Distance Performance* pada *RapidMiner*

Gambar 4 menunjukkan diagram alir yang menggambarkan proses pengelompokan data menggunakan algoritma *K-means Clustering* menggunakan aplikasi *Rapid Miner* dan penggunaan *Operator Distance Performance* untuk mengukur jarak antara *data point* dan *centroid cluster*. Diagram alir dimulai dengan data input, yang menggunakan file Excel. Data input kemudian dibaca dan diproses untuk persiapan pengelompokan. Langkah selanjutnya adalah menentukan nilai *k*, yang merupakan jumlah cluster yang diinginkan. Nilai *k* dapat ditentukan berdasarkan pengetahuan domain, analisis data awal, atau metode lain. Setelah nilai *k* ditentukan, algoritma *K-Means* akan menginisialisasi *centroid cluster*. *Centroid cluster* adalah titik pusat setiap *cluster*. *Centroid cluster* dapat diinisialisasi secara acak atau berdasarkan metode lain. Hasil akhir dari algoritma *K-Means* adalah data yang dikelompokkan menjadi *k cluster*. Setiap *cluster* berisi data yang serupa satu sama lain berdasarkan kriteria yang ditentukan oleh algoritma.

Pada penelitian ini, dilakukan proses dengan *Operator Cluster Distance Performance* untuk nilai *k=2* sampai dengan *k=8* dengan hasil evaluasi terlihat pada Tabel 2. *Operator Cluster Distance Performance* akan menghitung *Average Distance* untuk semua nilai *k* yang diberikan secara berulang-ulang. Angka-angka yang dihasilkan tersebut kemudian dicatat dan dimasukkan ke dalam tabel. Operator ini menghitung rata-rata jarak antara setiap data point dan centroid

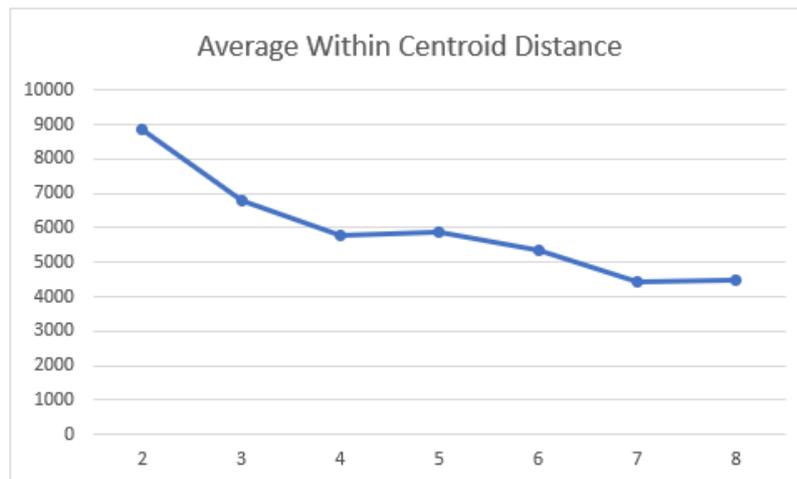
clusternya. Nilai yang lebih rendah menunjukkan *cluster* yang lebih rapat, menandakan kinerja *clustering* yang lebih baik.

Tabel 2. Nilai *Average Within Centroid Distance*

<i>k</i>	Average Distance
2	8.862
3	6.791
4	5.764
5	5.888
6	5.345
7	4.445
8	4.480

Tabel 2 menunjukkan nilai *Average Distance* tertinggi dicapai pada nilai $k=2$, yaitu sebesar 8.862. Hal ini menunjukkan bahwa ketika data dikelompokkan menjadi dua kelompok ($k=2$), jarak antar data dalam setiap kelompoknya paling besar dibandingkan dengan nilai k lainnya. Selanjutnya, nilai *Average Distance* terus menurun seiring dengan peningkatan nilai k . Pada nilai $k=3$, nilai *Average Distance* mencapai 6.791, diikuti dengan nilai sebesar 5.764 pada $k=4$, kemudian sedikit naik nilainya yaitu sebesar 5.888 pada $k=5$, nilai sebesar 5.345 pada $k=6$, nilai sebesar 4.445 pada $k=7$, dan nilai sebesar 4.480 pada $k=8$. Penurunan nilai *Average Distance* ini menunjukkan bahwa semakin banyak kelompok yang dibuat (nilai k semakin besar), semakin kecil jarak antar data dalam setiap kelompoknya.

Berdasarkan data pada tabel 2, kemudian dibuat grafik menggunakan *Microsoft Excel*. Grafik ini seperti terlihat pada gambar 5 dibawah ini. Gambar tersebut menggambarkan titik belok yang akan menentukan nilai k menggunakan *Elbow Method*. *Elbow method* adalah metode yang sering dipakai untuk menentukan jumlah cluster yang akan digunakan pada *K-Means Clustering*. Metode ini dapat menentukan nilai k yang paling ideal atau efektif berdasarkan data *Average Distance*.

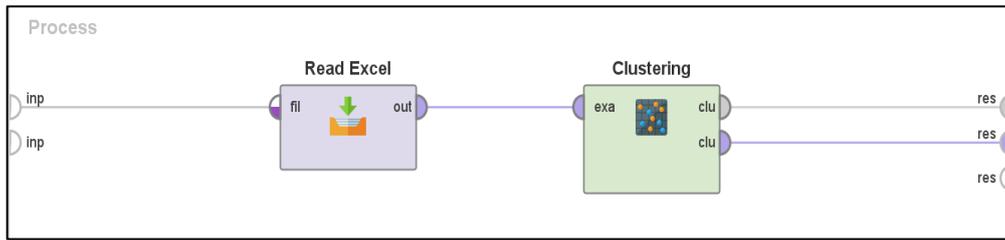


Gambar 5. Hasil Kurva *Elbow Method*

Gambar 5 menunjukkan grafik yang menggambarkan hubungan antara nilai k dan nilai *Average Within Centroid Distance*. Titik belok pada grafik menunjukkan nilai k yang optimal, yaitu $k = 4$. Hal ini dikarenakan nilai *average within centroid distance* turun mencapai titik belok dan setelah mencapai titik belok, nilai tidak berubah secara signifikan pada nilai k selanjutnya.

Berdasarkan nilai $k=4$ dilakukan kembali *Clustering* dengan algoritma *K-Means* menggunakan aplikasi *Rapid Miner* dan menghasilkan data per *Cluster* (*Cluster Model*) sebagai berikut:

Cluster 0 : 777 items Cluster 2 : 104 items Jumlah : 2.209 items
 Cluster 1 : 815 items Cluster 3 : 513 items



Gambar 6. Algoritma *K-Means Clustering* pada *RapidMiner*

Gambar 6 menunjukkan diagram alir yang menggambarkan proses pengelompokan data menggunakan algoritma *K-means Clustering* menggunakan aplikasi *Rapid Miner*. Proses ini dilakukan untuk $k=4$ dikarenakan nilai ini yang paling signifikan berdasarkan analisa *Elbow Method* pada pembahasan sebelumnya. Proses pengelompokan data dalam penelitian ini menghasilkan empat *cluster*, yaitu *Cluster 0*, *Cluster 1*, *Cluster 2* dan *Cluster 3*. *Cluster 0* memiliki jumlah data sebanyak 777 items. *Cluster 1* memiliki jumlah data paling banyak yaitu 815 items. *Cluster 2* memiliki jumlah data paling sedikit, yaitu sebesar 104 items dan *Cluster 3* memiliki jumlah data sebesar 513 items.

Tabel 3. *Centroid Table*

Attribute	Cluster_0	Cluster_1	Cluster_2	Cluster_3
Minggu ke	2.476	2.432	2.183	2.505
Jam Trx	14.547	18.947	6.442	10.838
Merchant Code	9.364	9.156	8.740	9.117
Nominal Trx	0.291	0.218	0.212	0.246
Interchange Fee	0.707	0.599	0.452	0.608

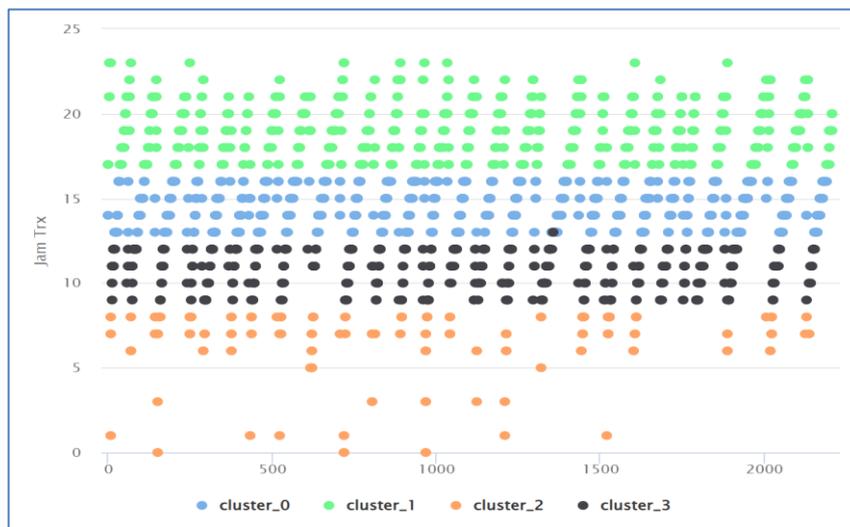
Tabel 3 menunjukkan nilai *centroid* untuk empat *cluster* yang berbeda. *Cluster 2* memiliki nilai *centroid* terendah untuk semua atribut, sedangkan *cluster 1* memiliki nilai *centroid* tertinggi. *Cluster 0* dan *Cluster 3* berada di antara *cluster 1* dan 2. Tabel *centroid* diatas juga dapat digunakan oleh Manajemen Bank Capital untuk analisis data transaksi kartu debit. Misalnya, tabel ini dapat digunakan untuk mengidentifikasi pola transaksi yang berbeda. Pada *Cluster 0* dapat mewakili transaksi yang dilakukan oleh nasabah dengan nilai nominal transaksi yang paling besar, *Cluster 3* dapat mewakili transaksi yang dilakukan oleh nasabah dengan nilai nominal transaksi sedang, serta *Cluster 1* dan *Cluster 2* dapat mewakili transaksi yang dilakukan oleh nasabah dengan nilai nominal transaksi kecil.

Penggunaan algoritma K-Means dalam penelitian ini bukan hanya untuk membagi data menjadi beberapa kelompok, tetapi juga untuk menyediakan analisis yang objektif dan terukur terkait pola transaksi nasabah. Metode ini memungkinkan identifikasi pola-pola tersembunyi yang sulit diungkapkan secara manual, seperti melalui atribut jam transaksi, kode merchant, dan nominal transaksi. Selain itu, algoritma ini didukung oleh evaluasi objektif menggunakan metode Elbow untuk menentukan jumlah cluster optimal dan metrik average within centroid distance yang memastikan kualitas pengelompokan. Pendekatan ini memberikan validasi tambahan yang tidak dapat dicapai hanya dengan membaca data secara manual dari aplikasi seperti Excel, sehingga hasil yang diperoleh lebih akurat dan dapat langsung diterapkan dalam strategi pengelolaan nasabah.

Dengan menggunakan algoritma Clustering K-Means, Bank Capital dapat melakukan pengelompokan data transaksi berdasarkan tanggal dan jam transaksi. Sebagai contoh, hasil pengelompokan menunjukkan bahwa cluster dengan nominal transaksi besar (*Cluster 0*) cenderung terjadi pada jam 14:00 hingga 16:00, sedangkan cluster dengan nominal kecil (*Cluster*

2) lebih sering terjadi pada pagi hari. Informasi ini memberikan wawasan tambahan mengenai kebiasaan transaksi nasabah pada waktu-waktu tertentu, seperti waktu puncak transaksi dan preferensi nasabah berdasarkan nominal transaksi. Wawasan ini dapat digunakan oleh manajemen Bank Capital untuk mengoptimalkan alokasi dana atau menawarkan produk yang lebih sesuai dengan kebutuhan nasabah berdasarkan pola transaksi mereka.

Selain hal di atas, tabel *centroid* ini juga dapat digunakan untuk memprediksi perilaku nasabah di masa depan. Misalnya, jika nasabah baru berada di *Cluster 0*, maka dapat diprediksi bahwa nasabah tersebut akan melakukan transaksi dengan nominal yang besar atau menghabiskan banyak uang di masa depan (Rizky Wijaya & Satriyo Wibowo, 2021). Hal ini dikarenakan pada *cluster 0* ini nasabah cenderung bertransaksi dengan nominal transaksi yang cukup besar dibandingkan dengan *cluster* yang lainnya. Selain itu dapat juga diketahui pola transaksi yang dilakukan nasabah berdasarkan jam transaksi.



Gambar 7. Visualisasi menggunakan *Plot Type Scatter/Bubble*

Penelitian ini memanfaatkan visualisasi data menggunakan *plot type scatter/bubble* seperti terlihat pada Gambar 7 untuk meningkatkan aksesibilitas dan pemahaman terhadap data. *Plot type scatter/bubble* pada aplikasi *Rapid Miner* merupakan salah satu metode visualisasi data yang digunakan untuk menampilkan data multidimensi dengan cara memetakan nilai dua variabel ke dalam dua dimensi (x dan y) pada sebuah *plot*.

KESIMPULAN

Analisis data transaksi debit Bank Capital dengan algoritma K-Means menunjukkan bahwa pengelompokan data optimal tercapai pada $k = 4$, berdasarkan nilai rata-rata jarak antar data dalam cluster (average within centroid distance) terendah, yaitu 5.764. Pengelompokan ini mengungkap pola-pola transaksi yang signifikan, seperti transaksi dengan nominal besar yang lebih sering terjadi pada jam 14:00 hingga 16:00, serta transaksi dengan nominal kecil yang lebih dominan pada pagi hari. Informasi ini memberikan wawasan yang dapat membantu manajemen Bank Capital dalam merancang strategi yang lebih tepat sasaran, seperti pengelolaan dana dan pengembangan layanan berdasarkan pola waktu dan nominal transaksi nasabah. Dengan demikian, hasil penelitian ini menunjukkan bahwa algoritma K-Means tidak hanya mampu mengelompokkan data, tetapi juga memberikan landasan bagi pengambilan keputusan yang berbasis data di sektor perbankan.

SARAN

Saran dalam penelitian ini mencakup beberapa keterbatasan yang dapat diperbaiki di masa depan. Penelitian ini belum mencakup pengelompokan nasabah berdasarkan bidang pekerjaan atau tingkat penghasilan, yang dapat memberikan wawasan lebih dalam mengenai perilaku transaksi. Selain itu, periode data yang digunakan terbatas pada satu bulan, sehingga belum sepenuhnya mencerminkan pola transaksi yang terjadi pada momen-momen tertentu, seperti hari raya, masa pendaftaran sekolah, atau momen penting lainnya. Untuk penelitian selanjutnya, disarankan menggunakan data transaksi dengan cakupan waktu yang lebih panjang, seperti satu tahun penuh, guna memperoleh gambaran yang lebih komprehensif tentang pola transaksi nasabah.

DAFTAR PUSTAKA

- Achir Ainun Yaumil, & Kusumaningrum Trias Madanika. (2021). Pengaruh penggunaan debit card, credit card, e-money, dan e-wallet terhadap pengeluaran konsumsi mahasiswa. *Jurnal Manajemen*, 13, 554–568.
- Al Indiani, N., Santoso, K. A., Anggraeni, D., Matematika, J., Mipa, F., & Jember Jl Kalimantan, U. (2022). Penentuan Lokasi Atm Bank Syariah Indonesia Di Wilayah Jember Kota Menggunakan K-Means Clustering (*Determining The Location of Bank Syariah Indonesia ATM In Jember County Using K-Means Clustering*). *SeNa-MaGeStiK 2022*, 75-84.
- Ardhianto, A., Relawanto, B., & Wibowo, A. (2021). Algoritme K-Means dalam Pengelompokan Kantor Cabang untuk Optimalisasi Manajemen Perbankan. *Jurnal Telematika*, Vol. 15 No.2, 71-76.
- A'yuni, Q., Nazir, A., Handayani, L., & Afrianty, I. (2023). Penerapan Algoritma K-Means Clustering untuk Mengetahui Pola Penerima Beasiswa Bank Indonesia (BI). *Journal of Computer System and Informatics (JoSYC)*, 4(3), 530–539. <https://doi.org/10.47065/josyc.v4i3.3343>
- Bilal, M., Rasyid, A., & Rasyid, A. (2022). Pengaruh Penggunaan Debit Card Dan E-Money Terhadap Perilaku Konsumtif. *Jurnal Ilmu Manajemen Saburai*, 8,111-122.
- Budiansyah Hasibuan, R., & Mahyuni, R. (2022). Penerapan Data Mining Clustering Dengan Menggunakan Algoritma K-Means Pada Data Nasabah Kredit Bermasalah PT. BPR Milala. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 5(1), 7–15. Retrieved from <https://ojs.trigunadharma.ac.id/>
- Defina Fatimah Setiti Alhamdani, Dianti Ananda Ayu, & Azhar Yufis. (2021). Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode K-Means Clustering. In *JISKa* (Vol. 6). 70-77.
- Harahap, B. (2020). Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung). *READY STAR – 2*, 394-403.
- Intan, K., Sari, S., Utami, W., Yudi, N., & Wijaya, A. (2024). Penerapan data mining dalam penentuan strategi marketing menggunakan algoritma k-means (studi kasus: pt bpr hoki). *Jurnal Mahasiswa Teknik Informatika* (Vol. 8), 535-542.

- Nyoman, I., & Guna, A. (2023). Analisis Pengelompokan Bank Umum Konvensional Di Indonesia Dengan Metode K-Means Clustering. Retrieved from <http://journal.unasmataram.ac.id/index.php/GARA>, 17(1), 93-99.
- Pradana, A., & Pratama, D. (2022). pengaruh penggunaan kartu debit dan uang elektronik (e-money) terhadap pengeluaran konsumsi mahasiswa. *Jurnal Ekonomi Islam*, (Vol. 22), 41-52.
- Purnama, V., & Arianto, D. B. (2024). Penerapan k-means clustering pada data pembayaran tagihan kartu kredit untuk menganalisis potensi fraud. *Journal of Information System and Computer*, 4(2), 52-57. Retrieved from <https://journal.unisnu.ac.id/JISTER/>
- Rizky Wijaya, M., & Satriyo Wibowo, G. (2021). Prosiding Seminar Nasional Universitas Ma Chung Customer Segmentation berdasarkan Usia, Jumlah Kredit dan Lama Kredit Nasabah di Bank XYZ menggunakan Model K-Means Clustering. *Prosiding Seminar Nasional Universitas Ma Chung, 2021*, 101–116. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29>
- Rohman, N., & Wibowo, A. (2024). Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall. *SINTECH Journal*, 7, 49–58. Retrieved from <https://doi.org/10.31598>
- Rojih, P., & Heikal, J. (2023). Clustering Pelanggan Menggunakan Algoritma K-Means Clustering di 7 MCC (Industry) Credit Card PT Bank CIMB Niaga. *Jurnal Penelitian Dan Pengabdian Masyarakat*, 02, 1924–1934. <https://doi.org/10.36418/comserva.v2i09.524>
- Sunata, H., Azrullah, F. J., & Rianto, Y. (2020). *Komparasi Tujuh Algoritma Identifikasi Fraud ATM Pada PT. Bank Central Asia Tbk* .7(3). 441-450. Retrieved from <http://jurnal.mdp.ac.id>
- Suwirya, I. P., Candiasa, I. M., & Dantes, G. R. (2022). Evaluation of ATM Location Placement Using the K-Means Clustering in BNI Denpasar Regional Office. *Journal of Computer Networks, Architecture and High Performance Computing*, 4(2), 158–168. <https://doi.org/10.47709/cnahpc.v4i2.1580>
- Tanu Wijaya Abe, & Subandi. (2024). Penerapan Metode Clustering Dengan Algoritma K-Means Pada Sistem Pendeteksi Pencucian Uang Perbankan Berbasis Web, *SENAFTI*, 3(2), 398-406.
- Wardhana Muhammad Rizki, Triayudi Agung, & Hayati Nur. (2021). Analisis Faktor Calon Nasabah PT.Bank Central Asia dalam Pembuatan Rekening Onlinemenggunakan Metode K-Means Clustering Studi Kasus Wisma Asia BCA. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 85–92.
- Wibowo, A., & Sasongko, R. (2022). Penerapan Data Mining Pada Suku Bunga Investasi Deposito Di Indonesia Menggunakan Metode K-Means Clustering Untuk Pengelompokan Profitabilitas (*Application of Data Mining on Deposit Investment Rate in Indonesia Using K-Means Clustering Method for Profitability Clustering*). *ASCARYA*, 2(1). <https://doi.org/10.53754/iscs.v2i1.369>