

# Comparison of Hierarchical, K-Means and DBSCAN Clustering Methods for Credit Card Customer Segmentation Analysis Based on Expenditure Level

Hafid Ramadhan <sup>1\*</sup>, Mohammad Rizal Abdan Kamaludin <sup>2\*</sup>, Muhammad Alfian Nasrullah <sup>3\*</sup>, Dwi Rolliawati <sup>4\*</sup>  
\* Sistem Informasi, UIN Sunan Ampel Surabaya

[h96219046@student.uinsby.ac.id](mailto:h96219046@student.uinsby.ac.id) <sup>1</sup>, [h76219026@student.uinsby.ac.id](mailto:h76219026@student.uinsby.ac.id) <sup>2</sup>, [h06219008@student.uinsby.ac.id](mailto:h06219008@student.uinsby.ac.id) <sup>3</sup>, [dwi\\_roll@uinsby.ac.id](mailto:dwi_roll@uinsby.ac.id) <sup>4</sup>,

## Article Info

### Article history:

Received 2023-06-30

Revised 2023-08-16

Accepted 2023-09-26

### Keyword:

Clustering,  
Credit Card,  
Comparison,  
Segmentation,  
Silhouette Coefficient.

## ABSTRACT

The amount of data from credit card users is increasing from year to year. Credit cards are an important need for people to make payments. The increasing number of credit card users is because it is considered more effective and efficient. The third method used today has a function to determine the effective outcome of credit card user scenarios. In this study, a comparison was made using the Hierarchical Clustering, K-Means and DBSCAN methods to determine the results of credit card customer segmentation analysis to be used as a market strategy. The results obtained based on the best silhouette coefficient score method is two cluster hierarchical clustering with 0.82322 score. Based on the best mean value customers are divided into two segments, and it is suggested to develop strategies for both segments.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Kartu kredit adalah salah satu inovasi digital era modern. Dengan menggunakan kartu kredit seseorang dapat melakukan transaksi pembayaran dengan mudah tanpa harus bertemu langsung dengan orang lain. Kartu kredit merupakan salah satu alat pembayaran yang digunakan sebagai pengganti uang tunai, yang kapan saja dapat ditukar dengan apapun yang diinginkan dimana ada cabang yang dapat menerima kartu kredit dari bank ataupun perusahaan yang mengeluarkannya [1]. Saat ini sudah banyak pengguna dari kartu kredit mulai dari para remaja hingga dewasa, untuk kegiatan sehari-hari hingga dalam urusan bisnis.

Penelitian ini akan menggunakan studi kasus kartu kredit. Kartu kredit menjadi faktor penting dari beberapa perusahaan untuk melakukan pembayaran, seperti pada karyawan dan mitra kerja mereka. Permasalahan dari banyaknya kartu kredit dapat berpengaruh terhadap strategi perusahaan atau pasar. Data yang diambil untuk studi kasus kartu kredit adalah data dari situs kumpulan data yaitu Kaggle. Data merupakan kumpulan fakta kasar mengenai orang, tempat, kejadian dan sesuatu yang penting yang didapatkan melalui pengamatan ataupun hasil pencarian di berbagai sumber. Dengan adanya data manusia dapat memperoleh dan mengelola informasi

agar dapat disampaikan kepada orang lain secara pasti. Pada perkembangannya, data menjadi faktor penting dari masa ke masa. Banyaknya data membuat manusia kesulitan dalam melakukan pengelolaan. Data mining atau penggalian data menjadi solusi untuk manusia melakukan ekstraksi terhadap database yang besar dengan menggunakan beberapa cara.

Pengguna kartu kredit sendiri telah mengalami peningkatan selama satu tahun kebelakang per bulan Juni kemarin. Dilansir dari website dataindonesia.id Bank Indonesia (BI) mencatat, jumlah kartu kredit yang beredar di Indonesia mencapai 16,58 juta unit pada Juni 2022 [2]. Jumlah tersebut naik 0,84% dibandingkan pada Juni 2021 yang sebesar 16.56 juta kartu kredit. Kondisi ini terjadi karena banyaknya masyarakat yang beralih ke pembayaran elektronik. Alasannya adalah uang elektronik dirasa lebih aman dari penularan Covid-19 karena minim kontak langsung, efisien, dan banyak promo. Nilai transaksi kartu kredit telah mencapai Rp149,18 miliar pada bulan keenam tahun ini. Jumlahnya naik 34,35% dibandingkan setahun sebelumnya yang sebesar Rp.19,81 triliun. Sedangkan, volume transaksi dari kartu kredit tercatat sebesar 27,93 juta kali pada Juni 2022. Jumlah itu meningkat sebesar 20,26%

dibandingkan periode yang sama tahun sebelumnya sebesar 23,22 juta kali. Dengan semakin banyaknya pengguna maka semakin banyak juga data yang ada dan harus dikelola.

Maka dari itu data mining menjadi hal yang penting untuk saat ini. Dalam penelitian sebelumnya [3], telah dilakukan pengolahan data penggunaan kartu kredit menggunakan metode K-Means Clustering dengan perbandingan tiga metode clustering lain meliputi Agglomerative Clustering, GMM dan DBSCAN. Dalam penelitian tersebut metode clustering dengan hasil terbaik dicapai K-Means dengan nilai silhouette score sebesar 0.207014. Hasil capaian silhouette score tiga metode lainnya tidak begitu baik, dengan hasil silhouette score paling rendah senilai -0.351371 pada metode DBSCAN. Hasil silhouette score tersebut dirasa kurang dan memiliki potensi untuk ditingkatkan dengan penyesuaian beberapa parameter.

Hierarchical clustering memiliki keunggulan dalam Penelitian ini menggunakan metode Clustering Hierarchical, K-Means dan DBSCAN. Metode ini digunakan karena sesuai dengan tujuan dari penelitian yaitu untuk melakukan segmentasi pelanggan kartu kredit berdasarkan tingkat pengeluaran. Penelitian ini juga akan melakukan perbandingan metode Hierarchical Clustering, K-Means dan DBSCAN untuk mengetahui metode mana yang paling sesuai dengan nilai silhouette score tertinggi dalam clustering data kartu kredit.

**II. METODE PENELITIAN**

Proses penelitian ini dimulai dengan pengumpulan data yang akan diolah. Dataset diambil dari sumber kaggle untuk selanjutnya diolah dengan data preprocessing. Setelah dilakukan preprocessing data kemudian dilanjutkan dengan memodelkan metode clustering menggunakan aplikasi KNIME. Setelah dilakukan pemodelan metode clustering dilakukan scoring untuk menilai keefektifan metode clustering dan mendapatkan nilai yang paling optimal.

**A. Sumber Data**

Data yang digunakan pada penelitian ini merupakan data sekunder yang diunduh dari situs kumpulan data bernama Kaggle. Data terpilih adalah dataset kartu kredit unggahan akun bernama Arjun Bhasin yang diunggah pada tahun 2017 [4]. Dataset ini memiliki record data sejumlah 8950 baris, dalam 18 kolom field data dictionary. 18 kolom field data dictionary dalam dataset kartu kredit dalam Tabel 1.

TABEL 1  
FIELD DATA DICTIONARY

No.	Nama Field	Deskripsi
1	CUSTID	Identitas pemegang kartu kredit
2	BALANCE	Jumlah saldo yang tersisa di akun untuk melakukan pembelian
3	BALANCEFREQUENCY	Seberapa banyak saldo diperbarui, skor antara 0 dan 1 (1 = sering diperbarui, 0 = tidak sering diperbarui)
4	PURCHASES	Jumlah dari pembelian yang dilakukan akun
5	ONEOFFPURCH	Jumlah dari pembelian maksimum

	ASES	dilakukan dalam sekali jalan
6	INSTALLMENTPURCHASES	Jumlah dari pembelian dilakukan secara cicilan
7	CASHADVANCE	Uang muka diberikan oleh pemegang kartu kredit
8	PURCHASEFREQUENCY	Seberapa sering pembelian dilakukan, skor antara 0 dan 1 (1 = sering dibeli, 0 = tidak sering dibeli)
9	ONEOFFPURCHASEFREQUENCY	Seberapa sering pembelian terjadi dalam sekali jalan (1 = sering dibeli, 0 = tidak sering dibeli)
10	PURCHASEINSTALLMENTSFREQUENCY	Seberapa sering pembelian dengan mencicil dilakukan (1 = sering dilakukan, 0 = tidak sering dilakukan)
11	CASHADVANCEFREQUENCY	Seberapa sering uang muka dibayarkan
12	CASHADVANCE	Jumlah transaksi yang dilakukan dengan "Cash in Advanced"
13	PURCHASESTRICT	Jumlah transaksi pembelian yang dilakukan
14	CREDITLIMIT	Limit Kartu Kredit untuk pemegang kartu
15	PAYMENTS	Jumlah Pembayaran yang dilakukan oleh pengguna
16	MINIMUM_PAYMENTS	Jumlah minimum pembayaran yang dilakukan pengguna
17	PRCFULLPAYMENT	Persentase pembayaran penuh yang dibayarkan pengguna
18	TENURE	Jangka waktu layanan kartu kredit untuk pengguna

**B. Data Preprocessing**

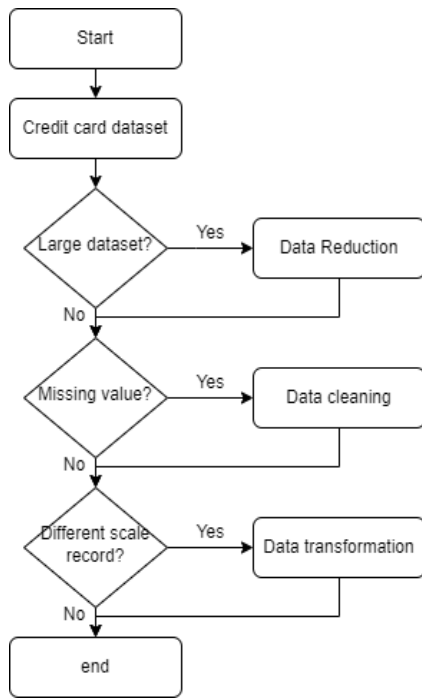
Data kartu kredit yang telah diperoleh kemudian dilakukan data preprocessing. Pada tahap preprocessing dilakukan penyiapan dokumen mentah menjadi dokumen atau representatif dokumen yang siap untuk diproses untuk langkah selanjutnya [5]. Dalam data preprocessing terdapat beberapa tahapan kerja, data yang digunakan pada penelitian ini perlu dilakukan beberapa tahap kerja preprocessing berikut pada gambar 1.

Tahap preprocessing pertama yang dilakukan pada dataset kartu kredit adalah reduction. Data reduction adalah suatu proses yang dilakukan untuk menghapus data pada attribute yang kurang dominan sehingga data dapat dikurangi, tetapi tetap menghasilkan data yang akurat [6]. Dikarenakan keterbatasan hardware untuk menjalankan iterasi pada beberapa metode, data yang dipilih adalah data dimulai pada baris pertama hingga baris ke-4475 atau setengah dari keseluruhan dataset kartu kredit.

Tahap berikutnya setelah didapatkan jumlah data yang sesuai adalah data cleaning. Cleaning adalah proses menghilangkan noise dan data yang tidak relevan atau inkonsisten disebut pembersihan data [7]. Dalam data yang digunakan dalam penelitian ini terdapat 151 value pada field "Minimum\_Payments" yang tidak memiliki nilai. Sejumlah value tidak bernilai tersebut diisi dengan metode mean. Metode mean dalam mengisi value yang tidak memiliki nilai menggunakan persamaan (1).

$$\bar{x} = \frac{\sum x_i}{n} \dots \dots \dots (1)$$

Keterangan:  
 $\bar{x}$  = nilai rata-rata  
 $x_i$  = nilai data ke-i  
 $n$  = banyaknya data



Gambar 1. Flowchart data preprocessing

Tahap kerja data preprocessing terakhir yang dilakukan pada data penelitian ini adalah data transformation. Pada tahap ini akan dilakukan normalisasi data, normalisasi merupakan proses dimana sebuah atribut numerik dirubah ke dalam skala yang ditentukan dengan jangkauan yang lebih kecil misalnya -1.0 sampai 1.0, atau 0.0 sampai 1.0 [8]. Normalisasi data diaplikasikan pada data kartu kredit dalam penelitian ini menggunakan teknik statistika “Z-Score Normalization” dengan persamaan (2).

$$Z = \frac{(x-\mu)}{\sigma} \dots\dots\dots(1)$$

Keterangan:  
 Z = z-score  
 x = nilai yang diuji  
 μ = rata-rata  
 σ = standar deviasi

C. Pemodelan

Proses pemodelan dari metode clustering data mining dalam penelitian ini menggunakan bantuan software KNIME. KNIME merupakan software yang memuat tools dalam mempelajari visualisasi data, machine learning, data mining dan sebagainya [9]. Dalam software KNIME metode clustering dapat dimodelkan dalam berbagai konfigurasi menggunakan node-node yang disediakan dalam dashboard software [10]. Setiap node memiliki berbagai macam kegunaan sesuai dengan yang dibutuhkan. Salah satu node utama adalah input yang berguna untuk memasukan dataset yang akan dimodelkan. Terdapat berbagai varian inputan dari csv serta input file reader untuk file secara umum.

Dalam pemodelan clustering, node yang dibutuhkan adalah file reader untuk membaca inputan dataset. Kemudian terdapat node missing value yang berfungsi untuk menambahkan data pada dataset yang kosong datanya. Selanjutnya terdapat node-node clustering mulai dari hierarchical, K-Means, DBSCAN yang berfungsi untuk mengolah data dengan metode-metode yang digunakan. Untuk memvisualisasikan datanya digunakan node scatter plot untuk melihat hasil sebaran data dari hasil olahan metode clustering.

1) Hierarchical Clustering

Hierarchical clustering merupakan metode analisis kelompok yang berusaha untuk membangun sebuah hierarki kelompok [11]. Mengelompokkan dua atau lebih objek yang memiliki nilai kemiripan yang paling dekat merupakan salah satu teknik yang dapat digunakan untuk clustering [12]. Setiap data dengan kemiripan akan memiliki hubungan hierarki dekat yang berakibat pada pembentukan cluster data. Bagan hierarki akan terus berbentuk hingga data saling terhubung dalam bagan hierarki secara keseluruhan [13].

Hierarchical clustering memiliki beberapa metode yang dapat digunakan, diantaranya: Average Linkage; Single Linkage; dan Complete Linkage Methods. Dalam penelitian ini metode average linkage atau juga disebut group average (GA) digunakan. Dalam metode group average, pencarian pasangan ditentukan dengan melihat rerata setiap nilai observasinya [14]. Perhitungan metode group average dapat diaplikasikan menggunakan persamaan (3) [15].

$$d_{(AB)C} = \frac{\sum_i \sum_j d_{ij}}{N_{AB}N_C} \dots\dots\dots(3)$$

Keterangan:  
 dij = jarak antar objek i dalam cluster (AB) dan objek j dalam cluster C.  
 NAB = jumlah item pada cluster AB  
 NC = jumlah item pada cluster C

Untuk visualisasi data hasil metode hierarchical clustering digunakan dendrogram yang didapat dari similarity matrix. Similarity matrix berisi tingkat kemiripan antar data yang diperoleh berdasarkan perhitungan jarak. Terdapat beberapa metode untuk menghitung jarak seperti manhattan distance, euclidean distance, mahalanobis distance dan sebagainya [16]. Penelitian ini menggunakan manhattan distance yang dapat dihitung menggunakan persamaan (4) [15].

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \dots\dots\dots(4)$$

Keterangan:  
 d = jarak antara x dan y  
 x = data pusat cluster  
 y = data pada atribut  
 n = jumlah data  
 i = setiap data  
 xi = data pada pusat cluster ke i  
 yi = data pada setiap data ke i

2) *K-Means Clustering*

K-Means Clustering adalah metode untuk mengelompokkan kumpulan objek sesuai atribut yang sama ke dalam beberapa grup. Ini menjelaskan suatu cluster oleh massa yang mewakili mean dari cluster. Pemodelan model clustering K-Means pada KNIME dimulai dengan node Row Filter untuk mereduksi data, setelah itu menggunakan node Normalizer untuk mentransformasi data, dilanjutkan dengan node K-Means untuk melakukan perhitungan persamaan K-Means. Setelah menggunakan node K-Means, digunakan serangkaian node (color manager, shape manager dan scatter plot) untuk menampilkan dendrogram hasil clustering dan node Silhouette Coefficient yang digunakan untuk mengetahui nilai performa clustering. Berikut ini rumus (persamaan 5) pengalokasian data ke cluster.

$$a_{il} = \begin{cases} 1 & d = \min \\ 0 & \text{lainnya} \end{cases} \{D(x_i, C_1)\} \dots\dots\dots(5)$$

Dimana  $a_{il}$  adalah nilai keanggotaan titik  $x_i$  ke pusat cluster  $C_1$ ,  $d$  adalah jarak terpendek dari data  $x_i$  ke  $K$  cluster setelah dibandingkan, dan  $C_1$  centroid (pusat cluster) ke-1 [17].

3) *DBSCAN Clustering*

Metode clustering menggunakan DBSCAN adalah pengelompokan yang disempurnakan sebagai pengidentifikasian berbasis kepadatan dalam ruang fitur yang dipisahkan oleh kelompok dengan kepadatan rendah [18]. Sehubungan dengan pengelompokan, kepadatan ditentukan oleh kelompok pengamatan yang berada di dalamnya [19]. Beberapa sifat yang diinginkan dari pengelompokan berbasis kepadatan termasuk kemampuan untuk menangani dan mengidentifikasi noise, menemukan cluster dengan bentuk yang berubah-ubah, dan penemuan jumlah cluster secara otomatis [20].

Pada metode clustering DBSCAN dibutuhkan identifikasi numerical distance nilai euclidean pada dataset. Euclidean Distance Matrix (EDM) merupakan metode atau langkah umum yang digunakan untuk mengukur jarak antara dua objek. Metode pengukuran ini sering dijumpai pada machine learning, jaringan sensor nirkabel dan masih banyak lagi [21]. EDM dapat dijabarkan dalam persamaan berikut.

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{in} - x_{jn})^2} \dots\dots\dots(6)$$

Setelah didapatkan nilai euclidean dari dataset selanjutnya adalah masuk perhitungan metode clustering DBSCAN. Untuk metode dan langkah perhitungannya adalah menentukan epsilon serta min poin lalu dimasukkan dalam persamaan jarak euclidean.

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \dots\dots\dots(7)$$

Dimana  $x_{ia}$  adalah sebuah variabel ke - a dari obyek  $i(i=1, \dots, n; a=1, \dots, p)$  lalu  $d_{ij}$  merupakan value dari euclidean distance [22].

D. *Evaluasi*

Setelah beberapa metode clustering yang dilakukan, perlu tahapan untuk menilai performa hasil jumlah cluster dan metode mana yang terbaik. Tahap tersebut merupakan tahap evaluasi, secara garis besar dapat dikatakan bahwa evaluasi merupakan pemberian nilai terhadap kualitas sesuatu [23]. Untuk mendapatkan nominal yang menjadi nilai untuk menentukan jumlah cluster dan metode yang terbaik, digunakan metode silhouette coefficient.

Silhouette coefficient adalah metode evaluasi cluster yang merupakan penggabungan metode cohesion dan separation. Cohesion diukur dengan melakukan perhitungan seluruh objek yang ada pada sebuah cluster dan separation diukur dengan melakukan perhitungan jarak rata-rata dari setiap objek pada sebuah cluster dengan cluster terdekatnya [24]. Silhouette coefficient didapatkan dengan menggunakan rumus persamaan (8).

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots(8)$$

Keterangan:

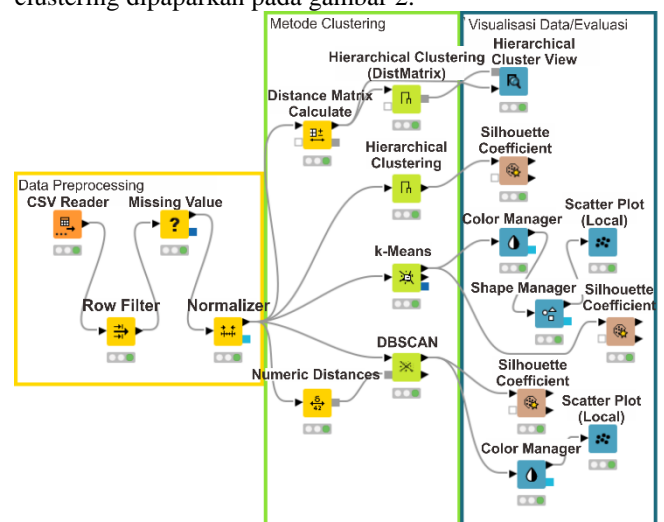
Si = silhouette coefficient

ai = nilai rata-rata objek

bi = nilai minimum rata-rata jarak objek [25]

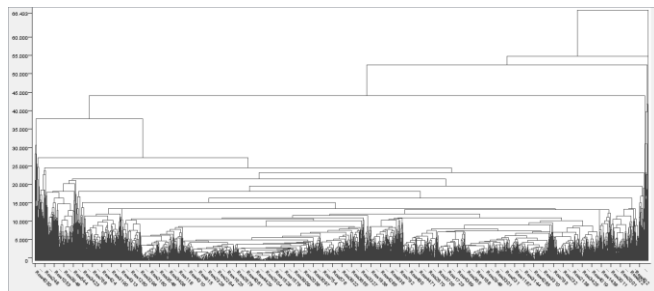
III. HASIL DAN PEMBAHASAN

Pemodelan ketiga metode clustering dimulai hierarchical clustering, K-Means hingga DBSCAN dilakukan dengan software knime. Masing-masing metode memiliki workflow dengan node berbeda sesuai dengan fungsinya. Pemodelan dibagi menjadi tiga fase yakni data preprocessing, metode clustering, dan visualisasi data/evaluasi. Pemodelan metode clustering dipaparkan pada gambar 2.



Gambar 2. Pemodelan metode clustering

Fase data preprocessing menggunakan node dimulai CSV reader kemudian row filter, missing Value, normalizer. Fase metode clustering hierarchichal dibagi menjadi dua untuk menampilkan dendrogram dan menentukan jumlah cluster. Dalam menampilkan dendrogram diperlukan node distance matrix calculate, hierarchichal clustering (distmatrix) dan hierarchichal cluster view. Dendrogram hasil hierarchichal clustering dilampirkan pada gambar 3.



Gambar 3. Dendrogram hierarchichal clustering data kartu kredit

Dendrogram hierarchichal clustering data kartu kredit pada gambar 3 menunjukkan anggota cluster dan jarak kombinasi cluster. Anggota cluster berada pada sumbu x dan jarak kombinasi cluster diskalakan pada sumbu y. Untuk mengetahui anggota cluster mana yang akan terbentuk ditentukan dengan jumlah clade terhitung dari atas dendrogram. Misal apabila dibentuk 2 cluster, maka cluster 1 adalah turunan dari clade terdekat sumbu 0 dan cluster 2 adalah turunan dari clade setara yang lain.

Untuk menentukan jumlah cluster digunakan node hierarchichal clustering. Pengujian hierarchichal clustering dilakukan sebanyak lima kali dengan menentukan jumlah cluster mulai dari dua hingga enam cluster. Untuk mendapatkan nilai performa dari setiap cluster digunakan node silhouette coefficient, didapatkan hasil pada tabel 3 berikut dimana nilai rata-rata keseluruhan silhouette coefficient terbaik dengan 2 cluster.

TABEL 3  
HASIL SILHOUETTE COEFFICIENT HIERARCHICAL CLUSTERING

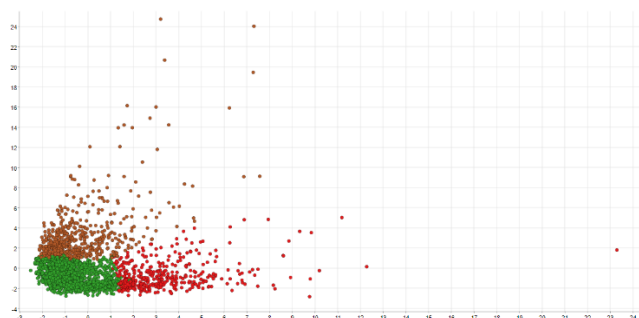
Jumlah Cluster	Overall Mean Silhouette Clustering
2	0.82322
3	0.82267
4	0.74542
5	0.67362
6	0.67267

Fase metode clustering untuk K-Means diperlukan node K-Means untuk menentukan jumlah cluster yang akan dibuat. Pengujian clustering K-Means dilakukan sebanyak lima kali dengan beberapa jumlah cluster yang sama dengan pengujian silhouette coefficient. Dengan node silhouette coefficient didapatkan nilai rata-rata keseluruhan silhouette coefficient yang dipaparkan pada tabel 4.

TABEL 4  
HASIL SILHOUETTE COEFFICIENT K-MEANS

Jumlah Cluster	Overall Mean Silhouette Clustering
2	0.21814
3	0.23577
4	0.20792
5	0.17692
6	0.17442

Pada tabel 4 ditunjukkan hasil performa terbaik metode clustering K-Means ada pada jumlah cluster 3. Untuk visualisasi data pada metode K-Means digunakan tiga node dimulai dari color manager, shape manager dan scatter plot. Pada scatter plot hasil clustering K-Means sumbu x dan y adalah jarak antar objek pada dua dimensi. Scatter plot clustering K-Means dengan jumlah cluster 3 digambarkan pada gambar 4.



Gambar 4. Scatter plot clustering K-Means data kartu kredit

Pemodelan model clustering metode DBSCAN pada KNIME dimulai dengan setelah node “normalizer” menggunakan node “numeric distance” yang menghasilkan nilai euclidean sebagai input node berikutnya. Node DBSCAN berfungsi untuk menjalankan metode DBSCAN berdasarkan input nilai euclidean. Kemudian data hasil node DBSCAN dapat divisualisasikan menggunakan node color manager untuk memberi warna lalu ditampilkan pada node scatter plot. Selanjutnya untuk melakukan validasi digunakan node “silhouette coefficient” untuk validasi dari metode DBSCAN. Setelah node DBSCAN kemudian dihubungkan dengan node entropy scorer yang nantinya akan menampilkan tabel quality. Output dari node “silhouette coefficient” berupa nilai overall mean silhouette coefficient yang tertera pada tabel 5.

TABEL 5  
HASIL SILHOUETTE COEFFICIENT DBSCAN

Jumlah Cluster	Overall Mean Silhouette Clustering
2	0.56872
3	0.34098
4	0.12541
5	0.02379
6	0.01268

Setelah mendapatkan hasil pengujian dari ketiga metode dengan beberapa jumlah cluster, dilakukan perbandingan hasil. Hasil overall mean silhouette coefficient ketiga metode

dibandingkan dan dipilih metode mana dan berapa jumlah cluster yang memiliki performa terbaik. Berikut merupakan tabel 6 hasil silhouette coefficient ketiga metode.

TABEL 6  
HASIL SILHOUETTE COEFFICIENT K-MEANS

Metode Clustering	Overall Mean Silhouette Clustering				
	2 Cluster	3 Cluster	4 Cluster	5 Cluster	6 Cluster
Hierarchical Clustering	0.82322	0.82267	0.74542	0.67362	0.67267
K-Means Clustering	0.21814	0.23577	0.20792	0.17692	0.17442
DBSCAN Clustering	0.56872	0.34098	0.12541	0.02379	0.01268

Tabel 6 hasil rata-rata keseluruhan silhouette coefficient ketiga metode menunjukkan bahwa metode hierarchical clustering memiliki rata-rata relatif lebih tinggi dibanding dua metode lainnya. Metode K-Means memiliki rata-rata nilai lebih rendah dibandingkan dengan dua metode lainnya. Secara keseluruhan jumlah cluster dengan nilai rata-rata tinggi ada pada dua dan tiga cluster. Dari nilai rata-rata didapatkan bahwa metode hierarchical clustering.

#### IV. KESIMPULAN

Telah dilakukan percobaan untuk analisa segmentasi pelanggan kartu kredit berdasarkan tingkat pengeluaran melalui perbandingan ketiga metode yaitu metode Clustering Hierarchical, K-Means dan DBSCAN. Setelah membandingkan ketiga metode yang memiliki fungsi untuk menentukan hasil yang efektif terhadap skenario pengguna kartu kredit. Diketahui bahwa berdasarkan silhouette score metode terbaik adalah Hierarchical dua cluster dengan nilai silhouette score sebesar 0.82322. Jadi dapat disarankan dua strategi untuk dua segmen pelanggan. Dengan begitu perusahaan kartu kredit dapat membuat strategi yang lebih baik untuk pengguna atau pelanggan kartu kredit kedepannya..

#### DAFTAR PUSTAKA

- [1] D. S. Kristianti, "Kartu Kredit Syariah Dan Perilaku Konsumtif Masyarakat," *AHKAM J. Ilmu Syariah*, vol. 14, no. 2, Jul 2014, doi: 10.15408/ajis.v14i2.1287.
- [2] S. Widi, "Jumlah Kartu Kredit Beredar Capai 16,58 Juta pada Juni 2022," *DataIndonesia.id*, 30 Agustus 2022. <https://dataindonesia.id/digital/detail/jumlah-kartu-kredit-beredar-capai-1658-juta-pada-juni-2022> (diakses 26 Desember 2022).
- [3] F. D. S. Alhamdani, A. A. Dianti, dan Y. Azhar, "Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode K-Means Clustering," *JISKA J. Inform. Sunan Kalijaga*, vol. 6, no. 2, Art. no. 2, Mei 2021, doi: 10.14421/jiska.2021.6.2.70-77.
- [4] A. Bhasin, "Credit Card Dataset for Clustering," 2017. <https://www.kaggle.com/datasets/arjunbhasin2013/ccdata> (diakses 5 Desember 2022).
- [5] B. Zaman dan E. Winarko, "Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia," *Indones. J. Comput. Cybern. Syst.* 5, vol. 2, hlm. 60-68, 2011.
- [6] G. A. Sandag, J. Leopold, dan V. F. Ong, "Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics," *CogITO Smart J.*, vol. 4, no. 1, Art. no. 1, Jun 2018, doi: 10.31154/cogito.v4i1.100.37-45.
- [7] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *J. Teknol. Rekayasa*, vol. 3, no. 1, hlm. 89, Jun 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.
- [8] H. Junaedi, H. Budianto, dan I. Maryati, "Data Transformation Pada Data Mining," hlm. 7, 2011.
- [9] S. Beisken, T. Meinel, B. Wiswedel, L. F. de Figueiredo, M. Berthold, dan C. Steinbeck, "KNIME-CDK: Workflow-driven cheminformatics," *BMC Bioinformatics*, vol. 14, no. 1, hlm. 257, Agu 2013, doi: 10.1186/1471-2105-14-257.
- [10] A. J. Jara, D. Genoud, dan Y. Bocchi, "Big data for smart cities with KNIME a real experience in the SmartSantander testbed," *Softw. Pract. Exp.*, vol. 45, no. 8, hlm. 1145-1160, 2015, doi: 10.1002/spe.2274.
- [11] L. Rahmawati, S. W. Sihwi, dan E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, FMIPA, Universitas Sebelas Maret)," *ITSMART J. Teknol. Dan Inf.*, vol. 3, no. 2, Art. no. 2, Agu 2016, doi: 10.20961/itsmart.v3i2.654.
- [12] A. Muhidin, "Analisa Metode Hierarchical Clustering Dan K-Mean Dengan Model Lrfmp Pada Segmentasi Pelanggan," *J. SIGMA*, vol. 8, no. 3, Art. no. 3, Sep 2017, doi: 10.37366/sigma.v8i3.134.
- [13] T. Alfina, B. Santosa, dan J. A. R. Hakim, "Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," vol. 1, 2012.
- [14] N. K. Zuhail, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," *STAINS Semin. Nas. Teknol. SAINS*, vol. 1, no. 1, Art. no. 1, Feb 2022.
- [15] W. Widyawati, W. L. Y. Saptomo, dan Y. R. W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *J. Ilm. SINUS*, vol. 18, no. 1, hlm. 75, Jan 2020, doi: 10.30646/sinus.v18i1.448.
- [16] K. Latifah, "Kombinasi Algoritma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang," *J. Inform. Upgris*, vol. 1, no. 1 Juni, Art. no. 1 Juni, 2015, doi: 10.26877/jiu.v1i1.
- [17] W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita," *J. Inform.*, vol. 15, no. 2, 2015.
- [18] A. Bryant dan K. Cios, "RNN-DBSCAN: A Density-Based Clustering Algorithm Using Reverse Nearest Neighbor Density Estimates," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, hlm. 1109-1121, Jun 2018, doi: 10.1109/TKDE.2017.2787640.
- [19] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, dan L. Shao, "Real-Time Superpixel Segmentation by DBSCAN Clustering Algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 12, hlm. 5933-5942, Des 2016, doi: 10.1109/TIP.2016.2616302.
- [20] J. Hou, H. Gao, dan X. Li, "DSets-DBSCAN: A Parameter-Free Clustering Algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 7, hlm. 3182-3193, Jul 2016, doi: 10.1109/TIP.2016.2559803.
- [21] M. I. P. Eriansya dan M. Syafrullah, "Implementasi Algoritma ST-DBSCAN dan K-MEANS Untuk Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Pulau Jawa Tahun 2014-2016 Berbasis Web Di Badan Pusat Statistik," *SKANIKA*, vol. 1, no. 3, Art. no. 3, Jul 2018.
- [22] R. Adha, N. Nurhaliza, U. Sholeha, dan M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN J. Sains Teknol. Dan Ind.*, vol. 18, no. 2, Art. no. 2, Jun 2021, doi: 10.24014/sitekin.v18i2.12469.
- [23] N. Purwanto, *Ilmu Pendidikan Teoritis dan Praktis*. Bandung: Remaja Karya, 2002.
- [24] E. Rendón, I. Abundez, A. Arizmendi, dan E. M. Quiroz, "Internal versus External cluster validation indexes," vol. 5, no. 1, 2011.
- [25] R. Hidayati, A. Zubair, A. H. Pratama, dan L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.Com*, vol. 20, no. 2, Art. no. 2, Mei 2021, doi: 10.33633/tc.v20i2.4556.