

Segmentasi Nasabah Kartu Kredit Berdasarkan Perilaku Penggunaan Kartu Kreditnya Menggunakan Algoritma K-Means

Ichwanul Muslim Karo Karo^{*1)}, Andreyas Yusmanto²⁾, dan Robi Setiawan³⁾

Fakultas Informatika, Fakultas Ilmu Hayati
Universitas Telkom, Universitas Surya
Bandung, Tangerang, Bandung

*ichwanulkarokaro@telkomuniversity.ac.id*¹, *andreyas.yusmanto18@student.surya.ac.id*²,
*robi.setiawan18@student.surya.ac.id*³

Abstract

The intensity of credit card customers in transacting has increased in the last 10 years in Indonesia. This is a challenge as well as an opportunity for the Bank. Customer segmentation information is very useful for reducing bad credit or increasing customer credit card limit capacity. This panel aims to segment credit card customers based on their credit card usage behavior with a clustering approach using the K-means algorithm. Meanwhile, the evaluation process of segmentation results uses a silhouette index. Based on the experimental results, the best number of clusters is six groups. The six groups are shopping hobbies, payment processing when due, payments in installments, withdrawing cash, buying expensive goods, and types that rarely use credit cards.

Keywords: K-Means, silhouette index.

Abstrak

Intensitas nasabah kartu kredit dalam bertransaksi meningkat dalam kurun 10 tahun terakhir di Indonesia. Hal tersebut menjadi tantangan sekaligus peluang bagi Bank. Informasi segmentasi nasabah sangat berguna untuk menekan kredit macet atau meningkatkan kapasitas limit kartu kredit nasabah. Penelitian ini bertujuan untuk mensegmentasikan nasabah kartu kredit berdasarkan perilaku penggunaan kartu kreditnya dengan pendekatan *clustering* menggunakan algoritma K-means. Sedangkan proses evaluasi hasil segmentasi menggunakan *silhouette index*. Berdasarkan hasil percobaan, diperoleh jumlah cluster terbaik adalah enam kelompok. Ke enam kelompok tersebut yakni hobi berbelanja, proses pembayaran saat jatuh tempo, pembayaran dengan mencicil, menarik uang tunai, membeli barang mahal, dan type yang jarang menggunakan kartu kredit.

Kata kunci: K-Means, *silhouette index*

1. Pendahuluan

Era perkembangan digital keuangan begitu cepat linear terhadap gaya hidup sebagian masyarakat dalam mengakses sebuah produk yang relative mahal. Kemudahan dalam pembayaranpun ditawarkan dalam bentuk kredit. Sehingga biasanya para pedagang memberikan alternatif pilihan berupa pembayaran dengan cara mencicil tiap periode. Salah satu metode pembayaran dengan cara mencicil adalah menggunakan kartu kredit. Penggunaan kartu kredit di Indonesia telah mengalami peningkatan dalam kurun waktu 10 tahun terakhir ini dan ditemukan banyaknya bank yang mulai berhati-hati dalam menawarkan limit kartu kredit akibat ketidakmampuan calon debitur untuk membayar hutang [1].

Dari sekian banyak faktor penentu debitur menggunakan kartu kredit, beberapa faktor yang menonjol antara lain: norma subjektif, kontrol perilaku persepsi, dan persepsi manfaat [2]. Ada juga faktor pembelian impulsif memiliki pengaruh pada para pengguna kartu kredit [3]. Untuk para mahasiswa yang menggunakan kartu kredit ditemukannya hubungan pengaruh orang tua, pengetahuan keuangan serta sikap mahasiswa itu sendiri dalam penggunaan kartu kredit tersebut [4]. Terdapat juga hubungan bahwa jumlah kartu yang dimiliki berhubungan negatif dengan usia dan berhubungan positif dengan tingkat pendapatan [5]. Di sisi lain ditemukan perilaku meningkatnya belanja kompulsif sebanding dengan orang yang memiliki gengsi yang tinggi dan sebanding dengan orang yang memiliki pemahaman tentang keuangan yang rendah [6].

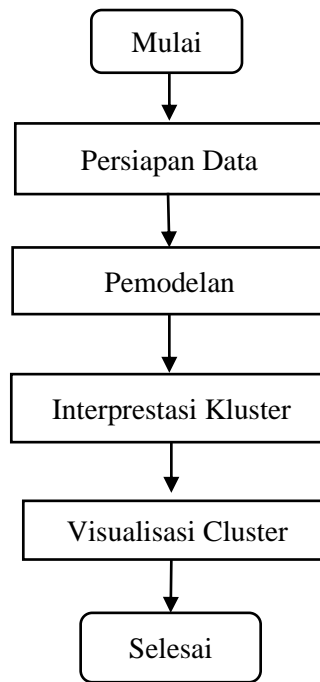
Terdapat beberapa penelitian terdahulu yang melatarbelakangi penelitian ini, antara lain:

1. Penelitian yang bertujuan untuk memprediksi kepuasan pelanggan pada layanan kartu kredit dengan menggunakan algoritma Random Forest [7].
2. Penelitian yang bertujuan untuk mengidentifikasi pelanggan kartu kredit yang berkualitas dan mengembangkan kualitas produk melalui pembuatan fitur yang sesuai dengan kebutuhan mereka untuk meningkatkan kepuasan pelanggan. Pada penelitian ini digunakan k-means dan algoritma C&RT [8].

3. Penelitian yang bertujuan untuk melakukan segmentasi pelanggan berdasarkan perilaku penggunaan kartu kredit di Afrika dengan menggunakan k-means clustering [9].
4. Penelitian yang bertujuan membantu manajemen bank dalam menilai klien kartu kredit menggunakan bidirectional LSTM neural networks dengan memodelkan dan memprediksi perilaku konsumen sehubungan dengan dua aspek: probabilitas pembayaran tidak terjawab tunggal dan berturut-turut untuk pelanggan kartu kredit [10].

2. Metode Penelitian

Pada penelitian terdapat beberapa tahapan, antara lain: tahap persiapan data, pemodelan dan evaluasi, interpretasi kluster dan visualisasi dari kluster tersebut (Gambar 1). Dataset ini diolah dengan menggunakan google colab.



Gambar 1 Kerangka Kerja Penelitian

A. Data

Data yang digunakan dalam penelitian ini dibuat pada tahun 2018 dan tersedia di situs <https://www.kaggle.com/arjunbhasin2013/ccdata>. Dataset ini terdiri dari 8950 baris (pemilik kartu kredit) dan 18 variabel perilaku (Tabel 1).

Tabel 1. Variabel Perilaku

| No | Nama Variabel | Deskripsi |
|----|-------------------------------|--------------------------------------------------------------------------------------------------------------|
| 1 | <i>Cust_id</i> | Identitas pemegang kartu kredit |
| 2 | <i>Balance</i> | Jumlah saldo yang tersisa di akun mereka untuk melakukan pembelian. |
| 3 | <i>Balance Frequency</i> | Intensitas saldo diperbaharui, skor antara 0 dan 1 (1 = sering diperbaharui, 0 = tidak sering diperbaharui). |
| 4 | <i>Purchases</i> | Jumlah pembelian yang dilakukan dari akun. |
| 5 | <i>One Off Purchases</i> | Jumlah pembelian maksimum yang dilakukan dalam sekali transaksi. |
| 6 | <i>Installments Purchases</i> | Jumlah pembelian yang dilakukan secara kredit. |

| | | |
|----|-----------------------------------------|---------------------------------------------------------------------------------------------------|
| 7 | <i>Cash Advance</i> | Uang tunai awal yang diberikan oleh pengguna. |
| 8 | <i>Purchases Frequency</i> | Intensitas pembelian dilakukan, skor antara 0 dan 1 (1 = sering dibeli, 0 = tidak sering dibeli). |
| 9 | <i>One Off Purchases Frequency</i> | Intensitas pembelian terjadi dalam sekali transaksi (1 = sering dibeli, 0 = tidak sering dibeli). |
| 10 | <i>Purchases Installments Frequency</i> | Intensitas pembelian secara kredit dilakukan (1 = sering dilakukan, 0 = tidak sering dilakukan). |
| 11 | <i>Cash Advance Frequency</i> | Intensitas uang tunai awal dibayarkan. |
| 12 | <i>Cash Advance Trx</i> | Jumlah transaksi yang dilakukan dengan tunai diawal |
| 13 | <i>Purchases Trx</i> | Jumlah transaksi pembelian yang dilakukan. |
| 14 | <i>Credit Limit</i> | Batas kartu kredit untuk pengguna. |
| 15 | <i>Payments</i> | Jumlah pembayaran yang dilakukan oleh pengguna. |
| 16 | <i>Minimum Payments</i> | Jumlah minimum pembayaran yang dilakukan oleh pengguna. |
| 17 | <i>Prc Full Payment</i> | Persentase pembayaran penuh yang dibayarkan oleh pengguna. |
| 18 | <i>Tenure</i> | Jangka waktu layanan kartu kredit untuk pengguna. |

B. Tahap Persiapan Data

Tahap persiapan data memegang peranan penting dalam proses segmentasi [11]. Salah satu tujuannya adalah untuk menstandarisasi dataset. Pada penelitian ini ada beberapa proses persiapan data dilakukan.

Missing Value : salah satu masalah yang dihadapi dalam dataset ialah komponen data yang hilang atau *missing value*. Pada dataset ini terdapat begitu banyak data yang hilang, sehingga diperlukan sebuah penanganan khusus untuk masalah ini. Metode paling umum yang digunakan untuk menggantikan *missing value* adalah dengan menggantikan nilai dengan nilai rata-rata atribut tersebut [12].

Terdapat begitu banyak nilai pencilan dalam dataset. Apabila pencilan dihilangkan, maka akan menyebabkan hilangnya begitu banyak *record* data implikasinya performansi model tidak akan maksimal [10]. Sehingga untuk mengatasi hal tersebut dibuat range untuk menangani nilai ekstrim [13].

C. Tahap Segmentasi

Proses segmentasi nasabah kartu kredit akan dilakukan dengan menggunakan pendekatan *Clustering*. Ada banyak jenis algoritma *Clustering*. Namun jenis algoritma *clustering* terbaik untuk menentukan *k* cluster adalah *Partitioning Clustering* [12]. Salah satu algoritma *partitioning clustering* yang terkenal adalah algoritma K-means.

Algoritma K-means merupakan salah satu metode cluster analysis non hirarki yang mempartisi objek yang kedalam satu atau lebih kelompok berdasarkan kemiripan karakteristiknya, sehingga objek yang mempunyai karakteristik yang lebih dekat dikelompokkan dalam satu cluster yang sama sedangkan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain. Dengan kata lain, algoritma K-Means bertujuan untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya. Pada algoritma pembelajaran mengelompokkan sendiri data-data yang menjadi *k* buah cluster tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam unsupervised learning. Proses menentukan *k* dibantu menggunakan *elbow method* [14]. Sedangkan proses berikutnya adalah evaluasi hasil segmentasi. Proses evaluasi menggunakan silhouette index (Si). Adapun tahapan tahapannya mengikuti [15].

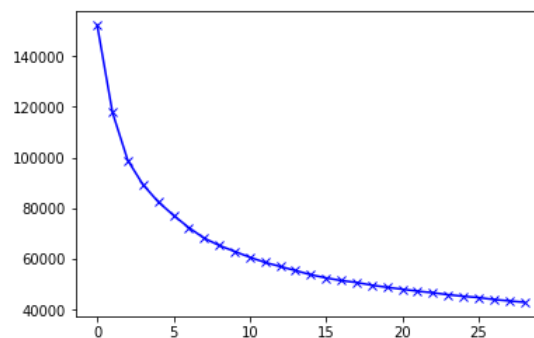
| Algoritma K-Means | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <i>Input</i> | |
| Data | |
| <i>Proses</i> | |
| <ol style="list-style-type: none"> 1. Tentukan k (jumlah cluster) 2. Tentukan titik awal centroid sebanyak k 3. Hitung jarak setiap data ke centroid 4. Masukkan setiap data ke cluster dengan jarak minimum 5. Update centroid dari masing masing cluster 6. Ulangi Langkah 3-5 hingga tidak terjadi perubahan centroid | |
| <i>Output</i> | |
| k dan masing masing anggota cluster | |

3. Hasil dan Pembahasan

Bagian ini menyajikan dan membahas hasil segmentasi nasabah kartu kredit, meliputi penentuan jumlah *cluster*, evaluasi jumlah cluster serta interpretasi cluster terbaik.

A. Penentuan Cluster

Setelah dilakukan tahap persiapan data, kami menguji jumlah cluster terbaik untuk algoritma K-means, dan diperoleh hasil uji elbow method (Gambar 2) dengan $k = 6$. Informasi ini menjadi acuan awal dalam proses *clustering*. Namun demikian akan diuji juga jumlah k lainnya,



Gambar 2 Hasil Metode Elbow

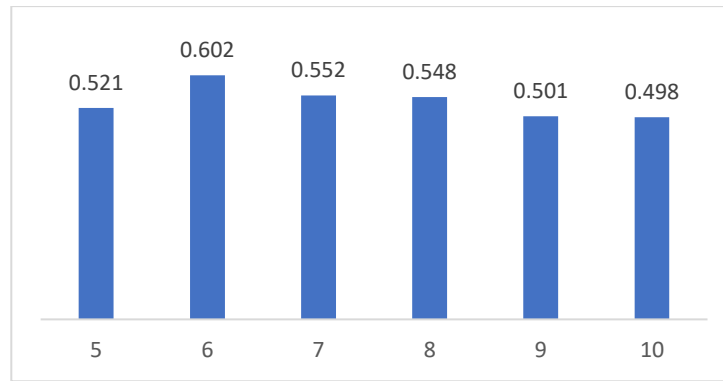
B. Evaluasi Jumlah Cluster

Bagian ini akan mengevaluasi setiap jumlah cluster yang di ujikan. Kriteria *cluster* terbaik ditentukan dan di interpretasikan berdasarkan nilai S_i (Tabel 2). Adapun jumlah k yang diujikan meliputi $k = 5,6,7,8,9,10$.

Tabel 2 Interpretasi Nilai S_i

| S_i | Deskripsi |
|---------------------|------------------------------------|
| $S_i > 0,71$ | Cluster yang diperoleh sangat kuat |
| $0,51 < S_i < 0,71$ | Data dapat diclusterkan |
| $0,25 < S_i < 0,5$ | Cluster yang diperoleh lemah |
| $S_i < 0,25$ | Tidak ditemukan cluster |

Berdasarkan hasil percobaan, jumlah cluster dengan nilai S_i tertinggi adalah $k = 6$ (Gambar 3), dengan kata lain jumlah cluster terbaik yang diperoleh adalah $k = 6$. Hasil ini sekaligus mengkonfirmasi dan menguatkan hasil interpretasi metode *elbow* yang memberikan rekomendasi k yang sama. Berdasarkan informasi Tabel 2, hasil $k = 6$ belum menghasilkan cluster yang kuat, data tersebut dapat diclusterkan namun anggota didalam cluster belum padu antara satu dengan yang lainnya. Walaupun demikian, jika diteruskan untuk mencari jumlah cluster yang lain, maka kualitas *cluster* yang dihasilkan belum tentu sebaik $k = 6$. Alasannya adalah pola jumlah $k > 6$ diperoleh S_i yang cenderung mengecil.

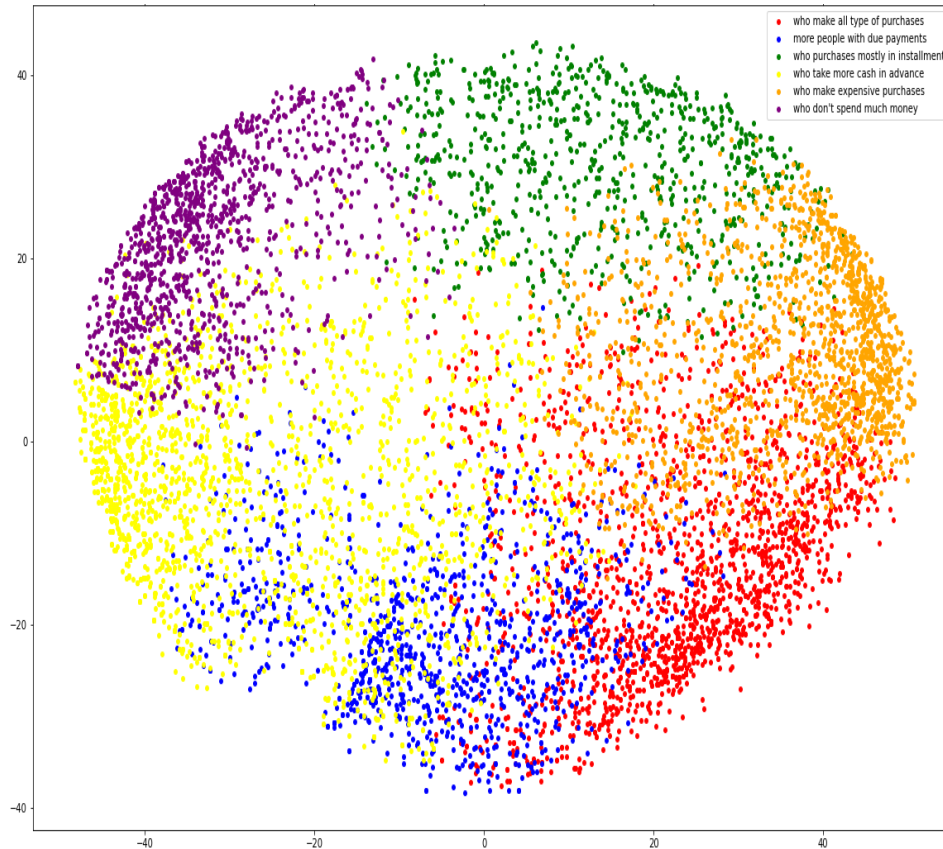


Gambar 3 Kualitas cluster k

C. Interpretasi Cluster Terbaik

Untuk memvisualisasikan hasil *cluster*, kami menggunakan metode *Principal Component Analysis (PCA)* untuk mentransformasikan data menjadi 2 dimensi [16]. Adapun visualisasi dapat dilihat pada Gambar .

- *Cluster* merah merupakan nasabah kartu kredit dengan perilaku yang melakukan semua jenis pembelian. Dengan kata lain type nasabah ini adalah suka berbelanja.
- *Cluster* biru tua merupakan nasabah kartu kredit dengan perilaku pembayaran jatuh tempo. Setiap penggunaan kartu kredit dibayar sesuai dengan waktu jatuh tempo.
- *Cluster* hijau merupakan nasabah dengan perilaku setiap pembelian sesuatu dilakukan dengan cara mencicil
- *Cluster* kuning merupakan nasabah dengan perilaku mempergunakan kartu kredit untuk menarik uang tunai. Dengan kata lain mengutang dengan kartu kredit.
- *Cluster* jingga merupakan nasabah dengan perilaku suka beli barang barang mahal dan mewah
- *Cluster* ungu merupakan nasabah dengan perilaku tidak banyak menghabiskan uang. Dengan kata lain type nasabah ini jarang menggunakan kartu kredit.



Gambar 4 Visualisasi Segmentasi Nasabah

4. Kesimpulan

Segmentasi nasabah kartu kredit dapat dilakukan dengan menggunakan pendekatan *clustering* dengan algoritma K-means. Dalam penelitian ini, dilakukan proses segmentasi nasabah berdasarkan perilaku penggunaan kartu kreditnya dengan algoritma k-means dan metode PCA dalam proses visualisasinya. Diperoleh jumlah *cluster* terbaik adalah 6, dengan interpretasi kualitas yang tidak buruk. Type enam kelompok nasabah yang diperoleh yakni hobi berbelanja, proses pembayaran saat jatuh tempo, pembayaran dengan mencicil, menarik uang tunai, membeli barang mahal, dan type yang jarang menggunakan kartu kredit. Kedepannya sangat memungkinkan dilakukan perbaikan dari sisi algoritma maupun penanganan variabel yang banyak untuk memperoleh hasil *cluster* yang baik.

5. Daftar Rujukan

- [1] S. Johan and C.E. Dewi, "Credit Limit of Unsecured Consumer Lending: Evidence from Micro Data," *Economics and Finance in Indonesia*, 67(1), 51–62, 2021.
- [2] N. Anastasia and S. Santoso, "Effects of Subjective Norms, Perceived Behavioral Control, Perceived Risk, and Perceived Usefulness towards Intention to Use Credit Cards in Surabaya, Indonesia," *SHS Web of Conferences*, 76, 01032, 2020.
- [3] F. Cuandra and Kelvin, "Analysis of influence of materialism on impulsive buying and compulsive buying with credit card use as mediation variable," 13(1), 7–16, 2021.
- [4] S. Kumar and L. Karlina, "Intention to Use Credit Card among College Students in Greater Jakarta," *JAAF (Journal of Applied Accounting and Finance)*, 4(1), 49–59, 2020.
- [5] K. Jung and M. Y. Kang, "Understanding credit card usage behavior of elderly korean consumers for sustainable growth: Implications for Korean credit card companies," *Sustainability (Switzerland)*, 13(7), 2021.
- [6] R. Khandelwal, A. Kolte, N. Veer and P. Sharma, "Compulsive Buying Behavior of Credit Card Users and Affecting Factors Such as Financial Knowledge, Prestige and Retention Time: A Cross-sectional Research," *Vision*, 2021.
- [7] M. K. Yaseen, M. Raheem and V. Sivakumar, "Credit Card Business in Malaysia: A Data Analytics Approach," *International Journal of Advanced Computer Science and Applications*, 11(12), 383–390, 2020.
- [8] L. Hassani and E. Taati, "Studying product quality by exploring credit card customers behavior via data mining techniques," *International Journal for Quality Research*, 14(1), 163–182, 2020.
- [9] Umuhzoza, E., Ntirushwamaboko, D., Awuah, J., & Birir, B. (2020). Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in africa. *SAIEE Africa Research Journal*, 111(3), 95-101, 2020.
- [10] M. Ala'raj, M. F. Abbod and M. Majdalawieh, "Modelling customers credit card behaviour using bidirectional LSTM neural networks," *Journal of Big Data*, 8(1), 2021.
- [11] Karo, I. M. K., & Huda, A. F. (2016). Spatial clustering for determining rescue shelter of flood disaster in South Bandung using CLARANS Algorithm with Polygon Dissimilarity Function. In 2016 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA) (pp. 70-75). IEEE,
- [12] Yadav, M. L., & Roychoudhury, B. (2018). Handling missing values: A study of popular imputation packages in R. *Knowledge-Based Systems*, 160, 104-118.
- [13] Rahmasari, A., & Noeryanti, N. (2021). PREDIKSI DATA SPASIAL YANG TIDAK TERSAMPEL DAN MENGANDUNG PENCILAN MENGGUNAKAN METODE ROBUST KRIGING. *Jurnal Statistika Industri dan Komputasi*, 6(2), 132-140.
- [14] Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering* (Vol. 336, No. 1, p. 012017). IOP Publishing.

- [15] Karo, I. M. K., MaulanaAdhinugraha, K., & Huda, A. F. (2017, November). A cluster validity for spatial clustering based on davies bouldin index and Polygon Dissimilarity function. In *2017 Second International Conference on Informatics and Computing (ICIC)* (pp. 1-6). IEEE.
- [16] Abdulhafedh, A. (2021). Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development*, 3(1), 12-30.