

Penerapan Algoritma *K-Means Clustering* untuk Segmentasi Nasabah Bank

Dudi Irawan¹, Guruh Wijaya², Taufiq Timur Warisaji³

¹Teknik Informatika, Universitas Muhammadiyah Jember, *dudi.irawan@unmuhjember.ac.id*

²Teknik Informatika, Universitas Muhammadiyah Jember, *guruh.wijaya@unmuhjember.ac.id*

³Teknik Informatika, Universitas Muhammadiyah Jember, *taufiqtimur@unmuhjember.ac.id*

Keywords:

Banking,
Cluster Analysis,
Customer Segmentation,
Data Mining,
K-Means Clustering,

ABSTRACT

Customer segmentation is an important strategy in the banking industry to understand customer characteristics and needs, thereby increasing the effectiveness of services and business decision making. In an increasingly competitive industry, segmentation allows banks to group customers based on factors such as demographics, transaction behavior, and income levels in order to develop more targeted marketing strategies. In addition, segmentation plays a role in risk management by identifying customers with high-risk profiles and supporting operational efficiency through optimal resource allocation. With this strategy, banks can increase customer satisfaction and loyalty and strengthen competitiveness in the digital era. This study applies the K-Means clustering algorithm to group customers based on demographic and financial factors, such as age, gender, income, marital status, and banking asset ownership. The data used consists of 600 customer records taken from online sources and processed using the unsupervised learning method in data mining. The clustering process is carried out with several k values to determine the optimal grouping. Validation is carried out using the Within-Cluster Sum of Squares (WCSS) to measure the quality of the segmentation results. The results of the study show that segmentation with values of $k = 2$ and $k = 4$ provides a clearer pattern than $k = 3$, where each cluster has different characteristics in terms of risk profile, product preferences, and level of involvement with banking services. With this approach, banks can improve marketing strategies, credit risk management, and service personalization based on the resulting segmentation. In addition, this study provides recommendations for the development of a data mining-based system that can assist in business decision making in the banking sector.

Kata Kunci

Analisis Klaster,
Data Mining,
K-Means Clustering,
Perbankan,
Segmentasi Nasabah,

ABSTRAK

Segmentasi nasabah merupakan strategi yang penting dalam industri perbankan untuk memahami karakteristik dan kebutuhan pelanggan, sehingga meningkatkan efektivitas layanan dan pengambilan keputusan bisnis. Dalam industri yang semakin kompetitif, segmentasi memungkinkan bank mengelompokkan nasabah berdasarkan faktor seperti demografi, perilaku transaksi, dan tingkat pendapatan guna menyusun strategi pemasaran yang lebih tepat sasaran. Selain itu, segmentasi berperan dalam pengelolaan risiko dengan mengidentifikasi nasabah yang memiliki profil risiko tinggi serta mendukung efisiensi operasional melalui alokasi sumber daya yang optimal. Dengan strategi ini, bank dapat meningkatkan kepuasan dan loyalitas nasabah serta memperkuat daya saing di era digital. Penelitian ini menerapkan algoritma *K-Means clustering* untuk mengelompokkan nasabah berdasarkan faktor demografi dan finansial, seperti usia, jenis kelamin, pendapatan, status pernikahan, dan kepemilikan aset perbankan. Data yang digunakan terdiri dari 600 *record* nasabah yang diambil dari sumber daring dan diproses menggunakan metode *unsupervised learning* dalam *data mining*. Proses *clustering* dilakukan dengan beberapa nilai k untuk menentukan pengelompokan yang optimal. Validasi dilakukan menggunakan *Within-Cluster Sum of Squares* (WCSS) untuk mengukur kualitas hasil segmentasi. Hasil penelitian menunjukkan bahwa segmentasi dengan nilai $k = 2$ dan $k = 4$ memberikan pola yang lebih jelas dibandingkan $k = 3$, di mana setiap *cluster* memiliki karakteristik yang berbeda dalam hal profil risiko, preferensi produk, dan tingkat keterlibatan dengan layanan perbankan. Dengan pendekatan ini, bank dapat meningkatkan strategi pemasaran, manajemen risiko kredit, serta personalisasi layanan berdasarkan segmentasi yang dihasilkan. Selain itu, penelitian ini memberikan rekomendasi untuk pengembangan sistem berbasis *data mining* yang dapat membantu dalam pengambilan keputusan bisnis di sektor perbankan.

Korespondensi Penulis:

Dudi Irawan,
Universitas Muhammadiyah Jember,
Jl. Karimata No 49 Jember
Telepon : +6285815308476
Email: dudi.irawan@unmuhjember.ac.id

Submitted : 25-03-2025; Accepted : 31-03-2025;
Published : 31-03-2025

Copyright (c) 2025 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Segmentasi nasabah merupakan strategi penting dalam industri perbankan untuk memahami karakteristik dan kebutuhan pelanggan secara lebih mendalam [1]. Segmentasi nasabah merupakan strategi krusial dalam perbankan untuk memahami karakteristik dan kebutuhan pelanggan, sehingga meningkatkan efektivitas layanan dan pengambilan keputusan bisnis. Dalam industri yang semakin kompetitif, segmentasi memungkinkan bank mengelompokkan nasabah berdasarkan faktor seperti demografi, perilaku transaksi, dan tingkat pendapatan guna menyusun strategi pemasaran yang lebih tepat sasaran. Selain itu, segmentasi berperan dalam pengelolaan risiko dengan mengidentifikasi nasabah yang memiliki profil risiko tinggi serta mendukung efisiensi operasional melalui alokasi sumber daya yang optimal. Dengan strategi ini, bank dapat meningkatkan kepuasan dan loyalitas nasabah serta memperkuat daya saing di era digital.

Salah satu pendekatan yang efektif dalam segmentasi ini adalah menggunakan teknik *clustering*, yang merupakan metode *unsupervised learning* dalam *data mining* [2]. *Clustering* memungkinkan bank untuk mengelompokkan nasabah berdasarkan kesamaan karakteristik tertentu, seperti pola transaksi, tingkat pendapatan, preferensi produk, dan risiko kredit, tanpa perlu adanya label awal pada data [3]. Dengan segmentasi berbasis *clustering*, bank dapat menyusun strategi pemasaran yang lebih terarah dan personalisasi [4]. Misalnya, nasabah dengan frekuensi transaksi tinggi pada layanan investasi dapat dikelompokkan ke dalam segmen investor potensial, sehingga bank dapat menawarkan produk investasi yang lebih sesuai. Hal ini meningkatkan efektivitas pemasaran dan meningkatkan tingkat konversi dari kampanye promosi.

Data mining merupakan suatu proses yang memanfaatkan teknik statistika, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi serta mengidentifikasi informasi yang bernilai dan pengetahuan yang relevan dari kumpulan data berskala besar [5]. *Data mining* dikategorikan berdasarkan jenis tugas yang dapat dilakukan, yang mencakup deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi [6]. *Clustering* merupakan salah satu task dalam *data mining* yang bertujuan untuk mengelompokkan data berdasarkan informasi yang ditemukan dan hubungan diantara data tersebut, atau dengan kata lain adalah pengelompokan berdasarkan karakteristik data [7]. Hasil akhir dari *clustering* adalah kelompok – kelompok atau *clusters* yang beranggotakan data, dimana setiap kelompok berisi obyek yang memiliki kemiripan karakteristik atau memiliki hubungan tertentu, dan setiap kelompok memiliki perbedaan karakteristik [8].

Pada umumnya teknik *clustering* dibedakan menjadi 2, yaitu berdasarkan konsep *hierarchical* dan berdasarkan konsep *partitioning* [9]. Konsep *Hierarchical* sendiri secara umum dibedakan menjadi *agglomerative* dan *divisive*. Sementara itu, metode *clustering* yang berbasis *partitioning* membagi data ke dalam sejumlah kelompok yang telah ditentukan sebelumnya berdasarkan kriteria tertentu [10]. Salah satu algoritma yang paling populer dalam kategori ini adalah *K-Means*. Algoritma *K-Means* bekerja dengan menentukan sejumlah pusat *cluster* atau *centroid* secara acak, lalu mengelompokkan setiap data ke dalam *cluster* terdekat berdasarkan jarak, biasanya menggunakan *Euclidean Distance* [11]. Selanjutnya, algoritma ini akan menghitung ulang posisi *centroid* berdasarkan rata-rata data dalam setiap *cluster*, lalu mengulangi proses tersebut hingga tidak ada perubahan signifikan dalam pembagian *cluster* atau sampai jumlah iterasi maksimum tercapai. Dengan sifatnya yang efisien dan mudah diterapkan, *K-Means* sering digunakan dalam berbagai aplikasi seperti pengelompokan pelanggan, analisis pola dalam data, serta segmentasi gambar [12].

Penelitian ini melakukan pengelompokan dengan dataset data nasabah melalui teknik data mining yaitu *clustering* yang secara khusus menrapkan kmeans yang dilanjutkan dengan validasi *cluster*. Adapun atribut data yang digunakan adalah faktor umur, jenis kelamin, domisili, pendapatan, status pernikahan, jumlah anak, kepemilikan mobil, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, dan status pembelian *Personal Equity Plan*. Dataset yang digunakan diambil secara online dari <http://facweb.cs.depaul.edu/>. Tujuan dari penelitian ini adalah menemukan kelompok nasabah berdasarkan faktor – faktor tersebut sehingga dapat digunakan untuk kepentingan penentuan kebijakan selanjutnya.

2. METODE PENELITIAN

Secara umum, tahapan dalam penelitian ini terdiri dari empat langkah utama. Tahap pertama adalah studi literatur untuk memperoleh landasan teoritis yang relevan. Selanjutnya, dilakukan proses pengumpulan dan pengolahan dataset guna memastikan kualitas serta kesesuaian data yang digunakan. Tahap berikutnya mencakup implementasi metode *clustering* sebagai bagian dari analisis data. Akhirnya, penelitian ini diakhiri dengan tahap

penarikan kesimpulan atau analisis hasil untuk mengevaluasi temuan yang diperoleh. Tahapan atau langkah – langkah penelitian ini secara umum dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Studi literatur merupakan tahap yang bertujuan untuk menelaah berbagai sumber referensi, seperti jurnal penelitian, makalah ilmiah, serta buku akademik yang relevan dengan topik penelitian. Langkah ini dilakukan untuk memperkaya pemahaman awal, mengidentifikasi teori yang mendukung, serta memastikan landasan konseptual yang kuat dalam pelaksanaan penelitian.

Data yang digunakan dalam penelitian ini merupakan data nasabah dari suatu institusi perbankan, yang diperoleh melalui situs web <http://facweb.cs.depaul.edu/>. Data tersebut dikategorikan sebagai data sekunder, karena telah mengalami proses pengolahan sebelumnya oleh pemilik situs *web* sebelum digunakan dalam penelitian ini. Data yang diperoleh berbentuk file .csv (*comma separated value*), jumlah *record* sebanyak 600 buah, yang mempunyai 11 atribut. Tabel 1 merupakan contoh data yang diperoleh. Dataset yang digunakan dalam penelitian ini terdiri dari data nasabah dengan sejumlah atribut yang mencerminkan karakteristik individu. Atribut-atribut tersebut mencakup usia, jenis kelamin, domisili, tingkat pendapatan, status pernikahan, jumlah anak, kepemilikan kendaraan, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, serta status partisipasi dalam *Personal Equity Plan*. Dataset ini diperoleh secara daring melalui situs *web* <http://facweb.cs.depaul.edu/>.

Tabel 1. Dataset

Id	Age	Sex	Region	Income	Married	children	car	save_act	current_act	mortgage	pep
ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	NO	NO	YES
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO
ID12105	57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO
ID12106	57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES
ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
ID12108	58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO
ID12109	37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO

Clustering akan menggunakan algoritma *K-Means* yang kemudian dilakukan validitas *cluster* menggunakan *within-cluster sum of squares* (WCSS). Secara umum prosedur metode *K-Means* adalah sebagai berikut :

1. Tentukan nilai *k* sebagai jumlah klaster yang akan dibentuk. Penentuan jumlah klaster *k* dilakukan berdasarkan beberapa faktor, termasuk pertimbangan teoritis dan konseptual yang digunakan sebagai dasar dalam menentukan jumlah klaster yang optimal [13].
2. Inisialisasi *k* *centroid* (titik pusat klaster) awal dilakukan secara acak. Pemilihan *centroid* awal ini dilakukan dengan memilih sejumlah objek secara acak sesuai dengan jumlah *k* klaster yang ditetapkan, yang selanjutnya digunakan sebagai dasar dalam perhitungan *centroid* untuk klaster ke-*i* pada iterasi berikutnya, menggunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots n$$

Dimana :

V = *centroid* pada *cluster*

X_i = obyek ke-*i*

N = banyaknya obyek atau jumlah obyek yang menjadi anggota *cluster*

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Kemudian hitung jarak antara objek dengan *centroid*, dalam penelitian ini menggunakan *Euclidian Distance*:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1, 2, 3, \dots n$$

Dimana;

xi : objek x ke-i

y: daya y ke-i

n: banyaknya objek

4. Mengelompokkan setiap objek ke dalam klaster berdasarkan kedekatannya dengan *centroid* terdekat.
5. Melakukan iterasi dan memperbarui posisi *centroid* menggunakan persamaan yang telah ditentukan.
6. Mengulangi langkah sebelumnya hingga posisi *centroid* yang baru tidak mengalami perubahan.

Proses pengelompokan titik dilakukan dengan membandingkan matriks kumpulan tugas pada iterasi sebelumnya dengan matriks kumpulan tugas pada iterasi saat ini. Jika hasil perbandingan menunjukkan kesamaan, maka algoritma *K-Means* telah mencapai konvergensi. Namun, jika terdapat perbedaan, proses iterasi perlu dilanjutkan hingga tercapai kondisi konvergen. Analisis hasil dilakukan dengan mendeskripsikan karakteristik klaster yang terbentuk, suatu proses yang dikenal sebagai profiling *cluster*. Hasil yang diharapkan dari analisis ini adalah memperoleh informasi yang relevan mengenai segmentasi nasabah berdasarkan atribut yang digunakan dalam penelitian.

3. HASIL DAN ANALISIS

Tahap pengolahan data merupakan proses persiapan dataset sebelum dilakukan segmentasi. Pada tahap ini, data diperiksa untuk memastikan tidak adanya duplikasi atau redundansi, serta mengidentifikasi kemungkinan adanya data yang tidak lengkap. Selain itu, evaluasi juga dilakukan terhadap kesebelas atribut yang tersedia untuk menentukan apakah seluruhnya dapat digunakan dalam analisis atau memerlukan penyesuaian lebih lanjut. Dari tabel 1 dapat kita lihat bahwa dataset terdiri dari 12 atribut, yaitu "Id", "age", "sex", "region", "income", "married", "children", "car", "save_act", "current_act", "mortgage", "pep". Setiap atribut dalam dataset merepresentasikan informasi tertentu, yaitu nomor ID, usia, jenis kelamin, domisili, tingkat pendapatan, status pernikahan, jumlah anak, kepemilikan kendaraan, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, serta status partisipasi dalam *Personal Equity Plan*, secara berurutan.

Dalam proses *clustering* maka semua nilai kolom "Id" tidak diperlukan karena memang atribut tersebut hanya merupakan informasi nomor urut identitas nasabah, sehingga semua data pada kolom "Id" dihapus dan potongan tabel data dapat dilihat pada Tabel 2.

Tabel 2. Dataset Hasil Pengolahan

age	Sex	region	income	Married	Children	car	save_act	current_act	mortgage	pep
48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	NO	NO	YES
40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO
51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO
23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO
57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO
57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES
22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO
37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO

Jumlah dataset setelah melalui tahap pengolahan data adalah tetap 600 *record*, sehingga keseluruhan data tersebut yang nantinya diproses dalam *clustering* pengelompokan nasabah. Tahap mining yang melakukan pengelompokan berdasarkan *clustering* dilakukan dengan menggunakan algoritma kmeans. Uji coba dilakukan dengan menggunakan nilai K sebanyak 3 buah yaitu dengan mengambil nilai K = 2 sampai dengan K = 4. Dengan menggunakan nilai K = 2, maka *cluster* yang terbentuk ada 2, *cluster* 1 berisi 366 *record* dan *cluster* 2 berisi 234 *record*, proses *clustering* berhenti sampai dengan iterasi ke-5 dan nilai *within cluster sum of squared errors* adalah 2456.0. Detail informasi tentang pusat *cluster* seperti disajikan pada tabel 3.

Tabel 3. Pusat *Cluster* K=2

Atribut	Cluster 1	Cluster 2
Age	0_34	52_max
Sex	MALE	FEMALE
Region	INNER_CITY	INNER_CITY
Income	0_24386	24387_43758
Married	YES	YES

Children	0	0
Car	NO	YES
save_act	YES	YES
current_act	YES	YES
Mortgage	NO	NO
Pep	NO	YES

Dari tabel 3, yang dapat dijadikan atribut pembeda adalah jenis kelamin (“sex”) atau atribut status Personal Equity Plan (“pep”), serta atribut kepemilikan mobil (“car”), karena ketiga atribut inilah yang mengelompokkan data dengan jelas. Dengan menggunakan nilai $K = 3$, maka *cluster* yang terbentuk ada 3, *cluster* 1 berisi 287 *record*, *cluster* 2 berisi 177 *record*, dan *cluster* 3 berisi 136 *record*, proses *clustering* berhenti sampai dengan iterasi ke-3 dan nilai *within cluster sum of squared errors* adalah 2440.0. Detail informasi tentang pusat *cluster* seperti disajikan pada tabel 4.

Tabel 4. Pusat *cluster* K=3

Atribut	Cluster 1	Cluster 2	Cluster 3
Age	0_34	52_max	52_max
Sex	MALE	FEMALE	FEMALE
region	INNER_CITY	INNER_CITY	INNER_CITY
income	0_24386	0_24386	24387_43758
married	NO	YES	YES
children	0	0	0
Car	NO	YES	YES
save_act	YES	YES	NO
current_act	YES	YES	YES
mortgage	NO	NO	NO
Pep	NO	NO	YES

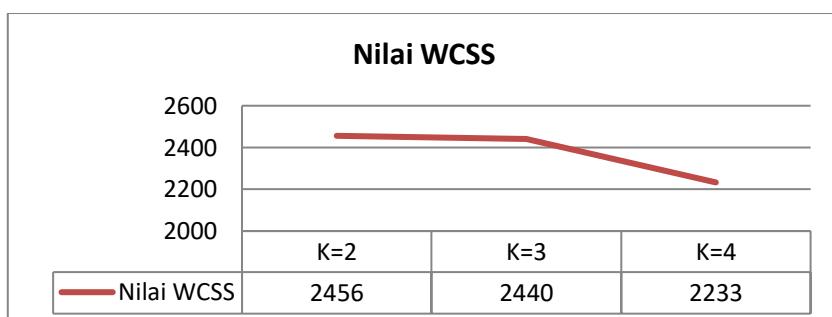
Dari tabel 4, terlihat bahwa *cluster* ini menunjukkan bahwa kelompok nasabah tidak mempunyai perbedaan yang signifikan dari segi domisili dan kepemilikan rekening giro, hal ini ditunjukkan dari atribut “region” dan “current_act” yang mempunyai pusat *cluster* yang sama untuk masing – masing *cluster*. Dengan menggunakan nilai $K = 4$, maka *cluster* yang terbentuk ada 4, *cluster* 1 berisi 224 *record*, *cluster* 2 berisi 148 *record*, *cluster* 3 berisi 109 *record*, dan *cluster* 4 berisi 119 *record*, proses *clustering* berhenti sampai dengan iterasi ke-3 dan nilai *within cluster sum of squared errors* adalah 2233.0. Detail informasi tentang pusat *cluster* seperti disajikan pada tabel 5.

Tabel 5. Pusat *cluster* K=4

Atribut	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Age	0_34	52_max	52_max	35_51
Sex	MALE	FEMALE	MALE	FEMALE
Region	INNER_CITY	INNER_CITY	INNER_CITY	TOWN
Income	0_24386	0_24386	24387_43758	24387_43758
Married	NO	YES	YES	YES
Children	0	2	0	0
Car	NO	YES	YES	YES
save_act	YES	YES	NO	YES
current_act	YES	YES	YES	YES
Mortgage	NO	NO	NO	NO
Pep	NO	NO	YES	NO

Dari tabel 5, terlihat bahwa *cluster* ini menunjukkan bahwa kelompok nasabah tidak mempunyai perbedaan yang signifikan dari segi kepemilikan rekening giro dan status hipotek, hal ini ditunjukkan dari atribut “current_act” dan “mortgage” yang mempunyai pusat *cluster* yang sama untuk masing – masing *cluster*. Akan tetapi pada atribut “sex” dan “income” memiliki masing – masing 2 pusat *cluster* yang berbeda, dan hal ini dapat dijadikan dasar sebagai pengelompokan. Atribut “sex” pada *cluster* 1 dan 3 sama – sama mempunyai pusat *cluster* MALE sedangkan pada *cluster* 2 dan 4 mempunyai pusat *cluster* FEMALE, sedangkan atribut “income” pada *cluster* 1 dan 2 sama – sama mempunyai pusat *cluster* 0_24386, sedangkan pada *cluster* 3 dan 4 sama – sama mempunyai pusat *cluster* 24387_43758.

Secara keseluruhan, pengelompokan nasabah dengan teknik *clustering* menggunakan algoritma *K-Means* dengan menggunakan uji $K = 2$ sampai $K = 4$ telah dilakukan dengan hasil yang bervariasi. Pada *clustering* menggunakan nilai $K = 2$ dan $K = 4$ dapat menunjukkan pola pengelompokan yang menarik, sedangkan dengan nilai $K = 3$ pola pengelompokan tidak dapat ditunjukkan dengan baik. Dari segi validasi, parameter yang digunakan adalah *within-cluster sum of squares* yaitu ukuran yang menilai seberapa dekat setiap anggota terhadap pusat *cluster*, sehingga semakin kecil nilai WCSS maka semakin baik dan valid *cluster – cluster* yang telah terbentuk. Masing – masing nilai WCSS untuk masing masing $K=2$ sampai $K=4$ dapat dilihat pada grafik dibawah ini.



Gambar 2. Nilai WCSS

Dari grafik ditunjukkan bahwa nilai WCSS berbanding terbalik dengan jumlah K yang digunakan, semakin tinggi nilai K yang digunakan maka nilai WCSS semakin turun, yang mempunyai arti bahwa jika nilai K semakin tinggi maka kedekatan antara masing – masing nilai terhadap pusat *cluster* semakin rendah yang menunjukkan bahwa *cluster* yang terbentuk memiliki validitas yang baik.

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa pengelompokan nasabah menggunakan algoritma *K-Means* dengan variasi jumlah klaster (K) antara 2 hingga 4 menghasilkan pola yang berbeda. Penggunaan $K = 2$ dan $K = 4$ mampu membentuk pola klaster yang lebih jelas dibandingkan dengan $K = 3$, yang tidak menunjukkan pola pengelompokan yang optimal. Validasi hasil klaster dilakukan menggunakan parameter *within-cluster sum of squares* (WCSS), yang mengukur kedekatan setiap anggota terhadap pusat klaster. Hasil analisis menunjukkan bahwa nilai WCSS menurun seiring dengan peningkatan jumlah K , yang mengindikasikan bahwa semakin besar jumlah klaster, semakin kecil jarak anggota terhadap pusat klaster, sehingga meningkatkan validitas hasil klasterisasi.

Berdasarkan hasil penelitian ini, terdapat beberapa saran untuk pengembangan lebih lanjut. Pertama, diperlukan pengembangan antarmuka pengguna yang lebih intuitif guna mempermudah operasional proses data mining, mengingat saat ini sistem yang digunakan belum memiliki fitur tersebut. Kedua, sistem dapat ditingkatkan dengan menambahkan fitur pendukung pengambilan keputusan, seperti identifikasi nasabah yang perlu diprioritaskan dalam layanan, klasifikasi nasabah dengan prospek rendah, serta visualisasi data berbasis peta geografis. Penambahan fitur-fitur ini diharapkan dapat meningkatkan efektivitas analisis data serta memberikan informasi yang lebih komprehensif dalam pengelolaan nasabah.

REFERENSI

- [1] H. Rizkyanto and F. L. Gaol, "Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and *K-Means* on Financial Industry," *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(4), 2023.
- [2] L. K. Kumar, D. Venkatesh, J. Katyayani, S. Sunkara and G. K., "A Study on Customer Segmentation for Banking Sector Through *Cluster* Analysis: Ethical Implications," *Communications on Applied Nonlinear Analysis* Vol.21 No.5s, 2024.

- [3] T. Kovács, A. Ko and A. Asemi, "Exploration of The Investment Patterns of Potential Retail Banking Customers Using Two-Stage *Cluster* Analysis," *Journal of Big Data*, 2021.
- [4] V. Djurisic, L. Kascelan, S. Rogic and B. Melovic, "Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method," *Applied Artificial Intelligence Volume 34 Issue 12*, 2020.
- [5] F. K. Nasser and S. F. Behadili, "A Review of Data Mining and Knowledge Discovery Approaches for Bioinformatics," *Iraqi Journal of Science Vol 63 No 7*, 2022.
- [6] S. M. Dol and D. P. M. Jawandhiya, "A Review of Data Mining in Education Sector," *Journal of Engineering Education Transformations Volume: 36, Issue: Special Issue 2*, pp. 13-22, 2023.
- [7] E. P. W. Mandala and D. E. Putri, "Data Mining Technique for Grouping Products Using *Clustering* Based on Association," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS) Vol 31, No 2*, 2023.
- [8] R. E. Abadi, M. J. Ershadi and S. T. A. Niaki, "A *Clustering* Approach For Data Quality Results of Research Information Systems," *Information Discovery and Delivery Volume 51 Issue 4*, 2022.
- [9] Z.-Y. Lim, L.-Y. Ong and M.-C. Leow, "A Review on *Clustering* Techniques: Creating Better User Experience for Online Roadshow," *Future Internet 13(9)*, p. 233, 2021.
- [10] H. Mittal, A. C. Pandey, M. Saraswat, S. Kumar, R. Pal and G. Modwel, "A Comprehensive Survey of Image Segmentation: *Clustering* Methods, Performance Parameters, and Benchmark Datasets," *Multimedia Tools and Applications Volume 81*, p. 35001–35026, 2022.
- [11] M. Capó, A. Pérez and J. A. Lozano, "An Efficient *K-Means Clustering* Algorithm For Tall Data," *Data Mining and Knowledge Discovery Volume 34*, p. 776–811, 2020.
- [12] C. Fang and H. Liu, "Research and Application of Improved *Clustering* Algorithm in Retail Customer Classification," *Symmetry Volume 13 Issue 10*, 2021.
- [13] A. Punhani, N. Faujdar, K. K. Mishra and M. Subramanian, "Binning-Based Silhouette Approach to Find the Optimal *Cluster* Using *K-Means*," *IEEE Access Volume 10*, 2022.