

Penerapan Metode K-Medoids untuk Segmentasi Nasabah pada Dataset Bank Marketing

Reysha Dwi Novaldo¹, Bahqri Fadillah², Muhamad Fahmi Maulana³, Mufidah Karimah⁴

¹⁻⁴ Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹reyshadwinovaldo@gmail.com, ²bahrifadillah8@gmail.com, ³fahmimaulana1301@gmail.com, ⁴dosen02829@unpam.ac.id

Abstrak—Perkembangan di bidang teknologi informasi telah meningkatkan penggunaan data untuk membantu proses pengambilan keputusan dalam industri perbankan. Salah satu aplikasinya adalah dalam segmentasi pelanggan untuk memperoleh pemahaman yang lebih mendalam mengenai karakteristik nasabah. Tujuan dari penelitian ini adalah untuk menerapkan algoritma K-Medoids dalam melakukan segmentasi nasabah menggunakan Dataset Bank Marketing. Proses penelitian mencakup pembersihan data, pemilihan fitur, transformasi logaritmik, normalisasi dengan StandardScaler, reduksi dimensi melalui Principal Component Analysis (PCA), proses pengelompokan dengan K-Medoids, dan evaluasi menggunakan Silhouette Score. Hasil dari pengujian mengindikasikan bahwa jumlah kluster yang paling optimal diperoleh pada $K=2$, dengan nilai Silhouette Score sebesar 0,6423, yang menunjukkan bahwa kualitas kluster tersebut cukup baik. Proses pengelompokan ini berhasil mengkategorikan nasabah ke dalam dua kelompok utama yang didasarkan pada karakteristik keuangan dan riwayat interaksi pemasaran. Perbedaan karakteristik antara kluster menyiratkan bahwa atribut balance, duration, pdays, dan previous berkontribusi dalam pembentukan kelompok nasabah. Temuan dari penelitian ini menegaskan bahwa algoritma K-Medoids merupakan alat yang efisien untuk segmentasi nasabah dan dapat membantu dalam merancang strategi pemasaran yang lebih terarah di sektor perbankan.

Kata Kunci: K-Medoids; Clustering; Segmentasi Nasabah; Bank Marketing; Data Mining

Abstract—*Developments in information technology have increased the use of data to assist decision-making processes in the banking industry. One application is in customer segmentation to gain a deeper understanding of customer characteristics. The purpose of this study is to apply the K-Medoids algorithm in customer segmentation using the Bank Marketing Dataset. The research process includes data cleaning, feature selection, logarithmic transformation, normalization with StandardScaler, dimensionality reduction through Principal Component Analysis (PCA), clustering with K-Medoids, and evaluation using the Silhouette Score. The test results indicate that the most optimal number of clusters is obtained at $K = 2$, with a Silhouette Score value of 0.6423, which indicates that the quality of the clusters is quite good. This clustering process successfully categorizes customers into two main groups based on financial characteristics and marketing interaction history. The differences in characteristics between clusters imply that the attributes balance, duration, pdays, and previous contribute to the formation of customer groups. The findings of this study confirm that the K-Medoids algorithm is an efficient tool for customer segmentation and can help in designing more targeted marketing strategies in the banking sector.*

Keywords: K-Medoids; Clustering; Customer Segmentation; Bank Marketing; Data Mining

1. PENDAHULUAN

Perubahan dalam teknologi informasi telah menyebabkan pertumbuhan signifikan pada volume data yang dihasilkan oleh berbagai bidang industri, terutama dalam dunia perbankan. Data yang disimpan dalam sistem bank tidak hanya berfungsi sebagai catatan transaksi, tetapi juga bisa digunakan untuk menciptakan informasi yang berharga dalam membantu proses pengambilan keputusan. Penggunaan data dalam skala besar melalui metode data mining memfasilitasi organisasi untuk mengidentifikasi pola, hubungan, dan karakteristik tertentu yang dapat dijadikan landasan dalam merumuskan strategi bisnis yang lebih efisien. (Karimah & Pahira, 2022).

Salah satu pemanfaatan data mining dalam sektor perbankan adalah segmentasi nasabah. Segmentasi nasabah merupakan proses pengelompokan pelanggan berdasarkan karakteristik yang dimiliki sehingga setiap kelompok memiliki kesamaan atribut dan perilaku tertentu. Melalui segmentasi yang tepat, perusahaan dapat menyusun strategi pemasaran yang lebih terarah dan sesuai dengan kebutuhan masing-masing kelompok pelanggan. Pendekatan ini telah banyak digunakan untuk meningkatkan efektivitas pemasaran serta membantu perusahaan memahami karakteristik

konsumennya secara lebih mendalam (Pelanggan & Masa, 2024). Dalam data mining, segmentasi nasabah umumnya dilakukan menggunakan teknik clustering. Clustering adalah teknik pembelajaran tanpa pengawasan yang bertujuan untuk mengelompokkan data berdasar kesamaan karakteristik yang dimiliki. Objek yang tergabung dalam satu kelompok memiliki tingkat kesamaan yang signifikan, sementara objek di kelompok lain menunjukkan karakteristik yang lebih bervariasi. Metode pengelompokan menjadi salah satu pilihan yang umum digunakan dalam eksplorasi data karena kemampuannya dalam mengidentifikasi struktur dan pola tersembunyi dalam kumpulan data yang besar (Rohman & Wibowo, 2024).

Algoritma clustering yang sering dipilih adalah K-Means karena proses perhitungannya yang mudah dan efisien. Namun, K-Means mengalami kesulitan saat menghadapi data yang memiliki nilai ekstrim (outlier) karena pusat kluster ditentukan berdasarkan rata-rata (centroid). Keberadaan outlier dapat mengubah posisi centroid yang berakibat pada hasil pengelompokan yang kurang akurat. Sebagai pilihan lain, K-Medoids memberikan pendekatan yang lebih kuat terhadap kehadiran outlier. Berbeda dengan K-Means, K-Medoids memakai objek data yang sebenarnya (medoid) sebagai pusat kluster, sehingga lebih konsisten dalam menggambarkan fitur-fitur data (Centroid et al., 2024). Pemanfaatan medoid menjadikan hasil pengelompokan lebih mencerminkan keadaan sebenarnya dan lebih tahan terhadap data yang tidak sesuai dengan pola yang umum. (R, 2023). Selain itu, algoritma K-Medoids juga telah digunakan dalam berbagai penelitian segmentasi pelanggan karena mampu menghasilkan pengelompokan yang lebih sesuai dengan karakteristik data yang dianalisis (Fikri et al., 2023).

Keberhasilan dalam proses pengelompokan tidak semata-mata bergantung pada algoritma yang diterapkan, namun juga pada mutu dari cluster yang terbentuk. Karena itu, diperlukan pendekatan evaluasi untuk menilai tingkat keterikatan dan pemisahan di antara cluster yang ada. Salah satu pendekatan evaluasi yang umum diterapkan adalah Skor Silhouette, karena dapat memberikan gambaran tentang kualitas hasil pengelompokan berdasarkan kedekatan objek dengan cluster yang sesuai dibandingkan dengan cluster yang berbeda (Ardana et al., 2024). Selain itu, validitas cluster juga menjadi faktor penting dalam memastikan bahwa kelompok yang terbentuk benar-benar merepresentasikan karakteristik data yang berbeda (Siregar, 2024). Dataset Bank Marketing adalah salah satu kumpulan data yang sering dimanfaatkan dalam studi pengolahan data karena berisi berbagai elemen yang merepresentasikan atribut nasabah, termasuk usia, jenis pekerjaan, status pernikahan, tingkat pendidikan, saldo akun, dan catatan interaksi pemasaran yang telah dilakukan oleh bank. Data ini memungkinkan proses pembagian nasabah untuk menemukan kelompok pelanggan yang memiliki kesamaan karakteristik yang dapat dijadikan landasan dalam merancang strategi pemasaran yang lebih efisien.

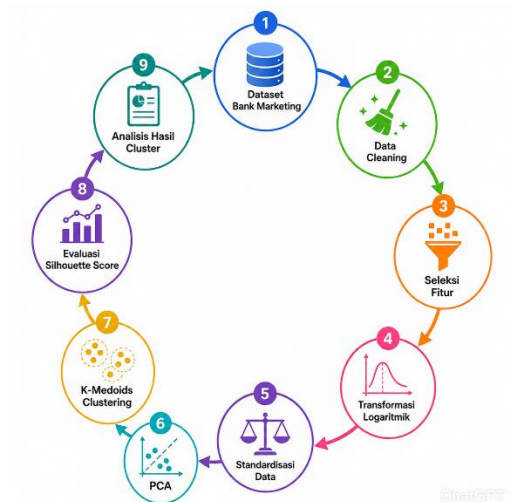
Berdasarkan studi literatur, sebagian besar penelitian pada dataset Bank Marketing lebih banyak berfokus pada pendekatan klasifikasi untuk memprediksi keberhasilan pemasaran produk deposito. Sementara itu, penelitian yang menerapkan algoritma K-Medoids untuk segmentasi nasabah masih relatif terbatas. Sebenarnya, hasil dari segmentasi pelanggan dapat menyajikan informasi krusial bagi bank untuk mengenali sifat-sifat konsumen dan merumuskan strategi promosi yang lebih efektif. Penelitian sebelumnya menunjukkan bahwa algoritma K-Medoids memiliki kemampuan yang baik dalam proses segmentasi pelanggan karena mampu menghasilkan kelompok yang representatif berdasarkan karakteristik data yang dimiliki (Centroid et al., 2024). Oleh karena itu, penelitian ini menerapkan algoritma K-Medoids pada dataset Bank Marketing untuk mengelompokkan nasabah berdasarkan karakteristik yang dimiliki. Hasil dari pemisahan data yang didapatkan diharapkan memfasilitasi lembaga perbankan untuk lebih mengenali ciri-ciri pelanggan dengan lebih baik dan mendukung pengembangan rencana pemasaran yang lebih efisien.

Penelitian ini bertujuan untuk menerapkan algoritma K-Medoids dalam proses segmentasi nasabah pada dataset Bank Marketing serta mengevaluasi kualitas cluster yang dihasilkan menggunakan metode *Silhouette Score*. Selain itu, penelitian ini juga bertujuan untuk menganalisis karakteristik masing-masing cluster yang terbentuk sehingga dapat memberikan rekomendasi strategi pemasaran yang sesuai bagi pihak perbankan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan data mining dengan teknik clustering untuk melakukan segmentasi nasabah berdasarkan karakteristik yang terdapat pada Dataset Bank Marketing. Metode yang digunakan adalah algoritma K-Medoids karena memiliki kemampuan yang lebih baik dalam menangani data yang mengandung nilai ekstrem (*outlier*) dibandingkan metode berbasis centroid seperti K-Means (Centroid et al., 2024). Tahapan penelitian meliputi pengumpulan data, *preprocessing* data, reduksi dimensi menggunakan Principal Component Analysis (PCA), proses clustering menggunakan algoritma K-Medoids, evaluasi cluster menggunakan Silhouette Score, serta analisis hasil cluster (Rohman & Wibowo, 2024). Secara umum, tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

2.2 Dataset Penelitian

Dataset yang diterapkan dalam studi ini adalah Bank Marketing Dataset yang sering dipakai dalam penelitian terkait data mining dan machine learning untuk memahami perilaku nasabah bank. Koleksi data ini mencakup 4.521 data nasabah dengan 17 atribut yang menjelaskan sifat demografis, situasi keuangan, dan rekam jejak komunikasi pemasaran yang dilakukan oleh bank. Atribut yang tersedia dalam dataset meliputi:

Tabel 1 Atribut Nasabah

No	Atribut
1	Age
2	Job
3	Marital
4	Education
5	Default
6	Balance
7	Housing
8	Loan
9	Contact
10	Day
11	Month
12	Duration
13	Campaign

14	Pdays
15	Previous
16	Poutcome
17	Y

Variabel target y tidak digunakan dalam penelitian ini karena metode clustering termasuk ke dalam kategori *unsupervised learning* yang tidak memerlukan label kelas dalam proses pengelompokan data.

2.3 Data Preprocessing

Tahap *preprocessing* dilakukan untuk memperbaiki mutu data sebelum clustering dilakukan. Langkah-langkah pra-pemrosesan meliputi beberapa tahapan sebagai berikut.

1. Data Cleaning

Pada tahap ini dilakukan evaluasi terhadap informasi yang tidak tepat serta penyesuaian terhadap nilai atribut yang spesifik. Selain itu, atribut target y dihapus karena tidak digunakan dalam proses clustering.

2. Seleksi Fitur

Tidak seluruh atribut pada dataset digunakan dalam proses clustering. Berdasarkan karakteristik data dan tujuan segmentasi nasabah, penelitian ini menggunakan empat atribut utama yaitu:

- a. Balance
- b. Duration
- c. Pdays
- d. Previous

Keempat atribut tersebut dipilih karena dianggap mampu merepresentasikan aktivitas dan riwayat interaksi nasabah dengan pihak bank dalam kegiatan pemasaran.

3. Transformasi Logaritmik

Data pada atribut *balance*, *duration*, dan *pdays* memiliki distribusi yang cenderung tidak normal dan mengandung nilai yang sangat bervariasi. Oleh karena itu dilakukan transformasi logaritmik (*log transformation*) untuk mengurangi tingkat kemencengan data (*skewness*) serta meminimalkan pengaruh nilai ekstrem terhadap proses *clustering*.

4. Normalisasi Data

Setelah dilakukan transformasi logaritmik, data dinormalisasi menggunakan metode *StandardScaler*. Proses ini dilakukan untuk memastikan bahwa setiap karakteristik memiliki rata-rata yang hampir sama dengan nol dan deviasi standar satu, sehingga tidak ada atribut yang mendominasi saat menghitung jarak antara data.

5. Reduksi Dimensi Menggunakan Principal Component Analysis (PCA)

Setelah tahap normalisasi, langkah selanjutnya adalah pengurangan dimensi melalui metode *Principal Component Analysis* (PCA). PCA digunakan untuk merangkum informasi dari beberapa atribut menjadi dua komponen utama yang tetap mempertahankan sebagian besar variasi data (Rohman & Wibowo, 2024). Hasil daripada PCA kemudian digunakan sebagai input dalam proses pengelompokan dengan bantuan algoritma K-Medoids.

2.4 Algoritma K-Medoids

K-Medoids merupakan algoritma clustering yang mengelompokkan objek data berdasarkan tingkat kemiripan tertentu dengan menggunakan objek data aktual (*medoid*) sebagai pusat cluster (Centroid et al., 2024). Berbeda dengan K-Means yang menggunakan nilai rata-rata (*centroid*) sebagai pusat cluster, K-Medoids memilih salah satu objek dalam cluster sebagai representasi kelompok sehingga lebih tahan terhadap pengaruh *outlier* (R, 2023).

Tahapan algoritma K-Medoids dalam penelitian ini adalah sebagai berikut:

1. Menentukan jumlah cluster (K).
2. Memilih objek awal sebagai medoid.
3. Menghitung jarak setiap data dengan medoid terdekat.
4. Mengelompokkan data berdasarkan jarak minimum.

5. Melakukan pertukaran medoid dengan objek lain untuk memperoleh biaya (*cost*) yang lebih kecil.
6. Mengulangi proses hingga tidak terjadi perubahan medoid.

Secara matematis, total biaya clustering dapat dinyatakan sebagai:

$$Cost = \sum_{i=1}^n d(x_i, m_j)$$

Keterangan:

x_i = data ke-i

m_j = medoid cluster ke-j

$d(x_i, m_j)$ = jarak antara data dan medoid

Tujuan utama algoritma adalah meminimalkan total biaya sehingga setiap cluster memiliki tingkat kemiripan internal yang tinggi (Afandi et al., 2022).

2.5 Evaluasi Cluster Menggunakan Silhouette Score

Evaluasi kualitas cluster dilakukan menggunakan metode **Silhouette Score**. Metode ini diterapkan untuk menilai sejauh mana kesamaan suatu data dengan cluster di mana data itu berada jika dibandingkan dengan cluster lainnya. (Ardana et al., 2024).

Nilai Silhouette Score dihitung menggunakan persamaan:

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

dengan:

$a(i)$ = rata-rata jarak data terhadap anggota dalam cluster yang sama

$b(i)$ = rata-rata jarak data terhadap cluster terdekat lainnya

Nilai Silhouette Score berada pada rentang -1 hingga 1. Semakin mendekati nilai 1 maka kualitas cluster semakin baik, sedangkan nilai yang mendekati 0 menunjukkan adanya tumpang tindih antar cluster (Yosia & Siregar, 2024).

Dalam penelitian ini pengujian dilakukan pada nilai $K = 2$ sampai $K = 10$. Nilai K yang menghasilkan Silhouette Score tertinggi dipilih sebagai jumlah cluster optimal.

2.6 Visualisasi Cluster Menggunakan PCA

Hasil clustering yang diperoleh divisualisasikan menggunakan *Principal Component Analysis* (PCA) dalam bentuk grafik dua dimensi. Visualisasi ini digunakan untuk melihat pola penyebaran data serta tingkat pemisahan antar cluster yang terbentuk (Rohman & Wibowo, 2024).

Selain visualisasi, dilakukan analisis karakteristik setiap cluster berdasarkan nilai rata-rata atribut pada masing-masing kelompok. Analisis ini bertujuan untuk mengidentifikasi karakteristik nasabah pada setiap cluster agar bisa dijadikan sebagai landasan dalam merancang strategi pemasaran yang lebih efisien dan terarah.

3. ANALISA DAN PEMBAHASAN

3.1 Hasil Preprocessing Data

Tahap awal penelitian dilakukan dengan proses preprocessing untuk mempersiapkan data sebelum dilakukan clustering. Dataset Bank Marketing yang digunakan terdiri atas 4.521 data nasabah dengan 17 atribut. Karena penelitian menggunakan pendekatan unsupervised learning, atribut target y tidak digunakan dalam proses analisis dan dihapus dari dataset. Selanjutnya dilakukan proses seleksi fitur untuk memilih atribut yang paling relevan dalam proses segmentasi nasabah. Berdasarkan karakteristik data dan tujuan penelitian, atribut yang digunakan meliputi *balance*, *duration*, *pdays*, dan *previous*. Keempat atribut tersebut dipilih karena mampu merepresentasikan kondisi finansial serta riwayat interaksi pemasaran nasabah dengan pihak bank.

Tahap berikutnya adalah transformasi logaritmik (*log transformation*) pada atribut *balance*, *duration*, dan *pdays* untuk mengurangi tingkat kemencengan distribusi data (*skewness*) serta meminimalkan pengaruh nilai ekstrem. Setelah itu dilakukan normalisasi menggunakan metode *StandardScaler* agar seluruh atribut berada pada skala yang seimbang. Data yang telah disesuaikan kemudian mengalami pengurangan dimensi dengan menggunakan metode *Principal Component Analysis* (PCA) untuk menjadi dua komponen utama. Hasil reduksi dimensi tersebut digunakan sebagai data masukan pada proses clustering menggunakan algoritma *K-Medoids*.

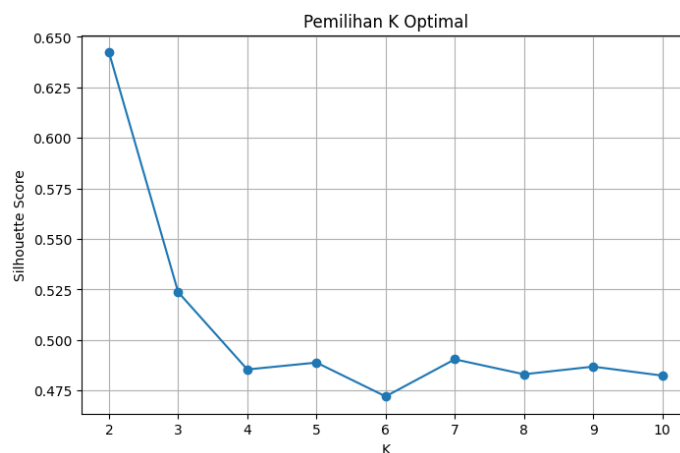
3.2 Penentuan Jumlah Cluster Optimal

Penetapan jumlah cluster dilakukan dengan cara menggunakan metode *Silhouette Score*. Pengujian dilaksanakan pada beragam jumlah kluster, yaitu $K = 2$ sampai $K = 10$. Nilai *Silhouette Score* berfungsi untuk menilai kualitas cluster yang dihasilkan.

Tabel 2 Hasil Pengujian *Silhouette Score*

Jumlah Cluster (K)	Silhouette Score
2	0.6423
3	0.5237
4	0.4854
5	0.4851
6	0.5054
7	0.4908
8	0.4946
9	0.4762
10	0.4896

Berdasarkan analisis yang dilakukan, skor *Silhouette* tertinggi ditemukan pada dua cluster, dengan nilai mencapai 0,6423. Maka, jumlah cluster yang diambil untuk penelitian ini adalah $K = 2$. Skor *Silhouette* sebesar 0,6423 menandakan bahwa kualitas cluster yang terbentuk adalah baik. Angka tersebut menunjukkan bahwa setiap objek dalam cluster memiliki kesamaan yang tinggi dengan anggota cluster yang sejenis, serta pemisahan yang jelas dari cluster lain. Dengan demikian, hasil pengelompokan yang didapat dapat dianggap mewakili karakteristik nasabah dalam dataset Bank Marketing.



Gambar 2 Grafik Pemilihan Jumlah Cluster Optimal Menggunakan *Silhouette Score*

3.3 Hasil Clustering Menggunakan *K-Medoids*

Setelah jumlah cluster yang terbaik ditentukan, langkah clustering dilanjutkan dengan menerapkan algoritma *K-Medoids* dengan nilai $K = 2$. Hasil dari pengelompokan memperlihatkan bahwa data pelanggan telah dipisahkan menjadi dua kelompok utama.

Tabel 3 Distribusi Jumlah Anggota Cluster

Cluster	Jumlah Data	Persentase
Cluster 0	812	17,96 %
Cluster 1	3709	82,04 %
Total	4521	100 %

Berdasarkan hasil clustering, informasi nasabah terbagi ke dalam dua kategori dengan sifat yang berbeda. Sebaran jumlah anggota dalam setiap kelompok menunjukkan bahwa ada satu kelompok yang memiliki lebih banyak anggota dibandingkan kelompok lainnya. Situasi ini menandakan terdapat variasi dalam pola karakteristik nasabah yang telah berhasil dikenali oleh algoritma K-Medoids.

3.4 Analisis Karakteristik Cluster

Karakteristik masing-masing cluster dianalisis berdasarkan nilai rata-rata atribut yang digunakan dalam proses *clustering*, yaitu *balance*, *duration*, *pdays*, dan *previous*.

Tabel 4 Total Nilai Karakteristik cluster

Variabel	Cluster 0	Cluster 1
Balance	1644.72	1374.04
Duration	273.19	261.94
Pdays	225.96	-1.00
Previous	3.01	0.00

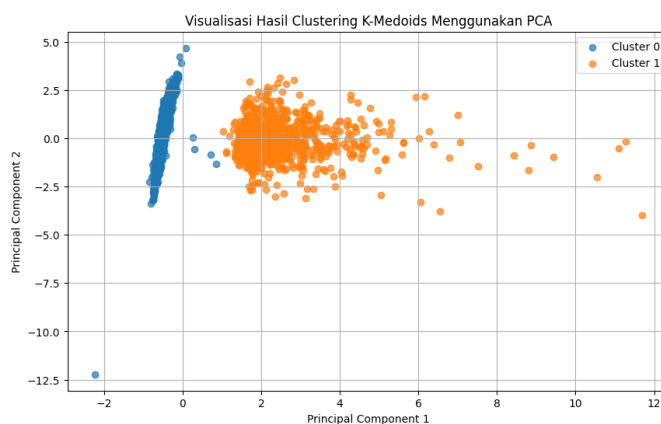
1. Cluster 0

Cluster 0 merupakan kelompok nasabah dengan karakteristik tertentu berdasarkan nilai rata-rata atribut yang dimiliki. Kelompok ini menunjukkan pola interaksi pemasaran yang berbeda dibandingkan cluster lainnya. Nilai rata-rata pada atribut *balance*, *duration*, *pdays*, dan *previous* menggambarkan perilaku serta riwayat komunikasi nasabah terhadap program pemasaran yang dilakukan oleh pihak bank. Dari perspektif pemasaran, cluster ini dapat dimanfaatkan sebagai landasan dalam merancang strategi promosi yang sesuai dengan ciri-ciri nasabah yang ada di dalam kluster tersebut.

2. Cluster 1

Cluster 1 menunjukkan ciri yang berbeda dibandingkan dengan Cluster 0. Perbedaan ini terlihat melalui rata-rata nilai atribut yang diterapkan dalam proses pengelompokan. Fluktuasi nilai pada atribut *pdays* dan *previous* mengindikasikan adanya perbedaan tingkat interaksi pemasaran yang telah dilakukan oleh bank kepada kelompok nasabah ini. Data tersebut bisa digunakan sebagai acuan dalam merancang strategi pemasaran yang lebih terfokus, seperti program untuk mempertahankan pelanggan atau aktivitas remarketing kepada nasabah yang sudah memiliki pengalaman interaksi sebelumnya.

3.5 Visualisasi Hasil Clustering



Gambar 3 Visualisasi Hasil Clustering K-Medoids Menggunakan PCA

Visualisasi hasil clustering dilakukan menggunakan metode *Principal Component Analysis* (PCA) untuk mengecilkan dimensi data menjadi dua komponen inti. Visualisasi yang dihasilkan memperlihatkan distribusi data ke dalam dua kelompok yang terbentuk setelah pengelompokan menggunakan algoritma K-Medoids. Dari visualisasi PCA, tampak bahwa data terdistribusi ke dalam dua kelompok utama yang relatif terpisah satu sama lain. Mayoritas anggota kelompok terkumpul di wilayah tertentu, sedangkan kelompok yang lain membentuk variasi yang berbeda. Pola ini mendukung temuan evaluasi Silhouette Score yang menunjukkan kualitas kelompok yang baik dengan nilai sebesar 0,6423.

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Medoids mampu melakukan segmentasi nasabah pada Dataset Bank Marketing dengan baik. Penggunaan seleksi fitur, transformasi logaritmik, normalisasi data, serta reduksi dimensi menggunakan PCA terbukti mampu meningkatkan kualitas cluster yang dihasilkan.

Evaluasi menggunakan Silhouette Score menghasilkan nilai sebesar 0,6423, yang menunjukkan bahwa kualitas cluster berada pada kategori baik. Nilai tersebut mengindikasikan bahwa setiap cluster memiliki tingkat kohesi yang tinggi dan tingkat separasi yang cukup jelas terhadap cluster lainnya. Dari perspektif bisnis, hasil segmentasi yang diperoleh dapat membantu pihak bank dalam memahami karakteristik nasabah secara lebih terstruktur. Informasi tersebut dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih efektif, efisien, dan tepat sasaran. Selain itu, hasil penelitian juga menunjukkan bahwa algoritma K-Medoids mampu menjadi alternatif metode clustering yang efektif untuk mengelompokkan nasabah berdasarkan karakteristik dan riwayat interaksi pemasaran yang dimiliki.

4. KESIMPULAN

4.1 Kesimpulan

Penelitian ini berhasil menerapkan algoritma K-Medoids untuk melakukan segmentasi pelanggan pada Database Pemasaran Perbankan. Proses penelitian dimulai dengan langkah-langkah preprocessing yang mencakup penghapusan atribut sasaran, pemilihan fitur, transformasi logaritmik, normalisasi data dengan menggunakan StandardScaler, dan juga reduksi dimensi dengan Principal Component Analysis (PCA). Langkah-langkah ini diambil untuk memperbaiki kualitas data sebelum proses pengelompokan dilaksanakan. Dari hasil pengujian menggunakan metode Silhouette Score, ditemukan bahwa jumlah cluster yang optimal adalah 2 cluster dengan nilai Silhouette Score sebesar 0,6423. Nilai tersebut menandakan bahwa cluster yang dihasilkan memiliki kualitas yang baik, di mana tingkat kohesi tinggi dan tingkat pemisahan antara cluster cukup jelas. Oleh karena itu, algoritma K-Medoids efektif dalam membuat kelompok data yang mencerminkan karakteristik pelanggan pada dataset yang digunakan. Hasil pengelompokan menunjukkan bahwa

data pelanggan berhasil dibagi menjadi dua kelompok utama yang memiliki karakteristik berbeda yang ditentukan oleh atribut yang dipakai dalam proses segmentasi, yaitu balance, duration, pdays, dan previous. Perbedaan karakteristik itu menegaskan bahwa riwayat interaksi pemasaran dan kondisi keuangan pelanggan menjadi faktor penting dalam pembentukan cluster.

Berdasarkan temuan penelitian, disimpulkan bahwa algoritma K-Medoids memiliki efektivitas dalam melakukan segmentasi nasabah dalam Dataset Bank Marketing. Hasil dari segmentasi ini dapat digunakan sebagai landasan untuk merumuskan strategi pemasaran yang lebih tepat, efisien, dan efektif. Di samping itu, nilai Silhouette Score yang mencapai 0,6423 menunjukkan bahwa metodologi preprocessing yang diterapkan berhasil meningkatkan kualitas cluster, sehingga menghasilkan segmentasi yang lebih baik dan lebih mudah untuk dipahami, mendukung proses pengambilan keputusan di sektor perbankan.

4.2 Saran

Berdasarkan hasil segmentasi yang diperoleh, pihak perbankan disarankan untuk menerapkan strategi pemasaran yang disesuaikan dengan karakteristik masing-masing kelompok nasabah. Nasabah yang memiliki tingkat interaksi pemasaran rendah dapat diberikan informasi produk secara lebih intensif dan terarah guna meningkatkan ketertarikan terhadap layanan perbankan yang ditawarkan. Sementara itu, nasabah yang memiliki riwayat interaksi pemasaran lebih tinggi dapat menjadi sasaran program pemeliharaan hubungan pelanggan (*customer retention*) melalui penawaran produk yang relevan dengan kebutuhan mereka. Dengan pendekatan pemasaran yang lebih tersegmentasi, diharapkan efektivitas komunikasi pemasaran dapat meningkat serta memberikan manfaat yang lebih optimal bagi nasabah maupun pihak bank.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penugasan jurnal yang berjudul “Penerapan Metode K-Medoids untuk Segmentasi Nasabah pada Dataset Bank Marketing” dapat diselesaikan dengan baik. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Ibu Mufidah Karimah, S.Kom., M.Kom. selaku dosen pengampu mata kuliah Data Mining yang telah memberikan arahan, bimbingan, masukan, serta motivasi selama proses penyusunan jurnal ini. Bimbingan yang diberikan sangat membantu penulis dalam memahami konsep data mining, proses analisis data, hingga penyusunan laporan penelitian secara sistematis dan ilmiah.

REFERENCES

- Ardana, C. H., Aldita, A., Aisyah, A., Khoyum, A., & Faisal, M. (2024). *Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering*. 9(1), 1–9.
- Centroid, I., Algorithm, O. K., Particle, U., & Optimization, S. (2024). *Optimasi Centroid Awal Algoritma K-Medoids Menggunakan Particle Swarm Optimization Untuk Segmentasi Customer*. 23(1), 221–232.
- Fikri, A., Hutabarat, B. F., & Khaira, U. (2023). *Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology*. 17(2), 228–239.
- Karimah, M., & Pahira, W. (2022). *Penerapan Algoritma Apriori Pada Prediksi Penjualan Tanaman Hias Bromelia (Studi Kasus : Karimah Flora)*. 7(1), 68–78.
- Pelanggan, S., & Masa, D. I. (2024). *INTI NUSA MANDIRI*. 18(2), 192–200.
- R, D. G. U. I. (2023). *Implementasi k - medoids dan model weighted-length recency frequency monetary (w-lrfm) untuk segmentasi pelanggan dilengkapi gui r 1,2,3*. 11, 429–438. <https://doi.org/10.14710/J.GAUSS.11.3.429-438>
- Rohman, N., & Wibowo, A. (2024). *Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall*. 7(1), 49–58.
- Siregar, B. (2024). *Comparative Analysis of K-Means and K-Medoids Algorithms for Product Sales Clustering and Customer*. 7(2), 360–370.