

Perbandingan Kinerja Algoritma *K-Means* dan DBSCAN dalam Segmentasi Nasabah Berdasarkan Data Pemasaran Bank

Tria Setyani¹, Helma Nopijani Heidy¹, Kevinda Sari^{1,*}, Yulia Indriani¹, Heni Sulistiani¹

¹Fakultas Teknik dan Ilmu Komputer, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹tria_setyani@teknokrat.ac.id, ²helma_nopijani_heidy@teknokrat.ac.id, ^{3,*}kevinda_sari@teknokrat.ac.id,

⁴yulia_indriani@teknokrat.ac.id, ⁵henisulistiani@teknokrat.ac.id

Email Penulis Korespondensi: kevinda_sari@teknokrat.ac.id

Abstrak-Segmentasi nasabah merupakan salah satu pendekatan penting dalam industri perbankan untuk meningkatkan efektivitas strategi pemasaran dan pengelolaan risiko. Penelitian ini memiliki tujuan untuk membandingkan performa dua algoritma clustering yang umum digunakan, yaitu *K-Means* dan DBSCAN, dalam melakukan segmentasi nasabah berdasarkan data pemasaran bank. Data yang digunakan berasal dari dataset pemasaran bank yang tersedia di Kaggle, mencakup atribut seperti usia, jenis pekerjaan, status perkawinan, tingkat pendidikan, saldo tabungan, serta riwayat kampanye sebelumnya. Proses analisis melibatkan pra-pemrosesan data, ekstraksi fitur, standardisasi data, dan implementasi algoritma clustering. Evaluasi hasil dilakukan menggunakan metrik Silhouette Score menunjukkan bahwa algoritma *K-Means* menghasilkan nilai Nilai Silhouette Score yang lebih tinggi serta Davies-Bouldin Index (DBI) yang lebih rendah dibandingkan DBSCAN menunjukkan bahwa *K-Means* mampu menghasilkan kluster yang lebih solid (kompak) dan memiliki batas pemisah yang lebih tegas antar kelompok data. Namun, DBSCAN menunjukkan keunggulan dalam kemampuan mendeteksi noise dan mengelompokkan data dengan bentuk tidak teratur. Dengan demikian, pemilihan algoritma terbaik sangat bergantung pada karakteristik data dan tujuan segmentasi. Untuk data pemasaran bank dengan distribusi relatif seragam, *K-Means* menjadi pilihan yang lebih optimal.

Kata Kunci: Segmentasi Nasabah; *Clustering*; *K-Means*; DBSCAN; Data Pemasaran Bank; *Unsupervised Learning*.

Abstract-Customer segmentation is one of the important approaches in the banking industry to improve the effectiveness of marketing strategies and risk management. This study aims to compare the performance of two commonly used clustering algorithms, namely *K-Means* and DBSCAN, in segmenting customers based on bank marketing data. The data used comes from the bank marketing dataset available on Kaggle, including attributes such as age, type of employment, marital status, education level, savings balance, and previous campaign history. The analysis process includes data pre-processing, feature extraction, data standardization, and implementation of the clustering algorithm. Evaluation of the results using the Silhouette Score metric shows that the *K-Means* algorithm produces a higher Silhouette Score value and a lower Davies-Bouldin Index (DBI) than DBSCAN indicating that *K-Means* is able to produce more solid (compact) clusters and has clearer boundaries between data groups. However, DBSCAN shows advantages in the ability to detect noise and group data with irregular shapes. Thus, the selection of the best algorithm is highly dependent on the characteristics of the data and the purpose of the segmentation. For bank marketing data with a relatively uniform distribution, *K-Means* is a more optimal choice.

Keywords: Customer Segmentation; Clustering; *K-Means*; DBSCAN; Bank Data Marketing; *Unsupervised Learning*.

1. PENDAHULUAN

Didalam era digital ini, industri perbankan menghadapi tantangan besar dalam mengelola dan memanfaatkan data nasabah dengan efektif. Seiring dengan pertumbuhan jumlah nasabah dan transaksi digital, volume data yang dihasilkan juga semakin besar. Data ini mencakup informasi dasar seperti usia, pekerjaan, dan pendapatan, serta sejarah transaksi, perilaku penggunaan produk bank, dan interaksi dengan layanan pelanggan. Mengelola dan menganalisis data ini dengan tepat dapat memberikan wawasan berharga bagi bank dalam mengambil keputusan strategis, meningkatkan efisiensi pemasaran, dan mengelola risiko dengan lebih baik. Salah satu pendekatan yang sering digunakan dalam mengelola data nasabah adalah segmentasi. Segmentasi nasabah adalah proses pengelompokan nasabah berdasarkan karakteristik yang serupa, sehingga bank dapat merancang strategi pemasaran yang lebih tepat dan efektif. Segmentasi nasabah dapat membantu bank mengidentifikasi kelompok nasabah yang memiliki potensi tinggi untuk menggunakan produk dan layanan tertentu, serta kelompok nasabah yang memiliki risiko lebih tinggi dalam hal kredit atau transaksi lainnya [1] [2].

Dalam beberapa tahun terakhir, metode data mining dan machine learning semakin banyak dimanfaatkan untuk melakukan segmentasi nasabah. Dua pendekatan yang umum digunakan adalah algoritma *K-Means* dan DBSCAN. *K-Means* adalah metode klusterisasi non-hierarkis yang berfungsi untuk membagi data ke dalam dua atau lebih kelompok, dengan tujuan mengelompokkan data yang memiliki kesamaan ke dalam kluster yang sama. Sementara itu, DBSCAN merupakan algoritma klustering yang didasarkan pada kepadatan data. Berdasarkan konsep kepadatan ini, DBSCAN mengklasifikasikan setiap data menjadi tiga kategori, yaitu titik inti (core), titik tepi (border), dan data yang dianggap sebagai gangguan atau tidak termasuk kluster manapun (noise) [3][4].

Menurut penelitian terbaru, kedua algoritma ini telah banyak digunakan dalam berbagai aplikasi segmentasi yang menunjukkan bahwa *K-Means* lebih efisien dalam mengelompokkan data dengan jarak yang seragam, sedangkan DBSCAN lebih fleksibel dalam mengelompokkan data dengan keragaman yang lebih tinggi. penelitian lain menunjukkan bahwa DBSCAN lebih efektif dalam mengidentifikasi kelompok dengan bentuk yang tidak teratur dan data outlier. Namun, hingga saat ini masih belum ditemukan kajian yang secara menyeluruh membandingkan performa kedua algoritma tersebut dalam melakukan segmentasi nasabah berdasarkan data pemasaran perbankan. [5] [6] [7].

Dalam konteks industri perbankan, segmentasi nasabah berdasarkan data pemasaran bank dapat membantu bank dalam merancang strategi pemasaran yang lebih tepat dan efektif. Data pemasaran bank mencakup informasi tentang nasabah seperti usia, pekerjaan, pendapatan, sejarah transaksi, dan interaksi dengan layanan pelanggan. Mengelompokkan nasabah berdasarkan karakteristik ini dapat membantu bank mengidentifikasi kelompok nasabah yang memiliki potensi tinggi untuk menggunakan produk dan layanan tertentu, serta kelompok nasabah yang memiliki risiko lebih tinggi dalam hal kredit atau transaksi lainnya. Oleh karena itu, studi ini bertujuan untuk melakukan perbandingan kinerja antara algoritma K-Means dan DBSCAN dalam proses segmentasi nasabah berdasarkan data pemasaran bank. Analisis difokuskan pada efektivitas masing-masing algoritma dalam mengelompokkan data nasabah serta mengevaluasi kemampuannya dalam mengenali kelompok nasabah dengan pola perilaku yang serupa.

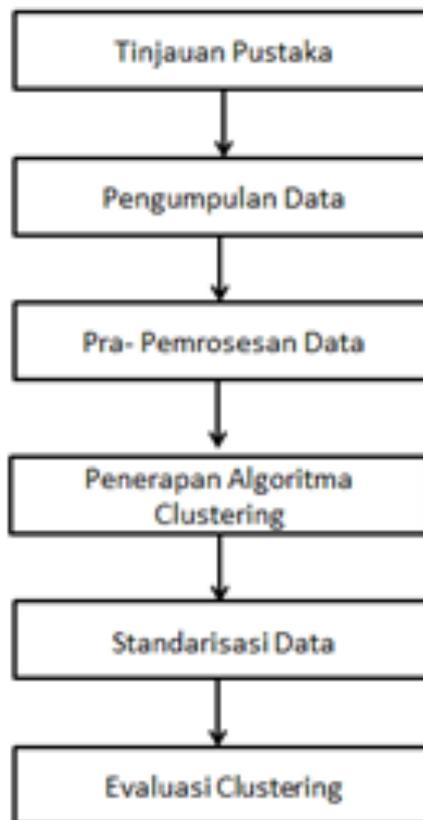
2. METODE PENELITIAN

2.1 Kerangka Dasar Penelitian

Kerangka dasar pada penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa algoritma K-Means dan DBSCAN dalam melakukan segmentasi nasabah berdasarkan data pemasaran bank. Penelitian ini memanfaatkan data yang diperoleh dari Kaggle, berupa dataset pemasaran bank yang mencakup berbagai atribut terkait profil dan interaksi nasabah, seperti usia, jenis pekerjaan, status perkawinan, tingkat pendidikan, saldo tabungan, kepemilikan pinjaman, serta hasil kampanye pemasaran sebelumnya. Pada penelitian ini penulis menggunakan metode pendekatan *Knowledge Discovery in Database (KDD)*. Pendekatan dalam penelitian ini mencakup studi literatur, pengumpulan data, tahap pra-pemrosesan, standarisasi data, serta penerapan teknik klusterisasi menggunakan algoritma K-Means dan DBSCAN. Proses evaluasi terhadap hasil klusterisasi dilakukan dengan memanfaatkan metrik Silhouette Score dan Davies-Bouldin Index (DBI)

2.2 Tahapan penelitian

Tahapan penelitian dapat dilihat pada gambar 1 dibawah ini.



Gambar 1. Tahapan Penelitian

2.3 Data Selection

Data yang digunakan dalam penelitian ini merupakan data pemasaran bank. Data diperoleh dari dataset Pemasaran Marketing Bank di Kaggle: <https://www.kaggle.com/datasets/rounakbanik/bank-marketing>. Data tersebut dipilih dan diseleksi terlebih dahulu guna memastikan kesesuaiannya serta menjaga kualitas sebelum dilibatkan dalam proses *clustering*. Berikut adalah tabel atribut dalam dataset pemasaran bank, dapat dilihat pada tabel 1:

Tabel 1. Atribut dataset

NO	ATRIBUT	DESKRIPSI
1	age	Usia nasabah
2	job	Jenis pekerjaan
3	marital	Status perkawinan
4	education	Tingkat pendidikan
5	default	Apakah nasabah memiliki kredit macet
6	balance	Saldo tabungan
7	housing	Apakah nasabah memiliki pinjaman perumahan
8	loan	Apakah nasabah memiliki pinjaman pribadi
9	contact	Metode kontak
10	day	Hari terakhir kontak
11	month	Bulan terakhir kontak
12	duration	Durasi kontak terakhir
13	campaign	Jumlah kontak selama kampanye
14	pdays	Hari sejak kontak terakhir
15	previous	Jumlah kontak sebelum kampanye
16	poutcome	Hasil kampanye sebelumnya
17	y	Target: apakah nasabah bedanggaan produk (deposito)

2.4 Data Preprocessing

Tahap ini mencakup pembersihan data dan penanganan outlier untuk memastikan data yang digunakan dalam penelitian dapat diproses dengan baik. Proses penanganan outlier hanya dilakukan pada data yang dikelompokkan menggunakan K-Means Clustering, sedangkan untuk DBSCAN, outlier dianggap sebagai bagian dari proses clustering.

2.4.1 Ekstraksi Fitur untuk Clustering

Dataset yang telah dipilih kemudian diproses untuk menghasilkan fitur-fitur yang relevan sebagai input dalam proses *clustering*. Setiap atribut pelanggan dievaluasi dan disiapkan agar sesuai dengan kebutuhan algoritma, seperti K-Means dan DBSCAN. Proses ekstraksi ini dilakukan bahasa Python digunakan dalam proses ini, dengan bantuan library Pandas dan NumPy. Berikut adalah fitur yang dipilih sebagai input untuk proses clustering dan yang tidak digunakan:

Tabel 2. Fitur-fitur Dataset

KATEGORI	FITUR YANG DIPILIH	ALASAN PEMILIHAN
Relevan	age, job, marital, education, balance	Mewakili karakteristik demografi dan finansial nasabah
	housing, loan	Menunjukkan perilaku kredit
	duration, campaign	Berhubungan dengan interaksi pemasaran
Tidak relevan	contact, day, month, pdays, previous	Tidak signifikan dalam segmentasi perilaku jangka panjang
	poutcome	Hanya hasil kampanye sebelumnya
	y	Variabel target (tidak digunakan dalam clustering)

2.4.2 Standardisasi Data

Data yang telah diproses kemudian distandarisasi untuk memastikan semua fitur berada pada skala yang sebanding. Teknik yang digunakan meliputi *Robust Scaler* dan *Square-root Transformation*, disesuaikan dengan distribusi masing-masing fitur.

a. Robust Scaler

Digunakan untuk data yang memiliki outlier tinggi.

Rumus:

$$X_{scaled} = (X - Q1) / (Q3 - Q1)$$

Di mana:

1. Q1 = Kuartil pertama
2. Q3 = Kuartil ketiga

b. Square-root Transformation

Digunakan untuk data dengan distribusi skew positif.

Rumus:

$$X_{\text{transformed}} = \text{sqrt}(X)$$

Standarisasi ini dilakukan agar tidak ada bias dalam pengelompokan karena perbedaan skala antar fitur.

Tabel 3. Hasil Standarisasi Data

ID	AGE (ENCODED)	JOB (ENCODED)	BALANCE (ENCODED)	HOUSING (ENCODED)	LOAN (ENCODED)	DURATION (ENCODED)	CAMPAIGN (ENCODED)	Y (ENCODED)
1	0.5	2	0.3	1	0	0.7	1	1
2	0.7	3	0.4	0	1	0.6	2	0
3	0.4	1	0.2	1	0	0.8		

Catatan: Angka telah melalui proses encoding dan standarisasi menggunakan Robust Scaler.

2.4.3 Clustering

Clustering adalah tahap penting dalam analisis data tanpa supervisi, Clustering merupakan salah satu teknik dalam data mining yang berfungsi untuk mengelompokkan sekumpulan data ke dalam beberapa kelompok tertentu (cluster) [8]. Setelah data melalui proses standarisasi, pengelompokan dilakukan dengan menggunakan dua metode clustering, yaitu K-Means dan DBSCAN. Penggunaan kedua metode ini bertujuan untuk menguji serta membandingkan kinerja masing-masing dalam menghasilkan kelompok data yang optimal [8].

2.4.4 Evaluasi Clustering

Evaluasi terhadap hasil *clustering* dilakukan dengan menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI). Kedua metrik ini digunakan untuk menentukan metode *clustering* yang paling efektif dalam segmentasi nasabah berdasarkan dataset pemasaran bank.

3. HASIL DAN PEMBAHASAN

Hasil *clustering* dari masing-masing dataset menunjukkan perbedaan efektivitas antara kedua algoritma. K-Means cenderung membentuk cluster yang lebih rapat namun sensitif terhadap outlier, sedangkan DBSCAN lebih adaptif dalam mengelola cluster dengan bentuk yang tidak teratur serta lebih tangguh terhadap noise dalam data.

3.1 K-Means

K-Means adalah algoritma klustering yang digunakan untuk memisahkan objek ke dalam kelompok yang saling eksklusif di bidang data mining [9] [10]. Metode ini populer untuk analisis kluster karena efisien dalam mempartisi data, terutama pada kumpulan data besar. Berbeda dengan algoritma klasifikasi yang bersifat terawasi, K-Means merupakan pembelajaran tak terawasi, sehingga susunan kluster baru dapat diketahui setelah proses dijalankan [11]. Karena jumlah kluster belum pasti, peneliti biasanya menentukan jumlah kelompok yang diinginkan terlebih dahulu lalu melakukan beberapa percobaan hingga ditemukan pembagian kluster yang paling tepat [12].

3.2 DBSCAN

DBSCAN merupakan algoritma yang efektif dalam mengidentifikasi segmen dengan bentuk tidak beraturan dan skala besar, karena menggunakan pendekatan berbasis kerapatan (density-based). Algoritma ini juga secara eksplisit memisahkan noise dari titik data yang termasuk dalam suatu segmen, sehingga mampu melakukan pengelompokan data dengan lebih akurat, terutama dalam situasi yang mengandung noise [13].

3.3 Evaluasi Clustering

Evaluasi hasil clustering dilakukan dengan memanfaatkan dua metrik, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Kedua metrik ini digunakan untuk menilai efektivitas metode clustering dalam mengelompokkan nasabah berdasarkan data pemasaran perbankan. Silhouette Score merupakan ukuran yang mengevaluasi sejauh mana suatu data point berada dalam klusternya sendiri dan seberapa jauh jaraknya dari kluster lain. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa kluster yang terbentuk memiliki kualitas yang lebih baik [14]. Rumus Silhouette Score:

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

Dengan keterangan sebagai berikut:

s(i) : Merupakan nilai Silhouette untuk data ke-i, yang menunjukkan seberapa baik titik tersebut cocok dengan klusternya.

a(i) : Menunjukkan rata-rata jarak antara titik data ke-i dengan seluruh titik lain dalam kluster yang sama (mengukur kohesi).

b(i) : Menunjukkan rata-rata jarak antara titik data ke-i dengan seluruh titik dalam kluster terdekat lainnya (mengukur pemisahan).

Setelah semua titik memiliki nilai s(i), maka dihitung rata-ratanya untuk mendapatkan Silhouette Score total:

Silhouette Score Total = $(1/n) * \sum s(i)$

dimana:

- n : jumlah total titik data

Interpretasi:

+1 : Titik sangat cocok dalam klusternya.

0 : Titik berada di antara dua kluster.

-1 : Titik kemungkinan salah dikelompokkan.

Davies-Bouldin Index (DBI) adalah metrik yang mengukur rasio antara dalam-kluster dan antar-kluster. Nilai DBI yang lebih rendah menunjukkan kluster yang lebih baik.

Rumus DBI:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i, c_j)} \right)$$

Dimana:

K : Jumlah kluster

S_i : Ketidakkompakan kluster i (rata-rata jarak titik ke pusat kluster)

d(c_i, c_j) : Jarak Euclidean antara pusat kluster i dan j

max(j≠i) : Untuk tiap kluster i, dipilih pasangan kluster j yang paling mirip.

Hasil evaluasi menunjukkan bahwa DBSCAN memiliki Silhouette Score yang lebih tinggi dan DBI yang lebih rendah dibandingkan dengan K-Means. Hal ini menunjukkan bahwa DBSCAN lebih efektif dalam mengelompokkan data nasabah dengan bentuk kluster yang tidak teratur dan mengandung noise.

3.4 Evaluasi Model Klustering

Penelitian ini menggunakan dua algoritma klustering yaitu K-Means dan DBSCAN untuk melakukan segmentasi nasabah berdasarkan data bank. Penilaian dilakukan menggunakan dua metrik evaluasi yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Tabel berikut menampilkan hasil evaluasi dari kedua algoritma:

Tabel 4. Hasil Evaluasi dari Kedua Algoritma

Algoritma	Silhouette Score	Davies-Bouldin Index
K-Means	0.231	1.295
DBSCAN	0.218	1.455

Dari hasil tersebut, K-Means menunjukkan performa yang lebih baik dari DBSCAN. Hal ini terlihat dari nilai Silhouette Score yang lebih tinggi dan DBI yang lebih rendah, yang menunjukkan bahwa kluster yang dihasilkan oleh K-Means lebih kompak dan terpisah dengan baik. Beberapa hasil penelitian terdahulu menunjukkan bahwa algoritma DBSCAN cenderung memerlukan waktu pemrosesan lebih lama pada kluster tertentu, sementara K-Means mampu menyelesaikan proses secara lebih cepat dalam kondisi serupa. Dalam hal kualitas segmentasi, K-Means unggul dengan nilai Silhouette Score sebesar 0,714917 dan Davies-Bouldin Index (DBI) sebesar 0,365776, dibandingkan DBSCAN yang mencatat Silhouette Score 0,699971 dan DBI sebesar 0,390784 [15] [16].

Secara fungsional, K-Means terbukti lebih efektif dalam membedakan pelanggan berdasarkan frekuensi dan nilai moneter, sedangkan DBSCAN memiliki keunggulan dalam mendeteksi outlier dengan interaksi dan nilai moneter yang tinggi. Temuan ini mengindikasikan bahwa K-Means lebih tepat diterapkan untuk segmentasi pelanggan Transjakarta, karena mampu mendukung perancangan strategi layanan yang lebih efisien dan peningkatan kepuasan pelanggan [17]. Dalam kasus lain, DBSCAN dengan parameter epsilon sebesar 11 dan 24 data sampel menghasilkan dua kluster dengan nilai Silhouette Score sebesar 0,2580 menunjukkan hasil yang kurang optimal, sedangkan K-Means dengan jumlah empat kluster—yang ditentukan melalui metode Elbow—menghasilkan Silhouette Score yang jauh lebih tinggi, yaitu 0,5697. Perbandingan tersebut memberikan dukungan bahwa K-Means lebih efektif dalam membentuk kluster yang terpisah dengan baik, memiliki struktur yang lebih rapi, dan tingkat kesamaan antar data dalam kluster yang lebih tinggi. Oleh karena itu, K-Means lebih direkomendasikan untuk digunakan dalam pengelompokan siswa SMP terbaik. [18] [19].

3.5 Interpretasi Kluster

Setelah proses clustering menggunakan algoritma K-Means atau DBSCAN selesai, data nasabah akan dikelompokkan ke dalam kluster tertentu berdasarkan kesamaan karakteristiknya. Berikut contoh sample data yang masuk ke dalam kluster A, B, dan C:

Tabel 5. Sampel Data

ID NASABAH	AGE	JOB	BALANCE	CLUSTER
1001	45	management	3500	A
1002	38	services	1200	B
1003	52	admin	800	C
1004	29	blue-collar	200	noise
1005	41	technician	5000	A

Keterangan :

Cluster A : Nasabah dengan saldo tinggi dan usia matang.

Cluster B : Nasabah menengah dengan saldo sedang.

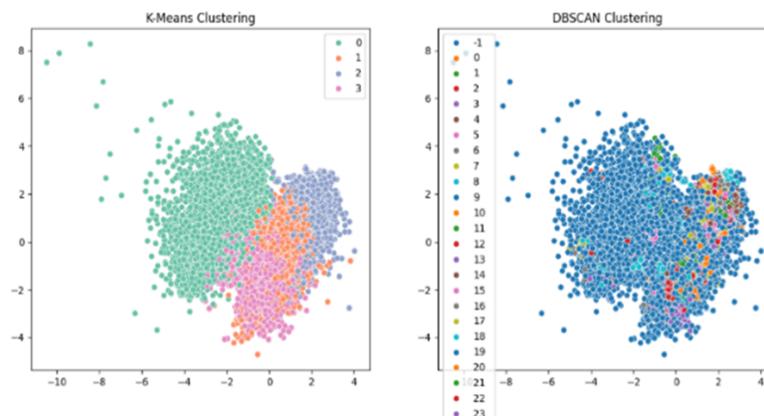
Cluster C : Nasabah lebih tua dengan saldo rendah.

Noise : Label -1 pada DBSCAN, menunjukkan outlier.

Visualisasi hasil kluster menggunakan PCA (Principal Component Analysis) menunjukkan bahwa K-Means menghasilkan kluster yang cukup jelas, sedangkan DBSCAN mendeteksi beberapa data sebagai *noise* (diberi label -1), yang umum terjadi dalam pendekatan density-based. Visualisasi hasil kluster menggunakan PCA membantu dalam melihat pola sebaran data serta mengevaluasi apakah kluster yang terbentuk memiliki batasan yang jelas dan kompak. Pada visualisasi ini, setiap titik merepresentasikan satu nasabah, dan warna titik menunjukkan kluster tempat nasabah tersebut dikelompokkan. Gambar visualisasi kluster PCA dapat dilihat pada gambar dibawah ini :

K-Means Silhouette Score: 0.046328989128304764
 K-Means Davies-Bouldin Index: 3.4399499425810562
 DBSCAN Silhouette Score: 0.3798865181702325
 DBSCAN Davies-Bouldin Index: 0.9510885534977848

Gambar 2. Hasil Visualisasi Cluster PCA



Gambar 3. Hasil Visualisasi Cluster PCA

4. KESIMPULAN

Penelitian ini dilakukan untuk mengevaluasi dan membandingkan kemampuan algoritma K-Means serta DBSCAN dalam melakukan segmentasi nasabah, dengan menggunakan data pemasaran yang berasal dari sektor perbankan. Berdasarkan hasil evaluasi menggunakan metrik Silhouette Score dan Davies-Bouldin Index, algoritma K-Means menunjukkan hasil yang lebih baik dalam hal pembentukan kluster yang kompak dan terpisah. Meskipun demikian, DBSCAN memiliki keunggulan dalam mengidentifikasi *outlier* dan membentuk kluster dengan bentuk yang tidak teratur, yang berguna pada data dengan distribusi yang kompleks. Dengan demikian, keberhasilan segmentasi sangat dipengaruhi oleh pemilihan algoritma yang sesuai dengan sifat data serta tujuan analisis yang ingin dicapai. Untuk data pemasaran bank dengan distribusi yang relatif seragam, K-Means dapat menjadi pilihan utama. Namun, jika terdapat banyak *noise* atau bentuk kluster yang tidak teratur, DBSCAN dapat memberikan hasil yang lebih adaptif. Sebagai pengembangan dari penelitian ini, disarankan agar peneliti berikutnya dapat mencoba menggunakan kombinasi beberapa metode *clustering* atau pendekatan *ensemble clustering*. Hal ini bertujuan untuk meningkatkan kualitas hasil segmentasi nasabah secara keseluruhan. Selain itu, metode *clustering* lain seperti OPTICS atau *Hierarchical Clustering* juga bisa diimplementasikan sebagai perbandingan tambahan terhadap K-Means dan DBSCAN. Untuk memperluas cakupan penelitian, disarankan pula untuk menggunakan data riil langsung dari institusi perbankan dengan jumlah sampel yang lebih besar. Dengan

demikian, representasi karakteristik nasabah akan lebih akurat dan relevan. Selain itu, evaluasi hasil segmentasi tidak hanya terbatas pada metrik teknis seperti Silhouette Score atau DBI, tetapi juga bisa dilengkapi dengan analisis bisnis seperti *Lifetime Value (LTV)*, tingkat *churn*, atau respons terhadap promosi tertentu. Dengan memadukan pendekatan yang lebih luas serta integrasi teknik analisis lanjutan, proses segmentasi nasabah dapat dilakukan dengan lebih tepat sasaran dan memberikan manfaat strategis bagi industri perbankan dalam merancang kebijakan pemasaran yang efektif.

REFERENCES

- [1] M. M. Putri, C. Dewi, E. Permata Siam, G. Asri Wijayanti, N. Aulia, and R. Nooraeni, "Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020 Komparasi DBSCAN dan K-Means Clustering pada Pengelompokan Status Desa di Jawa Tengah Tahun 2020," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 3, pp. 394–404, 2021, doi: 10.20956/j.v17i3.11704.
- [2] T. Setyani, A. F. Octaviansyah, and R. Andika, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Penjualan Handphone Dan Elektronik (Studi Kasus : Cv Rey Gasendra)," vol. 5, no. 1, pp. 18–22, 2024.
- [3] S. A. D. Budiman, D. Safitri, and D. Ispriyanti, "Perbandingan Metode K-Means Dan Metode Dbscan Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang," *J. Gaussian*, vol. 5, pp. 757–762, 2016, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [4] S. Ramadhani, M. Sihombing, and M. Simanjuntak, "Classification Of Diseases In Patients Based On Factors Environment Using The K-Means Algorithm At Puskesmas Subdistrict Selesai," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 1, pp. 279–286, 2023, doi: 10.59934/jaiea.v3i1.309.
- [5] P. Srivastava and D. S. Saxena, "Enhanced Customer : A Comparative Analysis of Leading Clustering Algorithms," vol. 10, no. 4, pp. 258–262, 2025.
- [6] F. M. Pranata, S. H. Wijoyo, and N. Y. Setiawan, "Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM," vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] M. Z. R. F. K. M. Razaul, "Optimizing Customer Segmentation in the Banking Sector: A Comparative Analysis of Machine Learning Algorithms Article," no. December, pp. 104–111, 2023, doi: 10.32996/jests.
- [8] A. W. W. Rizqina Rahmati, "Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia tahun 2019)," vol. 5, no. 2, pp. 73–80, 2021.
- [9] R. A. Widodo and A. Jananto, "Implementasi Data Mining dalam Clustering Menu Favorit Pada Cafe Anetos . Coffeebrunch Menggunakan Algoritme K-Means," *J. Ilm. Komput.*, vol. 19, pp. 710–724, 2023.
- [10] N. H. Ahsina, F. Fatimah, and F. Rachmawati, "Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," vol. 8, no. 3, 2022.
- [11] S. Syiva Multi Fani, Rukun Santoso, "Penerapan text mining untuk melakukan clustering data tweet akun blibli pada media sosial twitter menggunakan k-means clustering 1,2,3," vol. 10, pp. 583–593, 2021.
- [12] A. R. Dina Selvia, "Implementasi Metode K-Means Clustering Dengan Teknik Pengolahan Citra Untuk Mengidentifikasi Jenis Sepatu," pp. 361–366, 2025.
- [13] D. P. Agustino, I. G. B. A. Budaya, I. G. Harsemadi, I. K. Dharmendra, and I. M. S. A. Pande, "Comparison of the DBSCAN Algorithm and Affinity Propagation on Business Incubator Tenant Customer Segmentation," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 315–321, 2023, doi: 10.32736/sisfokom.v12i2.1682.
- [14] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster," vol. 06, no. 01, pp. 60–74, 2024.
- [15] B. S. Ashari, S. C. Otniel, and Rianto, "Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail," *J. Siliwangi*, vol. 5, no. 2, pp. 72–77, 2019, [Online]. Available: <http://jurnal.unsil.ac.id/index.php/jssainstek/article/view/1283>
- [16] I. T. Sarmini, Windiya Ma'arifah, "Sistem Pendukung Keputusan Berbasis K-Means untuk Evaluasi Keberhasilan Bisnis dan Nilai Perusahaan," vol. 04, 2024, doi: 10.21456/vol14iss4pp363-374.
- [17] A. Saputra and R. Yusuf, "Comparison of the DBSCAN and K-MEANS Algorithms in Segmenting Customers Using Public Transportation of Transjakarta Using the RFM Method Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM," vol. 4, no. October, pp. 1346–1361, 2024.
- [18] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, "Perbandingan algoritma clustering dbscan dan k-means dalam pengelompokan siswa terbaik Comparison of dbscan and k-means clustering algorithms in grouping the best students," vol. 5, pp. 301–309, 2024.
- [19] G. A. Dennis Alfa Imanuel, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut Rfm Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan Pada Toko Retail Online Visualization Of Customer Segmentation Based On Rfm Attributes Using K-Means Algorithm To Comprehend Cust," vol. 12, no. 2, pp. 283–292, 2025, doi: 10.25126/jtiik.2025128619.