Analisis Perbandingan Sensitivitas dan Kinerja K-Means dan DBSCAN Terhadap Variasi Noise

Rafael Austin[1]*, Ricky Dwi Putra[#2], Ardi[#3] , Fuzail Fazle Rabbi[#4], Lucky Pujiono Ws[#5], Vitri Tundjungsari[#6]
[#]Teknik Informatika,, Universitas Esa Unggul , Jakarta, Indoneisa
Korespodensi author *raf.austin05@gmail.com

Info Artikel

Diajukan: 28 Januari 2026
Diterima: 15 Juni 2026
Diterbitkan: 1 Juli 2026

Keywords:
unsupervised ; DBSCAN, K-Means ; Silhouette score ; Noise ; DBI

Kata Kunci:
unsupervised ; DBSCAN ; K-Means ; Silhouette score ; Noise ; DBI

Abstract

Inaccurate or noisy data presents a significant challenge in machine learning, particularly in unsupervised clustering tasks. This study evaluates the robustness and performance of two popular clustering algorithms, K-Means and DBSCAN, against various levels of Gaussian noise (5%, 10%, 20%, and 30%) injected into a customer dataset. Evaluation was conducted using Silhouette Score and Davies-Bouldin Index (DBI). Initial results indicated that DBSCAN performed slightly better with a Silhouette Score of 0.4817 compared to K-Means at 0.4101. However, after noise injection, K-Means demonstrated superior stability by maintaining more consistent cluster memberships, whereas DBSCAN was more sensitive to distance variations, leading to significant fluctuations in cluster assignments. The study concludes that K-Means is more reliable for datasets where cluster integrity must be preserved despite minor data irregularities.

Abstrak

Data yang tidak akurat atau mengandung noise merupakan tantangan signifikan dalam machine learning, terutama pada tugas pengelompokan (clustering) yang tidak diawasi (unsupervised). Penelitian ini bertujuan untuk menganalisis dan membandingkan tingkat ketahanan (robustness) dua algoritma clustering unsupervised, yaitu K-Means dan DBSCAN. terhadap variasi noise pada data. Proses dilakukan dengan mengevaluasi ketahanan (robustness) dan performa dua algoritma pengelompokan populer, yaitu K-Means dan DBSCAN, terhadap berbagai tingkat Gaussian noise (5%, 10%, 20%, dan 30%) yang disuntikkan ke dalam dataset pelanggan. Evaluasi dilakukan menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI). Hasil awal menunjukkan bahwa DBSCAN memiliki performa yang sedikit lebih baik dengan Silhouette Score 0,4817 dibandingkan K-Means sebesar 0,4101. Namun, setelah dilakukan penyuntikan noise, K-Means menunjukkan stabilitas yang lebih unggul dengan mempertahankan keanggotaan cluster yang lebih konsisten, sementara DBSCAN ditemukan lebih sensitif terhadap variasi jarak, yang menyebabkan fluktuasi signifikan pada penentuan cluster. Penelitian ini menyimpulkan bahwa K-Means lebih dapat diandalkan untuk dataset yang membutuhkan integritas struktur cluster tetap terjaga meskipun terdapat ketidakteraturan data yang kecil.

Cara mensitasi artikel:
R. Austin, R.D. Putra, Ardi, F.F.Rabbi, L.P.Ws, V.Tundjungsari. "Analisis Perbandingan Sensitivitas dan Kinerja K-Means dan DBSCAN Terhadap Variasi Noise *Jurnal Teknologi Informasi: Teori, Konsep, dan Implementasi (JTI-TKI), vol. 17, no. 1, pp. 11-16, Maret 2026, https://doi.org/10.36382/jti-tki.v17i1.627*

PENDAHULUAN

Ketahanan algoritma terhadap data noisy merupakan aspek penting dalam machine learning, tidak hanya untuk menurunkan persentase kesalahan klasifikasi tetapi juga agar algoritma dapat tetap melakukan klasifikasi pada data yang mengalami perubahan kecil [1]. Keberadaan data yang bising dalam kumpulan data dapat berdampak signifikan terhadap prediksi informasi yang bermakna [2]. Banyak studi telah menunjukkan bahwa noise dalam kumpulan data secara dramatis menyebabkan penurunan akurasi klasifikasi dan hasil prediksi yang buruk [2]. Terdapat dua pendekatan utama dalam menangani noise, yaitu data cleaning dan penggunaan robust learning technique [3]. Pembersihan data adalah proses mendeteksi dan memperbaiki (atau menghapus) data yang rusak atau tidak akurat, dan mengacu pada identifikasi bagian data yang tidak lengkap, salah, tidak akurat, atau tidak relevan [4] .Pada pendekatan data cleaning, identifikasi data noisy menjadi hal yang menantang karena kesalahan pada label dapat membuat proses eliminasi atau perbaikan sulit dilakukan . SementIara itu, robust learning techniques memungkinkan model untuk tetap menghasilkan prediksi yang stabil meskipun data mengandung noise [5].

Dalam konteks unsupervised learning, metode clustering menjadi salah satu pendekatan populer untuk pengelompokan data tanpa label [6]. Salah satu algoritma yang sering digunakan adalah K-Means, yang membutuhkan penentuan jumlah cluster secara manual menggunakan metode seperti Elbow dan Silhouette untuk memastikan kualitas pengelompokan [7], [8]. Metode Elbow menggunakan kurva distorsi (inertia) untuk menentukan titik optimal, sedangkan Silhouette mengukur kohesi dan pemisahan antar cluster [7], [8]. K-Means menginisiasi centroid secara acak dan melakukan iterasi hingga centroid tidak berubah, sehingga menghasilkan cluster yang relatif stabil pada data [8], [9].

Di sisi lain, algoritma DBSCAN bekerja berdasarkan kepadatan titik (density-connected) dan unggul dalam

Submitted 28 Januari 2026; Revised 10 Juni 2026; Accepted 15 Juni 2026

mendeteksi outlier serta menangani data dengan distribusi yang tidak beraturan [10]. Algoritma DBSCAN dapat digunakan untuk mengidentifikasi klaster dengan bentuk sembarang dalam suatu kumpulan data dan mendeteksi outlier secara efektif [11]. K-Means menggunakan centroid untuk membagi data menjadi k klaster, sedangkan DBScan mendefinisikan klaster berdasarkan kepadatan titik-titik dalam ruang data [12].Namun, sebagian besar algoritma clustering klasik tidak dirancang untuk robust terhadap data noisy atau tidak pasti [13]. Kesulitan utama dari metode ini adalah menemukan parameter yang optimal secara otomatis, terutama ketika data mengandung noise [14]. Oleh karena itu, penelitian ini akan mengeksplorasi pengaruh noise terhadap performa K-Means dan DBSCAN serta mengevaluasi ketahanan pengelompokan terhadap data yang berubah atau noisy.

METODE

A. Penggunaan Algoritma

Pada Penelitian ini digunakan dua algoritma dari machine learning yang bersifat unsupervised learning (pembelajaran tidak terawasi), yaitu algoritma yang ditujukan pada dataset yang tidak berlabel [6]. Kedua algoritma yang dipakai tersebut adalah K-Means dan DBSCAN karena perbedaan prinsip dimana K-Means adalah centroid-based dan DBSCAN adalah Density Based.

K-Means adalah algoritma pengelompokan (clustering) yang bekerja dengan membagi data ke dalam sejumlah cluster yang ditentukan sebelumnya. Algoritma ini menginisiasi centroid secara acak dan melakukan iterasi, memindahkan titik data ke cluster terdekat dan memperbarui posisi centroid hingga tidak terjadi perubahan lagi, sehingga menghasilkan cluster yang stabil [8], [9]. Untuk menentukan jumlah cluster yang tepat, K-Means biasanya dibantu dengan Elbow Method, yang menggunakan kurva distorsi (inertia) untuk menemukan titik optimal, dan Silhouette, yang mengukur kohesi dan pemisahan antar cluster [7], [8]. Metode ini cocok untuk data dengan cluster berbentuk bulat dan terpisah dengan jelas, namun cukup sensitif terhadap outlier dan noise.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma pengelompokan yang bekerja berdasarkan kepadatan titik dalam data. Algoritma ini mengidentifikasi area dengan kepadatan tinggi yang saling terhubung (density-connected) untuk membentuk cluster, sehingga dapat menangani distribusi data yang tidak beraturan dan mendeteksi outlier atau noise secara alami [10]. Berbeda dengan K-Means, DBSCAN tidak memerlukan jumlah cluster ditentukan sebelumnya, tetapi membutuhkan parameter eps (jarak maksimum untuk menghubungkan titik) dan minPts (jumlah minimum titik untuk membentuk cluster). Algoritma ini cocok untuk data dengan bentuk cluster arbitrer, namun hasil clustering bisa sensitif terhadap pemilihan parameter eps dan minPts [10], [13].

B. Injeksi Gaussian Noise

Penambahan Noise pada data akan menggunakan injeksi noise dengan metode gaussian Teknik injeksi noise ini memberikan gangguan acak pada data dengan distribusi normal (Gaussian) yang memiliki rata rata nol dan standar deviasi tertentu. Rumus untuk untuk injeksi Gaussian Noise adalah sebagai berikut

$$\mathbf{g(x) = x + n}$$

Dengan :

x = sebagai Data asli

n = sebagai variabel acak gaussian isotropik ($\varepsilon \sim N(0,\sigma 2I)$) (dengan σ = parameter, ε = nilai acak dari parameter yang ditentukan)

Injeksi Noise dengan metode Gaussian ini akan menyebabkan data akan menurun atau meningkat tergantung dari hasil variabel acak gaussian isotropik. Hal ini akan membantu analisis robustness kedua algoritma dalam pengukuran ketahanan terhadap data yang memiliki noise.

Variasi noise yang akan dilakukan untuk uji robustness adalah dengan persentase 5%, 10%, 20%, dan 30%. persentase ini akan dimasukan dalam rumus variabel acak gaussian isotropik untuk mengukur seberapa besar noise yang diberikan pada program.

C. Evaluasi dan Visualiasi

Untuk Pengukuran Kinerja Pada Penelitian ini akan menggunakan Sillhoutte Score dan DBI yang dibuktikan efesien pada penelitian sebelumnya [15].

Silhouette Score mengevaluasi seberapa baik data cocok dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilai Silhouette coefficient berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan cluster yang baik.

Davies-Bouldin menghasilkan nilai indeks untuk setiap jumlah cluster k, dengan grafik yang menunjukkan nilai DBI yang mendekati nilai 0 atau semakin kecil menunjukkan kualitas cluster yang lebih baik.[15] Untuk Visualisasi hasil digunakan scatter plot untuk menunjukan perubahan pengelompokan oleh algoritma. Line chart juga digunakan Line chart juga digunakan sebagai representasi perubahan evaluasi pada pemberian noise dalam data.

D. Dataset

Data set yang digunakan untuk penelitian ini berasal dari open-source kaggle. Data berisi 5 fitur yang akan menentukan clustering melalui algoritma. berikut adalah list dari fiturnya dan tipe datanya seperti dituliskan pada Tabel 1.

Tabel 1. Tipe Data pada Dataset

Nama Data	Tipe Data
CustomerID	Int
Gender	String
Age	Int

Annual Income (k)	Int
Spending Score (1-100)	Int

Pada salah satu fitur akan dilakukan encoding agar komputer dapat membaca dan seluruh fitur ini akan dipakai dalam pengelompokan clustering. Pada Gambar 1. Dituliskan potongan dataset.

```
CustomerID,Gender,Age,Annual Income (k$),Spending Score (1-100)
1,Male,19,15,39
2,Male,21,15,81
3,Female,20,16,6
4,Female,23,16,77
5,Female,31,17,40
6,Female,22,17,76
7,Female,35,18,6
8,Female,23,18,94
9,Male,64,19,3
10,Female,30,19,72
11,Male,67,19,14
12,Female,35,19,99
13,Female,58,20,15
14,Female,24,20,77
15,Male,37,20,13
16,Male,22,20,79
17,Female,35,21,35
18,Male,20,21,66
19,Male,52,23,29
20,Female,35,23,98
21,Male,35,24,35
22,Male,25,24,73
```

Gambar 1. Snippet dataset yang sudah dibuka

HASIL DAN PEMBAHASAN

Hasil percobaan yang telah dilakukan di laporkan serta dilakukan pembahasan terhadap data yang telah direkam.

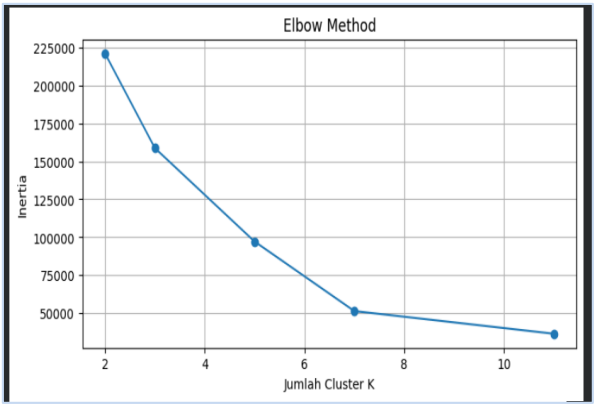
A. Pencarian K optimal

Sebelum dilakukannya perbandingan dan injeksi noise, dilakukan pengukuran sillhoutte score, inertia dan elbow method dengan tujuan mencari K optimal pada K-Means terhadap dataset yang dipakai.

Tabel 2. Hasil Pencarian K Optimal

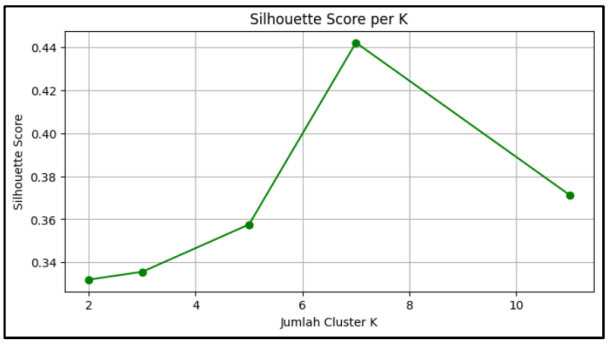
K	Inertia (elbow)	Sillhoutte Score
2	221087	0.331
3	158745	0.335
5	97211.8	0.357
7	51448.4	0.442
11	36521.1	0.371

Dari Hasil perhitungan program ditemukan hasil paliung optimal ditemukan pada K = 7 yang menghasilkan Sillhoutte score tertinggi. Untuk Proses selanjutnya akan memakai K = 7 sebagai variabel tetapnya. Gambar 2 disajikan hasil nilai K yang optimal dari analisis berdasarkan Grafik elbow method.



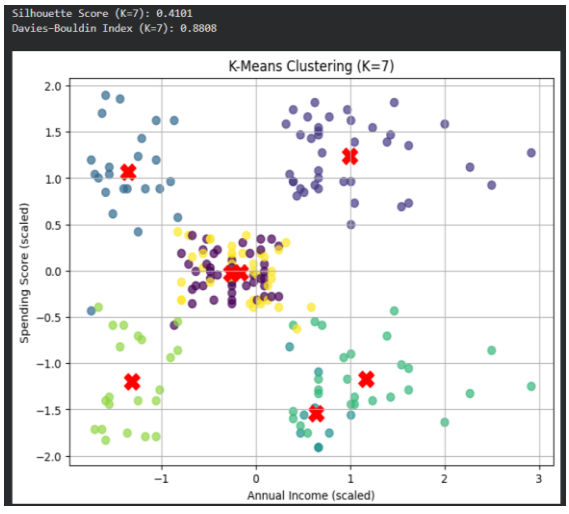
Gambar 2. Grafik elbow method

Gambar 3 disajikan hasil nilai K yang optimal dari analisis berdasarkan sillhoutte score



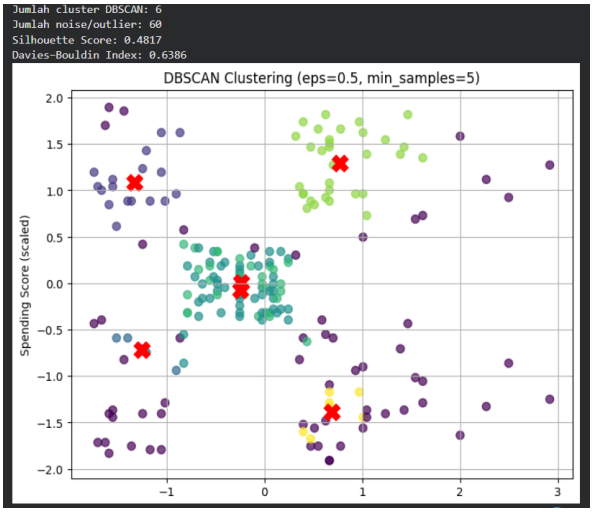
Gambar 3. Grafik sillhoutte score

B. Evaluasi Awal K-Means



Gambar 4. Scatter plot k-means k = 7

Pada gambar, warna warna berbeda menunjukan kelas berbeda sementara tanda silang merah menunjukan masing masing centroid kelas



Gambar 5. Scatter plot DBSCAN

Visualisasi pada gambar pertama menunjukkan hasil pengelompokan menggunakan algoritma K-Means dengan jumlah K=7, di mana setiap warna merepresentasikan kelas yang berbeda1. Tanda silang merah dalam grafik tersebut berfungsi untuk menunjukkan posisi centroid atau titik pusat dari masing-masing kelas yang terbentuk2. Melalui metode ini, data dibagi secara partisi ke dalam sejumlah cluster yang telah ditentukan sebelumnya berdasarkan jarak terdekat ke centroid3. Hasil evaluasi awal untuk model ini mencatatkan nilai Silhouette Score sebesar 0,4101 dan Davies-Bouldin Index (DBI) sebesar 0,88084.+3Sementara itu, gambar kedua menampilkan hasil clustering menggunakan algoritma DBSCAN dengan parameter epsilon sebesar 0,5 dan min_samples sebanyak 55. Berbeda dengan K-Means, DBSCAN berhasil mengidentifikasi 6 cluster dan mendeteksi adanya 60 titik data yang dianggap sebagai noise atau outlier6. Algoritma ini bekerja dengan mencari area kepadatan tinggi yang saling terhubung, sehingga lebih fleksibel dalam menangani distribusi data yang tidak beraturan7. Dari sisi performa awal, DBSCAN menunjukkan kualitas pengelompokan yang sedikit lebih baik dengan Silhouette Score yang lebih tinggi, yaitu 0,4817, dan nilai DBI yang lebih kecil sebesar 0,63868

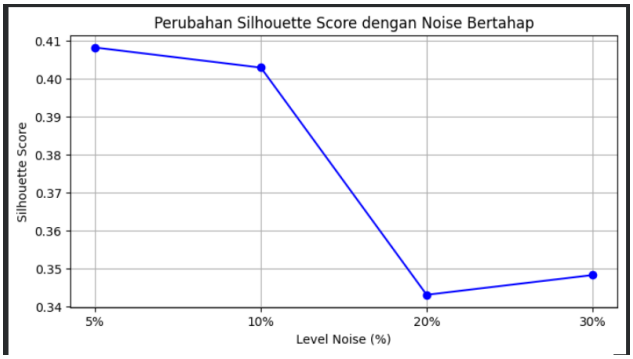
C. Gaussian Noise injection dan Evaluasi

Pada Tahap ini gaussian injection akan mulai dimasukan pada dataset lalu dilanjutkan dengan penggunaan algoritma lagi untuk menemukan perubahan apakah evaluasi akan berubah dan apakah beberapa titik akan berpindah kelas atau tidak. Semakin sedikit perubahan evaluasi dan semakin sedikit titik yang berubah kelas semakin bagus robustness algoritma tersebut.

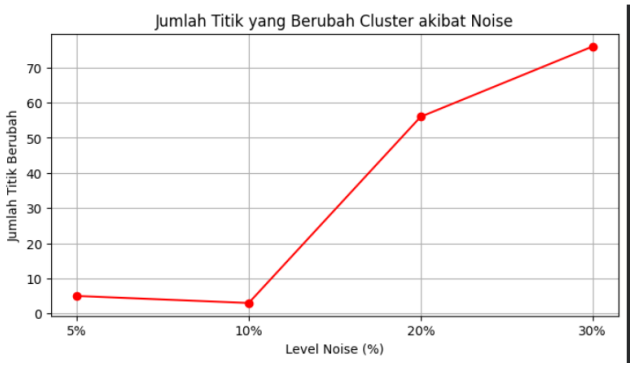
Perubahan Silhouette Score dan Cluster akibat Noise:

	Noise Level	Silhouette Score	Titik Berubah Cluster
0	5%	0.4082	5
1	10%	0.4029	3
2	20%	0.3430	56
3	30%	0.3482	76

Gambar 6. Perubahan evaluasi dan titik perubahan cluster



Gambar 7. Grafik perubahan silhoute score vs Noise %

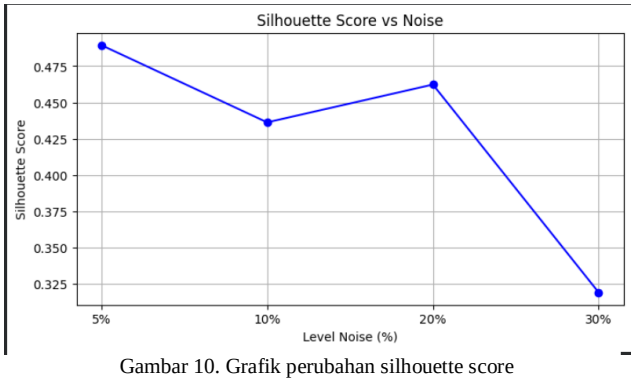


Gambar 8. Perubahan titik cluster yang berpindah kelas

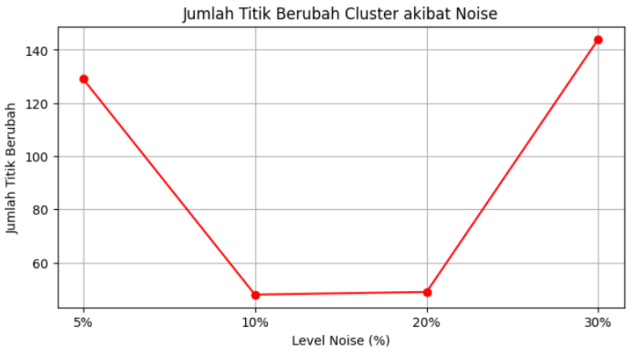
Berdasarkan Gambar 6, Gambar 7, dan Gambar 8, terlihat bahwa peningkatan tingkat noise berdampak signifikan terhadap kualitas dan stabilitas clustering. Nilai Silhouette Score cenderung menurun seiring bertambahnya noise, dengan penurunan paling tajam terjadi pada level 20%, yang menunjukkan berkurangnya kualitas pemisahan antar cluster. Di sisi lain, jumlah titik data yang mengalami perubahan cluster meningkat secara drastis pada noise yang lebih tinggi, menandakan bahwa keberadaan noise menyebabkan struktur cluster menjadi semakin tidak stabil.

	Noise Level	Silhouette Score	Titik Berubah Cluster	Jumlah Cluster
0	5%	0.4894	129	5
1	10%	0.4363	48	6
2	20%	0.4623	49	6
3	30%	0.3193	144	6

Gambar 9. Hasil evaluasi setelah injeksi gaussian noise



Gambar 10. Grafik perubahan silhouette score



Gambar 11. Perubahan jumlah titik yang berpindah kelas

Pada DBSCAN, hasil menunjukan terjadinya perubahan sillhoute score yang naik dan turun serta perubahan kelas yang yang lumayan banyak

Hasil perbandingan antara K-Means dan DBSCAN menunjukkan bahwa K-Means lebih tahan terhadap injeksi noise. Hal ini terlihat dari jumlah titik yang mengalami perubahan cluster, di mana K-Means mempertahankan sebagian besar titik dalam cluster aslinya meskipun noise bertambah hingga 30%. Sementara itu, DBSCAN cenderung menghasilkan lebih banyak titik yang berpindah cluster atau bahkan menjadi noise/outlier ketika data terganggu oleh noise. Hal ini menunjukkan bahwa struktur cluster K-Means lebih stabil terhadap variasi data kecil, sedangkan DBSCAN lebih sensitif terhadap perubahan jarak antar titik akibat noise, terutama karena metode berbasis kepadatan ini menganggap titik yang jauh dari cluster sebagai noise. Dengan demikian, untuk dataset dengan injeksi noise bertahap, K-Means menunjukkan kestabilan pengelompokan yang lebih tinggi, sehingga dapat diandalkan ketika integritas cluster perlu dijaga meskipun terjadi gangguan kecil pada data.

D. Ringkasan Temuan

Performa awal dari kedua algoritma menunjukkan perbedaan karakteristik yang menarik. Tanpa adanya noise, DBSCAN memberikan kualitas pengelompokan yang lebih baik dibandingkan K-Means, yang ditandai dengan nilai Silhouette Score lebih tinggi (0,4817 dibanding 0,4101) serta nilai Davies–Bouldin Index (DBI) yang lebih rendah (0,6386 dibanding 0,8808). Hal ini menunjukkan bahwa pada kondisi ideal tanpa gangguan, DBSCAN mampu menangkap struktur kepadatan data dengan lebih baik.

Namun, dari sisi stabilitas, K-Means terbukti lebih tahan terhadap gangguan pada data. Meskipun tingkat noise meningkat hingga 30%, jumlah titik yang berpindah cluster relatif sedikit, menunjukkan bahwa algoritma ini memiliki kemampuan mempertahankan cluster asli dengan cukup baik. Sebaliknya, DBSCAN menunjukkan sensitivitas yang tinggi terhadap perubahan jarak antar titik akibat noise. Banyak titik data yang akhirnya berpindah cluster atau bahkan dianggap sebagai noise/outlier ketika gangguan meningkat, menandakan bahwa algoritma ini lebih rentan terhadap fluktuasi kecil dalam dataset.

Dari segi metrik evaluasi, tren Silhouette Score juga memperlihatkan perbedaan karakteristik kedua algoritma. Pada K-Means, Silhouette Score cenderung menurun secara konstan seiring bertambahnya level noise, mencerminkan penurunan kualitas cluster secara gradual. Sedangkan pada DBSCAN, Silhouette Score mengalami fluktuasi yang tidak menentu, menunjukkan bahwa performa pengelompokan algoritma ini sangat dipengaruhi oleh posisi relatif titik-titik data dan sensitivitas terhadap parameter eps dan min_samples, terutama ketika noise diperkenalkan.

KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis dan membandingkan tingkat ketahanan (robustness) dua algoritma clustering unsupervised, yaitu K-Means dan DBSCAN, terhadap variasi noise pada data. Noise yang digunakan dalam penelitian ini berupa Gaussian noise dengan tingkat gangguan sebesar 5%, 10%, 20%, dan 30%. Evaluasi performa dilakukan menggunakan metrik Silhouette Score dan Davies–Bouldin Index (DBI) untuk mengukur kualitas serta stabilitas hasil pengelompokan.

Hasil eksperimen menunjukkan bahwa pada kondisi data tanpa noise, algoritma DBSCAN memiliki performa pengelompokan yang lebih baik dibandingkan K-Means. Hal ini ditunjukkan oleh nilai Silhouette Score yang lebih tinggi dan nilai DBI yang lebih rendah, yang menandakan bahwa DBSCAN mampu membentuk cluster yang lebih terpisah dan kompak pada data dengan struktur kepadatan yang jelas. Keunggulan ini menegaskan bahwa DBSCAN efektif digunakan pada dataset dengan distribusi data tidak beraturan dan keberadaan outlier.

Namun, ketika dilakukan injeksi Gaussian noise secara bertahap, karakteristik kedua algoritma menunjukkan perbedaan yang signifikan dari sisi stabilitas. K-Means terbukti memiliki ketahanan yang lebih baik terhadap gangguan data, ditunjukkan oleh jumlah titik data yang berpindah cluster relatif sedikit meskipun tingkat noise meningkat hingga 30%. Selain itu, penurunan nilai Silhouette Score pada K-Means terjadi secara bertahap dan konsisten, yang mengindikasikan degradasi performa yang masih dapat dikendalikan.

Sebaliknya, DBSCAN menunjukkan sensitivitas yang lebih tinggi terhadap noise. Perubahan kecil pada jarak antar titik akibat noise menyebabkan fluktuasi signifikan pada hasil clustering, baik dari segi nilai evaluasi maupun jumlah titik yang berpindah cluster atau diklasifikasikan sebagai noise. Fluktuasi Silhouette Score yang tidak stabil mencerminkan ketergantungan DBSCAN terhadap parameter eps dan min_samples, yang semakin sulit dipertahankan optimal ketika data mengalami gangguan.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa pemilihan algoritma clustering perlu mempertimbangkan kondisi dan karakteristik data. DBSCAN lebih unggul pada data yang bersih dan memiliki struktur kepadatan yang jelas, sedangkan K-Means lebih dapat diandalkan pada kondisi data yang mengandung noise dan membutuhkan kestabilan struktur cluster. Dengan demikian, penelitian ini menegaskan bahwa K-Means memiliki tingkat robustness yang lebih tinggi terhadap variasi noise dibandingkan DBSCAN, sehingga cocok digunakan pada skenario data dunia nyata yang rentan terhadap ketidakteraturan.

Penelitian selanjutnya disarankan untuk menguji variasi parameter algoritma, khususnya nilai eps dan min_samples pada DBSCAN, guna meningkatkan ketahanan terhadap noise. Selain itu, penggunaan teknik noise handling atau data preprocessing sebelum clustering dapat dipertimbangkan untuk meningkatkan stabilitas hasil pengelompokan. Pengujian pada dataset dengan karakteristik berbeda juga diperlukan agar hasil penelitian lebih general dan aplikatif.

REFERENSI

[1] J.-Y. Franceschi, A. Fawzi, and O. Fawzi, “Robustness of classifiers to uniform p and Gaussian noise,” 2018.

[2] S. Gupta and A. Gupta, “Dealing with noise problem in machine learning data-sets: A systematic review,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 466–474. doi: 10.1016/j.procs.2019.11.146.

[3] S. Pau, A. Perniciano, B. Pes, and D. Rubattu, “An Evaluation of Feature Selection Robustness on Class Noisy Data,” *Information (Switzerland)*, vol. 14, no. 8, Aug. 2023, doi: 10.3390/info14080438.

[4] Y. Sun and J. Zhang, “Distance-based Data Cleaning: A Survey (Technical Report),” Nov. 2020, [Online]. Available: http://arxiv.org/abs/2011.11176

[5] C.-H. Lin, E. L. Dyer, V. Muthukumar, and C. Kaushik, “The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective,” 2024.

[6] Tundjungsari Vitri, “Dasar Machine Learning”.

[7] P. Khairani Ritonga and M. S. Hasibuan, “Analisis Perbandingan Silhouette dengan Elbow pada Algoritma K-Means dan DBSCAN,” vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1027.

[8] P. Vania and B. Nurina Sari, “Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means,” *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 547–558, 2023, doi: 10.5281/zenodo.10081332.

[9] R. C. de Amorim and V. Makarenkov, “Improving clustering uality evaluation in noisy Gaussian mixtures,” Mar. 2025, [Online]. Available: http://arxiv.org/abs/2503.00379

[10] T. Setyani, Y. Indriani, M. Fadli, and E. R. Susanto, “Perbandingan Algoritma K-Means dan DBSCAN dalam Pengelompokkan Data Penjualan Elektronik,” *Progresif: Jurnal Ilmiah Komputer*, vol. 21, no. 2, p. 712, Aug. 2025, doi: 10.35889/progresif.v21i2.2775.

[11] C. H. Lin, K. C. Hsu, K. R. Johnson, M. Luby, and Y. C. Fann, “Applying density-based outlier identifications using multiple datasets for validation of stroke clinical outcomes,” *Int. J. Med. Inform.*, vol. 132, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103988.

[12] F. Andriyani and Y. Puspitarani, “Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews,” *SinkrOn*, vol. 7, no. 3, pp. 944–949, Jul. 2022, doi: 10.33395/sinkron.v7i3.11569.

[13] R. Adesunkanmi and R. Kumar, “Expectation Distance-based Distributional Clustering for Noise-Robustness,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2110.08871

[14] O. Ajmal, H. Arshad, M. A. Arshed, S. Ahmed, and S. Mumtaz, “Robust Parameter Optimisation of Noise-Tolerant Clustering for DENCLUE Using Differential Evolution,” *Mathematics*, vol. 12, no. 21, Nov. 2024, doi: 10.3390/math12213367.

[15] Shabila Ocktavia and Wahyu Tisno Atmojo, “Analisis Segmentasi Pelanggan Pembiayaan Berdasarkan Demografi Untuk Memprediksi Tingkat Kredit Menggunakan Algoritma K-Means”.