Laman web jurnal: https://ejournal.unama.ac.id/index.php/processor

Jurnal Processor

P-ISSN: 1907-6738 | E-ISSN: 2528-0082

Analisis Komparatif K-Means dan Dbscan pada Data Nilai Siswa SMP Bina Bangsa Surabaya

Nur Irvan Rizqi[1], M.Nur Hidayatulloh[2], Faisal Muttaqin[*3]

[1]Prodi Magister Teknologi Informasi UPN "Veteran" Jawa Timur.
[2]Prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Jawa Timur, Indonesia.
*Penulis Korespondensi, Email: faisalmuttaqin.if@upnjatim.ac.id

Abstrak− Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma K-Means dan DBSCAN dalam proses clustering data nilai siswa kelas VIII B SMP Bina Bangsa Surabaya. Data yang digunakan terdiri atas 24 siswa dengan tiga mata pelajaran utama, yaitu Matematika, Bahasa Indonesia, dan Ilmu Pengetahuan Alam (IPA). Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif, yang dilaksanakan melalui platform Google Colab untuk tahap pra-pemrosesan, normalisasi data, serta implementasi algoritma clustering. Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan tiga klaster optimal yang merepresentasikan tingkat kemampuan akademik rendah, sedang, dan tinggi, dengan nilai Silhouette Score tertinggi sebesar ±0,58 pada k = 3. Sementara itu, algoritma DBSCAN membentuk dua klaster utama, yaitu klaster Remedial yang terdiri dari 17 siswa dan klaster Tidak Remedial sebanyak 6 siswa, serta mendeteksi 2 data noise yang memiliki karakteristik nilai berbeda secara signifikan. Temuan ini menunjukkan bahwa K-Means lebih efektif dalam memetakan variasi kemampuan akademik secara bertingkat, sedangkan DBSCAN unggul dalam mengidentifikasi siswa yang memerlukan perhatian khusus melalui deteksi outlier. Hasil penelitian ini menegaskan bahwa penerapan metode clustering dapat menjadi alat bantu analisis yang efektif bagi guru dalam mengidentifikasi kebutuhan belajar siswa dan merancang intervensi pembelajaran yang lebih tepat sasaran.

Kata Kunci: K-Means; DBSCAN; Clustering; Nilai Siswa; Remedial; Analisis Data.

Abstract− This study aims to analyze and compare the performance of the K-Means and DBSCAN algorithms in clustering student academic score data of Class VIII B at SMP Bina Bangsa Surabaya. The dataset consists of 24 students and includes three core subjects, namely Mathematics, Indonesian Language, and Science. This research adopts a quantitative approach using a comparative experimental method, implemented through the Google Colab platform for data preprocessing, normalization, and clustering algorithm implementation. The results indicate that the K-Means algorithm produces three optimal clusters representing low, medium, and high academic performance levels, achieving the highest Silhouette Score of approximately 0.58 at k = 3. Meanwhile, the DBSCAN algorithm forms two main clusters, namely the Remedial cluster consisting of 17 students and the Non-Remedial cluster comprising 6 students, and also detects two noise data points with significantly different score characteristics. These findings suggest that K-Means is more effective in hierarchically mapping variations in academic performance, whereas DBSCAN excels in identifying students who require special attention through outlier detection. Overall, the results confirm that the application of clustering methods can serve as an effective analytical tool for teachers in identifying students' learning needs and designing more targeted instructional interventions.

Keywords: K-Means; DBSCAN; Clustering; Student Performance; Remedial; Data Analysis.

1. PENDAHULUAN

Pada era digital saat ini, perkembangan teknologi informasi dalam bidang pendidikan semakin pesat. Salah satu bentuk penerapannya adalah analisis data nilai siswa untuk mendukung pengambilan keputusan akademik(Soufitri, 2023). Melalui pengelompokan data nilai, sekolah dapat mengenali pola prestasi, merancang strategi pembelajaran yang tepat, dan memberikan perhatian khusus kepada siswa yang membutuhkan. Metode clustering, sebagai bagian dari data mining, menjadi teknik yang efektif untuk mengelompokkan data berdasarkan kemiripan karakteristik(Pronina & Piatykop, 2022a).

Dua algoritma yang sering digunakan dalam proses ini adalah *K-Means* dan DBSCAN (*Density-Based Spatial Clustering of Applications with Noise)*. Algoritma *K-Means* membagi data ke dalam sejumlah klaster berdasarkan kedekatan terhadap pusat klaster *(centroid*)(Du et al., 2021), sedangkan DBSCAN mengelompokkan data berdasarkan kepadatan titik dalam ruang. Keduanya memiliki kelebihan dan kekurangan yang bergantung pada sifat data yang dianalisis(Kulkarni & Burhanpurwala, 2024).

Berbagai penelitian telah membandingkan performa kedua algoritma tersebut. (Hasibuan et al., 2024) melakukan penelitian di SMP Negeri 3 Panyabungan untuk mengelompokkan siswa terbaik berdasarkan beberapa kriteria, dan hasilnya menunjukkan bahwa K-Means menghasilkan klaster yang lebih rapi dan terstruktur dengan

Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

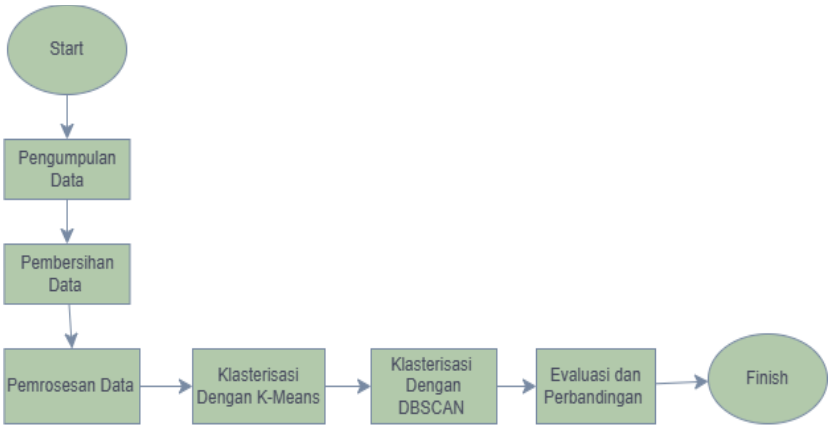
Silhouette Score 0,5697, lebih tinggi dibanding DBSCAN yang memperoleh 0,2580. Dalam penelitian lain (Kertanah et al., 2023) membandingkan kedua metode dalam pengelompokan hasil belajar kognitif siswa pada mata pelajaran fisika. Hasilnya, K-Means kembali menunjukkan kinerja lebih baik dengan Silhouette Score 0,43, sedangkan DBSCAN memperoleh 0,39. Temuan ini memperkuat bahwa K-Means lebih sesuai untuk data yang memiliki distribusi yang jelas. Namun, DBSCAN juga memiliki keunggulan tersendiri, terutama dalam mengenali klaster dengan kepadatan bervariasi serta mampu menangani data yang mengandung noise. Penelitian (Dewi et al., 2021) tentang pengelompokan status desa di Jawa Tengah menunjukkan bahwa DBSCAN lebih tepat digunakan dibanding K-Means karena mampu menyesuaikan diri dengan variasi kepadatan data.

Meskipun berbagai penelitian sebelumnya telah membahas penerapan algoritma K-Means dan DBSCAN dalam konteks pendidikan, sebagian besar studi tersebut masih berfokus pada dataset berskala besar atau konteks institusi tertentu. Penelitian mengenai penerapan dan perbandingan kedua algoritma pada data nilai siswa sekolah menengah pertama dengan jumlah data terbatas masih relatif minim, khususnya dalam konteks pengambilan keputusan pembelajaran berbasis data di tingkat sekolah. Oleh karena itu, penelitian ini menjadi penting untuk memberikan gambaran empiris mengenai efektivitas algoritma K-Means dan DBSCAN dalam mengelompokkan data nilai siswa dengan karakteristik heterogen dan jumlah data yang relatif kecil. Hasil penelitian ini diharapkan dapat menjadi dasar pertimbangan bagi pendidik dalam memilih metode clustering yang paling sesuai untuk mendukung identifikasi kebutuhan remedial dan pengayaan belajar secara lebih akurat dan efisien.

Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan DBSCAN dalam mengelompokkan data nilai siswa kelas VIII B di SMP Bina Bangsa Surabaya. Dengan menelaah keunggulan dan keterbatasan masing-masing algoritma terhadap karakteristik data nilai siswa(Jin, 2025), penelitian ini diharapkan dapat memberikan rekomendasi metode clustering yang paling tepat digunakan di lingkungan pendidikan, guna mendukung peningkatan kualitas pembelajaran dan prestasi siswa(Wang, 2024).

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif untuk membandingkan kinerja algoritma K-Means dan DBSCAN(Cahapin et al., 2023) dalam pengelompokan data nilai siswa kelas VIII B di SMP Bina Bangsa Surabaya. Data yang digunakan berupa nilai akademik dari 24 siswa, meliputi mata pelajaran inti seperti Matematika, Bahasa Indonesia, dan IPA. Pengolahan data dilakukan melalui platform Google *Colaboratory (Google Colab)* karena mendukung bahasa pemrograman Python serta berbagai pustaka data science seperti Pandas, Matplotlib, Scikit-learn, dan Seaborn(Oyelade et al., 2010). Data siswa dimasukkan dalam format CSV dan diolah menggunakan library Pandas(McKinney, 2011).



Gambar 1. Tahapan Penelitian

Proses penelitian diawali dengan pengumpulan data nilai siswa kelas VIII B SMP Bina Bangsa Surabaya, yang kemudian dilanjutkan dengan tahap pra-pemrosesan data meliputi pembersihan data dan normalisasi menggunakan Standard Scaler. proses clustering menggunakan algoritma K-Means, yang diawali dengan penentuan jumlah klaster optimal melalui analisis SSE dan Silhouette Score, kemudian dilanjutkan dengan visualisasi hasil klaster. Setelah itu, dilakukan clustering menggunakan algoritma DBSCAN dengan penentuan parameter eps dan min_samples, serta visualisasi hasil klaster menggunakan PCA.(Wang, 2024). Hasil pengelompokan dari kedua metode kemudian dievaluasi menggunakan Silhouette Score dan visualisasi klaster(Antonenko et al., 2012). Dengan membandingkan hasil tersebut, penelitian ini bertujuan untuk menentukan algoritma yang paling efektif dalam mengelompokkan data nilai siswa yang memiliki karakteristik beragam.

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis dan pembahasan mengenai penerapan serta perbandingan kinerja algoritma clustering K-Means dan DBSCAN dalam mengelompokkan data nilai siswa kelas VIII B di SMP Bina Bangsa Surabaya.

3.1 Proses Clustering dengan K-Means

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

Gambar 2. *Import* data *Library*

Tahapan awal dalam pelaksanaan proyek analisis data menggunakan *Google Colab* adalah melakukan proses impor pustaka-pustaka utama yang diperlukan. Pustaka pandas berfungsi untuk melakukan pengolahan dan manipulasi data, *numpy* digunakan dalam perhitungan dan operasi numerik, sedangkan *matplotlib.pyplot* dan *seaborn* dimanfaatkan untuk membangun visualisasi data yang informatif. Adapun *KMeans* dan *StandardScaler* dari pustaka *sklearn* digunakan dalam proses klasterisasi serta normalisasi data agar hasil analisis menjadi lebih akurat dan terstruktur.

```
[5]
from google.colab import files
uploaded = files.upload()

Choose Files DATA CLUSTERING.xlsx
DATA CLUSTERING.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 11793 bytes, last modified: 10/14/2025 - 100% done
Saving DATA CLUSTERING.xlsx to DATA CLUSTERING (1).xlsx
```

Gambar 3. *Import* data

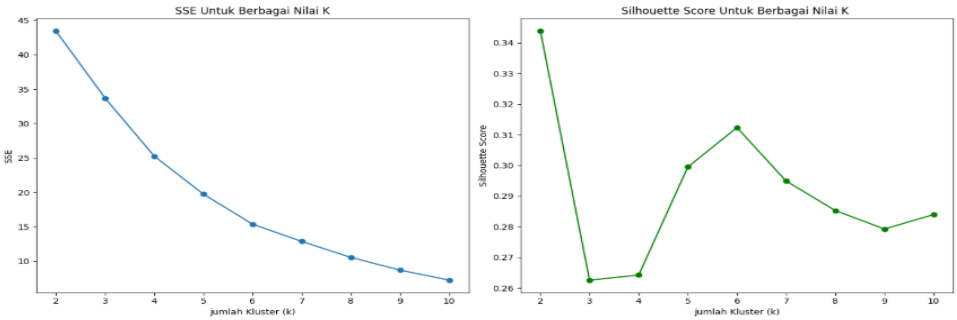
Tahap berikutnya adalah melakukan impor atau pembacaan data dari file Excel ke dalam lingkungan kerja *Google Colab*, sebagaimana ditunjukkan gambar 2. Langkah ini memiliki peran penting karena bertujuan untuk memuat dataset ke dalam sistem, sehingga data tersebut dapat diolah, dianalisis, dan divisualisasikan lebih lanjut menggunakan berbagai metode analisis data.

```
from sklearn.metrics import silhouette_score

K = []
SEE = []
Silhouette = []

for k in range(2, 11):
    K.append(k)
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(df_scaled)
    SEE.append(kmeans.inertia_)
    silhouette_avg = silhouette_score(df_scaled, kmeans.labels_)
    Silhouette.append(silhouette_avg)

plt.figure(figsize=(14, 6)) # Adjusted figure size for two subplots

# Plot SEE
plt.subplot(1, 2, 1) # Corrected subplot index
plt.plot(K, SEE, marker='o')
plt.title('SSE Untuk Berbagai Nilai K')
plt.xlabel('jumlah Kluster (k)')
plt.ylabel('SSE')

# Plot Silhouette Score
plt.subplot(1, 2, 2) # Corrected subplot index
plt.plot(K, Silhouette, marker='o', color='green')
plt.title('Silhouette Score Untuk Berbagai Nilai K ')
plt.xlabel('jumlah Kluster (k)')
plt.ylabel('Silhouette Score')

plt.tight_layout()
plt.show()
```

Gambar 4. Code *SSE* dan *Silhouette Score*

Dalam penerapan analisis klaster menggunakan algoritma seperti *K-Means*, salah satu tantangan utama yang sering dihadapi adalah menentukan jumlah klaster (k) yang paling sesuai atau optimal. Pemilihan nilai k yang tepat sangat penting karena akan memengaruhi kualitas hasil pengelompokan data. Untuk membantu proses penentuan nilai k yang ideal tersebut, biasanya digunakan dua metrik evaluasi utama, yaitu SSE (*Sum of Squared Errors)* dan *Silhouette Score*, yang masing-masing berfungsi untuk menilai seberapa baik data terbagi ke dalam setiap klaster.

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026



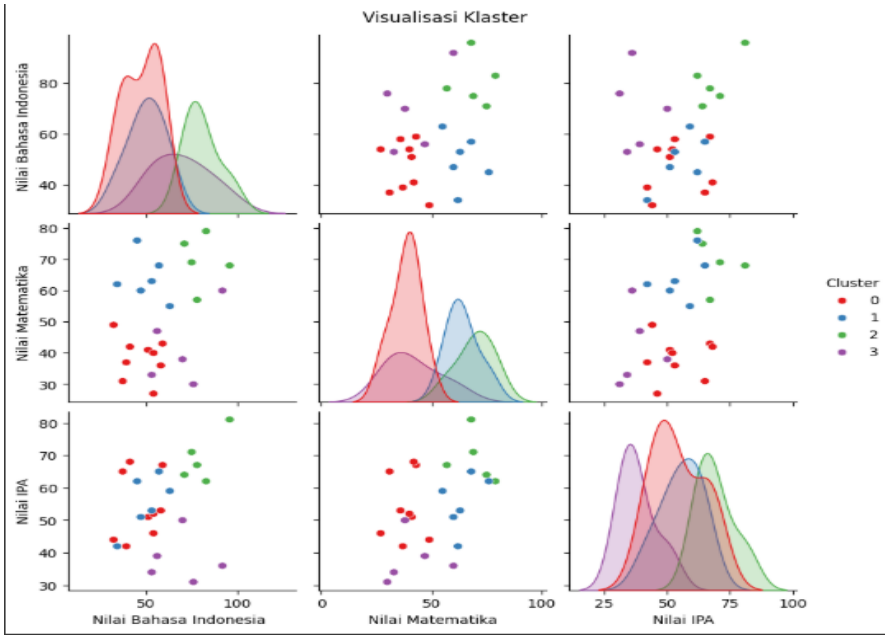
Gambar 5. Hasil *SSE* dan *Silhouette Score*

Berdasarkan hasil visualisasi grafik SSE dan *Silhouette Score* terhadap berbagai nilai jumlah klaster (k), dapat diinterpretasikan bahwa nilai k = 3 merupakan jumlah klaster yang paling optimal untuk data nilai siswa kelas VIII B di SMP Bina Bangsa Surabaya. Grafik SSE memperlihatkan adanya penurunan signifikan dari k = 2 ke k = 3, kemudian diikuti oleh penurunan yang lebih lambat pada nilai k berikutnya. Pola ini menunjukkan adanya titik siku *(elbow point*) pada k = 3, yang menandakan bahwa penambahan jumlah klaster setelah titik tersebut tidak lagi memberikan peningkatan signifikan terhadap kualitas klasterisasi.

Sementara itu, grafik *Silhouette Score* menunjukkan bahwa nilai tertinggi juga terjadi pada k = 3 (sekitar 0,58), yang menandakan bahwa pemisahan antar klaster pada nilai tersebut cukup jelas dan efisien. Dengan demikian, siswa dapat diklasifikasikan menjadi tiga kelompok utama, yaitu kelompok nilai tinggi, sedang, dan rendah. Hasil ini memberikan informasi penting bagi guru untuk menyusun strategi pembelajaran yang lebih terarah dan adaptif, sesuai dengan karakteristik kemampuan akademik siswa di kelas VIII B.

```
kmeans_final = KMeans(n_clusters=4, random_state=42, n_init=10)
kmeans_final.fit(df_scaled)
df['Cluster'] = kmeans_final.labels_

sns.pairplot(df, hue='Cluster',vars=['Nilai Bahasa Indonesia ', 'Nilai Matematika', 'Nilai IPA'] ,palette='Set1')
plt.suptitle('Visualisasi Klaster', y=1.02)
plt.show()
```

Gambar 6. *Code* Visualisasikan klaster



Gambar 7. Hasil Visualisasikan klaster

Gambar tersebut menampilkan hasil visualisasi klasterisasi terhadap siswa kelas VIII B, berdasarkan nilai dari tiga mata pelajaran utama, yaitu Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam (IPA). Proses pengelompokan dilakukan menggunakan algoritma clustering, kemungkinan besar *K-Means*, yang menghasilkan

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

tiga klaster utama dengan perbedaan tingkat kemampuan akademik siswa.Ketiga klaster tersebut ditunjukkan dengan warna yang berbeda, yakni:

1. Biru (Cluster 1) yang menunjukkan siswa dengan tingkat nilai tinggi
2. Hijau (Cluster 2) yang menggambarkan siswa dengan tingkat nilai sedang,
3. Ungu (Cluster 3) yang merepresentasikan siswa dengan tingkat nilai rendah.
4. Merah (Cluster 0) yang menggambarkan siswa dengan nilai sangat rendah

Dari visualisasi tersebut terlihat bahwa pola pengelompokan antar siswa cukup jelas dan terpisah. Klaster biru dominan berada pada area dengan nilai tinggi di ketiga mata pelajaran, klaster hijau terletak pada wilayah nilai rendah, sementara klaster merah menempati posisi di antara keduanya. Hal ini menunjukkan bahwa algoritma yang digunakan mampu mengidentifikasi serta membedakan tingkat kemampuan akademik siswa dengan baik berdasarkan distribusi nilai mereka.

Selanjutnya, grafik diagonal (*density plot)* memperlihatkan bahwa distribusi nilai siswa secara alami terbagi ke dalam tiga kelompok. Misalnya, pada mata pelajaran Matematika, siswa dalam klaster hijau memiliki nilai pada kisaran 20–45, klaster merah berada di rentang 50–70, dan klaster biru mendominasi nilai di atas 75. Pola yang serupa juga terlihat pada mata pelajaran Bahasa Indonesia dan IPA, memperkuat hasil bahwa klasterisasi berhasil menggambarkan variasi kemampuan dengan baik.

Selain itu, *scatterplot* antar variabel (bagian luar diagonal) menunjukkan adanya hubungan positif antar mata pelajaran, di mana siswa yang memiliki nilai tinggi pada satu mata pelajaran cenderung juga memperoleh nilai tinggi pada mata pelajaran lainnya. Temuan ini menegaskan bahwa hasil klasterisasi tidak hanya menggambarkan perbedaan kemampuan dalam satu bidang, melainkan mencerminkan tingkat prestasi akademik secara keseluruhan pada siswa kelas VIII B.

3.2 DBSCAN untuk Clustering Data

```
# Import library
import pandas as pd
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
```

Gambar 8. import Perpustakaan DBSCAN

Gambar **7** menunjukkan langkah awal penggunaan algoritma *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* dalam analisis data menggunakan Python. Pada tahap ini, pustaka **scikit-learn** diimpor untuk menyediakan fungsi pengelompokan data berbasis kepadatan. Langkah ini penting karena memungkinkan peneliti menjalankan DBSCAN untuk mengenali pola, kelompok data, dan anomali secara otomatis sesuai distribusi data yang dianalisis.

```
# Buat Datafrom dari data siswa (Bahasa indonesia, MTK, IPA)
data = {
'Siswa' : ['Siswa1', 'Siswa2', 'Siswa3', 'Siswa4','Siswa5', 'Siswa6','Siswa7','Siswa8','Siswa9','Siswa10','Siswa11','Siswa12','Siswa13','Siswa14','Siswa15','Siswa16','Siswa17','Siswa18','Siswa19','Siswa20','Siswa21','Siswa22','Siswa23','Siswa24'],
'Bahasa Indonesia' : [51,56,96,34,83,75,78,54,39,71,53,47,70,32,59,76,63,41,53,54,37,58,92,57,45],
'MTK' : [51,56,96,34,83,75,78,54,39,71,53,47,70,32,59,76,63,41,53,54,37,58,92,57,45],
'IPA' : [51,56,96,34,83,75,78,54,39,71,53,47,70,32,59,76,63,41,53,54,37,58,92,57,45]
}
df = pd.DataFrame(data)
```

Gambar 9. Input Nilai

Tahapan awal dalam pelaksanaan proyek analisis data menggunakan *Google Colab* adalah melakukan impor pustaka-pustaka utama yang diperlukan untuk mendukung proses pengolahan dan analisis data.

```
x = df[['Bahasa Indonesia','MTK', 'IPA']].values

scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)

dbscan = DBSCAN(eps=0.7, min_samples=3)
clusters = dbscan.fit_predict(x_scaled)

df['Clsuter']= clusters
```

Gambar 10. Code DBSCAN

Gambar 9 memperlihatkan tahapan implementasi algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise)* yang digunakan untuk melakukan pengelompokan data nilai siswa berdasarkan tiga mata pelajaran, yaitu Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam (IPA).

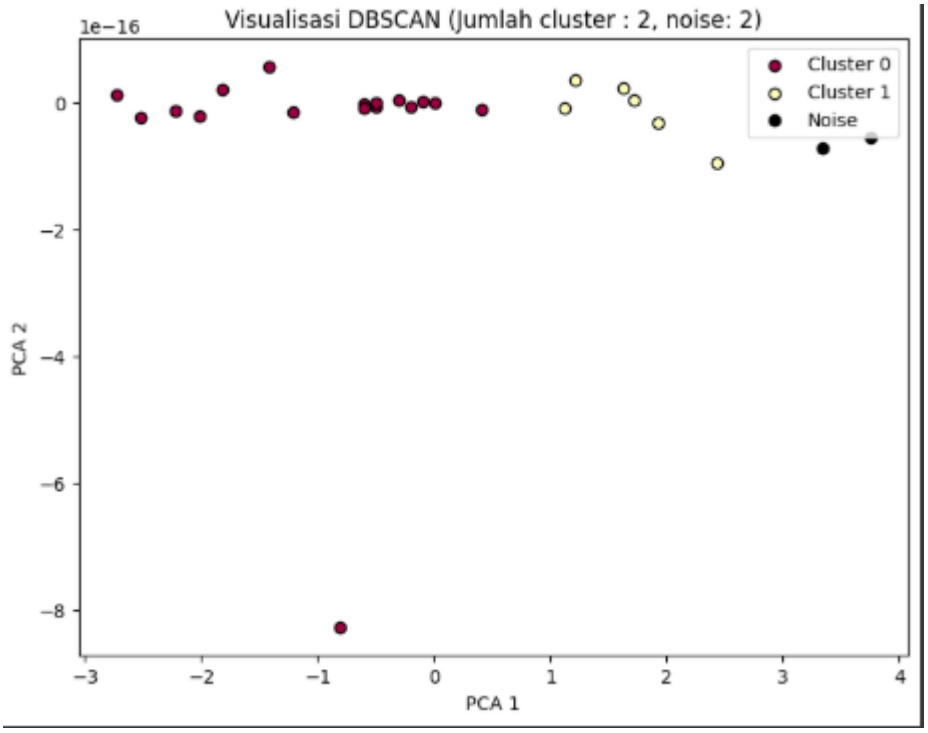
https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

Langkah pertama dilakukan dengan mengambil data nilai ketiga mata pelajaran tersebut dan menyimpannya ke dalam variabel X. Selanjutnya, data tersebut dinormalisasi menggunakan *StandardScaler* agar seluruh variabel memiliki skala yang seragam, karena keseragaman skala berperan penting dalam meningkatkan akurasi hasil klasterisasi. Setelah proses normalisasi, algoritma DBSCAN diterapkan dengan parameter eps = 0.7 dan min_samples = 3, yang berarti jarak maksimum antar titik dalam satu klaster adalah 0,7 dan minimal tiga data diperlukan untuk membentuk satu klaster.

Hasil dari proses klasterisasi ini kemudian disimpan dalam kolom baru bernama *'Cluster'* pada DataFrame. Melalui tahapan ini, sistem dapat mengelompokkan siswa secara otomatis berdasarkan kemiripan pola nilai akademik mereka, serta mendeteksi siswa yang tidak termasuk dalam kelompok mana pun *(outlier)*, yang dapat menjadi perhatian khusus dalam analisis prestasi belajar.

```
from sklearn.decomposition import PCA

# Visualisasi Hasil DBSCAN dengan PCA (2 dimensi)
pca = PCA(n_components=2)
x_pca = pca.fit_transform(x_scaled)

plt.figure(figsize=(8, 6))
unique_labels = set(clusters)
colors =[plt.cm.Spectral(each)for each in np.linspace(0, 1, len(unique_labels))] # Use np.linspace for color mapping

for label, color in zip(unique_labels, colors):
  if label == -1:
    # Black used for noise.
    color = [0, 0, 0, 1]

  class_member_mask = (clusters == label)

  # Plot the points in the current cluster
  xy = x_pca[class_member_mask]
  plt.scatter(xy[:, 0], xy[:, 1], c=[color], label=f'Cluster {label}' if label != -1 else 'Noise', edgecolors='k')

plt.title(f'Visualisasi DBSCAN (Jumlah cluster : {len(unique_labels) - (1 if -1 in unique_labels else 0)}, noise: {list(clusters).count(-1)})')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.legend()
plt.show()
```

Gambar 11. Code DBSCAM menggunakan PCA



Gambar 12. Hasil *Code DBSCAM* menggunakan *PCA*

Gambar 11 memperlihatkan hasil visualisasi penerapan algoritma DBSCAN *(Density-Based Spatial Clustering of Applications with Noise)* terhadap data nilai siswa. Berdasarkan hasil analisis, terbentuk dua klaster utama serta dua data yang teridentifikasi sebagai noise (outlier). Pada grafik, Cluster 0 direpresentasikan dengan titik berwarna merah, sedangkan Cluster 1 ditandai dengan warna kuning, dan data noise digambarkan dengan titik hitam. Sumbu horizontal (PCA 1) dan sumbu vertikal (PCA 2) merupakan hasil reduksi dimensi menggunakan *Principal*

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

Component Analysis (PCA), yang digunakan untuk memudahkan visualisasi pola klasterisasi dalam ruang dua dimensi. Dari hasil visualisasi terlihat bahwa DBSCAN mampu mengelompokkan data siswa berdasarkan kemiripan pola nilai akademik mereka, sekaligus mengidentifikasi beberapa siswa yang tidak sesuai dengan kelompok mana pun. Kehadiran dua data noise ini menunjukkan adanya siswa dengan karakteristik nilai yang berbeda secara signifikan dari mayoritas, sehingga memerlukan perhatian khusus dalam analisis prestasi belajar.

```
# Add cluster labels to the DataFrame
df['Cluster'] = clusters

# Buat kolom keterangan sesuai cluster
df['Keterangan'] = df ['Cluster']. apply(lambda x: 'Tidak Remedial'if x == 1 else ('Remedial' if x == 0 else 'Noise'))
# Tampilkan tabel dengan koiom cluster dan keterangan di sebelahnya
print(df[['siswa', 'Bahasa Indonesia', 'MTK', 'IPA', 'Cluster', 'Keterangan']])
```

Gambar 13. Keterangan Cluster

Gambar 12 memperlihatkan adanya penambahan tabel keterangan yang berfungsi untuk mempermudah pemahaman terhadap hasil pembentukan dua klaster. Pada tahap ini, klaster dengan nilai **0** diberi label “Remedial”, sedangkan klaster dengan nilai 1 diberi label “Tidak Remedial”. Penambahan tabel tersebut bertujuan untuk memperjelas interpretasi hasil proses klasterisasi, sehingga setiap kelompok siswa dapat dikenali dengan lebih mudah sesuai dengan kategori yang telah ditetapkan..

	siswa	Bahasa Indonesia	MTK	IPA	Cluster	Keterangan
0	Siswa1	51	51	51	0	Remedial
1	Siswa2	56	56	56	0	Remedial
2	Siswa3	96	96	96	-1	Noise
3	Siswa4	34	34	34	0	Remedial
4	Siswa5	83	83	83	1	Tidak Remedial
5	Siswa6	75	75	75	1	Tidak Remedial
6	Siswa7	78	78	78	1	Tidak Remedial
7	Siswa8	54	54	54	0	Remedial
8	Siswa8	39	39	39	0	Remedial
9	Siswa9	71	71	71	1	Tidak Remedial
10	Siswa10	53	53	53	0	Remedial
11	Siswa11	47	47	47	0	Remedial
12	Siswa12	70	70	70	1	Tidak Remedial
13	Siswa13	32	32	32	0	Remedial
14	Siswa14	59	59	59	0	Remedial
15	Siswa15	76	76	76	1	Tidak Remedial
16	Siswa16	63	63	63	0	Remedial
17	Siswa17	41	41	41	0	Remedial
18	Siswa18	53	53	53	0	Remedial
19	Siswa19	54	54	54	0	Remedial
20	Siswa20	37	37	37	0	Remedial
21	Siswa21	58	58	58	0	Remedial
22	Siswa22	92	92	92	-1	Noise
23	Siswa23	57	57	57	0	Remedial
24	Siswa24	45	45	45	0	Remedial

Gambar 14. Nilai Cluster 1

Berdasarkan keterangan gambar 13, Cluster 1 diidentifikasi sebagai kelompok yang terdiri dari siswa dengan nilai di atas 70, yang menunjukkan bahwa mereka termasuk dalam kategori “Tidak Remedial”. Hal ini mengindikasikan bahwa siswa dalam klaster ini memiliki pemahaman materi yang baik dan telah mencapai standar ketuntasan belajar yang ditetapkan.

```
# hitung jumlah berdasarkan keterangan
jumlah_keterangan = df['Keterangan'].value_counts()

print(jumlah_keterangan)
```

```
Keterangan
Remedial         17
Tidak Remedial    6
Noise             2
Name: count, dtype: int64
```

Gambar 15. Hasil Semua Cluster

Berdasarkan hasil analisis kode yang telah dijelaskan sebelumnya, dapat disimpulkan bahwa dari seluruh siswa kelas VIII B, terdapat 6 siswa yang termasuk dalam kategori “Tidak Remedial”, yang berarti mereka telah memenuhi standar kelulusan tanpa perlu mengikuti program perbaikan nilai. Sementara itu, 17 siswa lainnya tercatat sebagai “Remedial”, yang menunjukkan bahwa mereka perlu menjalani pembelajaran tambahan untuk memperbaiki atau meningkatkan hasil belajar agar dapat mencapai kriteria ketuntasan yang telah ditetapkan.

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

4.3. Intepretasi Hasil

Proses clustering terhadap data nilai siswa kelas VIII B SMP Bina Bangsa Surabaya dilakukan menggunakan dua metode yang umum digunakan, yaitu K-Means dan DBSCAN. Hasil analisis dari kedua metode tersebut memberikan gambaran yang jelas mengenai variasi kemampuan akademik siswa berdasarkan tiga mata pelajaran utama, yaitu Matematika, Bahasa Indonesia, dan IPA.

Pada penerapan metode K-Means, penentuan jumlah klaster yang paling optimal dilakukan melalui analisis terhadap SSE *(Sum of Squared Errors)* dan *Silhouette Score*. Berdasarkan hasil evaluasi tersebut, diperoleh bahwa tiga klaster merupakan jumlah yang paling tepat untuk menggambarkan pola pengelompokan siswa. Klaster-klaster tersebut mewakili tiga tingkatan kemampuan akademik, yaitu rendah, sedang, dan tinggi. Hasil analisis ini memberikan manfaat penting bagi guru dalam mengidentifikasi kelompok siswa berdasarkan tingkat kemampuan akademiknya. Melalui hasil pengelompokan tersebut, guru dapat lebih mudah mengenali siswa dengan nilai rendah yang memerlukan program remedial, siswa dengan nilai sedang yang masih memiliki peluang untuk ditingkatkan, serta siswa dengan nilai tinggi yang layak diberikan tantangan pembelajaran yang lebih lanjut. Temuan ini sejalan dengan hasil penelitian (Setyani et al., 2025) yang menyatakan bahwa algoritma *K-Means* efektif dalam mengelompokkan siswa berdasarkan prestasi akademik, sehingga membantu dalam identifikasi dini terhadap siswa yang berpotensi mengalami kesulitan belajar.

Visualisasi hasil clustering menunjukkan pola distribusi nilai yang alami dan konsisten, dengan adanya korelasi positif antar mata pelajaran. Hal ini berarti bahwa siswa yang memiliki nilai tinggi pada satu mata pelajaran cenderung juga memperoleh nilai tinggi pada mata pelajaran lainnya. Temuan ini memperkuat validitas model clustering, yang tidak hanya membedakan kemampuan siswa pada satu bidang saja, tetapi juga menggambarkan tingkat kemampuan akademik secara menyeluruh. Penelitian (Billan & Sutabri, 2025) juga menegaskan pentingnya memperhatikan korelasi antar variabel dalam proses clustering agar hasil pengelompokan menjadi lebih akurat dan relevan.

Selanjutnya, Justifikasi pemilihan parameter DBSCAN dilakukan dengan eps = 0.7 dan min samples = 3. Nilai eps ditentukan berdasarkan analisis jarak (k-distance graph) yang menunjukkan jarak maksimum antartitik yang masih termasuk dalam satu klaster. Nilai 0.7 dipilih karena menghasilkan klaster yang stabil tanpa terlalu banyak data noise, sedangkan min_samples = 3 disesuaikan dengan ukuran dataset (24 siswa). Kombinasi parameter ini mengacu pada praktik umum dalam penelitian sejenis (Hasibuan et al., 2024), (Dewi et al., 2021) yang menunjukkan keseimbangan antara deteksi outlier dan stabilitas klaster. penerapan metode DBSCAN memberikan pendekatan alternatif dalam proses pengelompokan data karena tidak memerlukan penentuan jumlah klaster di awal. Selain itu, DBSCAN mampu mendeteksi data yang menyimpang *(outlier atau noise)*, yang dalam penelitian ini ditemukan beberapa data dengan karakteristik tersebut. Metode ini menghasilkan dua klaster utama, yaitu Klaster 0 (Remedial) dan Klaster 1 (Tidak Remedial). Siswa yang termasuk dalam klaster Remedial umumnya memiliki nilai di bawah 70, yang menunjukkan perlunya intervensi pembelajaran tambahan agar mereka dapat mencapai standar ketuntasan minimal. Hasil ini sejalan dengan temuan (Sutramiani et al., 2024) yang menunjukkan bahwa algoritma DBSCAN efektif dalam mengidentifikasi kelompok siswa berdasarkan performa akademik, sekaligus membantu dalam mendeteksi kelompok siswa dengan kebutuhan pembelajaran yang berbeda. Hasil analisis akhir menunjukkan bahwa dari 24 siswa kelas VIII B, terdapat 6 siswa yang tergolong dalam klaster Tidak Remedial, sedangkan 17 siswa lainnya termasuk dalam klaster Remedial, dan 2 siswa teridentifikasi sebagai data noise atau berada di luar pola pengelompokan utama. Berdasarkan hasil tersebut, guru dapat lebih terarah dalam merancang strategi pembelajaran yang sesuai dengan kebutuhan masing-masing kelompok. Bagi kelompok siswa Remedial, guru dapat mengembangkan pendekatan pembelajaran yang lebih intensif dan adaptif, seperti pemberian materi pengayaan, bimbingan tambahan, atau metode pembelajaran yang lebih interaktif dan personal guna membantu mereka mencapai standar kompetensi yang ditetapkan. Selain itu, penggunaan metode clustering dalam konteks pendidikan ini sejalan dengan temuan (Pronina & Piatykop, 2022b) yang menyatakan bahwa penerapan teknik clustering dapat berperan penting dalam memprediksi kebutuhan program remedial, sehingga intervensi pembelajaran dapat dilakukan secara lebih tepat sasaran, efisien, dan efektif dalam meningkatkan capaian belajar siswa.

Secara keseluruhan, penerapan metode clustering terbukti sangat efektif dalam membantu pengelolaan kelas dengan tingkat kemampuan akademik yang beragam. Melalui pendekatan ini, guru dapat mengidentifikasi kebutuhan belajar setiap siswa secara lebih spesifik dan menyusun strategi pembelajaran yang sesuai dengan karakteristik masing-masing kelompok.

Pendekatan berbasis clustering ini tidak hanya meningkatkan efisiensi proses pembelajaran, tetapi juga mendorong peningkatan potensi keberhasilan akademik siswa secara menyeluruh. Temuan ini sejalan dengan hasil penelitian (Pecuchova & Drlik, 2024) yang menyatakan bahwa pengelompokan siswa menggunakan metode clustering dapat menjadi dasar dalam merancang intervensi pendidikan yang lebih terarah dan berdampak positif terhadap capaian belajar siswa

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *K-Means* dan DBSCAN pada data nilai siswa kelas VIII B SMP Bina Bangsa Surabaya memberikan hasil yang signifikan dalam mengelompokkan tingkat kemampuan akademik siswa. Algoritma *K-Means* berhasil membentuk tiga klaster utama yang merepresentasikan kategori kemampuan rendah, sedang, dan tinggi, dengan nilai Silhouette Score terbaik pada k = 3, yang menunjukkan kualitas pemisahan klaster yang baik.

Sementara itu, algoritma DBSCAN menghasilkan dua klaster utama, yaitu Remedial (17 siswa) dan Tidak Remedial (6 siswa), serta dua data noise yang menunjukkan adanya siswa dengan karakteristik nilai yang berbeda dari mayoritas. Hasil ini membuktikan bahwa DBSCAN mampu mendeteksi data yang menyimpang (outlier) dengan lebih efektif tanpa perlu menentukan jumlah klaster di awal.

Secara keseluruhan, kedua metode clustering terbukti efektif dalam memetakan variasi kemampuan akademik siswa, memberikan gambaran yang jelas tentang kebutuhan pembelajaran individual, serta mendukung guru dalam merancang strategi pembelajaran yang lebih adaptif dan tepat sasaran. Dengan demikian, penerapan metode ini berpotensi menjadi alat bantu analitik yang penting dalam pengambilan keputusan di bidang pendidikan, khususnya dalam pengelolaan kelas yang heterogen secara akademik.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan terima kasih yang tulus kepada pihak pemberi dana penelitian atas kepercayaan dan dukungan yang diberikan sehingga penelitian ini dapat terlaksana dengan baik. Bantuan finansial tersebut memiliki peran penting dalam mendukung setiap tahapan kegiatan penelitian hingga selesai.

Ucapan terima kasih juga penulis tujukan kepada Bapak Dr. Faisal Muttaqin, S.Kom., M.T., atas bimbingan, arahan, serta masukan berharga yang telah memberikan kontribusi besar terhadap proses penyusunan dan penyempurnaan penelitian ini.

Penulis menyadari bahwa laporan penelitian ini masih memiliki keterbatasan, sehingga kritik dan saran yang konstruktif sangat diharapkan demi penyempurnaan penelitian pada kesempatan berikutnya.

REFERENCES

[1] Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

[2] Billan, A. C., & Sutabri, T. (2025). Restorasi penjadwalan sumur minyak yang mengalami off-time menggunakan algoritma backtracking dalam upaya optimasi produksi. *Bulletin of Computer Science Research*, *5*(3), 228–234.

[3] Cahapin, E. L., Malabag, B. A., Santiago Jr, C. S., Reyes, J. L., Legaspi, G. S., & Adrales, K. L. (2023). Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms. *Bulletin of Electrical Engineering and Informatics*, *12*(6), 3647–3656.

[4] Dewi, C., Siam, E. P., Wijayanti, G. A., Putri, M., Aulia, N., & Nooraeni, R. (2021). Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020. *Jurnal Matematika, Statistika Dan Komputasi*, *17*(3), 394–404. https://doi.org/10.20956/j.v17i3.11704

[5] Du, H., Chen, S., Niu, H., & Li, Y. (2021). Application of DBSCAN clustering algorithm in evaluating students' learning status. *2021 17th International Conference on Computational Intelligence and Security (CIS)*, 372–376. https://ieeexplore.ieee.org/abstract/document/9701684/

[6] Hasibuan, M. S., Lubis, A. H., & Sari, M. N. (2024). Perbandingan algoritma clustering dbscan dan k-means dalam pengelompokan siswa terbaik. *INFOTECH: Jurnal Informatika & Teknologi*, *5*(2), 301–309.

[7] Jin, J. (2025). Student behavior patterns in vocational education big data based on clustering algorithm. *Discover Artificial Intelligence*, *5*(1), 197. https://doi.org/10.1007/s44163-025-00433-3

[8] Kertanah, K., Nurmayanti, W. P., Aini, S. R., Amrullah, L. Muh., & Sya'roni, M. (2023). Comparison of Algorithms K-Means and DBSCAN for Clustering Student Cognitive Learning Outcomes in Physics Subject. *Kappa Journal*, *7*(2), 251–255. https://doi.org/10.29408/kpj.v7i2.18428

[9] Kulkarni, O., & Burhanpurwala, A. (2024). A survey of advancements in DBSCAN clustering algorithms for big data. *2024 3rd International Conference on Power Electronics and IoT Applications in Renewable Energy and Its Control (PARC)*, 106–111. https://ieeexplore.ieee.org/abstract/document/10486339/

[10] McKinney, W. (2011). pandas: A foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, *14*(9), 1–9.

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026

[11] Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). *Application of k Means Clustering algorithm for prediction of Students Academic Performance* (No. arXiv:1002.2425). arXiv. https://doi.org/10.48550/arXiv.1002.2425
[12] Pecuchova, J., & Drlik, M. (2024). Enhancing the early student dropout prediction model through clustering analysis of students' digital traces. *IEEE Access*. https://ieeexplore.ieee.org/abstract/document/10736593/
[13] Pronina, O., & Piatykop, O. (2022a). Predicting Students' Academic Performance Based on the Cluster Analysis Method. In O. Ignatenko, V. Kharchenko, V. Kobets, H. Kravtsov, Y. Tarasich, V. Ermolayev, D. Esteban, V. Yakovyna, & A. Spivakovsky (Eds.), *ICTERI 2021 Workshops* (Vol. 1635, pp. 292–307). Springer International Publishing. https://doi.org/10.1007/978-3-031-14841-5_19
[14] Pronina, O., & Piatykop, O. (2022b). Predicting Students' Academic Performance Based on the Cluster Analysis Method. In O. Ignatenko, V. Kharchenko, V. Kobets, H. Kravtsov, Y. Tarasich, V. Ermolayev, D. Esteban, V. Yakovyna, & A. Spivakovsky (Eds.), *ICTERI 2021 Workshops* (Vol. 1635, pp. 292–307). Springer International Publishing. https://doi.org/10.1007/978-3-031-14841-5_19
[15] Setyani, T., Indriani, Y., Fadli, M., & Susanto, E. R. (2025). Perbandingan Algoritma K-Means dan DBSCAN dalam Pengelompokkan Data Penjualan Elektronik. *Progresif: Jurnal Ilmiah Komputer*, *21*(2), 712–721. https://doi.org/10.35889/progresif.v21i2.2775
[16] Soufitri, F. (2023). *Konsep sistem informasi*. PT Inovasi Pratama Internasional.
[17] Sutramiani, N. P., Arthana, I. M. T., Lampung, P. F., Aurelia, S., Fauzi, M., & Darma, I. W. A. S. (2024). The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover. *Journal of Information Systems Engineering and Business Intelligence*, *10*(1), 13–24. https://doi.org/10.20473/jisebi.10.1.13-24
[18] Wang, J. (2024). Predicting Student Performance Using Optimized DBSCAN-K-Means Clustering and RBF Neural Networks. *Informatica*, *48*(17), 33–46.

https://doi.org/10.33998/processor.2026.21.1.2636
Submitted: 20 Oktober 2025; Reviewed: 18 November 2025; Accepted; 23 Desember 2025; Published: 30 April 2026