Jurnal Elektronik Ilmu Komputer Udayana
Volume 13, No 4. May 2025

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasterisasi Siswa Unggulan Menggunakan Algoritma K-Means di SMAS Sultan Agung Puger

Mas'ud Hermansyah[a1], Mujiono[a2], Akas Bagus Setiawan[a3], M. Faiz Firdausi[b4], Iqbal Sabilirrasyad[b5]

[a]Jurusan Teknologi Informasi, Politeknik Negeri Jember
Jl. Mastrip, Sumbersari, Kec. Sumbersari, Kabupaten Jember, Jawa Timur, Indonesia
[1]mas_udhermansyah@polije.ac.id
[2]mujiono@polije.ac.id
[3]akasbagus_s@polije.ac.id

[b]Fakultas Sains, Teknologi dan Industri, Institut Teknologi dan Sains Mandala
Jl. Sumatra, Sumbersari, Kec. Sumbersari, Kabupaten Jember, Jawa Timur, Indonesia
[4]faizfirdausi@itsm.ac.id
[5]iqbal@itsm.ac.id

Abstract

Improving the quality of education does not only depend on academic aspects, but also needs to consider non-academic aspects such as extracurricular participation, social attitudes, and student attendance. Therefore, an analysis method is needed that is able to group students comprehensively based on these various indicators. This study aims to group class X students of SMAS Sultan Agung Puger based on academic and non-academic aspects using the K-Means Clustering algorithm. The data used include academic grades, involvement in extracurricular activities, achievement, social attitude values, and the number of absences. The data processing process is carried out through the pre-processing stage, data transformation, application of the K-Means algorithm, and evaluation of clustering results using the Davies-Bouldin Index (DBI) method. The results of the analysis show that the formation of 4 clusters is the optimal structure with a DBI value of 1.1601. Each cluster has different characteristics that reflect the level of student achievement in various aspects. This clustering provides useful information for schools in designing student development strategies based on group needs. These findings support the application of a data-based approach in decision making in the field of education.

Keywords: *K-Means, clustering, excellent students, Davies-Bouldin Index (DBI)*

1. Introduction

Proses identifikasi siswa unggulan di jenjang pendidikan menengah umumnya masih terbatas pada penilaian aspek akademik saja. Hal ini menyebabkan potensi siswa yang memiliki keunggulan dalam bidang non-akademik seperti prestasi ekstrakurikuler, keterampilan sosial, dan keaktifan dalam organisasi sekolah menjadi terabaikan [1]. Permasalahan ini juga dirasakan di SMAS Sultan Agung Puger, terutama dalam mengelompokkan siswa baru dari tingkat SMP ke dalam kategori pembinaan dan pengembangan sesuai potensi mereka. Tanpa sistem yang akurat dalam memetakan potensi siswa sejak awal masuk, pihak sekolah kesulitan dalam merancang strategi pembelajaran dan kegiatan yang tepat guna mendukung perkembangan siswa secara menyeluruh.

Sebagai upaya untuk mengatasi masalah tersebut, diperlukan sebuah pendekatan sistematis berbasis teknologi yang mampu mengelompokkan siswa berdasarkan kinerja akademik dan non-akademik secara objektif. Salah satu solusi yang dapat diterapkan adalah pemanfaatan *data mining* melalui metode *clustering*, dengan algoritma *K-Means* sebagai salah satu teknik populer [2]. Algoritma ini dapat digunakan untuk mengelompokkan siswa baru berdasarkan data seperti rata-rata nilai akademik, prestasi lomba, keterlibatan dalam ekstrakurikuler, kehadiran, hingga sikap sosial. Dengan melakukan klasterisasi ini, sekolah dapat memperoleh pemetaan siswa yang lebih adil dan menyeluruh, serta menyesuaikan program pembinaan yang sesuai dengan profil masing-masing kelompok siswa [3].

Beberapa penelitian sebelumnya menunjukkan keberhasilan penggunaan algoritma *K-Means* dalam dunia pendidikan. Penelitian oleh [4] memanfaatkan algoritma *K-Means* untuk klasterisasi performa siswa dalam pembelajaran Bahasa Indonesia. Studi serupa pernah dilakukan oleh [5] dengan algoritma yang sama melakukan analisis keterkaitan prestasi akademik (IPK) dengan aktivitas mahasiswa dalam organisasi kemahasiswaan. Peneliti lain yang turut mengkaji topik ini adalah [6], berhasil memanfaatkan *K-Means* dalam menentukan kelompok siswa berprestasi di SMK berdasarkan kombinasi nilai rapor dan tingkat kehadiran. Kajian sebelumnya yang relevan juga dilakukan oleh [7] mengelompokkan siswa berdasarkan aspek nilai akademik dan aspek penilaian prakarya. Sebagaimana diteliti oleh [8] penggunaan algoritma *K-Means* untuk mengelompokkan siswa ekstrakurikuler aktif dalam menyusun strategi pembelajaran akademik yang tepat. Meskipun demikian, mayoritas penelitian tersebut masih terbatas pada aspek akademik atau hanya sebagian variabel non-akademik.

Studi-studi yang telah dilakukan memperlihatkan masih adanya ruang kosong dalam kajian penelitian, yaitu belum banyak pendekatan yang secara eksplisit menggabungkan dimensi akademik dan non-akademik dalam proses klasterisasi siswa di jenjang pendidikan menengah, khususnya pada konteks siswa baru yang belum memiliki rekam jejak panjang di sekolah tersebut. Selain itu, belum banyak penelitian yang secara spesifik memanfaatkan data siswa baru untuk proses pengelompokan awal sebagai landasan pembinaan jangka panjang. Hal ini membuka peluang bagi penelitian lebih lanjut yang dapat memperluas pendekatan analisis data pendidikan dengan basis multidimensi.

Data dalam penelitian ini diperoleh dari siswa baru yang masuk ke SMAS Sultan Agung Puger, yang berasal dari berbagai latar belakang sekolah SMP. Data yang dianalisis meliputi rata-rata nilai akademik, jumlah dan jenis kegiatan ekstrakurikuler yang diikuti, jumlah sertifikat kejuaraan, nilai sikap atau sosial yang diberikan guru, serta persentase kehadiran siswa. Klasterisasi data tersebut berpotensi menjadi sarana pendukung bagi sekolah dalam merumuskan strategi pembinaan yang berbasis pada data, serta menghindari pengambilan keputusan subjektif dalam menentukan siswa yang masuk kategori unggulan [9].

Berdasarkan latar belakang tersebut, tujuan dari penelitian ini adalah untuk mengembangkan metode klasterisasi siswa unggulan berdasarkan kinerja akademik dan non-akademik menggunakan algoritma *K-Means*, dengan memanfaatkan data siswa baru di SMAS Sultan Agung Puger. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengelolaan pendidikan yang lebih cerdas, adil, dan berbasis data. Hasil dari penelitian ini juga diharapkan dapat menjadi referensi dalam pengembangan sistem informasi manajemen siswa yang lebih adaptif dan proaktif dalam mendukung potensi siswa sejak awal mereka masuk ke jenjang SMA.

2. Reseach Methods

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis teknik *data mining*, khususnya algoritma *K-Means Clustering*. Metode ini dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik multidimensional, sehingga memungkinkan identifikasi pola tersembunyi dalam data siswa yang berkaitan dengan aspek akademik dan non-akademik [2]. Algoritma *K-Means* dinilai efektif untuk menghasilkan klaster yang merepresentasikan kelompok siswa dengan karakteristik serupa, guna mendukung proses pengambilan keputusan yang lebih terarah dalam pengembangan potensi siswa [10].

Populasi dalam penelitian ini adalah seluruh peserta didik kelas 10 di SMAS Sultan Agung Puger pada tahun ajaran berjalan. Sampel yang digunakan berjumlah 70 siswa, yang merupakan siswa baru pada tingkat kelas 10. Teknik pengambilan sampel dilakukan secara purposif dengan pertimbangan bahwa siswa pada jenjang awal pendidikan menengah atas memiliki potensi beragam yang relevan untuk dianalisis, baik dari segi akademik maupun non-akademik, guna mendukung pengelompokan secara lebih komprehensif.

Tahapan penelitian meliputi proses pengumpulan data, praproses data, transformasi data, penerapan algoritma *K-Means*, dan analisis hasil klaster [3]. Data yang dikumpulkan meliputi rata-rata nilai akademik, jumlah keterlibatan dalam kegiatan ekstrakurikuler, jumlah prestasi akademik dan non-akademik, nilai sikap sosial, serta jumlah ketidakhadiran (absensi). Data yang telah terkumpul kemudian melalui tahap praproses yang mencakup pembersihan data dan normalisasi untuk memastikan keseragaman skala antar fitur. Selanjutnya, dilakukan penerapan algoritma *K-Means* untuk membentuk sejumlah klaster [11]. Secara umum, proses kerja algoritma K-Means terdiri dari beberapa tahapan utama sebagai berikut [12]:

1. Menentukan jumlah klaster (k):
   Jumlah klaster (k) ditentukan terlebih dahulu sebelum proses klasterisasi dimulai.
2. Inisialisasi centroid awal:
   Dipilih secara acak sebanyak k titik dari data sebagai pusat klaster awal *(centroid).*
3. Menghitung jarak setiap data ke *centroid*:
   Setiap data akan dihitung jaraknya terhadap semua centroid menggunakan metrik jarak tertentu, umumnya *Euclidean Distance*:
   $d(x,c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 + \cdots + (x_n - c_n)^2}$ (1)
   dimana :
   x adalah vektor data dan c adalah vektor *centroid*.
4. Pengelompokan data ke klaster terdekat:
   Data akan dimasukkan ke dalam klaster dengan jarak terdekat ke *centroid*.
5. Menghitung ulang posisi *centroid*:
   Centroid baru dihitung sebagai rata-rata dari semua data yang termasuk dalam masing-masing klaster.
6. Iterasi hingga *konvergen*:
   Langkah 3–5 diulangi hingga posisi *centroid* tidak lagi berubah secara signifikan, atau sampai mencapai batas iterasi tertentu.

Setelah proses klasterisasi dilakukan menggunakan algoritma *K-Means*, tahap selanjutnya adalah mengevaluasi kualitas hasil klaster yang terbentuk. Evaluasi ini dilakukan dengan menggunakan metode *Davies-Bouldin Index* (DBI), yang merupakan salah satu metrik evaluasi internal untuk mengukur seberapa baik pemisahan dan kepadatan antar klaster [13]. DBI menghitung rasio antara jarak antar klaster *(inter-cluster distance)* dan jarak dalam klaster *(intra-cluster distance)* untuk setiap pasang klaster. Nilai DBI yang lebih kecil menunjukkan bahwa klaster yang terbentuk memiliki pemisahan yang lebih baik dan distribusi data dalam setiap klaster lebih kompak, sehingga hasil klasterisasi dianggap lebih optimal [14]. Secara matematis, DBI didefinisikan sebagai [15]:

$$DBI = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq 1} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \quad (2)$$

$\sigma_i \; dan \; \sigma_j$ = rata-rata jarak dari setiap anggota klaster i dan j ke *centroid*-nya masing-masing

$d(c_i, c_j)$ = jarak antar *centroid* klaster i dan j

n = jumlah total klaster

Dalam konteks penelitian ini, DBI digunakan untuk menilai seberapa optimal pemilihan jumlah klaster (k) dalam proses *K-Means*. Nilai DBI yang lebih rendah menunjukkan bahwa pemilihan jumlah klaster tersebut menghasilkan struktur klaster yang lebih baik. Hasil pengujian ini kemudian menjadi dasar dalam pengambilan keputusan terhadap efektivitas klasterisasi yang telah dilakukan.

3. Result and Discussion

Bagian ini menyajikan hasil dari proses analisis data yang telah dilakukan menggunakan algoritma *K-Means*, serta pembahasan terhadap temuan yang diperoleh. Klasterisasi dilakukan dengan menerapkan metode *data mining* menggunakan data siswa yang mencerminkan aspek akademik dan non-akademik, dengan tujuan untuk mengelompokkan siswa ke dalam sejumlah klaster yang memiliki kesamaan karakteristik.

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data siswa kelas 10 SMAS Sultan Agung Puger yang diperoleh melalui dokumentasi dari pihak sekolah, khususnya dari bidang kurikulum. Jumlah sampel yang digunakan sebanyak 70 siswa, yang merupakan peserta didik baru pada tahun ajaran berjalan. Data yang dikumpulkan mencakup hasil akademik dan non-akademik siswa dari laporan belajar ketika masih di SMP sebagai dasar untuk proses klasterisasi. Data ini kemudian diolah dan dianalisis menggunakan algoritma *K-Means Clustering* untuk mengelompokkan siswa ke dalam beberapa klaster berdasarkan kesamaan karakteristik, guna mengidentifikasi siswa unggulan yang berpotensi untuk pembinaan lebih lanjut.

Tabel 1. Pengumpulan Data

ID_Siswa	RNA	JE	PA	PNA	NS	AB
S001	82.48	0	1	0	Baik Sekali	7

S002	79.31	3	0	2	Baik	6
S003	83.24	2	1	1	Baik Sekali	6
S004	87.62	0	3	0	Baik Sekali	3
S005	78.83	3	0	2	Baik	18
…	…	…	…	…	…	…
S069	80.46	1	0	0	Baik	14
S070	74.06	2	0	5	Baik	13

Data yang digunakan mencakup berbagai atribut penilaian yang mewakili kedua aspek tersebut, di antaranya nilai rata-rata mata pelajaran inti seperti rata-rata nilai akademik (RNA) dan prestasi akademik (PA) serta data non-akademik berupa keterlibatan dalam kegiatan ekstrakurikuler (JK), perolehan prestasi di luar kelas (PNA), nilai sikap sosial (NS), dan catatan kehadiran siswa (AB).

3.2. Praproses Data

Pada tahap preprocessing, dilakukan serangkaian prosedur untuk membersihkan dan mempersiapkan data sebelum dianalisis lebih lanjut. Data yang telah dipilih akan melalui proses validasi serta penyesuaian format agar tetap konsisten. Selama proses ini, entri data yang tidak memiliki nilai atau informasi yang diperlukan akan dihapus dari dataset untuk memastikan kualitas dan integritas data yang digunakan. Data hasil praproses disajikan pada Tabel 2.

Tabel 2. Praproses Data

ID_Siswa	RNA	JE	PA	PNA	NS	AB
S001	82.48	0	1	0	Baik Sekali	7
S002	79.31	3	0	2	Baik	6
S003	83.24	2	1	1	Baik Sekali	6
S004	87.62	0	3	0	Baik Sekali	3
S005	78.83	3	0	2	Baik	18
…	…	…	…	…	…	…
S069	80.46	1	0	0	Baik	14
S070	74.06	2	0	5	Baik	13

Seluruh data siswa kelas 10 SMAS Sultan Agung Puger telah memenuhi syarat kelengkapan, sehingga proses dapat dilanjutkan ke tahap berikutnya dalam alur data mining.

3.3. Transformasi Data

Dalam penerapan metode *K-Means Clustering* pada data tersebut, langkah awal yang perlu dilakukan adalah transformasi data, terutama terhadap atribut yang bersifat nominal, seperti data nilai sikap. Data nominal tersebut harus dikonversi terlebih dahulu ke dalam bentuk numerik agar dapat diproses oleh algoritma. Pada tahap ini, setiap kategori nilai sikap diberi kode numerik yang mewakili masing-masing kategori. Sebagai contoh, nilai sikap "Baik" dapat dikodekan dengan angka 1 dan "Baik Sekali" dengan angka 2. Rincian pengkodean tersebut disajikan pada Tabel 3.

Tabel 3. Hasil Transformasi Data

ID_Siswa	RNA	JE	PA	PNA	NS	AB
S001	82.48	0	1	0	2	7
S002	79.31	3	0	2	1	6
S003	83.24	2	1	1	2	6
S004	87.62	0	3	0	2	3
S005	78.83	3	0	2	1	18
…	…	…	…	…	…	…
S069	80.46	1	0	0	1	14
S070	74.06	2	0	5	1	13

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.4. K-Means Clustering

Setelah data berada dalam format yang sesuai, dilakukan pengelompokan menggunakan algoritma *K-Means Clustering*, yang berfungsi untuk mengidentifikasi pola dan membentuk kelompok siswa berdasarkan kemiripan karakteristik. Dalam pelaksanaan proses pengolahan data ini, penulis memanfaatkan platform *Google Colab* yang berbasis cloud karena fleksibilitas dan kemampuannya dalam menjalankan kode secara interaktif. Bahasa pemrograman *Python* digunakan sebagai alat utama, mengingat kemampuannya yang luas dalam manipulasi data, penerapan algoritma *machine learning*, serta penyajian visualisasi hasil analisis secara efisien.

```
[ ]  import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.cluster import KMeans
     from sklearn.metrics import davies_bouldin_score
     from sklearn.preprocessing import StandardScaler
     from sklearn.decomposition import PCA
```

Gambar 1. Kode Pustaka *Python*

Pada penelitian ini, digunakan beberapa pustaka pendukung berbasis *Python* untuk melakukan analisis dan visualisasi data. Diantarnya seperti yang terlihat pada Gambar 1.

a. *import pandas as pd*
   Pustaka *pandas*, yang disingkat sebagai *pd*, digunakan untuk keperluan pengelolaan dan manipulasi data dalam format tabel *(DataFrame)*, seperti membaca file berformat *CSV*, mengatur struktur kolom, melakukan penyaringan data, dan berbagai operasi lainnya.
b. *import numpy as np*
   Pustaka *numpy*, yang diimpor dengan penamaan *np*, digunakan untuk mendukung perhitungan matematis serta pengolahan *array* multidimensi secara efisien.
c. *import matplotlib.pyplot as plt*
   Modul *pyplot* dari pustaka *matplotlib* diimpor dan disingkat dengan nama *plt*. Modul ini digunakan untuk keperluan visualisasi data, seperti pembuatan grafik, diagram batang *(bar chart)*, diagram pencar *(scatter plot)*, dan jenis visualisasi lainnya.
d. *import seaborn as sns*
   Pustaka *seaborn*, yang diimpor dengan penamaan *sns*, digunakan untuk menghasilkan visualisasi statistik yang menarik dan informatif. Pustaka ini sering dimanfaatkan dalam pembuatan grafik seperti *heatmap, pairplot,* dan *boxplot* karena kemampuannya menyajikan data secara lebih estetis dan mudah dipahami.
e. *from sklearn.cluster import K-Means*
   Kelas *K-Means* dari pustaka *scikit-learn* digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan tingkat kesamaan antar fitur.
f. *from sklearn.metrics import davies_bouldin_score*
   Metrik *davies_bouldin_score* diimpor untuk mengevaluasi kualitas hasil klasterisasi. Nilai skor yang lebih rendah menunjukkan bahwa pemisahan antar klaster lebih optimal dan struktur klaster yang terbentuk lebih baik.
g. *from sklearn.preprocessing import StandardScaler*
   Mengimpor *StandardScaler* dari modul *sklearn.preprocessing* untuk menstandarisasi fitur data *(mean = 0, std = 1),* sehingga setiap fitur memiliki skala yang seragam sebelum proses clustering dilakukan.
h. *from sklearn.decomposition import PCA*
   Mengimpor PCA *(Principal Component Analysis)* dari *scikit-learn* untuk mengurangi jumlah dimensi data secara statistik, sambil tetap mempertahankan variansi. Metode ini sering diterapkan untuk visualisasi, seperti mengubah data dari dimensi 4D menjadi 2D agar dapat dipetakan.

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

     # Misalnya nama file: data_siswa.csv
     df = pd.read_csv('data_siswa.csv')

     # Tampilkan beberapa baris awal
     df.head()
```

Gambar 2. Kode Membaca File

Kode program pada Gambar 2 merupakan proses mengimpor pustaka *files* dari *Google Colab* untuk memungkinkan mengunggah *file* dari komputer. Fungsi *files.upload()* akan memunculkan dialog untuk

memilih file yang ingin diunggah. Setelah file diunggah, program membaca file *CSV* (misalnya "data_siswa.csv") menggunakan fungsi *pd.read_csv()* dari pustaka *pandas* dan menyimpannya dalam variabel *df* sebagai *DataFrame*. Selanjutnya, menggunakan *df.head(),* program menampilkan beberapa baris pertama dari data yang telah dibaca, sehingga dapat melihat isi awal dari dataset tersebut.

```
[ ]  # Hilangkan kolom non-numerik jika ada (misalnya nama)
     df_numeric = df.select_dtypes(include=[np.number])

     # Normalisasi fitur
     scaler = StandardScaler()
     X_scaled = scaler.fit_transform(df_numeric)
```

Gambar 3. Kode Menghilangkan Data Non-Numerik dan Melakukan Normalisasi Data

Gambar 3 memberikan informasi proses menghilangkan kolom yang tidak berisi angka, seperti kolom nama atau kategori, dengan menggunakan kode *df.select_dtypes(include=[np.number]).* Hal ini memastikan bahwa hanya data yang berupa angka yang diproses. Setelah itu, melakukan normalisasi pada data dengan menggunakan *StandardScaler*. Proses ini menghitung rata-rata dan *standar deviasi* dari data, lalu mengubah data sehingga memiliki rata-rata 0 dan *standar deviasi* 1. Tujuannya adalah agar data memiliki skala yang seragam dan siap digunakan dalam analisis lebih lanjut atau model pembelajaran mesin.

```
[ ]  dbi_scores = []
     k_values = range(2, 8)

     for k in k_values:
         kmeans = KMeans(n_clusters=k, random_state=42, n_init='auto')
         labels = kmeans.fit_predict(X_scaled)
         dbi = davies_bouldin_score(X_scaled, labels)
         dbi_scores.append(dbi)

     # Visualisasi DBI
     plt.figure(figsize=(8, 5))
     plt.plot(k_values, dbi_scores, marker='o')
     plt.title('Davies-Bouldin Index vs Jumlah Klaster')
     plt.xlabel('Jumlah Klaster (k)')
     plt.ylabel('DBI Score')
     plt.grid(True)
     plt.show()
```

Gambar 4. Kode Pemrosesan DBI *Score* dan Visualisasi Data

Kode pada Gambar 4 digunakan untuk mengevaluasi kualitas klaster yang dihasilkan oleh algoritma *K-Means* dengan berbagai jumlah klaster (k) dan menghitung skor DBI untuk setiap nilai k. Proses dimulai dengan mendefinisikan sebuah daftar kosong *dbi_scores* untuk menyimpan nilai DBI yang dihitung. Kemudian, dilakukan iterasi untuk setiap nilai k dalam rentang 2 hingga 7 *(k_values = range(2, 8)).* Untuk setiap nilai k, model *K-Means* diinisialisasi dengan jumlah klaster k, kemudian dilakukan proses klasterisasi pada data yang sudah dinormalisasi *(X_scaled).* Setelah itu, label hasil klasterisasi dihitung, dan skor DBI dihitung dengan menggunakan fungsi *davies_bouldin_score*, yang mengukur sejauh mana klaster-klaster yang terbentuk terpisah dengan baik. Nilai DBI tersebut disimpan dalam daftar *dbi_scores*. Terakhir, menampilkan visualisasi yang menggambarkan hubungan antara jumlah klaster dan nilai DBI, dengan grafik yang menunjukkan seberapa baik pembagian klaster berdasarkan nilai DBI yang dihasilkan untuk setiap k.

```
[ ]  # Menampilkan nilai DBI untuk tiap jumlah klaster
     print("Rincian Nilai Davies-Bouldin Index (DBI) untuk Setiap Jumlah Klaster:")
     for k, dbi in zip(k_values, dbi_scores):
         print(f"Jumlah Klaster = {k} | DBI = {dbi:.4f}")

     # Menentukan jumlah klaster optimal
     optimal_k = k_values[np.argmin(dbi_scores)]
     print(f"\nJumlah klaster optimal berdasarkan nilai DBI terendah: {optimal_k}")

     # Terapkan K-Means dengan jumlah klaster optimal
     kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init='auto')
     clusters = kmeans.fit_predict(X_scaled)

     # Tambahkan hasil klaster ke DataFrame
     df['Klaster'] = clusters
```

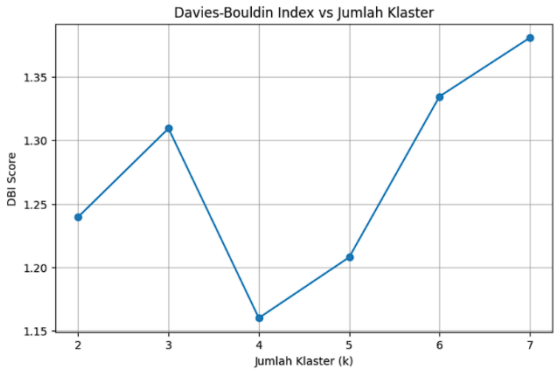
Gambar 5. Kode Menampilkan Nilai DBI dan Jumlah Klaster Optimal

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Gambar 5 menampilkan proses rincian nilai DBI untuk setiap jumlah klaster yang diuji. Nilai DBI dihitung untuk setiap jumlah klaster dalam *k_values*, dan setiap nilai DBI ditampilkan dalam format yang rapi dengan dua kolom: jumlah klaster dan nilai DBI. Kemudian, kode menentukan jumlah klaster optimal berdasarkan nilai DBI terendah. Nilai DBI yang lebih rendah menunjukkan klaster yang lebih baik, dan *optimal_k* diambil dari jumlah klaster yang memiliki nilai DBI terendah. Setelah itu, algoritma *K-Means* diterapkan dengan jumlah klaster optimal yang sudah ditemukan. *K-Means* ini digunakan untuk mengelompokkan data *(X_scaled)* ke dalam klaster-klaster yang ditentukan. Hasil klaster yang diperoleh disimpan dalam kolom baru bernama "Klaster" pada *DataFrame df*, sehingga setiap baris data akan terlabeli dengan klaster yang sesuai.

3.5. Analisis Hasil Klaster

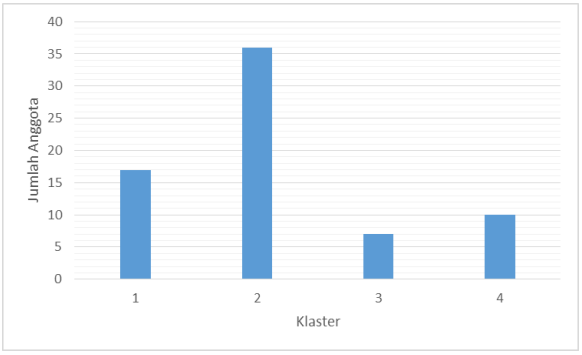
Setelah data dimasukkan dan konfigurasi klaster dilakukan menggunakan algoritma *K-Means*, pengujian dengan *Davies-Bouldin Index* (DBI) dilakukan sebanyak enam kali iterasi. Dari hasil pengujian tersebut, diperoleh empat klaster sebagai nilai klaster optimal. Visualisasi hasil pengujian ditampilkan pada Gambar 6.



Gambar 6. Visualisasi Hasil Pengujian DBI

Hasil evaluasi menunjukkan bahwa ketika jumlah klaster adalah dua, nilai DBI yang dihasilkan sebesar 1.2396, sedangkan untuk tiga klaster nilainya meningkat menjadi 1.3093. Pada jumlah empat klaster, diperoleh nilai DBI paling rendah yaitu 1.1601. Selanjutnya, lima klaster menghasilkan nilai DBI sebesar 1.2079, enam klaster sebesar 1.3345, dan tujuh klaster mencatat nilai tertinggi yaitu 1.3807. Berdasarkan hasil tersebut, jumlah klaster sebanyak empat dipilih sebagai jumlah yang paling optimal karena memiliki nilai DBI terendah. Nilai ini menunjukkan bahwa klaster yang terbentuk memiliki tingkat kekompakan yang baik serta pemisahan yang jelas antar klaster.

Berdasarkan hasil pengelompokan menggunakan algoritma *K-Means* dengan jumlah klaster optimal sebanyak empat, distribusi jumlah anggota pada masing-masing klaster menunjukkan variasi yang cukup signifikan. Visualisasi distribusi jumlah anggota klaster tersaji pada Gambar 7.



Gambar 7. Visualisasi Distribusi Jumlah Anggota Klaster

Klaster 2 memiliki jumlah anggota terbanyak, yaitu sebanyak 36 siswa. Selanjutnya, klaster 1 terdiri dari 17 siswa, diikuti oleh klaster 4 dengan 10 siswa, dan klaster 3 merupakan klaster dengan jumlah anggota paling sedikit, yaitu hanya 7 siswa. Perbedaan jumlah anggota ini mencerminkan keragaman karakteristik siswa berdasarkan data akademik dan non-akademik yang dianalisis.

Hasil dari proses klasterisasi dengan algoritma *K-Means* menghasilkan nilai rata-rata *(mean)* dari setiap fitur dalam masing-masing klaster. Nilai rata-rata ini merepresentasikan karakteristik umum dari

atribut-atribut yang digunakan dalam proses pengelompokan, seperti rata-rata nilai akademik, tingkat keikutsertaan dalam kegiatan ekstrakurikuler, prestasi non-akademik, sikap sosial, dan jumlah absensi. Setiap klaster memiliki ciri khas tersendiri yang tercermin dari nilai rata-rata tiap fitur, sehingga dapat digunakan untuk memahami perbedaan atau kekhususan antar kelompok siswa.

Tahapan ini bertujuan untuk mengenali profil dari masing-masing klaster, misalnya untuk mengidentifikasi klaster mana yang menunjukkan performa akademik paling tinggi atau tingkat partisipasi ekstrakurikuler yang paling aktif. Selain itu, langkah ini juga mempermudah proses perbandingan antar klaster guna mengetahui keunggulan dan kekuatan relatif dari masing-masing kelompok. Hasil analisis tersebut dapat dimanfaatkan dalam proses pengambilan keputusan, seperti menentukan siswa yang termasuk dalam kategori unggul atau menyusun strategi pembelajaran dan pembinaan yang sesuai dengan kebutuhan dan potensi masing-masing klaster. Detail rata-rata setiap fitur untuk masing-masing klaster ditampilkan dalam Tabel 4.

Tabel 4. Detail Rata-Rata Setiap Fitur Klaster

Klaster	RNA	JE	PA	PNA	NS	AB
1	81.8170	2.5294	1.0000	2.1764	1.9411	11.0000
2	76.9336	1.9166	0.1111	1.2222	1.0833	16.8333
3	86.6457	0.8571	4.2857	1.0000	2.0000	4.7142
4	82.0540	0.3000	1.2000	0.3000	2.0000	7.7000

Berdasarkan hasil klasterisasi yang ditampilkan, terlihat bahwa masing-masing klaster memiliki profil yang unik, mencerminkan kondisi siswa dari sisi akademik dan non-akademik. Klaster 1, dengan rata-rata nilai akademik yang cukup tinggi (81.8170) dan keikutsertaan ekstrakurikuler yang aktif (2.5294 kegiatan), menunjukkan adanya keseimbangan antara pencapaian akademik dan pengembangan diri di luar kelas. Nilai prestasi non-akademik dan sikap sosial yang baik mengindikasikan bahwa siswa dalam klaster ini memiliki motivasi belajar yang tinggi sekaligus mampu berinteraksi sosial dengan baik. Namun, tingkat absensi yang cukup tinggi (11.0000) bisa menjadi indikator bahwa meskipun mereka aktif dan berprestasi, ada kemungkinan kelelahan atau beban aktivitas yang berlebihan sehingga berdampak pada kehadiran. Data siswa yang termasuk dalam klaster 1 dapat dilihat secara rinci pada Tabel 5.

Tabel 5. Data Anggota Klaster 1

No.	ID_Siswa	RNA	JE	PA	PNA	NS	AB
1	S003	83.24	2	1	1	2	6
2	S017	81.57	3	1	2	2	18
3	S021	78.87	3	0	2	2	15
4	S027	81.88	2	1	2	2	2
5	S032	74.71	1	0	2	2	9
…	…	…	…	…	…	…	…
16	S057	84.06	2	2	2	2	12
17	S067	84.11	1	1	5	2	8

Klaster 2 menampilkan karakteristik siswa dengan nilai akademik dan non-akademik yang rendah, serta tingkat absensi yang paling tinggi (16.8333). Keterlibatan ekstrakurikuler yang sedang (1.9166) tidak cukup menstimulasi perkembangan siswa dalam aspek lainnya. Secara ilmiah, hal ini bisa dikaitkan dengan kurangnya keterlibatan siswa dalam kegiatan bermakna, yang dalam teori motivasi belajar, dapat menghambat pencapaian akademik dan sosial. Kelompok ini membutuhkan perhatian khusus, seperti bimbingan belajar dan program peningkatan motivasi. Informasi detail mengenai siswa yang termasuk ke dalam klaster 2 disajikan pada Tabel 6.

Tabel 6. Data Anggota Klaster 2

No.	ID_Siswa	RNA	JE	PA	PNA	NS	AB
1	S002	79.31	3	0	2	1	6
2	S005	78.83	3	0	2	1	18
3	S006	78.83	3	0	1	1	13
4	S008	77.65	2	0	3	1	14

p-ISSN: 2301-5373
e-ISSN: 2654-5101

5	S010	77.68	3	0	3	1	13
…	…	…	…	…	…	…	…
35	S069	80.46	1	0	0	1	14
36	S070	74.06	2	0	5	1	13

Klaster 3 memiliki siswa dengan nilai akademik tertinggi (86.6457) dan prestasi akademik paling menonjol (4.2857), namun tingkat keikutsertaan dalam kegiatan ekstrakurikuler rendah (0.8571) dan sikap sosial yang tidak terlalu menonjol. Hal ini menunjukkan tipe siswa yang cenderung fokus pada studi, namun kurang dalam pengembangan aspek sosial dan emosional. Dalam pendekatan pendidikan holistik, kondisi ini bisa menjadi catatan penting untuk mendorong siswa agar lebih aktif dalam interaksi sosial dan kegiatan pengembangan diri lainnya, agar tidak hanya unggul secara kognitif, tetapi juga afektif. Rincian data siswa yang menjadi bagian dari klaster 3 dapat dilihat pada Tabel 7.

Tabel 7. Data Anggota Klaster 3

No.	ID_Siswa	RNA	JE	PA	PNA	NS	AB
1	S004	87.62	0	3	0	2	3
2	S020	87.33	0	5	0	2	3
3	S030	84.26	2	5	1	2	3
4	S058	86.78	1	5	1	2	4
5	S060	85.02	1	4	2	2	8
6	S064	87.69	1	5	1	2	5
7	S066	87.82	1	3	2	2	7

Klaster 4 merupakan klaster dengan karakteristik menarik, yaitu memiliki nilai akademik yang tinggi (82.0540), sikap sosial yang paling tinggi (2.0000), dan absensi paling rendah (7.7000). Meskipun partisipasi ekstrakurikulernya rendah (0.300), siswa dalam klaster ini menunjukkan kedisiplinan dan kemampuan sosial yang sangat baik. Secara logis, kondisi ini bisa mencerminkan tipe siswa yang cenderung pendiam atau introver, tetapi memiliki kesadaran tinggi terhadap tanggung jawab pribadi. Dalam konteks pembinaan siswa unggulan, kelompok ini dapat difasilitasi untuk mengeksplorasi potensi melalui kegiatan yang sesuai dengan minat pribadi mereka. Data siswa yang masuk dalam klaster 4 ditampilkan secara rinci pada Tabel 8.

Tabel 8. Data Anggota Klaster 4

No.	ID_Siswa	RNA	JE	PA	PNA	NS	AB
1	S001	82.48	0	1	0	2	7
2	S007	83.84	0	1	0	2	14
3	S009	82.71	0	2	0	2	2
4	S012	81.21	1	1	1	2	10
5	S042	85.29	1	2	1	2	3
6	S050	84.66	0	2	0	2	8
7	S053	84.88	0	3	0	2	11
8	S056	74.02	0	0	0	2	7
9	S059	79.64	0	0	0	2	12
10	S063	81.81	1	0	1	2	3

Secara keseluruhan, hasil analisis ini memberikan landasan ilmiah untuk menyusun strategi pembinaan dan intervensi pendidikan yang tepat sasaran. Dengan memahami karakteristik masing-masing klaster, pihak sekolah dapat mengambil kebijakan yang berbasis data untuk meningkatkan prestasi siswa secara merata, baik dari aspek akademik maupun non-akademik.

4. Conclusion

Penelitian ini berhasil mengelompokkan siswa kelas 10 SMAS Sultan Agung Puger ke dalam empat klaster menggunakan algoritma *K-Means Clustering*. Pemilihan jumlah klaster didasarkan pada evaluasi nilai *Davies-Bouldin Index* (DBI), di mana nilai terendah diperoleh pada jumlah klaster 4 (DBI = 1.1601), yang menunjukkan kualitas pemisahan klaster yang baik. Hasil klasterisasi menunjukkan

bahwa masing-masing kelompok memiliki ciri khas yang berbeda, seperti klaster dengan nilai akademik tinggi namun minim keterlibatan ekstrakurikuler, serta klaster dengan keunggulan menyeluruh pada aspek akademik, non-akademik, sikap sosial, dan kehadiran. Informasi ini dapat dimanfaatkan oleh pihak sekolah untuk menyusun strategi pembinaan yang lebih tepat sasaran, baik dalam pengembangan potensi siswa unggul maupun pemberian intervensi kepada siswa yang membutuhkan perhatian khusus.

References

[1] A. Asmana, A. Y. Wijaya, and M. Martanto, “Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 552–559, 2022, doi: 10.36040/jati.v6i2.5236.

[2] B. Hasmaulina, M. Maryaningsih, and S. Sapri, “Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma,” *JUKOMIKA (Jurnal Ilmu Komput. dan Inform.*, vol. 4, no. 2, pp. 57–71, 2022, doi: 10.54650/jukomika.v4i2.368.

[3] P. S. Hasugia and J. R. Sagala, “Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 3, pp. 262–268, 2022, [Online]. Available: https://djournals.com/klik

[4] N. Bili, R. T. Abieno, and A. A. Pekuwali, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Performa Siswa Pada Pembelajaran Bahasa Indonesia (Studi Kasus : SD Inpres Waingapu 3),” in *SATI: Sustainable Agricultural Technology Innovation*, 2024, pp. 523–537.

[5] J. Irawan, A. N. Ryzkyani, R. A. Fauzan, A. J. Ramadhani, and M. Mutiah, “Analisis Hubungan Asosiasi Antara Indeks Prestasi Kumulatif (IPK) Dengan Kegiatan Organisasi Mahasiswa di Universitas Sultan Ageng Tirtayasa,” *J. Teknol. Pangan dan Ilmu Pertan.*, vol. 1, no. 4, pp. 14–23, 2023, [Online]. Available: https://doi.org/10.59581/jtpip-widyakarya.v1i4.1473

[6] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

[7] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, pp. 25–36, 2021, doi: 10.33365/jtk.v15i2.1162.

[8] M. A. Al Fauzie and A. J. Putra, “Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 642–648, 2023, doi: 10.30865/klik.v4i1.1116.

[9] S. Anwar, T. Suprapti, G. Dwilestari, and I. Ali, “Pengelompokkan Hasil Belajar Siswa dengan Metode Clustering K-Means,” *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.

[10] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, “Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means,” *J. Inf. dan Teknol.*, vol. 3, pp. 103–108, 2021, doi: 10.37034/jidt.v3i3.130.

[11] R. A. Farissa, R. Mayasari, and Y. Umaidah, “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.

[12] F. Rizki Ani, R. Kurniawan, and T. Suprapti, “Penerapan Algoritma K-Means Clustering Untuk Perencanaan Persediaan Stok Sepatu Di Toko Diomclothing,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3699–3707, 2024, doi: 10.36040/jati.v7i6.8233.

[13] F. Fathurrahman, S. Harini, and R. Kusumawati, “Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (Dbi),” *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023, doi: 10.36040/mnemonic.v6i2.6642.

[14] M. Hermansyah, R. A. Hamdan, F. Sidik, and A. Wibowo, “Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means,” *J. Ilmu Komput.*, vol. 13, no. 2, p. 8, 2020, doi: 10.24843/jik.2020.v13.i02.p06.

[15] L. Fimawahib and E. Rouza, “Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian,” *INOVTEK Polbeng - Seri Inform.*, vol. 6, no. 2, p. 234, 2021, doi: 10.35314/isi.v6i2.2096.

This page is intentionally left blank.