



Klasterisasi Siswa Berdasarkan Profil Akademik dan Karakteristik Belajar Menggunakan Algoritma K-Means untuk Mendukung Pembelajaran

Attaya Faiharani*, Baenil Huda, Fitria Nurapriani, April Lia Hananto

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Karawang
Jl. H.S. Ronggowaluyo Simabaya, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia

Email: ^{1,*}si22.attayafaiharani@mhs.ubpkarawang.ac.id, ²baenil88@ubpkarawang.ac.id,

³fitria.nurapriani@ubpkarawang.ac.id, ⁴aprilialia@ubpkarawang.ac.id

Email Penulis Korespondensi: si22.attayafaiharani@mhs.ubpkarawang.ac.id

Submitted: 31/03/2026; Accepted: 24/04/2026; Published: 26/04/2026

Abstrak—Pengelompokan siswa berdasarkan karakteristik akademik dan non-akademik penting dilakukan untuk mendukung penyusunan strategi pembinaan yang lebih tepat sasaran di sekolah. Permasalahan yang dihadapi adalah belum tersedianya pemetaan siswa berbasis data secara objektif sehingga program pengembangan masih bersifat umum. Penelitian ini bertujuan mengelompokkan siswa menggunakan algoritma K-Means Clustering berdasarkan profil akademik dan variabel pendukung lainnya, serta mengevaluasi kualitas cluster menggunakan silhouette coefficient. Tahapan penelitian meliputi preprocessing data, penentuan jumlah cluster optimal, proses clustering, dan evaluasi hasil pengelompokan. Hasil penelitian menunjukkan bahwa konfigurasi empat cluster digunakan sebagai hasil akhir dengan nilai silhouette score sebesar 0,1093, dengan distribusi anggota cluster terdiri atas 12, 4, 2, dan 2 siswa. Visualisasi menggunakan principal component analysis menunjukkan sebagian besar cluster terpisah dengan cukup jelas. Penelitian ini memberikan kontribusi berupa model pengelompokan siswa berbasis data yang dapat digunakan sebagai dasar rekomendasi pengembangan potensi para siswa sesuai karakteristik masing-masing kelompok.

Kata Kunci: K-Means clustering; data mining; silhouette coefficient; pengelompokan siswa; evaluasi cluster.

Abstract—Grouping students based on academic and non-academic characteristics is important to support the development of more targeted educational guidance strategies in schools. The main problem addressed in this study is the absence of objective data-based student mapping, which causes development programs to remain general and less targeted. This study aims to classify students using the K-Means clustering algorithm based on academic profiles and other supporting variables, and to evaluate cluster quality using the silhouette coefficient method. The research stages include data preprocessing, determining the optimal number of clusters, clustering using K-Means, and evaluating the clustering result. The results showed that four clusters were selected as the final configuration with a silhouette score of 0,1093, with cluster membership distributed into 12, 4, 2, and 2 students. Visualization using principal component analysis shows that most clusters are sufficiently well separated. This study contributes a data-driven student grouping model that can be used as a basis for recommending student potential development according to the characteristics of each group.

Keywords: K-Means Clustering; Data Mining; Silhouette Coefficient; Student Clustering; Cluster Evaluation.

1. PENDAHULUAN

Perkembangan sistem informasi dalam dunia pendidikan mendorong sekolah untuk mengelola data siswa dalam jumlah yang semakin besar dan beragam [1]. Data siswa tidak lagi terbatas pada nilai akademik semata, tetapi juga mencakup data non-akademik dan karakteristik belajar, seperti keterlibatan dalam kegiatan ekstrakurikuler, kemampuan kognitif, gaya belajar, minat dominan, serta capaian prestasi. Keseluruhan data tersebut membentuk profil siswa yang bersifat multidimensi dan berpotensi memberikan gambaran yang lebih komprehensif mengenai karakteristik belajar dan potensi akademik setiap siswa. Dalam konteks pendidikan menengah, ekstrakurikuler juga memiliki peran penting sebagai sarana pengembangan minat, bakat, kedisiplinan, dan karakteristik siswa yang secara tidak langsung dapat memengaruhi capaian akademik [2].

Namun demikian, pada praktiknya pemanfaatan data siswa di tingkat sekolah masih cenderung bersifat administratif dan deskriptif [3]. Data akademik, data aktivitas ekstrakurikuler, maupun data pendukung lainnya umumnya digunakan secara terpisah dan belum dianalisis secara terintegrasi untuk mengidentifikasi pola tertentu yang mendukung evaluasi pembelajaran maupun pengambilan keputusan akademik berbasis data [4]. Kondisi tersebut menyebabkan informasi penting yang sebenarnya terkandung dalam data siswa belum dimanfaatkan secara optimal untuk menghasilkan strategi pembinaan yang lebih tepat sasaran [5] [6].

Seiring dengan meningkatnya kompleksitas data siswa, diperlukan suatu pendekatan analitis yang mampu mengolah data multidimensi secara objektif [7]. Salah satu pendekatan yang banyak digunakan dalam pengolahan data pendidikan adalah data mining, karena mampu menemukan pola tersembunyi yang sulit diidentifikasi melalui analisis manual [8]. Dalam data mining, teknik clustering menjadi salah satu metode yang efektif untuk proses pengelompokan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan suatu label awal. Salah satu algoritma yang paling banyak digunakan adalah K-Means clustering karena memiliki proses komputasi yang relatif sederhana, efisien, dan mampu mengelompokkan data dalam jumlah besar berdasarkan centroid terdekat [9].

Berbagai penelitian sebelumnya menunjukkan bahwa algoritma K-Means telah banyak diterapkan dalam bidang pendidikan untuk berbagai tujuan analisis data [10]. Algoritma ini digunakan untuk menentukan para siswa

berprestasi, mengelompokan calon siswa baru, menganalisis hasil belajar, segmentasi kelas unggulan, hingga mendukung pengambilan keputusan akademik [11][12]. Penelitian sebelumnya juga menunjukkan bahwa K-Means mampu menghasilkan pola pengelompokan yang membantu sekolah memahami persebaran karakteristik siswa secara lebih sistematis [13]. Selain itu, beberapa penelitian telah menggabungkan K-Means dengan metode validasi elbow dan silhouette coefficient untuk menentukan jumlah cluster optimal dan mengevaluasi kualitas hasil pengelompokan. Penggunaan metode validasi tersebut terbukti dapat meningkatkan reliabilitas hasil cluster dan mengurangi subjektivitas dalam menentukan jumlah cluster [14].

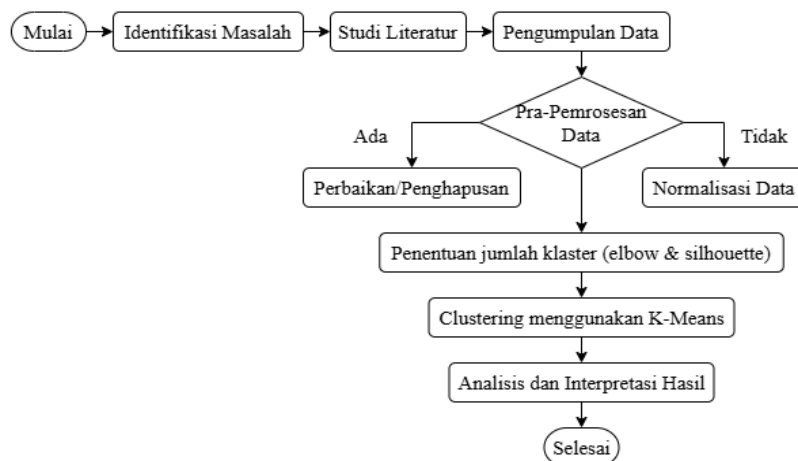
Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada satu jenis variabel atau aspek tertentu, terutama nilai akademik atau hasil belajar, sehingga belum sepenuhnya menggambarkan profil siswa secara menyeluruh [15]. Beberapa studi memang telah menggunakan variabel tambahan, namun masih terbatas pada kombinasi data tertentu dan belum mengintegrasikan berbagai dimensi profil siswa secara simultan dalam satu kerangka analisis clustering yang tervalidasi. Padahal dalam konteks pendidikan, karakteristik siswa tidak hanya ditentukan oleh capaian akademik, tetapi juga dipengaruhi oleh minat belajar, aktivitas ekstrakurikuler, kemampuan kognitif, dan prestasi non-akademik yang saling berkaitan dalam bentuk potensi siswa [16].

Keterbatasan tersebut menunjukkan adanya kebutuhan untuk melakukan proses pengelompokan siswa berdasarkan profil multidimensi yang lebih komprehensif. Integrasi berbagai variabel dalam satu proses clustering diharapkan dapat menghasilkan pola kelompok siswa yang lebih representatif dibandingkan pengelompokan berbasis satu dimensi data saja. Selain itu, penggunaan validasi cluster tetap diperlukan agar hasil pengelompokan tidak hanya terbentuk secara matematis, tetapi juga memiliki kualitas yang dapat dipertanggungjawabkan secara analitis. Berdasarkan kondisi tersebut, penelitian ini menerapkan algoritma K-Means untuk mengelompokan siswa berdasarkan profil akademik dan karakteristik belajar yang mencakup nilai akademik, aktivitas ekstrakurikuler, kemampuan kognitif, gaya belajar, minat dominan, dan prestasi siswa [17]. Untuk memastikan kualitas hasil clustering, penelitian ini menggunakan silhouette coefficient dalam evaluasi kualitas cluster yang terbentuk, serta Principal Component Analysis untuk membantu visualisasi distribusi cluster. Pendekatan ini dipilih agar proses pengelompokan tidak hanya menghasilkan jumlah cluster optimal, tetapi juga memberikan gambaran hubungan antar kelompok secara visual [18].

Kontribusi penelitian terletak pada penyusunan model pengelompokan siswa berbasis data multidimensi yang dapat memberikan gambaran lebih utuh mengenai karakteristik siswa dalam konteks pendidikan menengah. Hasil penelitian diharapkan dapat digunakan sebagai dasar pendukung evaluasi akademik, penyusunan strategi pembinaan siswa, serta pengambilan keputusan pendidikan yang lebih terarah berdasarkan karakteristik masing-masing kelompok siswa.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Flowchart Alur Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan yang disusun secara sistematis untuk menghasilkan proses analisis data yang terstruktur seperti yang ditunjukkan pada **Gambar 1** [19]. Tahapan penelitian dimulai dari identifikasi permasalahan terkait belum optimalnya pemanfaatan data siswa dalam proses evaluasi akademik di sekolah. Setelah permasalahan ditetapkan, tahap berikutnya adalah pengumpulan data siswa yang meliputi variabel akademik dan juga non-akademik yang relevan dengan karakteristik belajar.

Data yang diperoleh kemudian melalui tahap pra-pemrosesan untuk memastikan kualitas data sebelum dianalisis. Tahap ini mencakup pemeriksaan konsistensi data, transformasi variabel, serta normalisasi agar seluruh atribut memiliki rentang nilai yang sebanding. Setelah data siap digunakan, dilakukan proses clustering menggunakan algoritma K-Means untuk membentuk kelompok siswa berdasarkan tingkat kemiripan karakteristik.



Selanjutnya jumlah cluster optimal ditentukan dengan menggunakan metode elbow, kemudian kualitas hasil clustering dievaluasi menggunakan silhouette coefficient. Tahap akhir penelitian adalah analisis karakteristik setiap cluster dan interpretasi hasil untuk memperoleh gambaran profil kelompok siswa secara lebih komprehensif.

2.2 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan eksploratif berbasis data mining. Pendekatan kuantitatif digunakan karena analisis dilakukan terhadap data numerik yang berasal dari berbagai atribut siswa untuk menemukan pola pengelompokan secara objektif. Pendekatan eksploratif dipilih karena penelitian bertujuan mengidentifikasi struktur kelompok siswa tanpa menetapkan kategori awal sebelumnya. Dalam konteks ini, algoritma clustering digunakan sebagai teknik unsupervised learning untuk menemukan pola kesamaan antar data berdasarkan atribut multidimensi.

2.3 Objek dan Lokasi Penelitian

Objek penelitian ini adalah data siswa pada salah satu SMA Negeri di Kabupaten Karawang, yang mencakup data profil akademik dan karakteristik belajar siswa. Data yang digunakan berasal dari arsip sekolah dan data pendukung yang telah disesuaikan dalam bentuk dataset numerik. Seluruh data yang digunakan dalam penelitian ini telah melalui proses seleksi dan penyesuaian agar sesuai dengan kebutuhan analisis menggunakan metode K-Means Clustering.

2.4 Dataset dan Variabel Penelitian

2.4.1 Dataset Penelitian

Dataset yang digunakan terdiri dari 20 siswa yang dipilih sebagai sampel penelitian eksploratif. Jumlah data yang terbatas ini digunakan untuk mengidentifikasi pola awal karakteristik siswa pada data multidimensi dan tidak dimaksudkan untuk generalisasi populasi secara luas. Setiap data siswa direpresentasikan dalam bentuk atribut numerik yang telah disusun dalam satu dataset terintegrasi sehingga dapat diproses dalam algoritma clustering.

2.4.2 Variabel Penelitian

Variabel penelitian terdiri dari atribut akademik dan non-akademik yang dipilih berdasarkan relevansinya terhadap profil siswa. Variabel akademik meliputi nilai mata pelajaran inti, sedangkan non-akademik mencakup aktivitas ekstrakurikuler, kemampuan kognitif, gaya belajar, minat dominan, dan prestasi siswa. Pemilihan variabel dilakukan untuk membentuk representasi profil siswa secara multidimensi sehingga proses clustering tidak hanya menggambarkan performa akademik, tetapi juga karakteristik pendukung lainnya. Nilai nol pada atribut aktivitas ekstrakurikuler digunakan untuk menunjukkan ketidakterlibatan siswa dalam aktivitas tertentu dan bukan sebagai indikator performa rendah.

2.5 Pra-Pemrosesan Data

Proses awal sebelum pengolahan data bertujuan untuk memastikan bahwa data yang digunakan telah siap dan sesuai untuk proses clustering. Tahapan pra-pemrosesan meliputi pengecekan terhadap data yang kosong atau tidak valid, penyamaan skala data agar seluruh variabel memiliki rentang nilai yang sebanding, serta konversi seluruh data ke dalam format numerik sehingga dapat diproses menggunakan algoritma K-Means. Karena algoritma K-Means mudah terpengaruh perbedaan ukuran data, maka dinormalisasi menggunakan metode min-max normalization, dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Pada Persamaan (1), X' merupakan nilai hasil normalisasi, X menyatakan nilai data asli, X_{min} adalah nilai terkecil pada atribut, dan X_{max} adalah nilai terbesar pada atribut. Proses ini bertujuan mengubah rentang nilai data menjadi skala yang seragam sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses perhitungan jarak pada algoritma K-Means [20].

2.6 Penggunaan Algoritma K-Means untuk Proses Clustering

2.6.1 Penentuan Jumlah Cluster (k)

Jumlah cluster ditentukan menggunakan metode elbow dengan mengamati perubahan nilai inerti pada beberapa skenario jumlah cluster. Pemilihan jumlah cluster optimal dilakukan pada titik ketika penurunan inerti mulai melandai sehingga diperoleh jumlah cluster yang paling representatif terhadap struktur data.

2.6.2 Perhitungan Jarak

Dalam penelitian ini, perhitungan jarak dilakukan menggunakan metode euclidean distance yang ditunjukkan pada Persamaan (2).



$$D(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \tag{2}$$

Pada Persamaan (2), x_i menyatakan data ke- i , c_j merupakan centroid dari cluster ke- j , dan n adalah jumlah atribut yang digunakan dalam proses clustering. Perhitungan jarak ini digunakan untuk menentukan kedekatan antara data dengan centroid, sehingga data dapat dikelompokkan ke dalam cluster yang memiliki jarak paling minimum [21].

2.6.3 Proses Klusterisasi

Secara umum, proses klusterisasi pada algoritma K-Means dimulai dengan menentukan jumlah cluster (k) yang akan digunakan. Selanjutnya, centroid awal ditentukan secara acak sebagai pusat cluster sementara. Setiap data kemudian dihitung jaraknya terhadap seluruh centroid untuk menentukan kedekatan dengan masing-masing cluster. Berdasarkan hasil perhitungan tersebut, data kemudian dikelompokkan ke dalam cluster yang memiliki jarak terdekat. Setelah proses pengelompokan, centroid diperbarui berdasarkan rata-rata nilai anggota dalam setiap cluster. Proses ini diulangi secara iteratif hingga posisi centroid tidak mengalami perubahan atau telah mencapai kondisi stabil [22].

2.7 Validasi Cluster

Validasi cluster dilakukan menggunakan dua metode, yaitu elbow untuk melihat nilai WCSS, dan silhouette untuk mengukur kualitas pemisahan cluster. Nilai silhouette yang hampir mencapai 1 menunjukkan kualitas cluster yang lebih baik. Nilai silhouette dihitung dengan rumus:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

Pada Persamaan (3), $S(i)$ menyatakan nilai silhouette untuk data ke- i yang digunakan untuk mengukur kualitas pengelompokan. Nilai $a(i)$ merupakan jarak rata-rata antara data ke- i dengan seluruh data lain dalam cluster yang sama, yang menunjukkan tingkat kedekatan terhadap kelompoknya sendiri. Sementara itu, $b(i)$ adalah jarak rata-rata terkecil antara data ke- i dengan data pada cluster lain yang paling dekat (nearest cluster). Nilai maksimum dari $a(i)$ dan $b(i)$ digunakan sebagai faktor normalisasi sehingga nilai silhouette berada dalam rentang tertentu. Semakin mendekati nilai 1, maka kualitas cluster yang dihasilkan semakin baik karena menunjukkan pemisahan antar cluster yang lebih jelas [23].

2.8 Analisis dan Interpretasi Hasil

Hasil clustering dianalisis dengan mengidentifikasi karakteristik dominan pada setiap cluster berdasarkan nilai rata-rata atribut. Interpretasi dilakukan untuk memahami profil kelompok siswa, termasuk pada kecenderungan akademik, aktivitas non-akademik, dan minat dominan. Untuk visualisasi hasil clustering dilakukan menggunakan Principal Component Analysis (PCA) untuk membantu melihat distribusi cluster dalam ruang dua dimensi.

2.9 Alat dan Perangkat Penelitian

Penelitian menggunakan perangkat lunak berbasis bahasa pemrograman Python untuk pengolahan data dan proses clustering. Library yang digunakan meliputi pandas untuk pengelolaan data, scikit-learn untuk implementasi algoritma K-Means dan evaluasi silhouette coefficient, serta matplotlib untuk visualisasi data. Perangkat keras berupa komputer dengan spesifikasi yang mendukung pengolahan data numerik dan visualisasi hasil clustering. Secara keseluruhan, alat dan perangkat yang digunakan disajikan dalam **Tabel 1** berikut.

Tabel 1. Alat dan Perangkat Penelitian

No	Alat/Perangkat	Kegunaan
1	Laptop/PC	Media pengolahan dan analisis data
2	Google Colab	Lingkungan pemrograman berbasis Python
3	Python 3.10	Bahasa pemrograman untuk implementasi K-Means
4	Library Scikit-learn	Implementasi algoritma K-Means dan validasi
5	Microsoft Excel	Pengelolaan dan pengecekan awal data

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-Pemrosesan Data

Tahap awal penelitian ini dilakukan melalui pra-pemrosesan data terhadap dataset siswa yang digunakan sebagai objek analisis. Dataset terdiri atas 20 siswa dengan sejumlah variabel akademik dan non-akademik yang telah ditentukan pada tahap metodologi. Variabel-variabel tersebut meliputi nilai akademik, skor kemampuan kognitif, karakteristik belajar, serta indikator pendukung lainnya yang dianggap relevan dalam proses pengelompokan. Pra-

pemrosesan dilakukan untuk memastikan bahwa seluruh data berada dalam kondisi siap dianalisis menggunakan algoritma clustering. Tahapan yang dilakukan meliputi pengecekan kelengkapan data, transformasi variabel kategorikal menjadi numerik, serta normalisasi data menggunakan standarisasi agar setiap variabel memiliki skala yang sebanding. Proses normalisasi penting untuk dilakukan karena algoritma K-Means sangat sensitif terhadap perbedaan rentang nilai antarvariabel. Setelah seluruh data dinormalisasi, dataset siap digunakan dalam proses penentuan jumlah cluster dan pembentukan kelompok siswa.

3.2 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster dilakukan menggunakan silhouette coefficient untuk mengevaluasi kualitas pemisahan data pada beberapa alternatif jumlah cluster. Pengujian dilakukan pada rentang nilai $k = 2$ sampai $k = 6$, karena rentang tersebut dinilai paling relevan dengan jumlah data penelitian yang terbatas.

Tabel 2. Nilai Silhouette Coefficient pada Beberapa Alternatif Jumlah Cluster

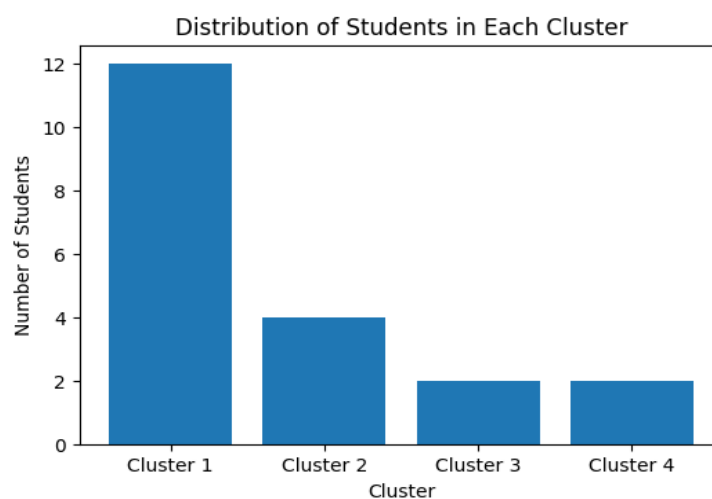
Jumlah Cluster (k)	Nilai Silhouette
2	0.450
3	0.410
4	0.378
5	0.406
6	0.416

Berdasarkan **Tabel 2**, nilai silhouette coefficient tertinggi diperoleh pada $k = 2$ dengan nilai 0.450. Nilai ini menunjukkan bahwa secara matematis konfigurasi dua cluster memberikan pemisahan antar kelompok yang paling jelas dibandingkan alternatif lainnya. Namun demikian, penggunaan dua cluster dianggap terlalu sederhana untuk menggambarkan variasi karakteristik siswa secara lebih rinci, karena tujuan penelitian ini tidak hanya memperoleh pemisahan data yang optimal secara numerik, tetapi juga menghasilkan pengelompokan yang cukup informatif untuk mendukung interpretasi karakteristik siswa berdasarkan profil akademik dan variabel pendukung lainnya.

Pada konfigurasi $k = 4$, nilai silhouette coefficient sebesar 0.378 masih menunjukkan kualitas cluster yang dapat diterima, dengan distribusi anggota yang lebih proporsional dibandingkan jumlah cluster yang lebih besar. Pemilihan empat cluster juga memungkinkan identifikasi variasi karakteristik siswa secara lebih seimbang tanpa menghasilkan cluster dengan jumlah anggota yang terlalu sedikit. Oleh karena itu, penelitian ini menetapkan empat cluster sebagai konfigurasi akhir, dengan mempertimbangkan keseimbangan antara kualitas pemisahan cluster, distribusi anggota kelompok, dan kebutuhan interpretasi hasil dalam konteks pendidikan.

3.3 Hasil Pengelompokan Siswa Menggunakan Algoritma K-Means

Setelah jumlah cluster ditetapkan sebanyak empat, proses clustering dilakukan menggunakan algoritma K-Means. Hasil pengelompokan menunjukkan bahwa seluruh data siswa berhasil terbagi ke dalam empat kelompok dengan distribusi anggota yang relatif lebih seimbang.

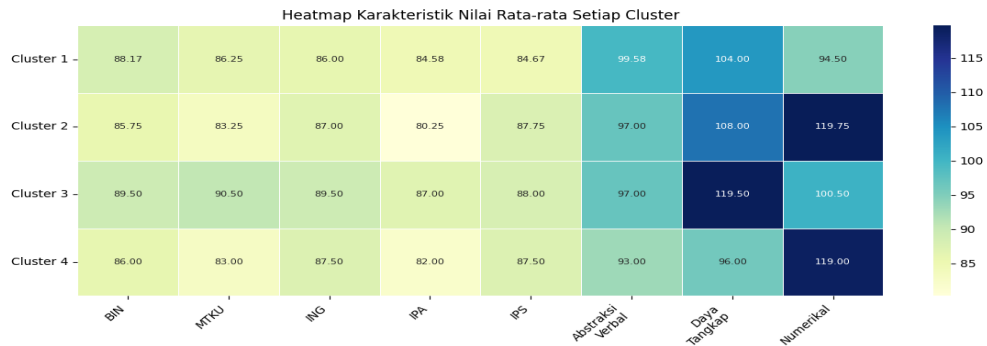


Gambar 2. Distribusi Jumlah Siswa pada Setiap Cluster

Distribusi jumlah siswa pada setiap cluster hasil pengelompokan menggunakan algoritma K-Means dengan empat cluster ditunjukkan pada **Gambar 2**. Berdasarkan hasil tersebut, distribusi anggota menunjukkan bahwa cluster 1 memiliki jumlah terbesar yaitu 12 siswa. Cluster 2 terdiri atas 4 siswa, sedangkan cluster 3 dan cluster 4 masing-masing terdiri atas 2 siswa. Distribusi ini menunjukkan bahwa sebagian besar siswa memiliki karakteristik yang relatif serupa sehingga terkonsentrasi pada satu cluster dominan, sementara sebagian lainnya membentuk kelompok dengan karakteristik yang lebih spesifik.

3.4 Analisis Karakteristik Setiap Cluster

Analisis karakteristik cluster dilakukan dengan menggunakan nilai rata-rata beberapa variabel utama untuk melihat perbedaan profil akademik antar kelompok siswa. Variabel yang digunakan dalam analisis meliputi nilai akademik inti dan indikator kemampuan kognitif.



Gambar 3. Heatmap Karakteristik Nilai Rata-Rata Setiap Cluster

Pola karakteristik rata-rata setiap cluster berdasarkan variabel akademik dan kognitif ditunjukkan pada **Gambar 3**. Berdasarkan heatmap, cluster 1 menunjukkan pola nilai yang relatif stabil pada sebagian besar variabel akademik, sehingga cluster ini dapat dianggap sebagai representasi kelompok mayoritas siswa. Cluster 2 memperlihatkan nilai numerikal yang relatif tinggi dibandingkan cluster lainnya, meskipun beberapa nilai akademik inti berada sedikit di bawah cluster dominan. Cluster 3 menunjukkan performa akademik paling tinggi pada beberapa mata pelajaran inti seperti Bahasa Indonesia, Matematika, dan Bahasa Inggris. Selain itu, cluster ini juga memiliki nilai daya tangkap tertinggi dibandingkan cluster lainnya.

Sementara itu, cluster 4 menunjukkan kombinasi nilai akademik yang cukup stabil dengan kemampuan numerikal tinggi, namun pada beberapa variabel akademik tertentu masih berada di bawah cluster 3. Secara umum, perbedaan antarcluster menunjukkan bahwa algoritma K-Means mampu mengidentifikasi variasi karakteristik siswa berdasarkan kombinasi performa akademik dan kemampuan kognitif.

3.5 Evaluasi Kualitas Cluster

Evaluasi kualitas cluster dilakukan menggunakan silhouette coefficient untuk mengukur tingkat kedekatan antar data dalam cluster yang sama dibandingkan dengan cluster lain. Pengujian dilakukan pada beberapa jumlah cluster, yaitu mulai dari $k = 2$ hingga $k = 10$.

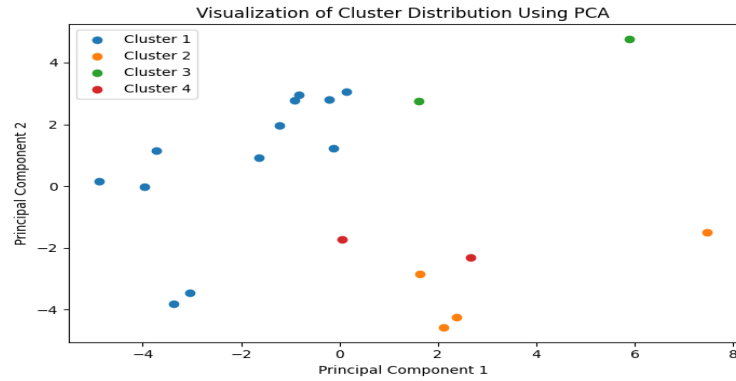
Tabel 3. Nilai Silhouette Score pada Berbagai Jumlah Cluster

Jumlah Cluster (k)	Silhouette Score
2	0.1358
3	0.1046
4	0.1093
5	0.0763
6	0.0551
7	0.0600
8	0.0760
9	0.0832
10	0.0709

Berdasarkan hasil evaluasi pada **Tabel 3**, nilai silhouette score tertinggi diperoleh pada konfigurasi dua cluster dengan nilai 0.1358. Namun, penggunaan dua cluster menghasilkan pembagian kelompok yang terlalu umum sehingga kurang mampu menggambarkan variasi karakteristik siswa secara lebih rinci. Pada konfigurasi empat cluster, diperoleh silhouette score sebesar 0.1093. Meskipun nilainya lebih rendah dibandingkan dua cluster, konfigurasi ini dipilih sebagai hasil akhir karena memberikan pembagian kelompok yang lebih informatif untuk kebutuhan analisis karakteristik siswa pada dataset dengan ukuran terbatas. Nilai silhouette score yang relatif rendah pada seluruh konfigurasi menunjukkan bahwa pemisahan antarcluster belum sepenuhnya kuat. Kondisi ini dipengaruhi oleh jumlah data penelitian yang hanya terdiri atas 20 siswa sehingga variasi data yang tersedia masih terbatas. Secara umum, hasil evaluasi ini menunjukkan bahwa clustering dalam penelitian lebih tepat digunakan sebagai analisis eksploratif untuk mengidentifikasi pola awal karakteristik siswa.

3.6 Visualisasi Sebaran Cluster

Visualisasi sebaran cluster dilakukan menggunakan principal component analysis (PCA) untuk mereduksi dimensi data ke dalam dua komponen utama sehingga pola distribusi cluster dapat diamati secara visual.



Gambar 4. Visualisasi Sebaran Cluster Menggunakan PCA

Visualisasi sebaran cluster ditunjukkan pada **Gambar 4**. Berdasarkan visualisasi PCA di atas, sebagian besar anggota cluster 1 terlihat berada dalam area yang berdekatan, menunjukkan karakteristik yang relatif homogen. Cluster lainnya tersebar pada posisi yang berbeda meskipun masih terhadap kedekatan antar beberapa titik data. Sebaran ini menunjukkan bahwa proses clustering mampu membentuk kelompok dengan pola distribusi tertentu, walaupun batas antarcluster belum sepenuhnya terpisah secara tegas. Kondisi tersebut sejalan dengan nilai silhouette score yang relatif rendah pada dataset penelitian.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan data siswa ke dalam empat cluster berdasarkan karakteristik akademik dan kemampuan kognitif yang dianalisis. Pemilihan empat cluster sebagai konfigurasi akhir dilakukan setelah mempertimbangkan hasil evaluasi silhouette coefficient dan distribusi anggota cluster yang dihasilkan. Berdasarkan hasil clustering, distribusi anggota cluster terdiri atas Cluster 1 sebanyak 12 siswa (60 %), Cluster 2 sebanyak 4 siswa (20 %), Cluster 3 sebanyak 2 siswa (10 %), dan Cluster 4 sebanyak 2 siswa (10 %). Distribusi ini menunjukkan bahwa sebagian besar siswa berada pada satu kelompok dominan dengan karakteristik siswa yang relatif serupa, sedangkan cluster lainnya merepresentasikan variasi karakteristik yang lebih spesifik.

Pemilihan empat cluster dilakukan meskipun nilai silhouette tertinggi diperoleh pada konfigurasi dua cluster sebesar 0,1358, sedangkan pada empat cluster nilai silhouette score sebesar 0,1093. Secara kuantitatif, nilai tersebut memang lebih rendah, namun empat cluster dipilih karena memberikan pembagian kelompok yang lebih rinci sehingga karakteristik antar kelompok siswa dapat diamati secara lebih informatif pada dataset terbatas. Analisis karakteristik cluster berdasarkan nilai rata-rata menunjukkan adanya perbedaan pola pada beberapa variabel akademik. Cluster tertentu memiliki kecenderungan nilai lebih tinggi pada mata pelajaran inti seperti matematika dan bahasa, sedangkan cluster lain menunjukkan nilai yang lebih menonjol pada aspek numerik tertentu. Perbedaan ini mengindikasikan bahwa proses clustering terbentuk dari kombinasi beberapa variabel, bukan hanya dipengaruhi oleh satu indikator tunggal.

Hasil evaluasi kualitas cluster menunjukkan bahwa seluruh konfigurasi cluster memiliki nilai silhouette score relatif rendah, yaitu berada pada rentang 0,0551 hingga 0,1358. Kondisi ini menunjukkan bahwa pemisahan antarcluster belum sepenuhnya kuat. Hal tersebut sangat dipengaruhi oleh ukuran dataset penelitian yang hanya terdiri atas 20 data, sehingga variasi yang tersedia masih terbatas. Visualisasi menggunakan principal component analysis juga menunjukkan bahwa sebagian titik data antarcluster masih berada pada posisi yang berdekatan. Kondisi tersebut sejalan dengan nilai silhouette score yang rendah, sehingga hasil clustering lebih tepat dipahami sebagai pola eksploratif awal dibandingkan pemisahan kelompok yang sepenuhnya tegas. Secara metodologis, hasil penelitian ini tetap menunjukkan bahwa algoritma K-Means dapat digunakan untuk memberikan gambaran awal mengenai pola pengelompokan siswa berdasarkan karakteristik akademik yang tersedia. Dalam konteks pendidikan, hasil clustering ini dapat dimanfaatkan sebagai dasar awal dalam identifikasi kelompok siswa dengan kecenderungan karakteristik belajar tertentu. Namun demikian, interpretasi hasil pada penelitian perlu untuk dipertimbangkan keterbatasan jumlah sampel. Hasil cluster pada penelitian ini tidak ditujukan untuk generalisasi luas, melainkan sebagai analisis awal pada data yang tersedia.

4. KESIMPULAN

Hasil penelitian mengenai penerapan algoritma K-Means Clustering dalam pengelompokan siswa berdasarkan profil akademik dan karakteristik belajar menunjukkan bahwa metode ini dapat digunakan untuk memberikan gambaran awal mengenai pola kemiripan data siswa berdasarkan variabel yang dianalisis. Proses penentuan jumlah cluster dilakukan dengan membandingkan nilai silhouette coefficient pada beberapa skenario jumlah cluster dari $k = 2$ hingga $k = 10$. Meskipun nilai silhouette score tertinggi diperoleh pada konfigurasi dua cluster sebesar 0,1358 penelitian ini menggunakan empat cluster sebagai hasil akhir karena distribusi anggota cluster dinilai lebih



proporsional untuk mendukung interpretasi karakteristik data pada dataset terbatas. Hasil pengelompokan ini menunjukkan distribusi cluster terdiri atas 12 siswa pada cluster pertama, 4 siswa pada cluster kedua, serta masing-masing 2 siswa pada cluster ketiga dan keempat. Perbedaan distribusi tersebut menunjukkan adanya satu kelompok dominan dengan pola karakteristik yang relatif serupa, sementara kelompok lainnya merepresentasikan variasi tertentu dalam data akademik siswa. Visualisasi menggunakan principal component analysis menunjukkan bahwa sebagian titik data antarcluster masih berada pada posisi yang berdekatan, yang sejalan dengan nilai silhouette score yang relatif rendah. Kondisi ini menunjukkan bahwa hasil clustering pada penelitian ini lebih tepat dipahami sebagai analisis eksploratif awal daripada pemisahan kelompok yang sepenuhnya kuat. Dengan jumlah sampel yang terbatas, yaitu 20 data siswa, hasil penelitian belum dapat digunakan untuk generalisasi luas, namun tetap dapat memberikan gambaran awal mengenai potensi penerapan pendekatan berbasis data dalam analisis karakteristik siswa.

REFERENCES

- [1] B. Huda, A. S. Amin, F. Nurapriani, and A. Damuri, "Aplikasi Monitoring Perkembangan Edukasi Anak Usia Dini Berbasis Web," *Jurnal Informatika Utama*, vol. 1, no. 1, Jun. 2023, doi: 10.55903/jitu.v1i1.70.
- [2] F. Nurapriani et al., "Meningkatkan Kemampuan Berpikir Kritis dan Komunikasi Matematik Mahasiswa Sistem Informasi dengan Pembelajaran Missouri Mathematics Project (MMP)," *RADIAN Journal: Research*, vol. 4, no. 3, Sep. 2025, doi: 10.35706/radian.v4i3.13152.
- [3] S. N. Adzra, F. N. Hasan, and A. Y. Kuntoro, "Penerapan Data Mining dalam Penilaian Kinerja Akademik Siswa/I SMP YPI Pulogadung dengan Metode K-Means Clustering," *Jurnal Ilmiah Informatika*, vol. 13, no. 2, 2025, doi: 10.33884/jif.v13i02.10396.
- [4] Z. Fatah and S. Maghfiroh, "Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTs Miftahul Ulum Bengkak," *JAMASTIKA: Jurnal Mahasiswa Teknik Informatika*, vol. 4, no. 2, Oct. 2025, doi: 10.35473/jamastika.v4i2.4527.
- [5] N. D. Rahayu, A. H. Anshor, I. Afriantoro, and A. Halim Anshor, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI: Jurnal Komputer dan Informatika*, vol. 6, no. 1, 2024, doi: 10.53842/juki.v6i1.474.
- [6] N. Suarna, N. Rahaningsih, and A. A. Suarna, "Optimalisasi Prestasi Akademik Siswa Melalui Pengelompokan Indeks Prestasi Dengan K-Means Clustering," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 2, May 2025, doi: 10.69916/jkbt.v4i2.321.
- [7] N. Wahyudi, Y. Ardilla, and N. P. Hastuti, "Educational Data Clustering Menggunakan K-Means pada Seleksi Penerimaan Peserta Didik Baru Madrasah Aliyah Negeri Unggulan," *Information System and Informatics Journal*, vol. 7, no. 2, Dec. 2022, doi: 10.29080/systemic.v7i2.1768.
- [8] D. A. Fajar Martha Subqi, "Data Mining Untuk Pemeliharaan Prediktif Mesin Produksi berdasarkan Database Kerusakan Mesin menggunakan Naïve Bayes Classifier," *Jurnal Ilmiah Komputasi*, vol. 20, no. 2, Jun. 2021, doi: 10.32409/jikstik.20.2.368.
- [9] L. Fabrianto, F. Riza, and N. M. Faizah, "Korelasi Antara Profil dan Nilai Akademis Siswa Dengan Menggunakan Algoritma K-Means," *JTIK: Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 4, Aug. 2021, doi: 10.25126/jtiik.202183034.
- [10] D. Oktario Dacwanda and Y. Nataliani, "Implementasi K-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. 2, Aug. 2021, doi: 10.24246/aiti.v18i2.125-138.
- [11] Awalia Rifqa, "Penerapan Metode K-Means Clustering untuk Pengelompokan Prestasi Siswa Menggunakan Orange Data Mining: Studi Kasus di MTs Muhammadiyah Tallo Makassar," *MAPLE: Mechatronics Journal in Professional and Entrepreneur*, vol. 6, no. 2, Dec. 2024, doi: 10.61141/maple.v6i2.579.
- [12] N. A. Rizki, K. Kurniawan, I. K. Hasan, and N. Sampe, "Implementasi Algoritma K-Means Untuk Mengelompokkan Mahasiswa Berdasarkan Sumber Belajarnya," *METIK: Media Teknologi Informasi dan Komputer*, vol. 7, no. 2, Dec. 2023, doi: 10.47002/metik.v7i2.584.
- [13] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru," *Jurnal Sistem Informasi Triguna Dharma*, vol. 2, no. 6, Nov. 2023, doi: 10.53513/jursi.v2i6.9148.
- [14] A. R. Lashiyanti, I. Rasyid Munthe, F. A. Nasution, and E. P. Korespondensi, "Optimisasi Klusterisasi Nilai Ujian Nasional Dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette," *JIKOMSI: Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 6, no. 1, pp. 14–20, Mar. 2023, Accessed: Apr. 2026. [Online]. Available: <https://ejournal.sisfokomtek.org/index.php/jikom/article/view/1550>.
- [15] K. Setiawan, Y. Yanti, and A. Saputry, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means di Pusat Pengembangan Anak Fajar Baru Cengkareng," *JTIK: Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, 2024, doi: 10.35870/jtik.v8i1.1426.
- [16] M. Milla, V. Aprila Kore Dima, and A. Purnami Setiawi, "Penerapan Algoritma K-Means untuk Mengidentifikasi Minat Belajar Siswa di Sekolah Dasar Negeri Puu Naga," *Jurnal Informatika dan Sains Teknologi*, vol. 3, no. 4, 2025, doi: 10.62951/modem.v3i4.630.
- [17] S. Rahmah, M. Jamil, A. Aslindah, A. Fawait, and Y. F. Saputra, "Pengelompokan Hasil Pembelajaran Mahasiswa dengan Algoritma K-Means Clustering," *JERKIN: Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, vol. 3, no. 4, pp. 4215–4221, Jun. 2025, doi: 10.31004/jerkin.v3i4.1217.
- [18] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *KOMPUTIKA: Jurnal Sistem Komputer*, vol. 11, no. 1, Jan. 2022, doi: 10.34010/komputika.v11i1.4350.



- [19] Y. N. Dewi, H. Rianto, C. Budihartanti, and F. W. Fibriany, “Penerapan Metode K-Means Dalam Menentukan Kelompok Pendalaman Materi Ujian Nasional,” *Journal of Information System, Applied, Management, Accounting and Research*, vol. 6, no. 1, Feb. 2022, doi: 10.52362/jisamar.v6i1.670.
- [20] A. Fadlil, I. Riadi, and Y. Mulyana, “Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan,” *JTIK: Jurnal Teknologi Informasi dan Komputer*, vol. 8, no. 2, pp. 352–366, Sep. 2022, doi: 10.37012/jtik.v8i2.1261.
- [21] R. Sudrajat, A. I. Hadiana, and M. Melina, “Evaluasi Kualitas Klaster Wilayah Rawan Bencana Menggunakan K-Means dengan Silhouette dan Elbow Method,” *Jurnal Algoritma*, vol. 22, no. 2, pp. 127–139, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2379.
- [22] N. Nurahman, A. Purwanto, and S. Mulyanto, “Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 337–350, Mar. 2022, doi: 10.30812/matrik.v21i2.1411.
- [23] M. Qusyairi, Zul Hidayatullah, and Arnita Sandi, “Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow,” *INFOTEK: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 500–510, Jul. 2024, doi: 10.29408/jit.v7i2.26375.