

ANALISIS KEAKTIFAN SISWA MENGGUNAKAN ALGORITMA K-MEANS DI SMAN 13 GARUT

Andi Maulana Yusup¹, Nana Suarna², Agus Bahtiar³, Rosidin⁴.

Program Studi Teknik Informatika¹²
Program Studi Sistem Informasi³⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
a7x3454@gmail.com

(*) Corresponding Author : a7x3454@gmail.com
Published : 15 Juli 2026

Abstract—Student activeness is an important indicator of the success of the learning process, which is not only assessed from academic aspects but also from attendance and participation in non-academic activities. This study aims to analyze student activeness using the K-Means *Clustering* algorithm on student data at SMA Negeri 13 Garut. The variables used in this study include gender, class, average academic score, attendance percentage, number of unexcused absences (alpha), and participation in extracurricular activities. The method used is K-Means *Clustering* with the determination of the optimal number of *clusters* using the *Elbow* Method. The results show that the optimal number of *clusters* is $K = 6$. The quality of *clustering* is evaluated using the *Silhouette score* and Davies-Bouldin Index (DBI), with values of 0.631 and 0.582, respectively. These values indicate that the resulting *clusters* have good separation and a relatively strong structure. The *clustering* results show that students can be grouped into six *clusters* with different characteristics, ranging from highly active students to those with moderate activity levels and those requiring special attention. This grouping provides a more comprehensive understanding of student activeness patterns based on multiple related aspects. Therefore, this study is expected to support decision-making for schools in designing more targeted student development strategies and improving the overall quality of the learning process.

Keywords: K-Means *Clustering*, Student Activeness, *Elbow* Method, *Silhouette score*, Davies-Bouldin Index

Abstrak—Keaktifan siswa merupakan salah satu indikator penting dalam keberhasilan proses pembelajaran, yang tidak hanya ditinjau dari aspek akademik, tetapi juga dari kehadiran dan keterlibatan dalam kegiatan non-akademik. Penelitian ini bertujuan untuk menganalisis keaktifan siswa menggunakan algoritma K-Means *Clustering* pada data siswa SMA Negeri 13 Garut. Variabel yang digunakan dalam penelitian ini meliputi jenis kelamin, kelas, nilai rata-rata akademik, persentase kehadiran, jumlah ketidakhadiran tanpa keterangan (alfa), dan jumlah keikutsertaan dalam kegiatan ekstrakurikuler. Metode yang digunakan adalah K-Means *Clustering* dengan penentuan jumlah *cluster* optimal menggunakan metode *Elbow*. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal adalah $K = 6$. Evaluasi kualitas *cluster* dilakukan menggunakan *Silhouette score* dan Davies-Bouldin Index (DBI), dengan hasil masing-masing sebesar 0,631 dan 0,582. Nilai tersebut menunjukkan bahwa *cluster* yang terbentuk memiliki tingkat pemisahan yang baik dan struktur yang cukup kuat. Hasil *clustering* menunjukkan bahwa siswa dapat dikelompokkan ke dalam enam kelompok dengan karakteristik yang berbeda, mulai dari siswa dengan keaktifan tinggi, sedang, hingga yang memerlukan perhatian khusus. Pengelompokan ini memberikan gambaran yang lebih komprehensif mengenai pola keaktifan siswa berdasarkan berbagai aspek yang saling berkaitan. Dengan demikian, penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan bagi pihak sekolah dalam menyusun strategi pembinaan siswa secara lebih tepat sasaran serta meningkatkan kualitas proses pembelajaran.

Kata Kunci: K-Means *Clustering*, Keaktifan Siswa, *Elbow* Method, *Silhouette score*, Davies-Bouldin Index

INTRODUCTION

Penelitian menurut [1] pendidikan merupakan salah satu pilar utama dalam

pembangunan sumber daya manusia yang berkualitas. Keberhasilan proses pendidikan tidak hanya ditentukan oleh kurikulum dan kompetensi

guru, tetapi juga sangat dipengaruhi oleh tingkat keaktifan dan partisipasi siswa dalam setiap kegiatan pembelajaran.

Penelitian menurut [2] Siswa yang aktif cenderung memiliki pemahaman lebih baik terhadap materi pelajaran, kemampuan berpikir kritis yang lebih tinggi, serta motivasi belajar yang lebih kuat dibandingkan dengan siswa yang pasif. Oleh karena itu, pemantauan dan analisis keaktifan siswa secara sistematis menjadi hal yang sangat penting bagi pengelola institusi pendidikan.

SMA Negeri 13 Garut sebagai salah satu sekolah menengah atas di Kabupaten Garut memiliki tanggung jawab besar dalam mencetak generasi penerus yang kompeten dan berkarakter. Sekolah ini menampung ratusan siswa dari berbagai latar belakang sosial dan akademis. Dalam kesehariannya, pihak sekolah menghadapi tantangan nyata dalam memantau keaktifan masing-masing siswa secara individual, mengingat jumlah siswa yang besar dan beragamnya indikator keaktifan yang perlu diperhatikan, mulai dari kehadiran, partisipasi kelas, keterlibatan dalam kegiatan ekstrakurikuler, hingga perolehan nilai akademik.

Penelitian menurut [3] data keaktifan siswa yang terus bertambah dari tahun ke tahun membutuhkan pendekatan analisis yang lebih terstruktur dan berbasis data. Pengelolaan data secara manual oleh guru dan staf sekolah tidak lagi mampu memberikan gambaran yang komprehensif dan akurat mengenai pola keaktifan siswa secara keseluruhan. Kondisi ini mendorong perlunya penerapan teknologi informasi dan teknik analisis data yang lebih canggih agar informasi keaktifan siswa dapat diolah secara efisien, tepat waktu, dan memberikan wawasan yang berarti bagi pengambilan keputusan pendidikan.

Penelitian menurut [4] *data mining* hadir sebagai solusi yang tepat untuk mengatasi permasalahan pengelolaan data pendidikan dalam skala besar. *Data mining* merupakan proses penemuan pola, korelasi, anomali, dan pengetahuan yang berguna dari sekumpulan data dalam jumlah besar menggunakan teknik-teknik statistik, kecerdasan buatan, dan *machine learning*. Dalam konteks pendidikan, *data mining* telah terbukti mampu memberikan kontribusi nyata dalam membantu institusi pendidikan memahami karakteristik siswanya secara lebih mendalam dan berbasis bukti empiris.

Menurut penelitian [5] salah satu teknik *data mining* yang paling banyak digunakan

dalam penelitian pendidikan adalah *clustering*, khususnya algoritma *K-Means*. Algoritma *K-Means* bekerja dengan cara mengelompokkan data ke dalam sejumlah *cluster* berdasarkan kemiripan atribut yang dimiliki, sehingga setiap anggota dalam satu *cluster* memiliki karakteristik yang lebih homogen dibandingkan dengan anggota dari *cluster* yang berbeda. Teknik ini sangat cocok diterapkan pada data siswa karena mampu mengidentifikasi pola-pola tersembunyi yang sulit terdeteksi melalui pengamatan biasa.

Penelitian yang dilakukan oleh [6] di SMP Negeri 1 Bilah Barat membuktikan bahwa algoritma *K-Means* dapat diterapkan secara efektif untuk menganalisis data absensi siswa. Dalam penelitian tersebut, data absensi diolah menggunakan metode *K-Means clustering* sehingga menghasilkan pengelompokan siswa berdasarkan tingkat kehadiran mereka. Hasilnya menunjukkan bahwa *clustering* berbasis *K-Means* mampu memberikan pola yang jelas dan bermakna terkait kondisi kehadiran siswa, yang selanjutnya dapat dijadikan dasar intervensi yang lebih terarah oleh pihak sekolah.

Selain kehadiran, keaktifan siswa dalam proses pembelajaran juga menjadi variabel penting yang perlu dianalisis. Dalam penelitiannya [7] mengimplementasikan algoritma *K-Means clustering* untuk melakukan seleksi siswa berprestasi berdasarkan tingkat keaktifan mereka dalam proses pembelajaran. Penelitian ini memperlihatkan bahwa dengan menggunakan *K-Means*, siswa dapat dikelompokkan ke dalam kategori berdasarkan berbagai indikator keaktifan seperti frekuensi bertanya, partisipasi diskusi, dan penyelesaian tugas. Hasil pengelompokan ini terbukti valid dan dapat digunakan sebagai referensi dalam program pembinaan siswa berprestasi.

Menurut penelitian [8] aspek kehadiran atau absensi merupakan salah satu indikator utama keaktifan siswa yang paling mudah diukur dan diverifikasi secara objektif.

Menurut penelitian [9] melakukan penelitian yang lebih komprehensif dengan mengidentifikasi pola belajar dan absensi siswa menggunakan teknik *clustering* yang dikombinasikan dengan analisis korelasi. Penelitian tersebut membuktikan bahwa terdapat keterkaitan yang signifikan antara pola absensi dengan pola belajar siswa. Dengan menggabungkan dua teknik analisis ini, pola keaktifan siswa dapat dipahami secara lebih holistik, sehingga intervensi yang dirancang pun menjadi lebih tepat sasaran dan efektif.

Identifikasi risiko akademik siswa secara dini juga merupakan salah satu manfaat strategis penerapan *K-Means* dalam analisis keaktifan.

Hasil penelitian [10] dalam studinya di SMPN 10 Palembang berhasil menggunakan metode *K-Means* untuk menganalisis risiko akademik siswa berdasarkan data keaktifan dan nilai akademik. Penelitian ini menemukan bahwa siswa yang tergolong dalam *cluster* risiko tinggi memiliki karakteristik keaktifan yang rendah dan perlu segera mendapatkan intervensi dari guru maupun konselor sekolah. Temuan ini menggarisbawahi pentingnya analisis berbasis data dalam sistem deteksi dini risiko akademik.

Implementasi *K-Means* berbasis web juga telah berhasil dikembangkan dalam penelitian [11] di SMK Yadika 3 Jakarta. Penelitian tersebut membangun sistem *clustering* berbasis web menggunakan algoritma *K-Means* yang dapat diakses secara online oleh pihak manajemen sekolah. Sistem ini terbukti mampu memudahkan proses analisis data siswa secara *real-time* tanpa memerlukan keahlian teknis yang mendalam dari pengguna. Hal ini membuktikan bahwa teknologi *K-Means* tidak hanya relevan secara akademis, tetapi juga memiliki potensi implementasi praktis yang tinggi di lingkungan sekolah.

Penentuan siswa berprestasi secara objektif juga menjadi salah satu fokus penelitian berbasis *K-Means*. [12] dalam penelitiannya di SMK Swasta TI Panca Dharma Stungkit menggunakan metode *K-Means* untuk menganalisis data siswa dan menentukan *cluster* siswa berprestasi berdasarkan berbagai variabel akademik dan non-akademik. Penelitian ini memperlihatkan bahwa penggunaan *K-Means* memberikan hasil pengelompokan yang konsisten dan dapat dipertanggungjawabkan secara ilmiah, sehingga proses seleksi siswa berprestasi menjadi lebih adil dan transparan.

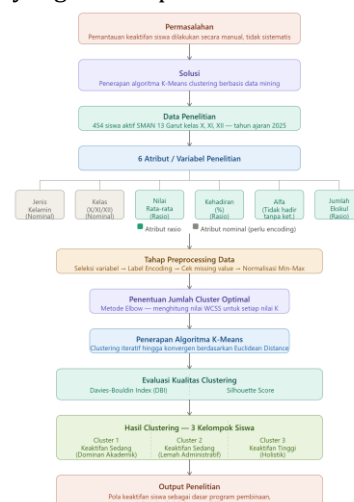
[13] dalam penelitiannya juga mengonfirmasi efektivitas *K-Means* dalam analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi. Penelitian ini memanfaatkan data nilai akademik sebagai variabel utama dan berhasil mengklasifikasikan siswa ke dalam *cluster* berprestasi tinggi, sedang, dan rendah. Hasil penelitian ini relevan secara langsung dengan kebutuhan SMA Negeri 13 Garut yang memerlukan metode objektif dan berbasis data untuk mengidentifikasi pola keaktifan dan capaian siswa secara

komprehensif, bukan hanya berdasarkan laporan guru semata.

Berdasarkan kajian dari berbagai penelitian terdahulu tersebut, terlihat jelas bahwa algoritma *K-Means* telah terbukti secara empiris sebagai metode yang handal, efisien, dan mudah diimplementasikan dalam berbagai konteks analisis data pendidikan di Indonesia. Meskipun demikian, apabila dicermati lebih dalam, terdapat kesenjangan yang cukup signifikan di antara penelitian-penelitian tersebut. Penelitian [6] dan [7] mengaitkan keaktifan siswa dengan proses pembelajaran, namun orientasinya lebih condong pada seleksi siswa berprestasi atau evaluasi hasil belajar, bukan pada analisis keaktifan secara holistik. Penelitian [9] dan [10], meskipun lebih komprehensif dengan mengombinasikan variabel absensi bersama aspek pola belajar dan risiko akademik, tetap belum menjadikan kehadiran sebagai indikator utama dalam mengukur keaktifan siswa. Di sisi lain, [11], [12], serta [13] lebih menitikberatkan pengelompokan siswa pada variabel nilai akademik dan prestasi, dengan absensi hanya berperan sebagai variabel pendukung. Dengan demikian, dapat disimpulkan bahwa penelitian-penelitian sebelumnya masih cenderung berfokus pada aspek prestasi akademik dan belum secara spesifik menganalisis keaktifan siswa secara multidimensi dengan data kehadiran sebagai variabel utamanya, sehingga terdapat celah penelitian yang nyata dan perlu diisi.

MATERIALS AND METHODS

Menghasilkan kategorisasi tingkat keaktifan siswa beserta karakteristik masing-masing kelompok, sehingga hasilnya dapat menjadi dasar bagi sekolah dalam merancang program pembinaan, intervensi dini, dan kebijakan pendidikan yang lebih tepat sasaran.



Gambar 1 kerangka berfikir

Menunjukkan alur lengkap penelitian yang bertujuan untuk mengelompokkan tingkat keaktifan siswa menggunakan metode *k-means clustering*. Proses dimulai dari permasalahan, yaitu pemantauan keaktifan siswa yang masih dilakukan secara manual dan tidak sistematis. Untuk mengatasi hal tersebut, diajukan solusi berupa penerapan algoritma *k-means* berbasis data mining.

Selanjutnya, ditentukan data penelitian yang digunakan, yaitu 454 siswa aktif di sman 13 garut dari kelas x, xi, dan xii pada tahun ajaran 2025. Data tersebut terdiri dari 6 atribut/variabel penelitian, yaitu jenis kelamin dan kelas (bersifat nominal sehingga perlu dilakukan encoding), serta nilai rata-rata, kehadiran, alfa (ketidakhadiran tanpa keterangan), dan jumlah ekstrakurikuler (bersifat rasio).

Tahapan berikutnya adalah preprocessing data, yang meliputi seleksi variabel, label encoding untuk data nominal, pengecekan missing value, serta normalisasi menggunakan metode min-max agar data berada dalam skala yang seragam. Setelah data siap, dilakukan penentuan jumlah cluster optimal menggunakan metode elbow dengan menghitung nilai wcss untuk setiap jumlah cluster (*k*).

Kemudian, algoritma *k-means* diterapkan untuk mengelompokkan data secara iteratif hingga mencapai kondisi konvergen berdasarkan perhitungan jarak euclidean. Hasil pengelompokan selanjutnya dievaluasi melalui evaluasi kualitas clustering menggunakan metrik davies-bouldin index (*dbi*) dan silhouette score untuk memastikan kualitas cluster yang terbentuk.

Dari proses tersebut diperoleh hasil clustering menjadi 3 kelompok siswa, yaitu: cluster 1 (keaktifan sedang dengan dominasi akademik), cluster 2 (keaktifan sedang namun lemah secara administratif), dan cluster 3 (keaktifan tinggi secara holistik). Akhirnya, penelitian ini menghasilkan output berupa pola keaktifan siswa yang dapat digunakan sebagai dasar dalam penyusunan program pembinaan, intervensi dini, serta kebijakan pendidikan yang lebih tepat sasaran.

Results and discussion

Setelah dilakukan penentuan jumlah *cluster* optimal menggunakan metode *elbow*, diperoleh nilai *k* = 6. Selanjutnya dilakukan proses *clustering* menggunakan algoritma *k-means* dengan jumlah *cluster* tersebut. Proses

clustering dilakukan dengan bantuan bahasa pemrograman python menggunakan library *scikit-learn*. Hasil dari proses ini adalah pengelompokan seluruh data siswa ke dalam enam *cluster* berdasarkan kemiripan karakteristik pada variabel yang digunakan.

Pada tahap awal, hasil *clustering* masih berupa label *cluster* yang dimulai dari angka 0 hingga 5. Untuk memudahkan interpretasi, dilakukan penyesuaian dengan menambahkan nilai +1, sehingga label *cluster* berubah menjadi *cluster 0*, *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4*, dan *cluster 5*.

Pengelompokan ini didasarkan pada kemiripan nilai dari variabel yang digunakan, yaitu nilai jenis kelamin, kelas, rata-rata akademik, kehadiran, alfa, dan jumlah ekstrakurikuler. Siswa yang memiliki karakteristik yang mirip akan berada dalam *cluster* yang sama.

Hasil pengelompokan ini selanjutnya akan dianalisis lebih lanjut untuk mengetahui jumlah anggota tiap *cluster* serta karakteristik masing-masing *cluster*.

Setelah proses *clustering* menggunakan algoritma *k-means* dengan jumlah *cluster* sebanyak 6 *cluster*, langkah selanjutnya adalah menghitung jumlah anggota pada masing-masing *cluster*. Perhitungan ini bertujuan untuk mengetahui distribusi data siswa dalam setiap kelompok yang terbentuk.

Berdasarkan hasil pengolahan data menggunakan python, diperoleh jumlah anggota tiap *cluster* setelah dilakukan penyesuaian label *cluster* (*cluster* + 1) sebagai berikut:

```
print("Jumlah anggota tiap Cluster (K=6, Fitur Tambahan, Label Disesuaikan):")
adjusted_cluster_counts_final = df_map['cluster_final_adjusted'].value_counts().sort_index()
print(adjusted_cluster_counts_final)
```

```
Jumlah anggota tiap Cluster (K=6, Fitur Tambahan, Label Disesuaikan):
cluster_final_adjusted
0      83
1      78
2      69
3      76
4      79
5      68
Name: count, dtype: int64
```

Gambar 2 hasil jumlah anggota percluster

Berdasarkan 2 dapat diketahui bahwa jumlah anggota pada masing-masing *cluster* adalah sebagai berikut: (a) *cluster 0* memiliki jumlah anggota sebanyak 83 siswa, (b) *cluster 1* sebanyak 78 siswa, (c) *cluster 2* sebanyak 69 siswa, (d) *cluster 3* sebanyak 76 siswa, (e) *cluster 4* sebanyak 79 siswa, dan (f) *cluster 5* sebanyak 68 siswa.

Dari distribusi tersebut, dapat dilihat bahwa jumlah anggota pada setiap *cluster* relatif seimbang, meskipun *cluster 0* memiliki jumlah anggota terbanyak dan *cluster 5* memiliki jumlah anggota paling sedikit. Perbedaan jumlah anggota antar *cluster* tidak terlalu signifikan, yang menunjukkan bahwa data siswa tersebar cukup merata ke dalam enam kelompok yang terbentuk.

Distribusi yang relatif seimbang ini mengindikasikan bahwa karakteristik siswa dalam dataset memiliki variasi yang cukup beragam dan tidak terpusat pada satu kelompok tertentu saja. Hal

ini memberikan dasar yang baik untuk melakukan analisis lebih lanjut terhadap karakteristik masing-masing *cluster* pada subbab berikutnya.

Setelah diketahui jumlah anggota pada masing-masing *cluster*, langkah selanjutnya adalah menganalisis nilai centroid atau rata-rata dari setiap variabel pada masing-masing *cluster*. Nilai *centroid* digunakan untuk mengetahui karakteristik utama dari setiap *cluster* yang terbentuk.

Perhitungan *centroid* dilakukan dengan menghitung rata-rata nilai dari setiap variabel berdasarkan kelompok *cluster* masing-masing. Hasil perhitungan *centroid* setelah penyesuaian label *cluster* (*cluster* + 1) disajikan pada gambar :

```

... Analisis Rata-rata per Cluster (Centroid K=6, Fitur Tambahan, Label Disesuaikan):
Nilai Rata-rata Kehadiran(%) Alfa \
Cluster_Final_Adjusted
0      82.891566      92.168675      1.216867
1      82.743590      91.871795      1.256410
2      82.463768      90.985587      1.275362
3      82.342105      90.815789      1.302632
4      81.974604      90.848101      1.291139
5      82.485294      91.073529      1.441376

Jumlah Ekskul Jenis Kelamin Kelas
Cluster_Final_Adjusted
0      1.397590      P      10
1      1.384615      L      11
2      1.428290      P      12
3      1.394737      P      11
4      1.367089      L      10
5      1.441176      L      12

```

Gambar 3 hasil analisis rata-rata percluster

Berdasarkan gambar 3 dapat dilakukan analisis karakteristik masing-masing *cluster* sebagai berikut:

cluster 0 memiliki nilai rata-rata akademik yang cukup tinggi (82.89) dengan tingkat kehadiran yang baik (92.17%). Jumlah alfa tergolong rendah (1.22) dan jumlah ekstrakurikuler sedang (1.39). Hal ini menunjukkan bahwa siswa dalam *cluster* ini tergolong aktif secara akademik dengan keaktifan non-akademik yang cukup seimbang.

cluster 1 memiliki nilai rata-rata akademik yang baik (82.74) dengan tingkat kehadiran tinggi (91.87%). Jumlah alfa relatif sedang (1.25) dan jumlah ekstrakurikuler juga sedang (1.38). Siswa pada *cluster* ini menunjukkan keseimbangan antara aktivitas akademik dan non-akademik dengan tingkat kedisiplinan yang cukup baik.

cluster 2 memiliki nilai rata-rata akademik yang cukup baik (82.46) dengan tingkat kehadiran (90.99%). Jumlah alfa (1.28) dan jumlah ekstrakurikuler (1.42) berada pada tingkat sedang. *Cluster* ini mencerminkan siswa dengan keaktifan yang cukup stabil namun tidak menonjol dibanding *cluster* lain.

cluster 3 memiliki nilai rata-rata akademik (82.34) dengan tingkat kehadiran (90.82%) yang sedikit lebih rendah dibanding *cluster* lain. Jumlah alfa (1.31) relatif lebih tinggi dan jumlah ekstrakurikuler (1.39) berada pada tingkat sedang. Hal ini menunjukkan bahwa siswa

dalam *cluster* ini memiliki kecenderungan keaktifan yang sedikit lebih rendah dibandingkan *cluster* lainnya.

cluster 4 memiliki nilai rata-rata akademik (81.97) dengan tingkat kehadiran (90.85%). Jumlah alfa (1.29) dan jumlah ekstrakurikuler (1.37) berada pada tingkat sedang. *Cluster* ini menunjukkan karakteristik siswa dengan keaktifan yang cukup, namun tidak terlalu dominan baik dari sisi akademik maupun non-akademik.

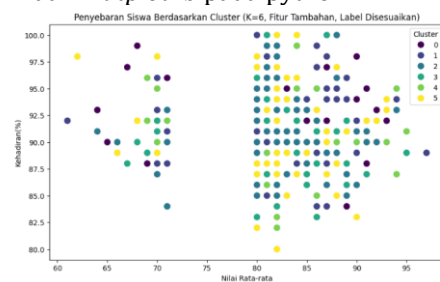
cluster 5 memiliki nilai rata-rata akademik (82.49) dengan tingkat kehadiran (91.07%). Namun, *cluster* ini memiliki jumlah alfa tertinggi (1.44) serta jumlah ekstrakurikuler tertinggi (1.44). Hal ini menunjukkan bahwa siswa dalam *cluster* ini cukup aktif dalam kegiatan non-akademik, namun memiliki kecenderungan tingkat ketidakhadiran tanpa keterangan yang lebih tinggi dibanding *cluster* lain.

Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa keenam *cluster* menunjukkan karakteristik yang relatif mirip dengan perbedaan yang tidak terlalu signifikan. Hal ini mengindikasikan bahwa data siswa memiliki pola yang cukup homogen, meskipun tetap terdapat variasi kecil pada tingkat keaktifan akademik maupun non-akademik. Oleh karena itu, pengelompokan ini lebih menggambarkan variasi tingkat keaktifan yang bersifat bertahap, bukan perbedaan yang sangat kontras antar *cluster*.

Analisis centroid ini menjadi dasar dalam proses interpretasi *cluster* yang akan dibahas lebih lanjut pada subbab berikutnya.

Visualisasi hasil *clustering* dilakukan untuk memberikan gambaran secara visual mengenai penyebaran data siswa berdasarkan *cluster* yang telah terbentuk. Pada penelitian ini, visualisasi dilakukan menggunakan grafik *scatter plot* dengan dua variabel utama, yaitu nilai rata-rata akademik sebagai sumbu x dan kehadiran (%) sebagai sumbu y.

Setiap titik pada grafik merepresentasikan satu siswa, sedangkan warna yang berbeda menunjukkan keanggotaan *cluster* masing-masing siswa. Visualisasi ini dibuat menggunakan *library seaborn* dan *matplotlib* pada python.



Gambar 4 hasil visualisasi data

Pada gambar 4 ditampilkan sebaran siswa sman 13 garut berdasarkan dua dimensi utama, yaitu nilai rata-rata (sumbu x) dan kehadiran (sumbu y), dengan menggunakan $k=6$.

Berdasarkan visualisasi tersebut, dapat ditarik beberapa poin analisis sebagai berikut:

Segmentasi performa akademik dan kedisiplinan: data menunjukkan adanya massa yang besar pada rentang nilai rata-rata 80 hingga 90 dengan tingkat kehadiran di atas 82%. Ini menunjukkan bahwa mayoritas siswa berada dalam kategori standar performa sekolah yang cukup tinggi. Penggunaan 6 *cluster* memungkinkan algoritma untuk memecah massa data yang padat ini menjadi sub-kelompok yang lebih detail berdasarkan fitur tambahan lainnya.

Identifikasi kelompok siswa "rentan": terlihat beberapa titik data yang tersebar di sisi kiri grafik (rentang nilai 60 hingga 75). Meskipun secara visual jumlahnya tidak sebanyak kelompok utama, algoritma k-means berhasil mengisolasi kelompok ini ke dalam *cluster* tersendiri. Hal ini sangat berguna bagi sekolah untuk mengidentifikasi siswa yang membutuhkan perhatian akademik lebih intensif.

Pola distribusi kehadiran: terdapat garis-garis horizontal yang konsisten pada tingkat kehadiran tertentu (seperti 90%, 92.5%, hingga 100%). Hal ini menunjukkan sifat data kehadiran yang cenderung diskrit (terukur tetap). *Cluster* yang terbentuk membagi titik-titik pada garis tersebut berdasarkan posisi nilai rata-ratanya, sehingga memisahkan mana siswa yang "hadir namun nilai rendah" dengan siswa "hadir dan nilai tinggi".

Kualitas pemisahan *cluster*: meskipun terdapat beberapa *cluster* yang terlihat saling bertumpuk (terutama pada area padat di tengah), hal ini disebabkan oleh visualisasi yang terbatas pada dua dimensi (2d). Secara matematis, sebagaimana dibuktikan pada nilai *silhouette score* 0,631, setiap *cluster* sebenarnya terpisah dengan baik jika ditinjau dari seluruh dimensi fitur yang digunakan dalam perhitungan.

Kesimpulan visualisasi: gambar 4.9 memperkuat hasil analisis statistik sebelumnya. Dengan $k=6$, sekolah tidak hanya mendapatkan gambaran umum, tetapi dapat memetakan secara spesifik enam profil perilaku siswa yang berbeda. Visualisasi ini memudahkan pengambilan keputusan dalam memberikan *reward* bagi *cluster* unggulan maupun intervensi bagi *cluster* yang memiliki nilai atau kehadiran di bawah rata-rata.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis keaktifan siswa menggunakan algoritma K-Means pada siswa SMA Negeri 13 Garut, maka dapat disimpulkan sebagai berikut:

1. Penerapan Algoritma K-Means
Algoritma K-Means berhasil diterapkan untuk mengelompokkan 454 siswa SMA Negeri 13 Garut berdasarkan enam variabel, yaitu jenis kelamin, kelas, nilai rata-rata akademik, persentase kehadiran, jumlah ketidakhadiran tanpa keterangan (alfa), dan jumlah kegiatan ekstrakurikuler. Proses *clustering* mampu mengelompokkan siswa secara objektif berdasarkan kemiripan karakteristik tanpa memerlukan label awal, sehingga menghasilkan pengelompokan yang berbasis data.
2. Jumlah Cluster Optimal
Penentuan jumlah *cluster* optimal menggunakan metode *Elbow* menunjukkan bahwa nilai K terbaik adalah $K = 6$, yang ditandai dengan titik siku (*Elbow point*) pada grafik WCSS. Hasil ini diperkuat dengan evaluasi menggunakan *Silhouette score* dan Davies-Bouldin Index (DBI), di mana nilai terbaik juga diperoleh pada $K = 6$.
3. Kualitas Hasil Clustering
Hasil evaluasi menunjukkan bahwa model *clustering* memiliki kualitas yang baik, dengan:
 - a) *Silhouette score* sebesar 0,631, yang menunjukkan struktur *cluster* yang kuat (strong structure), dengan tingkat kohesi dan separasi yang tinggi.
 - b) Davies-Bouldin Index (DBI) sebesar 0,582, yang menunjukkan bahwa *cluster* yang terbentuk memiliki tingkat pemisahan yang baik dan minim overlap antar *cluster*.

REFERENCE

- [1] M. Solihin and M. Hamid, "Enhancing Student Engagement in Indonesian Language Learning for Grade IV at MI Nurussyifa Timbanuh Through Active Learning Strategies: A Classroom Action Research," *JURNAL Studi Tindakan Edukatif (JSTE)*, vol. 1, no. 5, pp. 1928–1934, 2025.
- [2] F. Rahmandani, M. R. Hamzah, T. Handayani, and S. Fatimah, "Penerapan Problem Based Learning (PBL) dalam peningkatan keaktifan dan motivasi belajar peserta didik pada mata pelajaran Pendidikan Pancasila

- di SMAN 2 Batu," *Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI)*, vol. 4, no. 3, pp. 1016–1027, 2024.
- [3] N. A. Dwijaksana and A. H. Muhammad, "Pemanfaatan Teknologi Learning Analytics Dalam Mendeteksi Pola Belajar Untuk Meningkatkan Kualitas Pembelajaran di SMK Negeri 1 Kabupaten Sorong," *Jurnal Pendidikan UNIGA*, vol. 19, no. 1, pp. 122–128, 2025.
- [4] E. Saputri, "Teknik dan aplikasi data mining di Indonesia: tinjauan literatur satu dekade (2015-2024)," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 2, pp. 138–149, 2025.
- [5] B. A. Prayudha, R. Kurniawan, Y. Wijaya, and U. Hayati, "Algoritma K-Means Untuk Meningkatkan Model Klasterisasi Data Siswa SMK Samudra Nusantara Kabupaten Cirebon Berdasarkan Nilai Akademik," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 1314–1321, 2025.
- [6] S. R. F. Sinurat, M. Nasution, and R. M. Ah, "Analisis Data Mining Absensi Siswa SMP Negeri 1 Bilah Barat Dengan Metode Algoritma K-Means Clustering," *INFORMATIKA*, vol. 12, no. 2, pp. 292–303, 2024.
- [7] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022.
- [8] H. Pramudya, A. F. Cobantoro, and J. Karaman, "Implementasi Algoritma Selection Sort Dalam Sistem Absensi Siswa Untuk Pengurutan Keaktifan Berdasarkan Kehadiran," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 2887–2895, 2025.
- [9] C. Renita, T. N. Padilah, and M. A. E. Wahyu, "IDENTIFIKASI POLA BELAJAR DAN ABSENSI SISWA DENGAN TEKNIK CLUSTERING DAN ANALISIS KORELASI," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 10286–10292, 2024.
- [10] R. Nabilatulrahmah *et al.*, "ANALISIS RISIKO AKADEMIK SISWA DENGAN METODE K-MEANS: STUDI KASUS SMPN 10 PALEMBANG," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6766–6773, 2025.
- [11] Y. Ahmad and T. Fatimah, "IMPLEMENTASI TEKNIK CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS PADA SMK YADIKA 3 JAKARTA BERBASIS WEB," 2024.
- [12] D. Apriandi, R. M. Sari, and M. I. Sarif, "Analisis clustering untuk menentukan siswa berprestasi di SMK Swasta TI Panca Dharma Stungkit menggunakan metode K-Means," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1117–1129, 2024.
- [13] E. A. Saputra and Y. Nataliani, "Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, pp. 424–439, 2021.