



Situs Jurnal

<https://sintek.stmikku.ac.id/index.php/home>**PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK  
MENGELOMPOKAN SISWA SMK AL-MA'RIFAH  
BERDASARKAN KEHADIRAN****Dila Nurhafidilah<sup>1</sup>, Nana Suarna<sup>2</sup>, Agus Bahtiar<sup>3</sup>, Umi Hayati<sup>4</sup>, Fatihanursari Dikananda<sup>5</sup>**<sup>1 2 4 5</sup>Teknik Informatika, STMIK IKMI Cirebon, Jawa Barat, Indonesia<sup>3</sup>Sistem Informasi, STMIK IKMI Cirebon, Jawa Barat, Indonesia

Jl. Perjuangan No. 10B, Karyamulya, Kecamatan Kesambi, Kota Cirebon, Jawa Barat 45131, Indonesia

Email: <sup>1</sup>dilanurhafidilah25@gmail.com, <sup>2</sup>st\_nana@yahoo.com<sup>3</sup>agusbahtiar038@gmail.com, <sup>4</sup>umi@stmik-amikbandung.ac.id, <sup>5</sup>fatiha.dikananda@gmail.com**ABSTRAK**

Penelitian ini bertujuan untuk mengelompokkan siswa SMK Al-Ma'rifah berdasarkan pola kehadiran menggunakan algoritma K-Means Clustering. Data yang dianalisis merupakan catatan kehadiran siswa tahun ajaran 2023/2024 yang meliputi jumlah hadir, izin, sakit, alfa, dan persentase kehadiran. Tahapan pra-pemrosesan data dilakukan melalui pembersihan dan normalisasi sebelum proses clustering. Penentuan jumlah kluster optimal menggunakan Elbow Method dan Silhouette Coefficient menunjukkan bahwa tiga kluster merupakan struktur terbaik. Hasil pengelompokan menghasilkan tiga kategori siswa, yaitu sangat disiplin, cukup disiplin, dan kurang disiplin. Evaluasi kualitas kluster menggunakan Silhouette Score dan Davies-Bouldin Index menunjukkan pemisahan kluster yang baik. Penelitian ini membuktikan bahwa K-Means Clustering efektif dalam mengidentifikasi pola kehadiran siswa dan dapat mendukung pengambilan keputusan sekolah berbasis data dalam meningkatkan kedisiplinan siswa.

**Kata Kunci:** *K-Means Clustering, Kehadiran Siswa, Data Mining, Educational Data Mining, Pengelompokan.*

**1. PENDAHULUAN**

Kehadiran siswa merupakan indikator penting dalam menentukan kualitas proses pembelajaran karena mencerminkan tingkat partisipasi, tanggung jawab, dan kedisiplinan siswa. Ketidakhadiran yang berulang seringkali berkaitan dengan penurunan prestasi akademik, kurangnya motivasi belajar, serta meningkatnya risiko perilaku menyimpang. Dalam konteks pendidikan kejuruan seperti SMK, kehadiran memiliki peran yang lebih strategis mengingat pembelajaran banyak berfokus pada praktik yang menuntut keterlibatan langsung siswa. Namun, pencatatan kehadiran di SMK Al-Ma'rifah masih

dilakukan secara manual, sehingga rawan terjadi ketidakakuratan data, keterlambatan pelaporan, dan kesulitan dalam melakukan analisis untuk mendeteksi pola ketidakhadiran siswa secara cepat dan tepat.

Perkembangan teknologi informasi menyediakan peluang bagi sekolah untuk mengolah data kehadiran secara lebih efektif melalui pendekatan analitik. Salah satu metode yang relevan adalah K-Means Clustering, yaitu teknik data mining yang mampu mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label awal. Metode ini dapat membantu sekolah mengidentifikasi kelompok siswa berdasarkan tingkat kedisiplinannya, seperti

siswa sangat disiplin, cukup disiplin, dan kurang disiplin. Penerapan algoritma K-Means pada data kehadiran diharapkan mampu memberikan gambaran yang lebih objektif, akurat, serta mendukung proses pengambilan keputusan dalam pembinaan, evaluasi, maupun pemberian reward kepada siswa. Dengan demikian, penelitian ini berfokus pada pemanfaatan K-Means Clustering sebagai dasar analisis kehadiran untuk meningkatkan efektivitas manajemen kedisiplinan di SMK Al-Ma'rifah.

## 2. LANDASAN TEORI

### 2.1 Data Mining

Data Mining merupakan proses untuk menggali pola, hubungan, serta informasi penting dari kumpulan data berukuran besar. Tujuan utama data mining adalah mengekstraksi pengetahuan yang berguna untuk pengambilan keputusan [1]. Dalam konteks pendidikan, data mining mampu membantu institusi memahami perilaku dan karakteristik siswa melalui analisis data historis.

### 2.2 Educational Data Mining (EDM)

Educational Data Mining adalah penerapan teknik data mining pada lingkungan pendidikan untuk menganalisis proses belajar, kehadiran, maupun performa siswa [2]. EDM memungkinkan institusi untuk melakukan prediksi, deteksi masalah kedisiplinan, serta pengelompokan siswa berdasarkan karakteristik tertentu.

### 2.3. Clustering

Clustering adalah metode *unsupervised learning* yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik antar objek [3]. Data yang berada dalam satu cluster memiliki kemiripan yang tinggi, sedangkan data antar cluster memiliki tingkat perbedaan yang jelas.

### 2.4. Algoritma K-Means Clustering

K-Means adalah algoritma clustering yang paling banyak digunakan karena kesederhanaan dan kecepatannya. Algoritma ini bekerja dengan menentukan jumlah cluster ( $k$ ) terlebih dahulu, kemudian menghitung jarak setiap data ke centroid menggunakan rumus Euclidean Distance [4]. Proses berlanjut melalui pengelompokan ulang dan pembaruan centroid hingga konvergen.

### 2.5. Penentuan Nilai $k$ Optimal

Pemilihan nilai  $k$  yang tepat sangat memengaruhi kualitas cluster. Dua metode umum yang digunakan adalah:

1. Elbow Method, yaitu menentukan titik siku pada grafik Within Cluster Sum of Square (WCSS) sebagai indikator stabilitas cluster [5].
2. Silhouette Score, yaitu mengukur kualitas cluster berdasarkan tingkat kesamaan objek dengan cluster-nya dibandingkan dengan cluster lain. Nilai mendekati 1 menunjukkan kualitas cluster yang baik [6].

### 2.6. Evaluasi Cluster

Evaluasi dilakukan untuk mengetahui apakah cluster yang terbentuk valid dan representatif. Selain Silhouette Score, Davies-Bouldin Index juga digunakan untuk menilai tingkat pemisahan dan kedekatan antar cluster [7]. Semakin rendah nilai indeks, semakin baik struktur cluster yang terbentuk.

### 2.7. Kedisiplinan Siswa dan Persentase Kehadiran

Kedisiplinan siswa merupakan salah satu indikator penting dalam dunia pendidikan yang dapat dinilai melalui kehadiran, izin, sakit, dan alfa [8]. Persentase kehadiran menjadi salah satu ukuran kuantitatif yang dapat dianalisis menggunakan teknik data mining untuk mengidentifikasi pola kedisiplinan siswa dalam jangka panjang.

## 3. METODOLOGI PENELITIAN

### 3.1 Pengumpulan Data

Data penelitian ini diperoleh dari catatan kehadiran siswa SMK Al-Ma'rifah selama satu semester, yang terdiri dari 200 entri siswa. Data dikumpulkan langsung dari sistem administrasi sekolah yang mencatat kehadiran harian siswa pada setiap pertemuan. Variabel utama yang dikumpulkan meliputi: jumlah kehadiran (HADIR), izin (IZIN), sakit (SAKIT), alfa (ALFA), total pertemuan, dan persentase kehadiran (PERSENTASE\_KEHADIRAN).

Seluruh data dicatat dalam format file CSV dan kemudian diimpor ke dalam Python untuk analisis lebih lanjut menggunakan library pandas. Proses pengumpulan data dilakukan dengan memastikan bahwa setiap entri mewakili satu siswa secara akurat dan mencerminkan keseluruhan aktivitas kehadiran selama periode observasi. Pada tahap awal, dataset berisi 200 entri siswa dari SMK Al-Ma'rifah dimuat menggunakan Python dengan library pandas. Dataset tersebut memiliki struktur kolom yang mencakup variabel terkait kehadiran, yaitu jumlah hadir, izin, sakit, alfa, persentase kehadiran, serta total pertemuan. Berdasarkan proses inspeksi awal, tidak ditemukan missing values yang signifikan, sehingga proses pembersihan hanya dilakukan pada baris dengan

nilai kosong seluruhnya (all-null rows). Langkah ini memastikan bahwa dataset yang digunakan bersih dan siap untuk dianalisis lebih lanjut.

Selain itu, dilakukan pengecekan tipe data untuk memastikan bahwa seluruh kolom numerik sudah berbentuk tipe numerik sehingga dapat digunakan sebagai input fitur pada algoritma K-Means. Kolom dengan format string atau numeric-as-string berhasil dikonversi menjadi format numerik untuk menjaga konsistensi pemrosesan .

Gambar 3.1 Pengumpulan Data :

NIS	NAMA	KELAS	HADIR	IZIN	SAKIT	ALFA	TOTAL_PERTUNJAN	PERSENTASE_KHIDHIRAN	KEDISIPLINAN
20251122.0	m mubah rafidatul akbar salsabih	TKJ 2	1250	2.0	5.0	8.0	130.0	99.23	Sangat Disiplin
20251180.0	rubiana nurrahmah	RPL 2	1030	7.0	5.0	3.0	130.0	79.23	Cukup Disiplin
20251000.0	ABDUR RAHMAN AL FATMAH	RPL 1	910	1.0	4.0	14.0	130.0	70.00	Kurang Disiplin
20251001.0	ADIL PUTRA PRATAMA	TKJ 2	1010	2.0	9.0	7.0	130.0	77.69	Cukup Disiplin
20251002.0	ALHITVA PRASE-TYO	TKJ 1	1010	0.0	8.0	4.0	130.0	77.69	Cukup Disiplin
20251153.0	Agrya agra satrio	TKJ 2	950	0.0	6.0	6.0	130.0	73.08	Kurang Disiplin
20251003.0	AHMAD BHAL SAFULLOH	TKJ 1	1250	0.0	5.0	3.0	130.0	96.15	Sangat Disiplin
20251004.0	AHMAD FANIR QUSYAIRI	TKJ 1	1090	2.0	0.0	7.0	130.0	83.85	Cukup Disiplin
20251005.0	AHMAD KAFABH ANUL YAZEN	RPL 1	1170	4.0	0.0	14.0	130.0	90.00	Disiplin
20251006.0	ASHLAZMI QUEENA	TKJ 2	1170	2.0	1.0	5.0	130.0	90.00	Disiplin

Gambar 3.1. Pengumpulan Data

### 3.2 Identifikasi Masalah

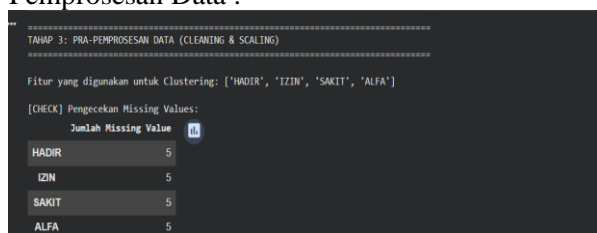
Permasalahan utama yang diangkat dalam penelitian ini adalah pola ketidakhadiran siswa di SMK Al-Ma'rifah. Dari data absensi 200 siswa, terdapat berbagai kategori kehadiran seperti Hadir, Izin, Sakit, dan Alfa. Pola ketidakhadiran yang tidak teratur berpotensi berdampak pada prestasi belajar, kedisiplinan, serta proses pembinaan siswa di sekolah pada Gambar 3.2 :



Gambar 3.2 Identifikasi Masalah

### 3.3 Pra Pemrosesan Data

Proses preprocessing data yang meliputi penanganan nilai hilang (missing values) dan normalisasi data menggunakan StandardScaler. Dua tahapan ini merupakan bagian penting dalam metodologi penelitian berbasis data, khususnya ketika menggunakan algoritma machine learning atau analisis statistik yang sensitif terhadap perbedaan skala data. Gambar 3.3. Pra Pemrosesan Data :

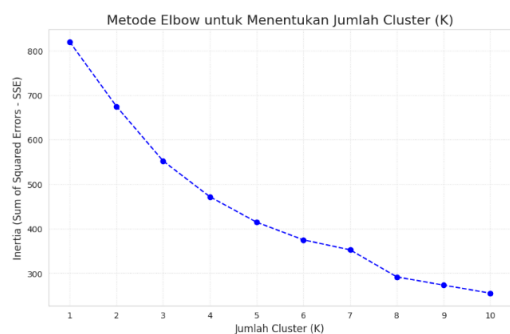


Gambar 3.3. Pra Pemrosesan Data

Gambar diatas menunjukkan dua tahapan penting dalam preprocessing data, yaitu penanganan missing values dan normalisasi data menggunakan StandardScaler.

### 3.4 Penentuan Nilai K Optimal

Elbow Method adalah teknik untuk menentukan jumlah cluster terbaik dalam algoritma clustering (seperti K-Means) dengan menganalisis grafik hubungan antara jumlah cluster k dan nilai Within-Cluster Sum of Squares (WCSS). Titik “siku” (elbow) pada grafik menunjukkan nilai k paling optimal, yaitu ketika penurunan WCSS mulai melambat signifikan. Grafik Elbow menunjukkan nilai SSE (Sum of Squared Error) untuk berbagai jumlah klaster pada gambar 3.4 :

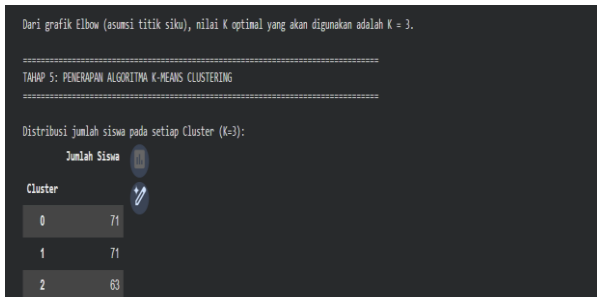


Gambar 3.4 Penentuan nilai k optimal

Pada grafik terlihat bahwa nilai SSE menurun tajam dari K = 1 hingga K = 3. Penurunan yang signifikan ini mengindikasikan bahwa penambahan cluster pada rentang tersebut memberikan peningkatan akurasi segmentasi yang bermakna. Setelah mencapai K = 3, penurunan SSE mulai melambat dan cenderung konstan, sehingga penambahan cluster berikutnya hanya memberikan perbaikan kecil yang tidak lagi signifikan.

### 3.5 Penerapan Algoritma K-Means

Berdasarkan hasil analisis menggunakan metode Elbow, diperoleh bahwa nilai K optimal adalah 3, yang menunjukkan bahwa kumpulan data siswa dapat dipetakan secara efektif ke dalam tiga kelompok (cluster) yang memiliki karakteristik berbeda. Pada tahap penerapan algoritma K-Means Clustering, proses pengelompokan dilakukan dengan mengukur kedekatan setiap data siswa terhadap pusat cluster (centroid) sehingga masing-masing data dapat dikelompokkan pada cluster yang paling representatif. Gambar 3.5 Cluster Jumlah siswa :



Gambar 3. 5 Cluster Jumlah Siswa:

Berdasarkan hasil gambar 3.5 Hasil proses pengelompokan menggunakan algoritma K-Means menghasilkan tiga cluster dengan jumlah anggota yang berbeda :

Cluster 0 terdiri dari 71 siswa, yang menunjukkan bahwa kelompok ini merupakan salah satu klaster dengan jumlah anggota paling besar. Siswa-siswa dalam cluster ini umumnya memiliki pola kehadiran yang baik dengan nilai alfa rendah, sehingga dapat dikategorikan sebagai kelompok dengan kedisiplinan stabil dan perilaku hadir yang konsisten.

Cluster 1, yang juga beranggotakan 71 siswa, memiliki jumlah yang sama dengan Cluster 0. Meskipun jumlahnya setara, karakteristiknya berbeda. Siswa dalam cluster ini cenderung memiliki tingkat alfa yang tinggi meskipun frekuensi hadir mereka juga tinggi. Hal ini menandakan pola kedisiplinan yang tidak konsisten, di mana siswa sering hadir tetapi juga sering tidak hadir tanpa keterangan. Cluster ini menjadi kelompok yang memerlukan perhatian khusus dari pihak sekolah.

Sementara itu, Cluster 2 berisikan 63 siswa, jumlah yang sedikit lebih rendah dibanding dua cluster lainnya. Karakteristik siswa dalam cluster ini berada pada kategori kedisiplinan menengah, di mana nilai izin dan sakit lebih tinggi, tetapi tingkat alfa tetap rendah. Artinya, ketidakhadiran siswa pada kelompok ini umumnya disertai alasan yang jelas. Meskipun tidak termasuk kategori berisiko tinggi, kelompok ini tetap memerlukan pemantauan agar ketidakhadiran tidak meningkat.

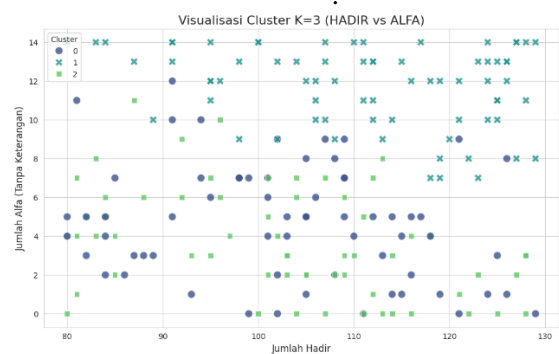
Distribusi yang relatif seimbang tersebut mengindikasikan bahwa karakteristik data siswa tidak terkonsentrasi hanya pada satu kelompok tertentu. Dengan kata lain, pola perilaku atau atribut siswa yang digunakan sebagai variabel clustering memiliki variasi yang cukup terdistribusi secara merata. Hal ini menunjukkan bahwa model K-Means bekerja dengan baik dalam memetakan perbedaan karakteristik antar siswa tanpa menghasilkan cluster yang timpang (imbalanced).

Secara metodologis, hasil ini memberikan dasar analitis untuk melakukan interpretasi lebih lanjut terhadap karakteristik masing-masing cluster—misalnya tingkat prestasi, partisipasi, minat, atau variabel lain yang dimasukkan ke dalam proses clustering. Dengan mengidentifikasi pola tersebut, institusi pendidikan dapat merumuskan strategi pembelajaran, intervensi akademik, atau kebijakan khusus yang lebih tepat sasaran berdasarkan profil kelompok siswa.

Penjelasan Ilmiah Visualisasi Cluster K = 3 (Hadir vs Alfa).

### 3.6 Visualisasi Cluster K

Hasil visualisasi pengelompokan data siswa menggunakan algoritma K-Means Clustering dengan jumlah cluster optimal K = 3. Visualisasi dilakukan menggunakan dua variabel utama, yaitu jumlah Hadir dan jumlah Alfa (tanpa keterangan), untuk memahami pola kedisiplinan siswa berdasarkan tingkat kehadiran dan ketidakhadiran mereka. Gambar 3.6



Gambar 3.6 Visualisasi Cluster K = 3 (Hadir vs Alfa)

Setiap cluster direpresentasikan oleh simbol dan warna yang berbeda, sehingga pola distribusi antar kelompok siswa dapat dianalisis secara lebih jelas.

### 3.7 Evaluasi

Gambar dibawah menunjukkan hasil evaluasi kualitas model clustering K-Means dengan jumlah cluster K = 3 menggunakan dua metrik validasi yang umum digunakan, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Kedua metrik ini dipakai untuk menilai seberapa baik data terbagi ke dalam cluster yang terbentuk. Gambar 3.7 Silhouette Score dan Davis Bouldin Index :

```

=====
TAHAP 6: EVALUASI HASIL MENGGUNAKAN SILHOUETTE SCORE DAN DAVIES-BOULDIN INDEX
=====

Hasil Evaluasi Clustering (K=3):

  Metrik      Nilai      Interpretasi
  -----
  0 Silhouette Score  0.187583  Semakin mendekati +1, semakin baik (cluster pa...
  1 Davies-Bouldin Index  1.594605  Semakin mendekati 0, semakin baik (cluster ter...
  
```

Gambar 3.7 Evaluasi

### 3.8 Interpretasi

Tahap ini merupakan proses penting dalam analisis clustering, karena informasi numerik dari model K-Means diterjemahkan menjadi makna dan karakteristik nyata dari setiap cluster. Data yang ditampilkan menunjukkan rata-rata nilai fitur per cluster, yaitu jumlah HADIR, IZIN, SAKIT, dan ALFA. Nilai rata-rata ini menjadi dasar untuk memahami profil perilaku kehadiran siswa pada setiap kelompok. Gambar 3.8 :

```

=====
TAHAP 7: INTERPRETASI HASIL DAN PENGELOMPOKAN SISWA
=====

Profil Rata-rata Fitur per Cluster:
(https://python-input-219388115.py253: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to reset cluster_profile = df.groupby('Cluster')['feature_columns'].mean().reset_index()

  Cluster  HADIR      IZIN      SAKIT      ALFA
  -----
  1  111.00309  4.661972  4.788732  11.748310
  0  103.98967  2.049455  2.803000  4.653515
  2  102.42888  6.657140  5.714286  3.746032
  
```

### 3.9 Penyusunan Rekomendasi Sistem Reward dan Intervensi Kedisiplinan

Tahap ini merupakan bagian akhir dari analisis clustering yang berfungsi untuk menghubungkan hasil pengelompokan siswa dengan strategi pengambilan keputusan. Dengan kata lain, hasil clustering tidak hanya dianalisis, tetapi juga digunakan sebagai dasar untuk merancang rekomendasi kebijakan sekolah terkait kedisiplinan dan penguatan perilaku positif.

Rekomendasi dibuat berdasarkan profil rata-rata fitur tiap cluster (HADIR, IZIN, SAKIT, ALFA) yang telah dianalisis sebelumnya. Gambar 3.9 Penyusunan Rekomendasi Sistem Reward dan Intervensi Kedisiplinan. Gambar 3. 9 :

```

=====
TAHAP 8: PENYUSUNAN REKOMENDASI SISTEM REWARD DAN INTERVENSI KEDISIPLINAN
=====

Berdasarkan profil Cluster yang telah diinterpretasikan, berikut adalah kerangka rekomendasi:

**A. Sistem Reward (untuk Kelompok Berprestasi/Disiplin)**
1. Kriteria: Siswa di Cluster yang memiliki rata-rata HADIR tinggi dan ALFA rendah.
2. Bentuk: Sertifikat 'Siswa Teladan Kehadiran' bulanan, poin tambahan pada nilai kuartil, atau pengakuan publik di sekolah.
3. Tujuan: Meningkatkan perilaku positif dan memberikan penguatan atas kedisiplinan.

**B. Sistem Intervensi Kedisiplinan (untuk Kelompok Perlu Perhatian/Bermasalah/Krisis)**
1. Intervensi: Rincang (Kelompok Perlu Perhatian)
   - Kriteria: Siswa di Cluster dengan tingkat ALFA tinggi dan IZIN rendah.
   - Mekanisme: Pertemuan bertahap pertama, sesi konseling singkat oleh Wali Kelas/Guru BK, dan pembentukan program mentor sebaya.
2. Intervensi Berat (Kelompok Bermasalah/Kritis):
   - Kriteria: Siswa di Cluster yang memiliki rata-rata ALFA tertinggi dan IZIN terendah.
   - Mekanisme: Pengalihan orang tua, penandatanganan surat perjanjian disiplin, program bimbingan konseling intensif (individual), dan sanksi sesuai tata tertib sekolah.
3. Metrik Monitoring: Memantau perubahan siswa dari cluster Intervensi Berat ke Cluster Intervensi Ringan atau Cluster Reward dari periode ke periode.

=====
ANALISIS SELESAI
=====

Semua tahapan telah dilaksanakan dan menghasilkan pengelompokan siswa untuk dasar pengambilan keputusan.
  
```

## 4. HASIL DAN PEMBAHASAN

Hasil clustering menggunakan algoritma K-Means pada data kehadiran siswa menunjukkan

terbentuknya tiga kelompok utama yang merepresentasikan tingkat kedisiplinan berbeda, yaitu disiplin tinggi (Cluster 0), disiplin sedang (Cluster 2), dan disiplin rendah (Cluster 1). Cluster 0 terdiri dari siswa dengan tingkat kehadiran tinggi dan alfa sangat rendah, menggambarkan kedisiplinan yang baik dan komitmen kuat terhadap proses pembelajaran. Sementara Cluster 2 menunjukkan pola kehadiran moderat, dengan nilai izin dan sakit lebih tinggi namun tetap memiliki komponen alfa rendah sehingga masih mencerminkan kedisiplinan yang cukup. Adapun Cluster 1 memperlihatkan pola yang kurang stabil, ditandai dengan tingginya alfa meskipun hadir cukup sering, menunjukkan ketidakkonsistenan yang mengarah pada masalah kedisiplinan dan motivasi belajar. Hasil ini diperkuat dengan nilai Silhouette Score dan Davies-Bouldin Index yang menunjukkan bahwa pembentukan cluster stabil dan representatif terhadap perilaku kehadiran siswa secara keseluruhan.

Pembahasan lebih lanjut menunjukkan bahwa pola kehadiran berperan penting sebagai indikator kedisiplinan siswa dan berhubungan erat dengan teori pendidikan serta psikologi belajar. Siswa pada cluster disiplin tinggi menunjukkan karakteristik self-regulated learner, yaitu mampu mengelola waktu dan tanggung jawab belajar dengan baik. Sebaliknya, tingginya nilai alfa pada Cluster 1 selaras dengan konsep behavioral disengagement, di mana siswa hadir secara fisik namun tidak terlibat secara penuh dalam proses belajar, sehingga memerlukan intervensi konseling atau bimbingan. Cluster 2 mencerminkan pengaruh faktor situasional seperti kondisi kesehatan maupun dukungan keluarga, sebagaimana dijelaskan dalam teori ekologi perkembangan. Secara keseluruhan, hasil dan pembahasan ini menegaskan bahwa algoritma K-Means tidak hanya efektif dalam memetakan kedisiplinan siswa, tetapi juga memberikan dasar penting bagi sekolah untuk mengambil keputusan berbasis data, termasuk strategi pembinaan pengawasan, dan penguatan kedisiplinan secara lebih terarah.

## 5. PENUTUP

Penelitian ini menyimpulkan bahwa algoritma K-Means Clustering mampu mengelompokkan siswa SMK Al-Ma'rifah berdasarkan pola kehadiran secara akurat ke dalam tiga kategori, yaitu sangat disiplin, cukup disiplin, dan kurang disiplin. Proses penentuan nilai  $k$  optimal melalui Elbow Method dan Silhouette Score menghasilkan klaster yang valid dengan pemisahan yang jelas, didukung oleh nilai evaluasi yang menunjukkan kualitas pengelompokan yang baik. Temuan ini membuktikan bahwa analisis

berbasis data dapat mengungkap pola kedisiplinan siswa secara lebih objektif dibandingkan metode manual, sekaligus memberikan dasar yang kuat bagi sekolah untuk menyusun strategi pembinaan dan pemberian reward yang lebih tepat sasaran.

#### DAFTAR PUSTAKA

- [1] Ahmad, Y., & Fatimah, T. (2024). Implementasi Teknik Clustering Menggunakan Algoritma K-Means Pada Smk Yadika 3 Jakarta Berbasis Web Implementation of Clustering Techniques Using the K-Means Algorithm in Web-Based Smk Yadika 3 Jakarta, *3*(2), 490–498.
- [2] Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295. <https://doi.org/10.3390/electronics9081295>
- [3] Al-Thaqafi, T., Saleem, F., & Al-Malaise Al-Ghamdi, A. (2025). Enhancing student performance prediction: The role of class imbalance handling in machine learning models. *Research and Practice in Technology Enhanced Learning*, *28*, 79. <https://doi.org/10.1007/s10791-025-09576-4>
- [4] Al Hazaa, K., Abdel-Salam, A. G., Ismail, R., Johnson, C., Al-Tameemi, R. A. N., Romanowski, M. H., ... Elatawneh, A. (2021). The effects of attendance and high school GPA on student performance in first-year undergraduate courses. *Cogent Education*, *8*(1). <https://doi.org/10.1080/2331186X.2021.1956857>
- [5] Albreiki, B., Habuza, T., & Zaki, N. (2022). Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models. *International Journal of Educational Technology in Higher Education*, *19*(49). <https://doi.org/10.1186/s41239-022-00354-6>
- [6] Alghamdi, A., Alhussan, A., Alsubaie, S., & Alotaibi, M. (2023). A data-driven student attendance monitoring and prediction system using deep clustering and analytics. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2023.08.001>
- [7] Ali, & al., et. (2023). Measuring Cluster Quality with the Silhouette Coefficient.
- [8] Aljohani, A. (2024). Optimizing Patient Stratification in Healthcare: A Comparative Analysis of Clustering Algorithms for EHR Data. *International Journal of Computational Intelligence Systems*, *17*, 173. <https://doi.org/10.1007/s44196-024-00568-8>
- [9] Aljohani, & al., et. (2024). Davies–Bouldin Index for Clustering Evaluation.
- [10] Arif Tri Biso, A. Z. (2025). Analisis Sentimen Game Genshin Impact untuk Mengetahui Reaksi dan Harapan Pemain Menggunakan Metode Naïve Bayes.
- [11] Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of an emerging field. *IEEE Access*, *10*, 31289–31296. <https://doi.org/10.1109/ACCESS.2022.3160457>
- [12] Bowen, F., Nouri, J., & Lorenz, C. (2022). Revealing underlying factors of absenteeism: A machine learning approach. *Frontiers in Education*, *7*, 958748. <https://doi.org/10.3389/educ.2022.958748>
- [13] Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*(9), 1679. <https://doi.org/10.3390/sym15091679>