

Penerapan K-Means Clustering untuk Pemetaan Pola Pelanggaran Kehadiran Siswa SMK Hidayatus Sholihin

Muhamad Abdul Aziz[1*], Widodo[2], Saifulloh Azhar[3]
Program Studi Informatika, Institut Teknologi Al Mahrusiyah, Kota Kediri, Indonesia
Email: ajizusongep@gmail.com[1], widodoido7@gmail.com[2], azharnian@gmail.com[3]

Abstrak

Data absensi *online* berpotensi mendukung pengambilan keputusan di sekolah, namun evaluasi konvensional umumnya sebatas rekapitulasi persentase kumulatif yang belum mampu membedakan pola spesifik pelanggaran kesiswaan. Penelitian ini bertujuan menganalisis pola pelanggaran kehadiran siswa SMK Hidayatus Sholihin berdasarkan data perilaku indisipliner berupa keterlambatan dan ketidakhadiran tanpa keterangan (alpa). Penelitian menerapkan kerangka *Knowledge Discovery in Databases* (KDD) terhadap data 63 siswa kelas X pada semester ganjil tahun 2025. Proses pengolahan data dilakukan menggunakan *Altair AI Studio* dengan algoritma K-Means *Clustering*, sedangkan penentuan jumlah klaster menggunakan *Elbow Method* dan validasi dilakukan melalui Silhouette Score dengan skor sebesar 0,502. Hasil penelitian menunjukkan bahwa data dapat dipetakan secara stabil ke dalam tiga klaster, yaitu *Cluster* 0 sebanyak 16 siswa dengan pola "Membolos Akut", *Cluster* 1 sebanyak 12 siswa dengan pola "Rentan Terlambat Sedang", dan *Cluster* 2 sebanyak 35 siswa dengan pola "Pelanggaran Keterlambatan Akut". Hasil klasterisasi tersebut menjadi dasar penyusunan rekomendasi strategi intervensi yang lebih terarah untuk mendukung penegakan tata tertib dan pembinaan kedisiplinan siswa.

Kata Kunci: Absensi Online; Altair AI Studio; Kehadiran Siswa; K-Means Clustering

ABSTRACT

Online attendance data has the potential to support decision-making in schools; however, conventional evaluations are generally limited to summarizing cumulative percentages, which are unable to distinguish specific patterns of student misconduct. This study aims to analyze patterns of attendance violations among students at SMK Hidayatus Sholihin based on data regarding disciplinary behavior, specifically tardiness and unexcused absences. The study applied the Knowledge Discovery in Databases (KDD) framework to data from 63 10th-grade students during the odd-numbered semester of 2025. Data processing was conducted using Altair AI Studio with the K-Means Clustering algorithm, while the number of clusters was determined using the Elbow Method, and validation was performed via the Silhouette Score, which yielded a score of 0.502. The results showed that the data could be reliably mapped into three clusters: Cluster 0, consisting of 16 students exhibiting a pattern of "Acute Truancy"; Cluster 1, consisting of 12 students exhibiting a pattern of "Moderate Tendency to Be Late"; and Cluster 2, consisting of 35 students exhibiting a pattern of "Acute Tardiness Violations." These clustering results serve as the basis for formulating more targeted intervention strategy recommendations to support the enforcement of school rules and the fostering of student discipline.

Keywords: *Online Attendance; Altair AI Studio; Student Attendance; K-Means Clustering*

1. PENDAHULUAN

Kehadiran siswa di Sekolah Menengah Kejuruan (SMK) menjadi salah satu aspek yang penting dalam mendukung proses pembelajaran. Kehadiran yang teratur dan tepat waktu tidak hanya berkaitan dengan kewajiban akademik, tetapi juga berhubungan dengan pembentukan sikap disiplin dan tanggung jawab yang diperlukan ketika siswa memasuki dunia kerja. Hal ini relevan bagi pendidikan kejuruan yang berorientasi pada kesiapan lulusan menghadapi kebutuhan industri. Karena itu, keterlambatan yang terjadi berulang kali maupun ketidakhadiran tanpa keterangan (alpa) perlu diperhatikan sebagai bagian dari evaluasi kedisiplinan siswa, terutama karena dapat mengurangi keterlibatan mereka dalam pembelajaran dan kegiatan praktikum (Alfiansyah et al., 2025).

Pemanfaatan teknologi informasi dalam pengelolaan data kehadiran memberikan peluang bagi sekolah untuk melakukan pencatatan dan pemantauan secara lebih terstruktur. Salah satunya melalui penggunaan absensi *online* yang dapat membantu sekolah mencatat, memantau, dan mengevaluasi kehadiran siswa. SMK Hidayatus Sholihin telah menggunakan sistem absensi *online* sebagai bagian dari digitalisasi layanan akademik sekaligus upaya mendukung kedisiplinan siswa. Meskipun demikian, penggunaan sistem digital tidak secara otomatis menyelesaikan persoalan kehadiran. Masih ditemukan siswa yang datang terlambat, tidak hadir tanpa keterangan, atau memiliki pola kehadiran yang tidak konsisten. Permasalahan utama yang menjadi practical gap terletak pada evaluasi kehadiran yang selama ini hanya dilakukan dengan melihat jumlah atau persentase ketidakhadiran secara akumulatif. Data yang dihasilkan dari sistem tersebut sebenarnya dapat memberikan informasi yang lebih luas apabila tidak hanya digunakan sebagai catatan administratif, tetapi juga dianalisis untuk melihat pola pelanggaran kehadiran siswa (Dewi et al., 2022).

Selama ini, evaluasi kehadiran sering dilakukan dengan melihat jumlah atau persentase ketidakhadiran. Pendekatan tersebut dapat memberikan gambaran umum, tetapi belum tentu mampu menunjukkan pola pelanggaran setiap siswa (Mawarni & Budi, 2022). Dua siswa dengan jumlah pelanggaran yang sama, misalnya, dapat memiliki karakteristik yang berbeda. Salah satunya mungkin lebih sering terlambat, sementara siswa lainnya lebih banyak tidak hadir tanpa keterangan. Perbedaan tersebut perlu diperhatikan karena bentuk pembinaan yang diberikan dapat disesuaikan dengan pola pelanggaran yang dimiliki. Tanpa pemetaan yang lebih rinci, informasi dalam data absensi cenderung berhenti pada angka dan belum sepenuhnya dapat digunakan untuk mendukung pengambilan keputusan.

Evaluasi kehadiran siswa pada penelitian-penelitian sebelumnya umumnya masih dilakukan menggunakan pendekatan konvensional, yaitu sebatas menghitung akumulasi jumlah atau persentase ketidakhadiran bulanan. Pendekatan konvensional tersebut memiliki keterbatasan mendasar karena hanya memberikan gambaran akumulatif secara umum, namun tidak mampu menunjukkan pola serta karakteristik pelanggaran spesifik dari setiap siswa. Dua orang siswa dengan total angka ketidakhadiran yang sama, misalnya, dapat memiliki karakteristik perilaku yang sangat berbeda—seorang siswa mungkin lebih sering datang terlambat, sementara siswa lainnya lebih banyak tidak hadir tanpa keterangan (alpa). Akibat dari keterbatasan evaluasi konvensional yang memukul rata angka tersebut, informasi dalam data absensi cenderung berhenti sebagai catatan administratif mentah belaka dan sebelumnya dapat dimaksimalkan untuk mendukung pengambilan keputusan pembinaan kesiswaan yang presisi.

Salah satu pendekatan yang dapat digunakan untuk memperoleh pemetaan tersebut adalah data mining melalui *unsupervised learning* dengan metode K-Means *Clustering*. Metode *clustering* mengelompokkan data yang belum memiliki label berdasarkan kemiripan karakteristiknya. K-Means dipilih karena memiliki proses yang relatif sederhana, efisien dalam pengolahan data, dan menghasilkan *centroid* yang dapat membantu menjelaskan karakteristik setiap kelompok (Purnama Sari & Hanif Batubara, 2021) . Dengan menerapkan metode ini pada data absensi, siswa dapat dikelompokkan berdasarkan pola keterlambatan dan ketidakhadiran tanpa bergantung pada penilaian subjektif. Hasil pengelompokan selanjutnya dapat menjadi informasi pendukung bagi sekolah dalam menentukan bentuk pembinaan dan mengidentifikasi siswa yang perlu mendapatkan perhatian lebih awal.

Penerapan K-Means *Clustering* pada data pendidikan dan kehadiran telah dilakukan dalam beberapa penelitian. Penggunaaan K-Means untuk mengelompokkan data absensi siswa SMP ke dalam kategori rajin dan tidak rajin sebagai bagian dari evaluasi pembelajaran (Raja et al., 2024). Fahrillah Fahrillah dan Zaehol Fatah (2025) menerapkan K-Means untuk memetakan kehadiran mahasiswa pada semester ganjil dan menunjukkan bahwa hasil pengelompokan dapat divisualisasikan untuk membantu interpretasi data. (Hendrastuty, 2024) menggunakan K-Means dalam evaluasi hasil pembelajaran dengan mengelompokkan prestasi siswa ke dalam dua klaster dan mengevaluasinya menggunakan *Silhouette Score*. (Elisa Br Tarigan & Megaria, 2024; Yaomulfurqqan & Solichin, 2023) juga menerapkan teknik *clustering* untuk mengidentifikasi pola belajar dan absensi siswa. Temuan dari penelitian-penelitian tersebut

menunjukkan bahwa K-Means dapat digunakan untuk menemukan pola pada data pendidikan yang dapat mendukung proses evaluasi dan pengambilan keputusan.

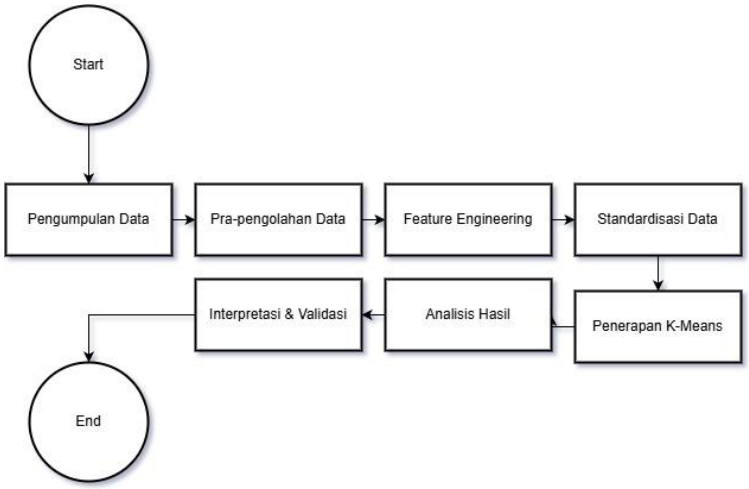
Berdasarkan kondisi tersebut, penelitian ini menganalisis pola kedisiplinan siswa di SMK Hidayatus Sholihin dengan menerapkan algoritma K-Means *Clustering* pada data absensi online. Pengelompokan dilakukan untuk memperoleh karakteristik kelompok siswa berdasarkan perilaku indisipliner dan selanjutnya menjadi dasar penyusunan rekomendasi intervensi yang sesuai dengan karakteristik setiap kelompok. Penelitian dibatasi pada data absensi *online* siswa kelas X dari kompetensi keahlian Desain Pemodelan dan Informasi Bangunan (DPB), Teknik Komputer dan Jaringan (TKJ), serta Teknik Otomotif (TO) pada semester ganjil tahun 2025, yaitu Juli hingga Desember 2025. Variabel yang digunakan terdiri atas dua indikator numerik, yaitu frekuensi keterlambatan (Terlambat) dan jumlah ketidakhadiran tanpa keterangan (Alpa). Pengolahan dan visualisasi data dilakukan menggunakan Altair AI Studio tanpa membandingkan hasilnya dengan algoritma *clustering* lainnya.

Secara teoretis, penelitian ini diharapkan dapat memperkaya kajian mengenai pemanfaatan *data-driven decision making* dalam pengelolaan data pendidikan, khususnya dalam analisis kedisiplinan siswa. Dari sisi praktis, hasil penelitian dapat digunakan sebagai informasi pendukung bagi guru Bimbingan dan Konseling (BK), wali kelas, dan manajemen sekolah untuk mengenali kelompok siswa yang membutuhkan perhatian serta menentukan bentuk pembinaan yang lebih sesuai dengan pola pelanggarannya.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik *data mining* melalui metode *unsupervised learning* menggunakan algoritma K-Means *Clustering* (Ghahramani, 2004). Algoritma tersebut digunakan untuk memetakan pola kedisiplinan siswa berdasarkan data kehadiran. Seperti pada Gambar 1 seluruh proses pengolahan data mengikuti kerangka kerja *Knowledge Discovery in Databases* (KDD), mulai dari pemilihan data hingga evaluasi dan interpretasi hasil klasterisasi (Waruwu, 2024).

Penelitian dilaksanakan di SMK Hidayatus Sholihin, Kediri, Jawa Timur. Pemilihan lokasi didasarkan pada telah diterapkannya sistem absensi *online* dalam pencatatan kehadiran siswa. Penelitian berlangsung selama empat bulan, yaitu November 2025 hingga Februari 2026.



Gambar 1. Alur Penelitian

2.1. Subjek Penelitian

Subjek penelitian terdiri atas 63 siswa kelas X dari tiga kompetensi keahlian, yaitu Desain Pemodelan dan Informasi Bangunan (DPB), Teknik Komputer dan Jaringan (TKJ), serta Teknik Otomotif (TO). Data yang dianalisis merupakan rekapitulasi perilaku kehadiran siswa selama semester ganjil tahun ajaran 2025, yaitu Juli hingga Desember 2025. Karena seluruh data siswa yang tersedia digunakan dalam analisis, penelitian ini menggunakan keseluruhan data sebagai sampel.

2.2. Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui tiga metode, yaitu:

a. Dokumentasi digital (*data extraction*)

Data mentah rekapitulasi presensi diekstrak (*export*) dari basis data pada sistem informasi absensi online sekolah. Sebelum dataset absensi online dimasukkan ke dalam algoritma, identifikasi terhadap arsitektur tipe data dari masing-masing variabel pelanggaran wajib dilakukan secara mendetail. Langkah ini bertujuan untuk memastikan kelayakan komputasi geometri jarak saat proses iterasi dijalankan di dalam perangkat lunak Altair AI Studio. Adapun spesifikasi teknis operasional dari kedua atribut utama perilaku indisipliner yang dianalisis dalam penelitian ini dijabarkan secara terstruktur pada Tabel 1.

Tabel 1. Matriks Spesifikasi Teknis Atribut Indisipliner Kesiswaan

No	Nama Atribut	Simbol	Tipe Data	Satuan	Batas Minimum (X_{min})	Batas Maksimal (X_{max})	Definisi Operasional Atribut
1	Terlambat	X_1	Integer	Kali	0	75	Akumulasi frekuensi ketidaktepatan waktu siswa tiba di sekolah melewati

No	Nama Atribut	Simbol	Tipe Data	Satuan	Batas Minimum (X_{min})	Batas Maksimal (X_{max})	Definisi Operasional Atribut
							batas toleransi pagi hari.
2	Alpa	X_2	Integer	Hari	0	129	Jumlah total ketidakhadiran siswa di dalam kelas tanpa adanya keterangan sah atau izin resmi.

b. Wawancara (interview)

Wawancara dilakukan dengan staf kesiswaan dan operator IT sekolah untuk memperoleh informasi mengenai penerapan tata tertib serta batas toleransi keterlambatan yang digunakan dalam pencatatan kehadiran siswa.

c. Studi pustaka (literature review)

Studi pustaka dilakukan dengan mengumpulkan referensi dari jurnal ilmiah nasional serta dokumentasi yang berkaitan dengan metode data *mining*, K-Means *Clustering*, dan perangkat lunak analisis data yang digunakan dalam penelitian.

Analisis dibatasi pada dua variabel numerik yang merepresentasikan perilaku indisipliner siswa, yaitu Terlambat (X_1) dan Alpa (X_2). Variabel Terlambat menunjukkan frekuensi keterlambatan masuk sekolah dalam satuan kali dengan rentang nilai 0–75. Sementara itu, variabel Alpa menunjukkan jumlah ketidakhadiran tanpa keterangan yang sah dalam satuan hari dengan rentang nilai 0–129.

Pengolahan data dilakukan menggunakan laptop HP Victus 15 dengan prosesor Intel Core i5-12500H dan RAM 16 GB yang menjalankan sistem operasi Windows 11. Microsoft Excel digunakan untuk penataan dan pemeriksaan awal data, sedangkan Altair AI Studio versi 2026.0 digunakan sebagai platform utama untuk proses pengolahan dan analisis K-Means *Clustering*.

2.3. Teknik Analisis Data

2.3.1 Pemilihan Data (Data Selection)

Tahap ini dilakukan dengan memilih data absensi dari 63 siswa kelas X dan menentukan atribut yang digunakan dalam analisis, yaitu Terlambat (X_1) dan Alpa (X_2). Kedua atribut tersebut dipilih karena secara langsung merepresentasikan perilaku indisipliner yang menjadi fokus penelitian. Pada Tabel 3.3 menunjukkan sampel data baris pertama berupa data terlambat 11, Alpa 118 dari total selama 6 bulan terakhir, pada periode Juli – Desember 2025.

Tabel 2. Sampel Dataset

Nama Siswa	Terlambat	Alpa
Achmad	11	118
Ridlo	6	123
Mikael	44	46
Ferdi	55	49
Galih	51	74
Iksan	59	61
Lutfi	45	81
Harun Al Rosid	26	46
Rangga	31	96
Nata	30	50

2.3.2 Pra-Pemrosesan Data (Data pre-processing/cleaning)

Pada tahap ini dilakukan pemeriksaan dan pembersihan data untuk memastikan validitas dan keutuhan dataset sebelum klasterisasi. Nilai kosong (missing values) pada atribut Terlambat dan Alpa dikonfirmasi dan diverifikasi secara operasional sebagai kondisi tidak adanya rekam pelanggaran yang dicatat oleh sistem absensi online selama periode Juli–Desember 2025, sehingga direpresentasikan secara objektif dengan nilai numerik 0. Selanjutnya data duplikat berdasarkan nomor induk siswa dihapus untuk memastikan setiap siswa hanya direpresentasikan satu kali. Selain itu, rekam presensi yang berada di luar kalender akademik, seperti hari libur nasional dan periode dispensasi, dieliminasi agar tidak memengaruhi hasil pengelompokan.

2.3.3 Transformasi Data (Data Transformation)

Karena kedua variabel memiliki rentang nilai yang berbeda, dilakukan normalisasi menggunakan metode *Min-Max Normalization* agar nilai setiap atribut berada pada rentang 0,0 hingga 1,0 (Marmay et al., 2026). Normalisasi diperlukan untuk mengurangi pengaruh perbedaan skala terhadap proses perhitungan jarak pada algoritma K-Means. Persamaan normalisasi dirumuskan pada Persamaan (1). Di mana (X') sebagai nilai hasil normalisasi, (X) sebagai nilai asli, sedangkan X_{min} dan X_{max} masing-masing menunjukkan nilai minimum dan maksimum pada atribut yang dinormalisasi.

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

2.3.4 Penggalian Data (Data Mining)

Proses pengelompokan data dilakukan menggunakan algoritma K-Means *Clustering* pada *Altair AI Studio*. Untuk mengatasi kelemahan mendasar K-Means terkait sensitivitas penentuan *centroid* awal (*random Initialization*). proses klasterisasi dikonfigurasi

menggunakan skema inisialisasi K-Means++ yang mengombinasikan pengujian berulang dengan nilai random seed terintegrasi (*multiple random runs*). Pendekatan ini menjamin ketahanan (robustness) model, sehingga koordinat centroid awal ditempatkan secara optimal berdasarkan sebaran probabilitas jarak antar-data. Penentuan jumlah klaster terlebih dahulu dilakukan menggunakan *Elbow Method* dengan mengamati perubahan nilai *Within Cluster Sum of Squares* (*WCSS*) (Sholeh & Aeni, 2023). Berdasarkan hasil pengujian, jumlah klaster yang digunakan adalah ($K = 3$).

Pada proses K-Means, *centroid* awal ditentukan oleh sistem, kemudian setiap data siswa ditempatkan ke dalam klaster berdasarkan jarak terdekat terhadap *centroid*. Jarak antar objek dihitung menggunakan *Euclidean Distance* dengan Persamaan(2):

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (2)$$

$D(x, y)$ sebagai jarak *Euclidean*, x_1 dan x_2 sebagai koordinat data siswa (Terlambat dan Alpa), serta y_1 dan y_2 sebagai koordinat *centroid*. Setelah setiap data ditempatkan pada klaster terdekat, posisi *centroid* dihitung kembali berdasarkan nilai rata-rata anggota pada masing-masing klaster. Proses tersebut dilakukan secara berulang hingga posisi *centroid* dan keanggotaan klaster mencapai kondisi konvergen.

2.3.5 Evaluasi dan Interpretasi (Evaluation and Interpretation)

Hasil klasterisasi dievaluasi menggunakan *Silhouette Score* sebagai metrik validasi internal untuk mengukur kualitas pemisahan dan kekompakan klaster. Setelah diperoleh struktur klaster yang sesuai, koordinat final *centroid* dianalisis untuk mengidentifikasi karakteristik masing-masing kelompok. Karakteristik tersebut kemudian digunakan untuk menyusun klasifikasi pola perilaku siswa dan menjadi dasar dalam merumuskan rekomendasi strategi intervensi yang sesuai bagi sekolah.

3. HASIL DAN PEMBAHASAN

Analisis pada penelitian ini dilakukan pada data absensi online 63 siswa kelas X SMK Hidayatus Sholihin selama periode Juli hingga Desember 2025. Data tersebut diolah menggunakan tahapan *Knowledge Discovery in Databases* (KDD) dengan bantuan Altair AI Studio versi 2026.0. Analisis difokuskan pada dua indikator kedisiplinan siswa, yaitu frekuensi keterlambatan dan jumlah ketidakhadiran tanpa keterangan.

3.1. Konfigurasi Proses pada Altair AI Studio

Sebelum proses *clustering* dilakukan, seluruh tahapan pengolahan data disusun pada Process Canvas di Altair AI Studio. Proses tersebut terdiri atas beberapa operator yang digunakan secara berurutan, mulai dari memasukkan dataset hingga melakukan pengelompokan menggunakan K-Means. Operator *Retrieve* digunakan untuk memasukkan dataset absensi ke dalam proses analisis. Atribut nama siswa dan nomor induk tetap disimpan sebagai informasi identitas, tetapi tidak digunakan dalam perhitungan *clustering*. Selanjutnya, operator *Select Attributes* digunakan untuk memilih dua variabel yang menjadi dasar analisis, yaitu Terlambat (X_1) dan Alpa (X_2).

Data yang telah dipilih kemudian dinormalisasi menggunakan operator *Normalize* dengan metode $Min - Max$. Normalisasi dilakukan agar kedua variabel memiliki skala yang sama, yaitu antara 0,0 dan 1,0. Setelah itu, data diproses menggunakan operator *Clustering* dengan algoritma K-Means, jumlah klaster $K = 3$, dan pengukuran jarak menggunakan *Euclidean Distance*.

3.2. Transformasi Data dengan Normalisasi $Min - Max$

Data awal menunjukkan adanya perbedaan rentang yang cukup besar antara kedua variabel. Nilai Terlambat berada pada rentang 0–75 kali, sedangkan nilai Alpa berada pada rentang 0–129 hari. Jika digunakan secara langsung, perbedaan skala tersebut dapat menyebabkan variabel Alpa memberikan pengaruh yang lebih besar dalam perhitungan jarak. Untuk mengatasi hal tersebut, kedua variabel dinormalisasi menggunakan metode $Min - Max$. Nilai minimum dari masing-masing variabel diubah menjadi 0,0, sedangkan nilai maksimum menjadi 1,0. Dengan demikian, Terlambat dan Alpa berada pada skala yang sama sebelum digunakan dalam proses K-Means. Persamaan normalisasi yang digunakan adalah pada Persamaan(3) berikut. X' merupakan nilai setelah normalisasi, X merupakan nilai asli, sedangkan X_{min} dan X_{max} masing-masing merupakan nilai minimum dan maksimum dari atribut yang digunakan.

$$X^1 = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

3.3. Proses Iterasi K-Means

Setelah data dinormalisasi, proses K-Means dilakukan untuk mengelompokkan 63 siswa ke dalam tiga klaster. Proses pengelompokan berlangsung melalui beberapa iterasi. Pada setiap iterasi, posisi *centroid* diperbarui berdasarkan anggota klaster yang terbentuk hingga tidak terjadi perubahan lagi pada hasil pengelompokan.

Tabel 3. Perubahan Koordinat *Centroid* pada Setiap Iterasi

Iterasi	*Cluster 0*	*Cluster 1*	*Cluster 2*	Kondisi
Iterasi 1	0,147 ; 0,880	0,347 ; 0,038	0,587 ; 0,098	Inisialisasi centroid
Iterasi 2	0,210 ; 0,810	0,335 ; 0,085	0,690 ; 0,150	Posisi centroid berubah
Iterasi 3	0,255 ; 0,775	0,345 ; 0,095	0,742 ; 0,185	Perubahan mulai mengecil
Iterasi 4	0,267 ; 0,760	0,349 ; 0,099	0,755 ; 0,199	Konvergen

Pada Tabel 3 menunjukkan bahwa iterasi pertama, posisi *centroid* masih merupakan titik awal yang digunakan oleh algoritma. Setelah data siswa ditempatkan berdasarkan jarak terdekat, posisi *centroid* berubah pada iterasi berikutnya. Perubahan tersebut terus mengecil hingga iterasi keempat, ketika nilai *centroid* tidak lagi berubah. Kondisi tersebut menunjukkan bahwa proses K-Means telah mencapai konvergensi dan pembagian anggota klaster telah stabil.

3.4. Hasil K-Means Clustering

Sebelum menentukan jumlah klaster, pengujian dilakukan menggunakan *Elbow Method* dengan melihat perubahan nilai *Within Cluster Sum of Squares* (WCSS). Hasil pengujian menunjukkan titik siku pada $K = 3$, sehingga tiga klaster digunakan dalam proses selanjutnya. Kualitas hasil pengelompokan kemudian dievaluasi menggunakan *Silhouette Score*. Hasil pengujian menghasilkan nilai Silhouette Score sebesar 0,5016 (atau 0,502), yang berada pada rentang nilai positif. Nilai ini mengonfirmasi secara kuantitatif bahwa struktur 3 klaster yang terbentuk memiliki tingkat kerapatan di dalam kelompok (*cohesion*) yang kuat serta keterpisahan antar-kelompok (*separation*) yang jelas dan stabil.

3.5. Analisis Final Centroid

Nilai final *centroid* digunakan untuk memahami karakteristik setiap klaster. Karena data telah dinormalisasi, nilai *centroid* berada pada rentang 0 sampai 1. Semakin tinggi nilai centroid pada suatu variabel, semakin tinggi pula kecenderungan nilai variabel tersebut pada anggota klaster.

Tabel 4. Nilai *Final Centroid* Setiap Klaster

Klaster	Terlambat	Alpa	Pola Kedisiplinan	Jumlah
Cluster 0	0,267	0,760	Membolos Akut	16 siswa
Cluster 1	0,349	0,099	Rentan Terlambat	12 siswa
Cluster 2	0,755	0,199	Pelanggaran Keterlambatan Akut	35 siswa

Pada Tabel 4 menampilkan nilai final centroid dari setiap klaster, berdasarkan nilai tersebut, Cluster 0 memiliki nilai Alpa paling tinggi, yaitu 0,760, sedangkan nilai Terlambat berada pada angka 0,267. Karakteristik ini menunjukkan bahwa kelompok tersebut lebih banyak ditandai oleh

ketidakhadiran tanpa keterangan dibandingkan keterlambatan. Sebanyak 16 siswa berada dalam kelompok ini sehingga pola yang terbentuk dapat dikategorikan sebagai "Membolos Akut".

Cluster 1 memiliki nilai Terlambat sebesar 0,349 dan Alpa sebesar 0,099. Nilai Alpa yang relatif rendah menunjukkan bahwa siswa dalam kelompok ini memiliki frekuensi ketidakhadiran tanpa keterangan yang lebih rendah dibandingkan dua kelompok lainnya. Namun, keterlambatan masih cukup terlihat. Dengan 12 siswa di dalamnya, kelompok ini dikategorikan sebagai "Rentan Terlambat".

Sementara itu, *Cluster 2* merupakan kelompok dengan jumlah anggota terbesar, yaitu 35 siswa. Nilai Terlambat pada klaster ini mencapai 0,755, sedangkan nilai Alpa hanya 0,199. Pola tersebut menunjukkan bahwa karakteristik utama kelompok ini adalah keterlambatan, bukan ketidakhadiran tanpa keterangan. Oleh karena itu, kelompok ini dikategorikan sebagai "Pelanggaran Keterlambatan Akut".

Hasil tersebut memperlihatkan bahwa jumlah pelanggaran saja belum cukup untuk menggambarkan karakteristik kedisiplinan siswa. Dua bentuk pelanggaran yang digunakan dalam penelitian dapat menghasilkan pola yang berbeda. *Cluster 0* lebih berkaitan dengan ketidakhadiran tanpa keterangan, sedangkan Cluster 2 lebih banyak menunjukkan masalah keterlambatan. Perbedaan tersebut menjadi informasi penting karena bentuk pembinaan dapat disesuaikan dengan karakteristik masing-masing kelompok.

3.6. Implikasi Hasil Klasterisasi

Hasil pengelompokan dapat dimanfaatkan sebagai informasi tambahan dalam proses pembinaan siswa. Data yang sebelumnya hanya menunjukkan jumlah keterlambatan dan ketidakhadiran dapat digunakan untuk melihat kelompok siswa dengan karakteristik pelanggaran yang berbeda.

Hasil klasterisasi yang diperoleh pada penelitian ini memberikan konfirmasi ilmiah sekaligus menawarkan kebaruan (*novelty*) jika dibandingkan dengan temuan pada penelitian-penelitian terdahulu. Secara umum, keberhasilan pemetaan 63 siswa ke dalam tiga kelompok kedisiplinan sejalan dengan temuan (Hendrastuty, 2024; Mawarni & Budi, 2022; Raja et al., 2024), yang menunjukkan bahwa algoritma K-Means Clustering sangat efektif untuk mengelompokkan data absensi dan perilaku siswa secara objektif tanpa bergantung pada penilaian subjektif. Penggunaan metrik Silhouette Score yang menghasilkan nilai 0,502 juga memperkuat hasil studi (Hendrastuty, 2024; Sholeh & Aeni, 2023) yang mengonfirmasi bahwa pengujian validitas internal mampu membuktikan kerapatan di dalam kelompok (cohesion) dan keterpisahan antarkelompok (separation) secara terukur.

Hasil pemisahan atribut ini mengungkap fakta baru di lapangan bahwa pola pelanggaran siswa ternyata bersifat asimetris. *Cluster 2* (Pelanggaran Keterlambatan Akut) mendominasi populasi dengan

35 siswa, sedangkan Cluster 0 (Membolos Akut) terdiri atas 16 siswa. Perbedaan pola ini menunjukkan bahwa faktor penyebab keterlambatan masuk sekolah di pagi hari bersifat kolektif dan dipengaruhi oleh kendala eksternal (seperti ketersediaan transportasi dan jarak tempuh), sementara perilaku membolos (alpa) lebih dipengaruhi oleh faktor internal dan motivasi individu siswa. Temuan komparatif ini mempertegas bahwa evaluasi kehadiran berbasis persentase kumulatif konvensional tidak lagi memadai, sehingga pembinaan kesiswaan oleh guru Bimbingan Konseling (BK) harus disesuaikan secara presisi dengan karakteristik spesifik dari masing-masing kelompok pelanggaran.

Dengan demikian, hasil clustering tidak hanya menunjukkan pembagian siswa ke dalam tiga kelompok, tetapi juga memberikan gambaran mengenai karakteristik pelanggaran pada masing-masing kelompok. Informasi tersebut dapat digunakan sebagai salah satu dasar dalam menentukan prioritas pembinaan, sehingga pendekatan yang diberikan tidak harus sama untuk seluruh siswa.

4. PENUTUP

Simpulan dan Saran

Penerapan algoritma K-Means *Clustering* dalam kerangka *Knowledge Discovery in Databases* (*KDD*) dengan bantuan *Altair AI Studio* berhasil memetakan data absensi 63 siswa kelas X SMK Hidayatus Sholihin ke dalam tiga kelompok pola pelanggaran kehadiran yang konvergen pada iterasi keempat. Penentuan jumlah klaster dilakukan menggunakan *Elbow Method* dan menghasilkan K=3, sedangkan kualitas pengelompokan dievaluasi menggunakan Silhouette Score dengan skor sebesar 0,502. Proses iterasi K-Means mencapai kondisi konvergen pada iterasi keempat. Hasil klasterisasi menunjukkan tiga pola kedisiplinan yang berbeda, yaitu Cluster 0 yang terdiri atas 16 siswa dengan nilai centroid Alpa sebesar 0,760 dan dikategorikan sebagai "Pola Membolos Akut", Cluster 1 yang terdiri atas 12 siswa dengan nilai Alpa sebesar 0,099 dan Terlambat sebesar 0,349 sehingga dikategorikan sebagai "Pola Kedisiplinan Sedang/Rentan Terlambat", serta *Cluster* 2 yang merupakan kelompok terbesar dengan 35 siswa dan memiliki nilai centroid Terlambat sebesar 0,755 sehingga dikategorikan sebagai "Pola Pelanggaran Keterlambatan Akut". Hasil tersebut menunjukkan bahwa data absensi dapat memberikan gambaran yang lebih spesifik mengenai karakteristik pelanggaran siswa dibandingkan hanya melihat jumlah ketidakhadiran secara keseluruhan. Pemetaan ini dapat menjadi informasi pendukung bagi sekolah dalam menentukan bentuk pembinaan sesuai dengan karakteristik masing-masing kelompok. *Cluster* 0 dapat menjadi prioritas dalam pembinaan terkait ketidakhadiran melalui koordinasi dengan orang tua dan konseling, *Cluster* 1 dapat diarahkan pada pembiasaan dan pengelolaan waktu untuk mengurangi keterlambatan,

sedangkan *Cluster* 2 dapat menjadi perhatian dalam pemantauan kedatangan dan pembinaan kedisiplinan waktu. Penelitian ini berhasil memberikan gambaran eksploratif mengenai variasi pola pelanggaran siswa yang tidak dapat ditangkap oleh evaluasi persentase akumulatif konvensional. Rekomendasi bentuk pembinaan yang dirumuskan seperti kunjungan rumah untuk *Cluster* 0 dan pengetatan pengawasan gerbang untuk *Cluster* 2 merupakan usulan konseptual berbasis karakteristik klaster yang berpotensi menjadi pertimbangan awal bagi Bimbingan Konseling (BK) dan manajemen sekolah. Keterbatasan utama penelitian ini terletak pada belum dilakukannya uji validasi empiris di lapangan untuk mengukur efektivitas nyata dari rekomendasi intervensi tersebut terhadap penurunan angka pelanggaran. Oleh karena itu, bagi pihak sekolah dan peneliti selanjutnya, disarankan untuk melakukan pengujian serta evaluasi dampak secara berkala terhadap implementasi strategi pembinaan tersebut di lapangan. Selain itu, pengembangan pemodelan klaster ini menjadi Sistem Pendukung Keputusan (DSS) yang terintegrasi penuh dalam sistem informasi sekolah dapat dijadikan sebagai prospek penelitian dan pengembangan (*research and development*) pada studi mendatang.

DAFTAR PUSTAKA

Alfiansyah, R., Fitriana, S., & Dian, M. A. P. (2025). Tingkat Kedisiplinan Belajar Siswa Kelas XI SMAN 1 Jakenan. *Jurnal Sarjana Ilmu Pendidikan*, *5*(2), 40–46. https://doi.org/10.26877/jsip.v5i2.23539

Dewi, F. P., Aryni, P. S., & Umaidah, Y. (2022). Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *7*(2), 111–121. https://doi.org/10.14421/jiska.2022.7.2.111-121

Elisa Br Tarigan, S., & Megaria, P. (2024). Penerapan Data Mining untuk Persediaan Barang pada Toko Cahaya Kasih dengan Metode K-Means. In *JDSP Jurnal Data Science Penusa JDSP 2024* (Vol. 1, Number 2). https://journal.pelitanusantara.ac.id/index.php/jdsppg.691

Fahrillah Fahrillah, & Zaehol Fatah. (2025). PENGELOMPOKAN DATA NILAI SISWA MADRASAH TA'HILIYAH MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Riset Sistem Informasi*, *2*(1), 53–59. https://doi.org/10.69714/0v1pkz05

Ghahramani, Z. (2004). *Unsupervised Learning* (pp. 72–112). https://doi.org/10.1007/978-3-540-28650-9_5

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Marmay, R., Luthvina, R., Ulandari, O., Studi, P., Logistik, M., Agro, I., Rekayasa, T., Energi, B., Politeknik, T., & Padang, A. (2026). Implementasi Algoritma K-Means dalam

Pengelompokan Data Harga Laptop. In *Journal Scientific of Mandalika (jsm) e-ISSN* (Vol. 7, Number 1).

Mawarni, Q. I., & Budi, E. S. (2022). Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *3*(4), 522. https://doi.org/10.30865/json.v3i4.4242

Purnama Sari, I., & Hanif Batubara, I. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering for Grouping Students' Abilities in Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)*, *2*(1), 139–144. https://doi.org/10.30596/jcositte.v2i1.6504

Raja, S., Sinurat, F., Nasution, M., Ah, R. M., Informasi, S., Sains, F., Teknologi, D., & Labuhanbatu, U. (2024). *Analisis Data Mining Absensi Siswa SMP Negeri 1 Bilah Barat Dengan Metode Algoritma K-Means Cluestering*. *12*(2).

Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, *8*(1), 56. https://doi.org/10.30998/string.v8i1.16388

Waruwu, M. (2024). Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan. *Jurnal Ilmiah Profesi Pendidikan*, *9*(2), 1220–1230. https://doi.org/10.29303/jipp.v9i2.2141

Yaomulfurqqan, A., & Solichin, A. (2023). *PENGELOMPOKAN NILAI SISWA DI SMPN 177 JAKARTA SELATAN MENGGUNAKAN ALGORITMA K-MEANS* (Vol. 2, Number 2).