JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 2, Mei 2026, pp. 674 - 686
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/171

Grouping of Student Attendance Discipline Levels Based on Daily Attendance Data Using the K-Means Algorithm

Syahwa Mutiara Putri[1], Anita Rahmawati[2], Alfina Chintya Bella*[3], Damar Aji Gurowo[4], Muhammad Arifin[5]
[1,2,3,4,5] Fakultas Teknik, Universitas Muria Kudus
Jl. Lingkar Utara UMK No. 01, Kode Pos PO BOX 53, Kudus, Jawa Tengah, Indonesia
e-mail: 202353005@std.umk.ac.id[1], 202353006@std.umk.ac.id[2], 202353008@std.umk.ac.id*[3], 202353030@std.umk.ac.id[4], arifin.m@umk.ac.id[5]

ARTICLE INFO

Article History:

Received: 20-03-2026
Revised: 23-04-2026
Accepted: 14-05-2026
Published: 21-05-2026

ABSTRACT

Student attendance and discipline are important aspects of the learning process because they reflect participation, responsibility, and compliance with school rules. However, attendance data are often treated only as administrative records and are rarely analyzed to identify recurring patterns of lateness. This study focuses on grouping student lateness patterns from daily attendance data using the K-Means algorithm to obtain a more structured and objective overview of student attendance discipline. The research employed a data mining approach comprising data collection, preprocessing, data transformation, normalization, and clustering. The dataset comprised 678 daily attendance records collected over 47 weeks. After recapitulation, the data yielded 54 lateness-duration categories, totaling 1,738 lateness incidents. The frequency of lateness incidents in each category was the primary variable in this study. Prior to the clustering process, the data were normalized using StandardScaler. The data were subsequently divided into three clusters: High, Medium, and Low. The findings indicate that lateness durations of 1–7 minutes were classified into the High cluster with 1,007 incidents, making it the most dominant lateness pattern. Meanwhile, the Medium cluster consisted of lateness durations ranging from 8 to 13 minutes with 425 incidents, whereas the Low cluster included durations of 14 to 63 minutes with a total of 306 incidents. This study offers a clustering-based perspective on interpreting attendance records and suggests that schools should focus not only on severe lateness but also on frequent short delays through earlier monitoring and targeted disciplinary interventions.

Keywords: *Daily Attendance Data, Lateness Patterns, K-Means Clustering, Student Attendance Discipline, Educational Data Mining.*

Abstrak

Kedisiplinan kehadiran siswa merupakan aspek penting dalam proses pembelajaran karena mencerminkan partisipasi, tanggung jawab, dan kepatuhan terhadap aturan sekolah. Namun, data absensi sering kali hanya diperlakukan sebagai catatan administratif dan jarang dianalisis untuk mengidentifikasi pola keterlambatan yang berulang. Penelitian ini bertujuan untuk mengelompokkan pola keterlambatan siswa berdasarkan data absensi harian menggunakan algoritma *K-Means* guna memberikan gambaran yang lebih terstruktur dan

DOI: https://doi.org/10.51903/bxrc5648

p-ISSN : 2803-1507 e-ISSN : 2803-1531

objektif tentang disiplin kehadiran. Penelitian ini menggunakan pendekatan data mining yang meliputi pengumpulan data, *preprocessing*, transformasi data, normalisasi, dan *clustering*. *Dataset* yang digunakan terdiri atas 678 catatan absensi harian selama 47 minggu. Setelah direkapitulasi, data menghasilkan 54 kategori durasi keterlambatan dengan total 1.738 kejadian keterlambatan. Variabel utama yang dianalisis adalah frekuensi kejadian keterlambatan pada setiap kategori, dan *StandardScaler* digunakan sebelum proses *clustering*. Data kemudian dikelompokkan ke dalam tiga *cluster*, yaitu *High, Medium,* dan *Low*. Hasil penelitian menunjukkan bahwa keterlambatan singkat berdurasi 1-7 menit membentuk *cluster High* dengan total 1.007 kejadian, yang menunjukkan bahwa keterlambatan ringan yang berulang merupakan pola paling dominan. *Cluster Medium* mencakup keterlambatan 8-13 menit dengan 425 kejadian, sedangkan *Cluster Low* mencakup keterlambatan 14-63 menit dengan 306 kejadian. Penelitian ini memberikan kontribusi berupa sudut pandang berbasis *clustering* dalam menafsirkan data absensi dan berimplikasi bahwa sekolah perlu memberi perhatian tidak hanya pada keterlambatan berat, tetapi juga pada keterlambatan singkat yang sering berulang melalui pemantauan lebih dini dan intervensi disiplin yang lebih tepat sasaran.

Kata Kunci: *Data Absensi Harian, Pola Keterlambatan, Clustering K-Means, Disiplin Kehadiran Siswa, Data Mining Pendidikan.*

1. PENDAHULUAN

Pembentukan karakter siswa tidak hanya dipengaruhi oleh lingkungan keluarga, tetapi juga oleh lingkungan sekolah sebagai tempat pendidikan lanjutan. Di sekolah, siswa dapat terpengaruh oleh berbagai perilaku negatif yang berpotensi memengaruhi kondisi mental dan perkembangan diri mereka. Karena itu, kedisiplinan perlu diterapkan dengan baik karena menjadi salah satu faktor yang mendukung keberhasilan siswa di masa depan. Sikap disiplin dapat terlihat dari kepatuhan siswa terhadap tata tertib sekolah, seperti mengikuti proses pembelajaran dengan tertib serta datang tepat waktu [1]. Kehadiran siswa merupakan indikator penting dalam menentukan kualitas proses pembelajaran karena mencerminkan tingkat partisipasi, tanggung jawab, dan kedisiplinan siswa. Ketidakhadiran yang berulang sering kali berkaitan dengan penurunan prestasi akademik, kurangnya motivasi belajar, serta meningkatnya risiko perilaku menyimpang [2].

Dalam pengelolaan data absensi, keterlambatan umumnya hanya dilihat dari jumlah siswa yang terlambat datang. Padahal, cara tersebut belum sepenuhnya mampu menggambarkan tingkat kedisiplinan kehadiran siswa secara menyeluruh. Seorang siswa dapat tercatat terlambat beberapa kali, tetapi tetap memiliki pola kehadiran yang teratur. Sebaliknya, ada siswa yang jarang terlambat, tetapi waktu kedatangannya tidak konsisten dan cenderung berubah-ubah. Oleh karena itu, analisis data absensi tidak cukup hanya berfokus pada jumlah keterlambatan, tetapi juga perlu memperhatikan pola kehadiran siswa secara lebih menyeluruh [3].

Kedisiplinan kehadiran siswa dapat dianalisis melalui berbagai indikator, seperti frekuensi keterlambatan, durasi keterlambatan, dan konsistensi waktu kedatangan. Namun, penelitian ini menyesuaikan analisis dengan struktur data absensi yang tersedia, yaitu kategori durasi keterlambatan dalam satuan menit dan jumlah kejadian keterlambatan pada setiap kategori. Dengan demikian, fokus penelitian diarahkan pada pengelompokan kategori durasi keterlambatan berdasarkan frekuensi kemunculannya [4].

Dalam konteks penelitian ini, data absensi harian digunakan untuk melihat pola keterlambatan berdasarkan durasi dan jumlah kejadian. Pengelompokan tidak dilakukan secara individual terhadap siswa, melainkan berdasarkan kategori durasi keterlambatan. Pendekatan ini dipilih agar hasil *clustering* sesuai dengan bentuk data yang tersedia dan dapat menunjukkan kategori keterlambatan yang paling dominan selama periode pengamatan.

Penelitian ini berfokus pada pengelompokan kategori durasi keterlambatan siswa berdasarkan data absensi harian menggunakan algoritma *K-Means*. Variabel yang digunakan dalam penelitian ini disesuaikan dengan struktur data yang tersedia, yaitu durasi keterlambatan dalam satuan menit dan jumlah kejadian keterlambatan pada setiap kategori durasi. Variabel minggu digunakan sebagai informasi periode pengamatan untuk melihat perubahan keterlambatan dari waktu ke waktu, sedangkan variabel jumlah digunakan sebagai dasar utama dalam proses clustering. Pemanfaatan data absensi siswa dengan teknik data mining dapat digunakan untuk mengidentifikasi pola kehadiran serta memberikan rekomendasi dalam meningkatkan disiplin sekolah [5].

Selain itu, penerapan algoritma *K-Means* dalam bidang pendidikan dapat digunakan untuk menemukan pola pada data siswa dan menghasilkan informasi yang bermanfaat bagi pihak sekolah dalam pengambilan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

keputusan [6]. Penelitian lain juga menunjukkan bahwa penerapan *K-Means* pada data pendidikan perlu melalui tahapan penyiapan data, prapemrosesan, penghilangan outlier, dan normalisasi agar hasil pengelompokan lebih terarah [7]. Dengan demikian, penelitian ini tidak mengelompokkan siswa secara individual, melainkan mengelompokkan kategori durasi keterlambatan berdasarkan frekuensi kemunculannya. Penyesuaian ini dilakukan agar konsep penelitian, variabel yang digunakan, dan implementasi metode memiliki kesesuaian yang jelas.

2. TINJAUAN PUSTAKA

Data mining dalam bidang pendidikan dapat digunakan untuk mengolah data akademik maupun data administrasi sekolah menjadi informasi yang lebih terstruktur guna mendukung pengambilan keputusan [8], [9], [10]. Salah satu data administrasi yang dapat dimanfaatkan adalah data absensi, karena data tersebut berkaitan dengan pola kehadiran, keterlambatan, dan kedisiplinan siswa. Dalam literatur yang membahas data absensi siswa [11], dinyatakan bahwa data absensi dapat digunakan untuk menggambarkan pola kehadiran siswa secara lebih objektif. Selain itu, [12] menunjukkan bahwa data absensi yang dikombinasikan dengan variabel akademik dapat dianalisis menggunakan *K-Means* dan analisis korelasi untuk memahami karakteristik siswa serta mendukung penyusunan strategi pembelajaran. Pemanfaatan data kehadiran juga dapat membantu sekolah dalam memantau tingkat kehadiran dan kedisiplinan siswa secara lebih sistematis [1]. Selain itu, penggunaan atribut hadir, izin, sakit, alfa, dan persentase kehadiran dapat menghasilkan kelompok kedisiplinan siswa, seperti disiplin tinggi, sedang, dan rendah [2].

Dalam konteks analisis data siswa, algoritma K-Means banyak diterapkan sebagai metode clustering yang mengelompokkan objek berdasarkan kemiripan karakteristik antar data. Pada bidang pendidikan, metode ini dapat dimanfaatkan untuk menganalisis berbagai data siswa, seperti tingkat kehadiran, hasil akademik, maupun pola pembelajaran. Dengan pengolahan data kehadiran menggunakan *K-Means*, sekolah dapat mengenali kelompok siswa tertentu sehingga lebih mudah menentukan tindak lanjut atau perhatian khusus yang diperlukan [13]. Pada konteks lain, *K-Means* juga digunakan untuk menilai kedisiplinan siswa berdasarkan data yang tersedia, sehingga hasil pengelompokan dapat memberikan gambaran awal mengenai karakteristik kedisiplinan siswa [14].

Penggunaan algoritma *K-Means* dalam pengolahan data pendidikan masih dinilai relevan berdasarkan beberapa penelitian terbaru. Metode ini dapat diterapkan untuk mengelompokkan siswa ke dalam beberapa kategori dengan memanfaatkan atribut seperti tingkat kehadiran, nilai akademik, psikomotorik, dan afektif sebagai dasar pengelompokan data [15]. Dalam bidang pendidikan, algoritma *K-Means* juga dimanfaatkan untuk proses pengelompokan peserta didik berdasarkan karakteristik pembelajaran yang dimiliki. Pada penelitian tersebut, metode ini digunakan sebagai teknik *clustering* untuk menganalisis data siswa [16], *K-Means* digunakan untuk mengelompokkan *learners* ke dalam kategori *advanced*, *average*, dan *slow learners* dengan dukungan proses *scaling*, *Elbow Method*, serta evaluasi kualitas *clustering*. Dalam konteks penempatan siswa, *K-Means* dapat digunakan untuk membentuk kelompok siswa yang lebih seimbang berdasarkan karakteristik tertentu sehingga hasilnya dapat mendukung pengambilan keputusan sekolah [17]. Pada penelitian terhadap data siswa SMK, penentuan jumlah *cluster* dilakukan terlebih dahulu menggunakan metode *Elbow* sebelum proses *clustering* dengan algoritma *K-Means* diterapkan [18].

Penerapan *K-Means* memerlukan validasi jumlah *cluster* dan proses pengolahan data yang tepat agar hasil *clustering* dapat diinterpretasikan dengan baik. Tahapan seperti seleksi data, prapemrosesan, transformasi, dan normalisasi diperlukan karena *K-Means* bekerja berdasarkan jarak antar data. Jika data tidak disiapkan dengan baik, hasil pengelompokan dapat menjadi kurang stabil. Selain itu, jumlah *cluster* pada *K-Means* perlu ditentukan sebelum proses *clustering* dijalankan. Dasar penentuan jumlah *cluster* dapat diperkuat dengan menggunakan metode evaluasi seperti *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index* [19]. Pada penelitian *clustering*, penggunaan kombinasi metode *Elbow*, *Silhouette Score*, dan *Davies-Bouldin Index* juga diterapkan untuk menentukan jumlah *cluster* yang lebih valid dan dapat dipertanggungjawabkan.

Berdasarkan sintesis literatur tersebut, penelitian terdahulu dapat dikelompokkan ke dalam tiga arah utama. Pertama, penelitian yang menggunakan data absensi atau kehadiran untuk mengidentifikasi pola kedisiplinan siswa. Kedua, penelitian yang menerapkan *K-Means* pada data pendidikan untuk mengelompokkan siswa berdasarkan karakteristik akademik, keaktifan, atau penempatan siswa. Ketiga, penelitian yang menekankan pentingnya evaluasi jumlah *cluster* agar hasil *K-Means* tidak hanya didasarkan pada asumsi peneliti. Berbeda dari penelitian sebelumnya, penelitian ini berfokus pada pengelompokan kategori durasi keterlambatan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

berdasarkan jumlah kejadian pada data absensi harian. Dengan demikian, penelitian ini tidak mengelompokkan siswa secara individual, melainkan mengidentifikasi kategori durasi keterlambatan yang paling dominan sehingga hasil analisis lebih sesuai dengan struktur data dan tujuan penelitian.

3. METODOLOGI PENELITIAN

3.1. *Tahapan Penelitan*

Tahapan penelitian dalam studi ini disusun secara sistematis untuk memastikan proses analisis berjalan terarah dan menghasilkan keluaran yang sesuai dengan tujuan penelitian. Penelitian diawali dengan identifikasi masalah terkait belum adanya pengelompokan tingkat kedisiplinan siswa berdasarkan data keterlambatan secara terstruktur. Selanjutnya dilakukan pengumpulan data, *preprocessing*, transformasi data, hingga penerapan metode *clustering*. Pendekatan bertahap ini sesuai dengan konsep *data mining* yang menekankan proses sistematis dalam menemukan pola dari data (*KDD – Knowledge Discovery in Database*) [20].

Tahap berikutnya adalah penerapan metode *K-Means Clustering* untuk mengelompokkan kategori durasi keterlambatan ke dalam tingkat frekuensi tinggi, sedang, dan rendah. Setelah itu dilakukan pengujian hasil clustering untuk memastikan kesesuaian pola data yang terbentuk. Pendekatan ini umum digunakan dalam penelitian *clustering* pada bidang pendidikan untuk membantu pengambilan keputusan berbasis data [21].

3.2. *Pengumpulan Data*

Pengumpulan data dalam penelitian ini dilakukan menggunakan metode dokumentasi dan observasi. Data diperoleh dari arsip absensi siswa dalam bentuk file *Microsoft Excel* yang berisi informasi keterlambatan. Teknik dokumentasi sering digunakan dalam penelitian *data mining* karena mampu menyediakan data historis yang terstruktur dan siap diolah [22].

Data yang digunakan merupakan data sekunder yang berasal dari sistem pencatatan absensi. Namun demikian, keterbatasan data perlu diperhatikan karena hanya berfokus pada keterlambatan tanpa mempertimbangkan variabel lain seperti kehadiran penuh atau izin. Hal ini dapat mempengaruhi kualitas analisis dan interpretasi hasil *clustering* [14].

3.3. *Analisis Data*

Pada penelitian ini, analisis data keterlambatan dilakukan menggunakan metode *clustering* dalam pendekatan data mining dengan memanfaatkan jumlah kejadian pada setiap kategori durasi sebagai dasar pengelompokan. *Clustering* sendiri merupakan teknik untuk mengelompokkan data yang memiliki karakteristik serupa ke dalam kelompok yang sama agar data lebih mudah dianalisis dan mendukung proses pengambilan keputusan [23].

Analisis dalam penelitian ini juga dapat dijelaskan menggunakan pendekatan *5W1H*, yaitu *Why* untuk mengetahui tujuan analisis, *Who* sebagai objek penelitian yaitu siswa, *What* berupa data keterlambatan, *Where* dan *When* berdasarkan data absensi pada periode tertentu, serta *How* menggunakan algoritma *K-Means*. Meskipun metode ini efektif dalam menemukan pola, hasil analisis sangat bergantung pada kualitas data dan jumlah variabel yang digunakan [24].

3.4. *Validasi Jumlah Cluster*

Sebelum algoritma *K-Means* dijalankan, terlebih dahulu dilakukan penentuan jumlah *cluster* sebagai tahap awal penelitian. Untuk mendapatkan jumlah *cluster* yang paling sesuai, proses evaluasi dilakukan menggunakan metode *Elbow*, *Silhouette Score*, dan *Davies-Bouldin Index* [25]. Analisis perubahan nilai *Sum of Squared Error* (SSE) atau inertia pada beberapa nilai K dilakukan menggunakan metode *Elbow*. Jumlah cluster dipilih pada titik ketika penurunan nilai SSE mulai melambat atau membentuk pola yang lebih stabil [26]. Pada tahap *evaluasi clustering, Silhouette Score* digunakan untuk mengukur kesesuaian data terhadap cluster yang ditempatinya dengan melihat kedekatan data dalam *cluster* yang sama serta perbedaannya dengan *cluster* lain. Sementara itu, *Davies-Bouldin Index* digunakan untuk menilai kualitas *cluster* berdasarkan tingkat kemiripan data di dalam *cluster* dan jarak antar cluster. Semakin kecil nilai *Davies-Bouldin Index*, maka kualitas *cluster* yang dihasilkan dianggap semakin baik. Oleh karena itu, metode *Elbow, Silhouette Score*, dan *Davies-Bouldin Index* dapat digunakan sebagai acuan dalam menentukan jumlah *cluster* yang paling tepat pada algoritma *K-Means* [19].

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Evaluasi jumlah *cluster* dengan kombinasi *Elbow, Silhouette*, dan *Davies-Bouldin Index* juga digunakan dalam penelitian clustering untuk memperkuat pemilihan jumlah cluster [27]. Untuk memperoleh jumlah cluster yang lebih tepat, dapat digunakan metode evaluasi seperti *Elbow Method, Silhouette Score*, dan *Davies-Bouldin Index* sebagai dasar dalam proses penentuan *cluster*. Selain itu, metrik evaluasi seperti *Silhouette Coefficient* dan *Davies-Bouldin Index* juga banyak dimanfaatkan pada penelitian *clustering* di bidang pendidikan untuk mengukur kualitas pemisahan antar kelompok data yang dihasilkan [28]. Analisis data dilakukan dengan mengelompokkan kategori durasi keterlambatan berdasarkan jumlah kejadian pada masing-masing kategori. Variabel utama yang digunakan dalam proses *clustering* adalah jumlah kejadian keterlambatan, sedangkan durasi keterlambatan digunakan sebagai identitas kategori dan minggu digunakan sebagai informasi periode pengamatan.

4. HASIL DAN PEMBAHASAN

4.1. *Deskripsi dan Pra-pemprosesan Data*

Penelitian ini menggunakan data absensi harian yang memuat informasi keterlambatan siswa. Data yang dianalisis berjumlah 678 baris dan terdiri atas tiga variabel, yaitu Minggu, Keterlambatan, dan Jumlah. Variabel Minggu menunjukkan periode pengamatan dalam bentuk mingguan, variabel Keterlambatan menunjukkan lama keterlambatan dalam satuan menit, sedangkan variabel Jumlah menunjukkan banyaknya kejadian keterlambatan pada setiap kategori. Data tersebut digunakan untuk melihat pola keterlambatan yang muncul selama periode pengamatan, kemudian dikelompokkan menggunakan algoritma *K-Means*. Secara umum, data menunjukkan bahwa keterlambatan siswa memiliki variasi yang cukup beragam. Terdapat keterlambatan dengan durasi sangat singkat, seperti 1 menit, dan terdapat pula keterlambatan dengan durasi yang lebih panjang. Variasi ini menunjukkan bahwa data memiliki keragaman yang cukup untuk dianalisis lebih lanjut. Dengan adanya perbedaan durasi dan frekuensi keterlambatan, pola keterlambatan dapat diidentifikasi dengan lebih jelas.

Sebelum dilakukan analisis, data terlebih dahulu melalui tahap *pra*-pemrosesan. Pada tahap ini dilakukan pembersihan nama kolom agar format data menjadi lebih rapi dan mudah diproses. Selanjutnya, variabel Minggu diubah ke dalam format periode mingguan agar data dapat diurutkan berdasarkan waktu. Variabel Jumlah tetap digunakan dalam bentuk numerik karena menjadi dasar dalam proses perhitungan dan pengelompokan data. Hasil pra-pemrosesan menunjukkan bahwa seluruh data dapat digunakan dengan baik. Jumlah data tetap 678 baris, yang berarti tidak ada data yang hilang selama proses pembersihan. Dengan demikian, data telah siap untuk dianalisis pada tahap berikutnya. Setelah data dinyatakan siap, langkah selanjutnya adalah melakukan rekapitulasi untuk melihat pola keterlambatan yang muncul secara keseluruhan.

4.2. *Validasi Jumlah Cluster*

Penentuan jumlah *cluster* dilakukan terlebih dahulu sebelum proses *clustering* dimulai. Pada tahap ini digunakan metode *Elbow, Silhouette Score*, dan *Davies-Bouldin Index* untuk mengevaluasi beberapa nilai K sehingga dapat diperoleh jumlah *cluster* yang paling sesuai dengan pola data. Data yang diuji berasal dari variabel jumlah keterlambatan pada tiap kategori durasi yang telah melalui proses normalisasi menggunakan *StandardScaler*. Hasil pengujian jumlah cluster tersebut ditampilkan pada Tabel 1.

Tabel 1. Hasil Validasi Jumlah Cluster

Jumlah Cluster	SSE/Inertia	Silhouette Score	Davies-Bouldin Index
2	7,835	0,822	0,337
3	2,498	0,809	0,353
4	1,425	0,762	0,424
5	0,868	0,767	0,361
6	0,538	0,735	0,369

Berdasarkan Tabel 1, nilai SSE/Inertia mengalami penurunan cukup besar dari K=2 ke K=3, yaitu dari 7,835 menjadi 2,498. Setelah K=3, penurunan nilai SSE masih terjadi, tetapi cenderung lebih kecil sehingga K=3 dapat dipertimbangkan sebagai titik *elbow*. Nilai *Silhouette Score* tertinggi terdapat pada K=2, yaitu 0,822, sedangkan K=3 memperoleh nilai 0,809. Meskipun nilai *Silhouette Score* pada K=3 sedikit lebih rendah dibandingkan K=2, nilai tersebut masih menunjukkan kualitas pemisahan *cluster* yang baik. Sementara itu, nilai *Davies-Bouldin Index* pada K=2 adalah 0,337 dan pada K=3 adalah 0,353. Perbedaan nilai tersebut

p-ISSN : 2803-1507 e-ISSN : 2803-1531

relatif kecil, sehingga K=3 tetap dapat digunakan karena menghasilkan pembagian *cluster* yang lebih informatif dan sesuai dengan tujuan penelitian, yaitu membedakan kategori keterlambatan ke dalam kelompok frekuensi tinggi, sedang, dan rendah. Dengan demikian, jumlah *cluster* sebanyak tiga dipilih karena memberikan keseimbangan antara hasil evaluasi model dan kemudahan interpretasi hasil *clustering*.

4.3. *Perhitungan Manual K-Means Clustering pada Data Keterlambatan*

Pada penelitian ini, metode *K-Means Clustering* digunakan untuk mengelompokkan kategori keterlambatan berdasarkan jumlah kejadian keterlambatan. Data yang digunakan merupakan hasil rekapitulasi dari 47 minggu pengamatan. Setelah seluruh data direkapitulasi, diperoleh 54 kategori durasi keterlambatan dengan total 1.738 kejadian.

Tahap awal dilakukan dengan menghitung total jumlah kejadian pada setiap kategori keterlambatan. Total keterlambatan untuk kategori ke-K dihitung menggunakan Rumus (1).

$$J_k = \sum_{t=1}^{T} j_{kt} \tag{1}$$

dengan J_k adalah total jumlah keterlambatan pada kategori ke-k, j_{tk} adalah jumlah keterlambatan kategori ke-k pada minggu ke-t, dan n adalah jumlah minggu pengamatan.

Berdasarkan proses validasi menggunakan metode *Elbow*, *Silhouette Score*, dan *Davies-Bouldin Index*, jumlah *cluster* yang dipilih dalam penelitian ini adalah tiga *cluster*. Penentuan nilai K=3 dilakukan karena nilai SSE atau Inertia mengalami penurunan yang cukup besar sampai pada K=3. Selain itu, hasil pengelompokan juga lebih mudah dipahami karena dapat dibagi menjadi kategori keterlambatan tinggi, sedang, dan rendah.

Perhitungan jarak antara data dengan *centroid* pada penelitian ini menggunakan metode *Squared Euclidean Distance*. Karena variabel yang digunakan hanya satu, yaitu jumlah keterlambatan, perhitungan jarak antara data dan *centroid* dilakukan menggunakan Rumus (2).

$$D(J_k, c_j) = (J_k - c_j)^2 \tag{2}$$

Setiap data kemudian dimasukkan ke cluster yang memiliki jarak paling kecil terhadap *centroid*. Proses penentuan *cluster* menggunakan Rumus (3).

$$C_k = \arg\min_j D(J_k, c_j) \tag{3}$$

Centroid baru dihitung berdasarkan rata-rata seluruh anggota pada masing-masing *cluster* menggunakan Rumus (4).

$$c_j = \frac{1}{n_j} \sum_{k \in C_j} J_k \tag{4}$$

dengan c_j adalah *centroid cluster* ke-j, n_j adalah jumlah anggota cluster ke-j, C_j adalah himpunan anggota pada cluster ke-j, dan J_k adalah total jumlah keterlambatan pada kategori ke-k.

Pada proses awal, *centroid* dipilih dari data hasil rekapitulasi, yaitu 13, 134, dan 97. Nilai tersebut digunakan sebagai pusat awal *Cluster Low*, *High*, dan *Medium* menggunakan Rumus (5).

$$c_{Low}^{(0)} = 13, c_{High}^{(0)} = 134, c_{Medium}^{(0)} = 97 \tag{5}$$

Sebagai contoh, kategori keterlambatan 1 menit memiliki jumlah kejadian sebanyak 167. Jarak kategori tersebut terhadap *centroid* awal dihitung menggunakan Rumus (6).

$$\begin{aligned} D(167, c_{Low}^{(0)}) &= (167 - 13)^2 = 23716 \\ D(167, c_{High}^{(0)}) &= (167 - 134)^2 = 1089 \\ D(167, c_{Medium}^{(0)}) &= (167 - 97)^2 = 4900 \end{aligned} \tag{6}$$

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Karena jarak terkecil adalah 1089, maka kategori keterlambatan 1 menit masuk ke *cluster High*. Proses yang sama dilakukan pada seluruh kategori keterlambatan hingga tidak ada lagi data yang berpindah *cluster*.

Proses penentuan *cluster* dilakukan pada seluruh kategori, dan setelah iterasi diperoleh *centroid* akhir menggunakan Rumus (7).

$$c_{Low} = 7{,}46, c_{Medium} = 70{,}83, c_{High} = 143{,}86 \quad (7)$$

Hasil akhir *clustering* menunjukkan bahwa *cluster High* (kategori keterlambatan 1 sampai 7 menit dengan total 1.007 kejadian) menggunakan Rumus (8).

$$C_{High} = \{1,2,3,4,5,6,7\} \quad (8)$$
$$\sum J_{High} = 1007$$

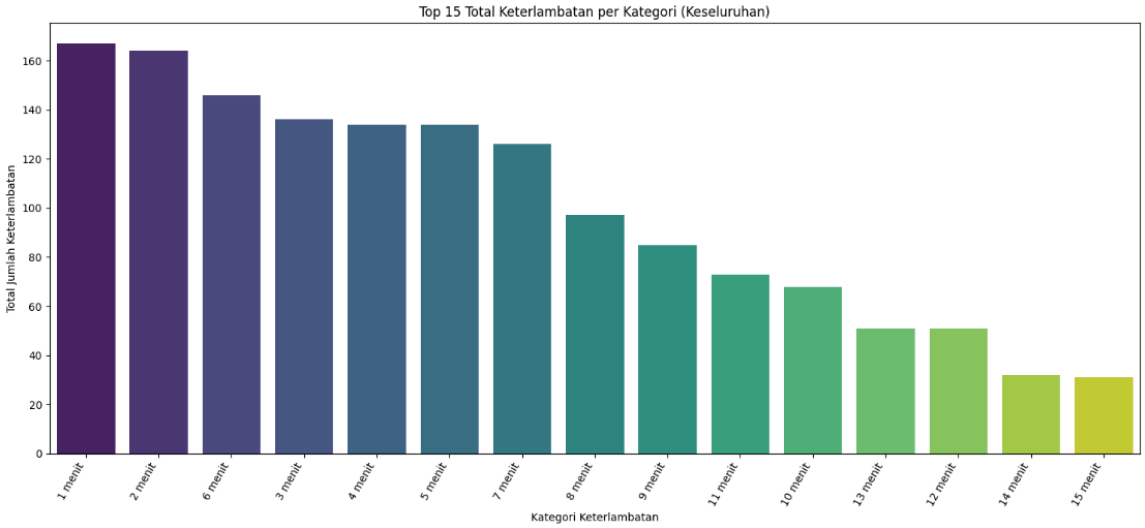
Cluster Medium (kategori keterlambatan 8 sampai 13 menit dengan total 425 kejadian) menggunakan Rumus (9).

$$C_{Medium} = \{8,9,10,11,12,13\} \quad (9)$$
$$\sum J_{Medium} = 425$$

Cluster Low (kategori keterlambatan mulai dari 14 sampai 63 menit dengan total 306 kejadian) menggunakan Rumus (10).

$$C_{Low} = \{14,15,16, \dots ,63\} \quad (10)$$
$$\sum J_{Low} = 306$$

Hasil tersebut menunjukkan bahwa keterlambatan berdurasi pendek merupakan pola yang paling dominan pada data penelitian. Kategori keterlambatan 1 sampai 7 menit masuk ke dalam *cluster High* karena memiliki jumlah kejadian yang paling besar. Kategori 8 sampai 13 menit masuk ke dalam *cluster Medium*, sedangkan kategori 14 sampai 63 menit masuk ke dalam *cluster Low*. Dengan demikian, masalah utama dalam data bukan terletak pada keterlambatan yang sangat lama, melainkan pada tingginya frekuensi keterlambatan singkat yang terjadi berulang kali.



Gambar 1. Top 15 Total Keterlambatan per Kategori (Keseluruhan)

4.4. *Rekapitulasi dan Visualisasi Pola Keterlambatan*

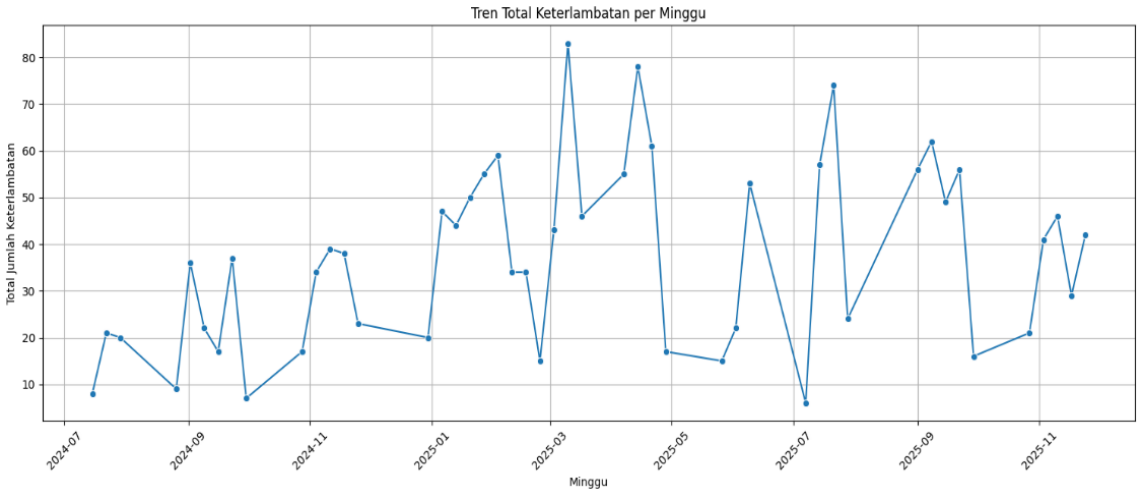
Setelah data selesai diproses, langkah berikutnya adalah merekap total keterlambatan berdasarkan kategorinya. Rekapitulasi ini dilakukan untuk mengetahui kategori keterlambatan yang paling sering terjadi selama periode pengamatan. Hasil pengolahan menunjukkan bahwa keterlambatan dengan durasi pendek

p-ISSN : 2803-1507 e-ISSN : 2803-1531

merupakan kategori yang paling dominan dalam data. Berdasarkan hasil rekapitulasi, kategori keterlambatan tertinggi adalah 1 menit sebanyak 167 kejadian, diikuti 2 menit sebanyak 164 kejadian, 3 menit sebanyak 136 kejadian, 4 menit sebanyak 134 kejadian, serta 5 menit sebanyak 134 kejadian. Selain itu, keterlambatan 7 menit juga masih tergolong tinggi, yaitu sebanyak 126 kejadian. Hasil ini menunjukkan bahwa keterlambatan singkat merupakan bentuk keterlambatan yang paling sering terjadi.

Setelah kategori 1 sampai 7 menit, jumlah keterlambatan mulai menurun. Keterlambatan 8 menit tercatat sebanyak 97 kejadian, 9 menit sebanyak 85 kejadian, 11 menit sebanyak 73 kejadian, 10 menit sebanyak 68 kejadian, serta 12 menit dan 13 menit masing-masing sebanyak 51 kejadian. Selanjutnya, keterlambatan 14 menit tercatat sebanyak 32 kejadian dan 15 menit sebanyak 31 kejadian. Pola ini menunjukkan bahwa semakin lama durasi keterlambatan, semakin kecil jumlah kejadiannya. Hasil rekapitulasi tersebut diperkuat melalui visualisasi data. Pada Gambar 1 terlihat bahwa kategori keterlambatan 1 menit, 2 menit, 6 menit, 3 menit, 4 menit, 5 menit, dan 7 menit memiliki jumlah kejadian paling tinggi dibandingkan kategori lainnya. Grafik ini memperjelas bahwa distribusi data lebih banyak terkonsentrasi pada keterlambatan berdurasi pendek.

Selain dilihat berdasarkan kategori, keterlambatan juga divisualisasikan berdasarkan minggu pengamatan. Pada Gambar 2 terlihat bahwa jumlah keterlambatan berubah-ubah dari minggu ke minggu. Pada beberapa periode jumlah keterlambatan berada pada tingkat rendah, sedangkan pada periode lain meningkat cukup tajam. Salah satu puncak tertinggi tampak terjadi sekitar Maret 2025 dengan jumlah keterlambatan lebih dari 80 kejadian dalam satu minggu. Peningkatan juga terlihat pada beberapa minggu lain, seperti sekitar April 2025, Juli 2025, dan September 2025. Berdasarkan grafik tren mingguan tersebut, dapat dipahami bahwa keterlambatan siswa tidak terjadi dalam pola yang tetap. Jumlah keterlambatan cenderung berfluktuasi selama periode pengamatan. Hal ini menunjukkan bahwa kondisi kedisiplinan kehadiran siswa dapat berubah dari waktu ke waktu, sehingga perlu dianalisis secara menyeluruh, bukan hanya dilihat pada satu minggu tertentu saja.



Gambar 2. Tren Total Keterlambatan per Minggu

4.5. *Penerapan Algoritma K-Means*

Setelah pola umum keterlambatan diketahui, tahap berikutnya adalah menerapkan algoritma *K-Means*. Pada penelitian ini, pengelompokan dilakukan terhadap jumlah keterlambatan pada setiap kategori. Tujuan dari proses ini adalah untuk menyederhanakan data yang cukup banyak menjadi beberapa kelompok yang lebih mudah dipahami.

Sebelum algoritma dijalankan, variabel Jumlah terlebih dahulu dinormalisasi menggunakan *StandardScaler*. Tahap normalisasi dilakukan agar data memiliki skala yang seragam. Langkah ini penting karena algoritma *K-Means* bekerja berdasarkan jarak antar data, sehingga perbedaan skala nilai dapat mempengaruhi hasil pengelompokan. Dengan normalisasi, proses *clustering* dapat dilakukan dengan lebih seimbang.

Dalam penelitian ini, jumlah *cluster* ditetapkan sebanyak 3 *cluster*. Ketiga *cluster* tersebut kemudian diberi label *High* (Tinggi), *Medium* (Sedang), dan *Low* (Rendah). Pemberian label ini didasarkan pada tingkat

p-ISSN : 2803-1507 e-ISSN : 2803-1531

frekuensi keterlambatan pada masing-masing kelompok. *Cluster* dengan frekuensi tertinggi diberi label tinggi, *cluster* dengan frekuensi menengah diberi label sedang, dan *cluster* dengan frekuensi terendah diberi label rendah.

Melalui penerapan *K-Means*, data yang awalnya terdiri dari banyak kategori keterlambatan dapat diringkas menjadi tiga kelompok utama. Proses ini membantu peneliti dalam melihat pola keterlambatan dengan lebih terstruktur. Dengan demikian, hasil pengelompokan tidak hanya menunjukkan banyaknya keterlambatan, tetapi juga memperlihatkan tingkat dominasi masing-masing kategori berdasarkan frekuensi kejadiannya.

4.6. *Hasil Pengelompokan dan Karakteristik Tiap Cluster*

Hasil pengelompokan menunjukkan bahwa data keterlambatan terbagi menjadi tiga *cluster*, yaitu *High* (Tinggi), *Medium* (Sedang), dan *Low* (Rendah). Setiap *cluster* memiliki karakteristik yang berbeda berdasarkan jumlah kejadian keterlambatan. *Cluster High* (Tinggi) berisi kategori keterlambatan dengan frekuensi paling besar. Kategori yang masuk ke dalam *cluster* ini adalah 1 menit, 2 menit, 3 menit, 4 menit, 5 menit, 6 menit, dan 7 menit. Rentang nilai pada *cluster* ini berada pada kisaran 126 sampai 167 kejadian. Hasil ini menunjukkan bahwa keterlambatan singkat merupakan kelompok yang paling dominan dalam data.

Cluster Medium (Sedang) terdiri atas kategori keterlambatan 8 menit, 9 menit, 10 menit, 11 menit, 12 menit, dan 13 menit. Rentang nilai pada *cluster* ini berada pada kisaran 51 sampai 97 kejadian. Meskipun jumlahnya tidak sebesar *cluster* tinggi, *cluster* sedang tetap menunjukkan frekuensi keterlambatan yang cukup sering terjadi selama periode pengamatan. Sementara itu, *Cluster Low* (Rendah) mencakup kategori keterlambatan mulai dari 14 menit sampai 63 menit. *Cluster* ini memiliki jumlah kategori paling banyak, tetapi frekuensi setiap kategorinya relatif kecil. Rentang nilainya berada pada kisaran 1 sampai 32 kejadian. Hal ini menunjukkan bahwa keterlambatan berdurasi lebih lama memang ada dalam data, tetapi jumlah kemunculannya jauh lebih sedikit dibandingkan keterlambatan singkat.

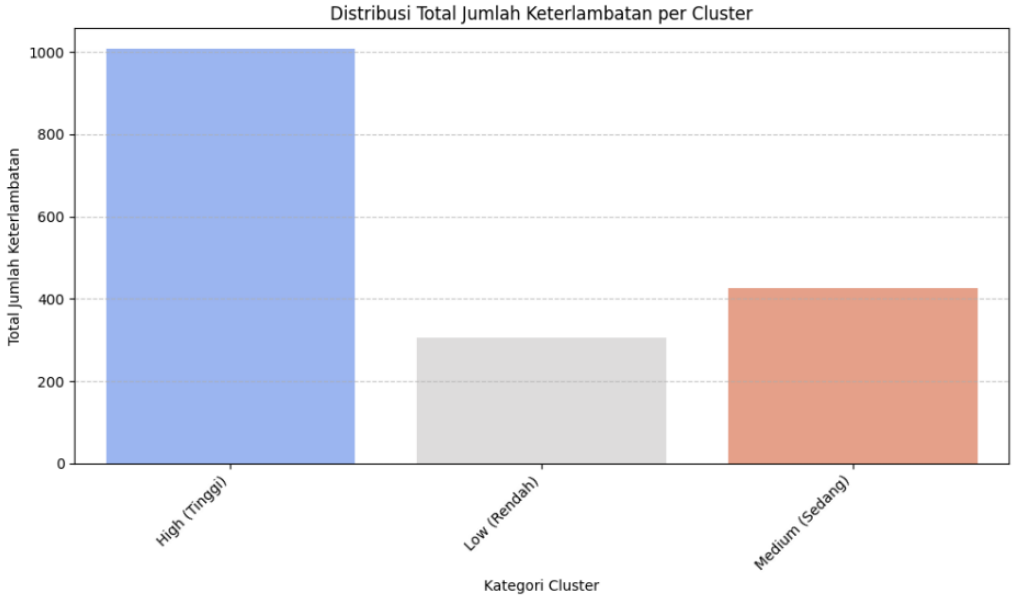
Jika dilihat dari total jumlah keterlambatan pada setiap *cluster*, *cluster High* memiliki total paling besar, yaitu sekitar 1007 kejadian. Setelah itu terdapat *cluster Medium* sekitar 425 kejadian, sedangkan *cluster Low* sekitar 306 kejadian. Hasil ini menunjukkan bahwa sebagian besar kejadian keterlambatan berada pada kelompok keterlambatan singkat. Perbandingan total tiap *cluster* dapat dilihat pada Gambar 3. sedangkan rincian pembagian kategori keterlambatan pada masing-masing cluster disajikan secara lengkap pada Tabel 2.

Tabel 1. Hasil Clustering Kategori Keterlambatan

No	Keterlambatan	Jumlah	Cluster Label
1	1 menit	167	High (Tinggi)
2	2 menit	164	High (Tinggi)
3	6 menit	146	High (Tinggi)
4	3 menit	136	High (Tinggi)
5	4 menit	134	High (Tinggi)
6	5 menit	134	High (Tinggi)
7	7 menit	126	High (Tinggi)
8	8 menit	97	Medium (Sedang)
9	9 menit	85	Medium (Sedang)
10	11 menit	73	Medium (Sedang)
11	10 menit	68	Medium (Sedang)
12	13 menit	51	Medium (Sedang)
13	12 menit	51	Medium (Sedang)
14	14 menit	32	Low (Rendah)
15	15 menit	31	Low (Rendah)
16	16 menit	27	Low (Rendah)
17	17 menit	25	Low (Rendah)
18	31 menit	18	Low (Rendah)
19	21 menit	13	Low (Rendah)
20	29 menit	13	Low (Rendah)
21	24 menit	11	Low (Rendah)
22	20 menit	11	Low (Rendah)
23	18 menit	10	Low (Rendah)
24	19 menit	9	Low (Rendah)
25	35 menit	8	Low (Rendah)
26	32 menit	7	Low (Rendah)

p-ISSN : 2803-1507 e-ISSN : 2803-1531

27	23 menit	7	Low (Rendah)
28	25 menit	7	Low (Rendah)
29	26 menit	6	Low (Rendah)
30	22 menit	6	Low (Rendah)
31	42 menit	6	Low (Rendah)
32	36 menit	6	Low (Rendah)
33	40 menit	6	Low (Rendah)
34	28 menit	5	Low (Rendah)
35	33 menit	5	Low (Rendah)
36	41 menit	4	Low (Rendah)
37	43 menit	4	Low (Rendah)
38	30 menit	4	Low (Rendah)
39	27 menit	3	Low (Rendah)
40	34 menit	3	Low (Rendah)
41	39 menit	3	Low (Rendah)
42	37 menit	2	Low (Rendah)
43	38 menit	2	Low (Rendah)
44	45 menit	2	Low (Rendah)
45	47 menit	1	Low (Rendah)
46	44 menit	1	Low (Rendah)
47	52 menit	1	Low (Rendah)
48	51 menit	1	Low (Rendah)
49	49 menit	1	Low (Rendah)
50	48 menit	1	Low (Rendah)
51	56 menit	1	Low (Rendah)
52	54 menit	1	Low (Rendah)
53	53 menit	1	Low (Rendah)
54	63 menit	1	Low (Rendah)



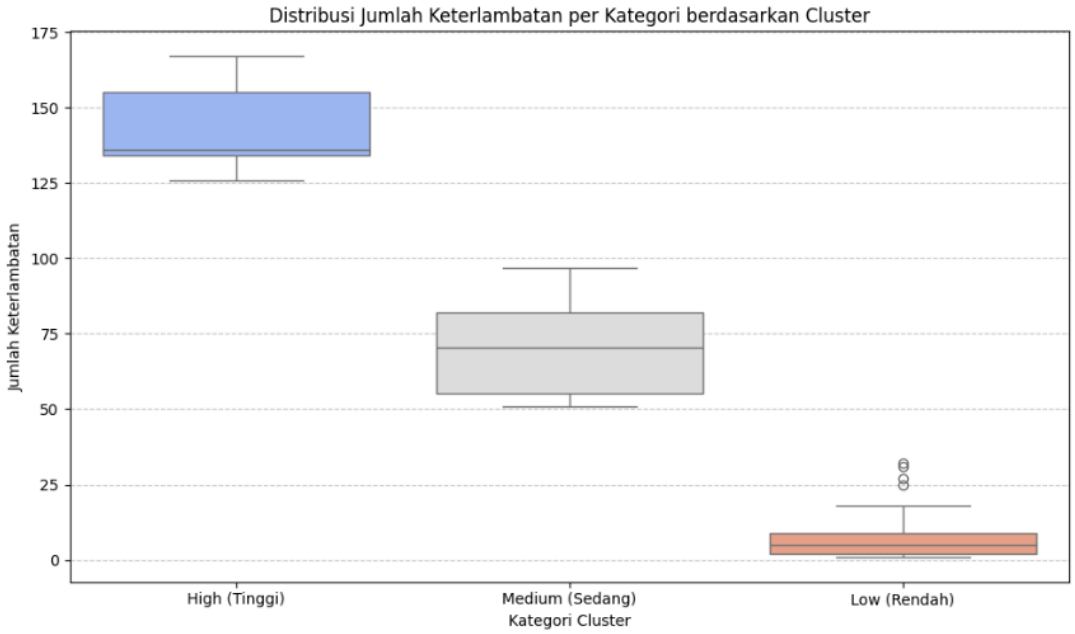
Gambar 3. Distribusi Total Jumlah Keterlambatan per Cluster

Karakteristik tiap *cluster* juga dapat dilihat melalui *boxplot* distribusi jumlah keterlambatan per kategori. Pada *cluster High*, sebaran data berada pada kisaran nilai yang tinggi dengan median sekitar 135. Hal ini menunjukkan bahwa kategori dalam *cluster* tinggi memiliki jumlah kejadian yang besar dan relatif terkonsentrasi. Pada *cluster Medium*, median berada di sekitar 70, dengan sebaran data yang berada di antara *cluster* tinggi dan *cluster* rendah. Sementara itu, *cluster Low* memiliki median yang jauh lebih kecil, dengan sebagian besar data berada pada kisaran rendah, meskipun masih terdapat beberapa titik pencilan pada kisaran sekitar 25 sampai 32.

Berdasarkan *boxplot* pada Gambar 4, terlihat bahwa perbedaan antar *cluster* cukup jelas. *Cluster* tinggi menampung kategori dengan jumlah kejadian besar, *cluster* sedang berada pada tingkat menengah,

p-ISSN : 2803-1507 e-ISSN : 2803-1531

sedangkan *cluster* rendah berisi kategori dengan jumlah kejadian kecil. Hasil ini menunjukkan bahwa algoritma *K-Means* mampu memisahkan data sesuai dengan karakteristik frekuensi keterlambatan yang dimiliki.



Gambar 4. Distribusi Jumlah Keterlambatan per Kategori berdasarkan Cluster

5. DISKUSI

Berdasarkan hasil pengelompokan, dapat dipahami bahwa bentuk keterlambatan yang paling dominan dalam data ini adalah keterlambatan ringan, terutama pada rentang 1 sampai 7 menit. Temuan ini menunjukkan bahwa masalah kedisiplinan kehadiran tidak selalu muncul dalam bentuk keterlambatan yang sangat lama. Justru masalah yang paling sering terjadi adalah kebiasaan datang sedikit terlambat yang berlangsung berulang kali. Hasil ini sejalan dengan penelitian [14] yang menyatakan bahwa metode *K-Means* mampu mengidentifikasi pola kedisiplinan siswa berdasarkan karakteristik perilaku yang sering muncul dalam aktivitas sekolah.

Dominannya *cluster High* menunjukkan bahwa sebagian besar kejadian keterlambatan berada pada kategori yang singkat. Jika dikaitkan dengan kondisi di sekolah, hasil ini memperlihatkan bahwa siswa pada umumnya tetap hadir, tetapi belum sepenuhnya tepat waktu. Dengan kata lain, bentuk ketidakdisiplinan yang lebih menonjol dalam data ini adalah kebiasaan datang sedikit terlambat. Walaupun durasinya pendek, frekuensi yang tinggi menunjukkan bahwa masalah ini tetap perlu mendapat perhatian. Temuan ini juga memiliki kesamaan dengan penelitian [29] yang menjelaskan bahwa pelanggaran kedisiplinan ringan sering kali menjadi kategori dengan jumlah paling dominan karena dianggap sebagai pelanggaran kecil oleh siswa sehingga sering diulangi.

Cluster Medium menunjukkan adanya kelompok keterlambatan yang lebih serius, yaitu sekitar 8 sampai 13 menit. Meskipun jumlahnya tidak sebesar *cluster* tinggi, kelompok ini tetap penting karena menunjukkan bahwa sebagian siswa mengalami keterlambatan yang lebih nyata. Jika tidak ditangani sejak awal, keterlambatan pada kelompok ini dapat berkembang menjadi keterlambatan yang lebih berat. Sementara itu, *cluster Low* menunjukkan adanya keterlambatan dengan durasi yang lebih panjang, yaitu mulai 14 menit sampai 63 menit. Frekuensi pada kelompok ini memang lebih kecil, tetapi bukan berarti dapat diabaikan. Keterlambatan berdurasi panjang dapat menunjukkan adanya masalah yang lebih serius. Dalam kondisi nyata, hal ini dapat berkaitan dengan faktor jarak rumah, transportasi, kurangnya pengawasan, atau rendahnya kesadaran siswa terhadap aturan sekolah. Hasil ini memiliki sedikit perbedaan dengan penelitian [14] yang menemukan bahwa faktor perilaku siswa lebih dominan dibanding faktor eksternal. Perbedaan tersebut kemungkinan terjadi karena karakteristik data yang digunakan berbeda, di mana penelitian ini lebih berfokus pada data keterlambatan kehadiran sehingga faktor lingkungan dan akses menuju sekolah menjadi lebih

p-ISSN : 2803-1507 e-ISSN : 2803-1531

berpengaruh terhadap hasil *clustering*.

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *K-Means* mampu mengelompokkan data keterlambatan siswa ke dalam tiga kategori utama sehingga data menjadi lebih mudah dipahami, yaitu kelompok keterlambatan tinggi, sedang, dan rendah. Dengan adanya pengelompokan ini, pihak sekolah dapat melihat pola keterlambatan secara lebih terstruktur. Oleh karena itu, data absensi tidak hanya berfungsi sebagai catatan administrasi, tetapi juga dapat digunakan untuk menggambarkan pola kedisiplinan kehadiran siswa berdasarkan keterlambatan yang terjadi.

Berdasarkan hasil *clustering*, pihak sekolah dapat menyusun strategi penanganan yang berbeda untuk setiap kelompok keterlambatan. *Cluster High* menunjukkan perlunya pemantauan lebih awal terhadap keterlambatan singkat yang sering berulang, misalnya melalui pengawasan kedatangan pada menit-menit awal sebelum jam pelajaran dimulai. *Cluster* Medium dapat dijadikan dasar untuk pembinaan siswa yang mulai menunjukkan kecenderungan keterlambatan lebih serius. Sementara itu, *Cluster Low* meskipun frekuensinya lebih kecil tetap perlu diperhatikan karena dapat berkaitan dengan faktor eksternal seperti jarak rumah, transportasi, atau kondisi pribadi siswa. Dengan demikian, hasil clustering tidak hanya berfungsi sebagai deskripsi data, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan sekolah dalam merancang intervensi kedisiplinan yang lebih tepat sasaran.

6. KESIMPULAN DAN SARAN

6.1. *Kesimpulan*

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa data absensi harian, khususnya keterlambatan siswa, memiliki pola yang beragam dan dapat dianalisis secara lebih mendalam menggunakan algoritma *K-Means*. Proses *pra*-pemrosesan data yang meliputi pembersihan, transformasi, dan normalisasi berhasil menyiapkan data sehingga layak untuk dilakukan analisis *clustering*.

Hasil pengelompokan menggunakan *K-Means* dengan tiga cluster menunjukkan bahwa kategori durasi keterlambatan terbagi menjadi tiga kelompok utama, yaitu *High* (tinggi), *Medium* (sedang), dan *Low* (rendah). *Cluster High* didominasi oleh kategori keterlambatan singkat berdurasi 1–7 menit dengan frekuensi paling tinggi, yang menunjukkan bahwa bentuk ketidakdisiplinan yang paling sering terjadi adalah keterlambatan ringan namun berulang. *Cluster Medium* mencakup kategori keterlambatan 8–13 menit dengan frekuensi sedang, sedangkan *Cluster Low* mencakup kategori keterlambatan lebih lama, yaitu 14–63 menit, dengan frekuensi yang relatif rendah.

Temuan ini menunjukkan bahwa meskipun keterlambatan panjang jumlahnya sedikit, keterlambatan singkat yang sering terjadi justru menjadi masalah utama dalam kedisiplinan kehadiran siswa. Dengan demikian, analisis menggunakan *K-Means* mampu menyederhanakan data yang kompleks menjadi informasi yang lebih mudah dipahami dan memberikan gambaran yang lebih objektif terkait pola kedisiplinan kehadiran siswa.

6.2. *Saran*

Berdasarkan hasil penelitian, pihak sekolah diharapkan dapat memanfaatkan hasil pengelompokan keterlambatan sebagai acuan dalam meningkatkan kedisiplinan siswa. Penanganan sebaiknya tidak hanya difokuskan pada keterlambatan dengan durasi lama, tetapi juga pada keterlambatan singkat yang terjadi secara berulang. Selain itu, diperlukan langkah pencegahan melalui peningkatan pengawasan serta pemberian motivasi kepada siswa. Pengembangan sistem absensi berbasis data juga perlu dilakukan agar dapat mendukung pengambilan keputusan yang lebih efektif.

DAFTAR PUSTAKA

[1] F. Agustina, R. Kurniawan, and T. Suprapti, "Peningkatan Efisiensi Pemantauan Kehadiran Siswa Melalui Analisis K-Means Clustering di Sekolah Menengah Pertama Negeri 3 Rancaekek, Kabupaten Bandung," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 45–52, Mar. 2024, doi: 10.36499/jinrpl.v6i1.9842.

[2] D. Nurhafidilah, N. Suarna, A. Bahtiar, U. Hayati, and F. Dikananda, "Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Siswa SMK Al-Ma'rifah Berdasarkan Kehadiran," *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, vol. 4, no. 2, pp. 112–120, Des. 2024, [Online]. Available: https://sintek.stmikku.ac.id/index.php/home/article/view/148

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[3] R. D. Parinduri, "Penerapan K-Means Clustering pada Absensi Mahasiswa Semester Ganjil untuk Mengelompokkan Kehadiran Mahasiswa," *JUSTEK: Jurnal Sains dan Teknologi*, vol. 6, no. 2, pp. 214–222, Des. 2023, doi: 10.556442/justek.v6i2.341.

[4] D. Gustian and M. Y. Ismatulloh, "Analisis Pemetaan Tingkat Kehadiran Siswa dengan Metode K-Means Clustering," *Jurnal Teknologi Komputer dan Sistem Informasi (JTKSI)*, vol. 7, no. 1, pp. 28–35, Jan. 2024, doi: 10.56327/jtksi.v7i1.1542.

[5] I. Triana and M. Hasanuddin, "Analisis Data Absensi Siswa Menggunakan Data Mining untuk Meningkatkan Disiplin Sekolah," *Journal of Data Science and Informatics Engineering*, vol. 1, no. 2, pp. 56–62, Jul. 2026, doi: 10.64803/jodsie.v1i2.29.

[6] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.

[7] T. Widyanti, S. S. Hilabi, A. Hananto, Tukino, and E. Novalia, "Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi," *Jurnal Pendidikan Tambusai*, vol. 7, no. 3, pp. 24810–24819, Des. 2023, doi: 10.31004/jptam.v7i3.10652.

[8] H. Arifin, S. Suendri, and I. A. Sinaga, "Attendance Monitoring System Using Web-Based Location-Based Service at the Lumban Dolok Village Office," *Jurnal Ilmiah Sistem Informasi*, vol. 5, no. 2, pp. 382–400, May 2026, doi: 10.51903/k3j1e109.

[9] D. Santi, R. D. Harahap, and S. U. Siregar, "Scientific Attitude Analysis on Students in Class XI-IPA at SMA Negeri 2 Bilah Hulu Regarding Human Blood Circulation System Material," *Journal of Management and Informatics*, vol. 2, no. 1, pp. 01–07, Apr. 2023, doi: 10.22216/jcc.2019.v4i3.4.

[10] M. Word and F. D. Silalahi, "Mandiri UNBK Try Out Information System Web-Based Using Linear Congruent Method at Madrasah Working Group (KKM) MTs N 2 Demak," *Journal of Technology Informatics and Engineering*, vol. 2, no. 1, pp. 56–73, May 2023, doi: 10.51903/jtie.v2i1.176.

[11] S. Raja et al., "Analisis Data Mining Absensi Siswa SMP Negeri 1 Bilah Barat dengan Metode Algoritma K-Means Clustering," *Jurnal Cyber Tech*, vol. 12, no. 2, pp. 45–54, Nov. 2024, doi: 10.53513/jct.v12i2.8943.

[12] C. Renita, T. N. Padilah, and M. A. E. Wahyu, "Identifikasi Pola Belajar dan Absensi Siswa dengan Teknik Clustering dan Analisis Korelasi," *Jurnal Informatika dan Ilmu Komputer (JIKO)*, vol. 4, no. 1, pp. 34–42, Feb. 2024, doi: 10.58602/jiko.v4i1.102.

[13] Y. Ahmad and T. Fatimah, "Implementasi Teknik Clustering Menggunakan Algoritma K-Means pada SMK Yadika 3 Jakarta Berbasis Web," *Jurnal Komputer Bisnis dan Manajemen*, vol. 5, no. 2, pp. 89–96, Agu. 2024, doi: 10.59134/jkbm.v5i2.612.

[14] Q. I. Mawarni and E. S. B. Setiawan, "Implementasi Algoritma K-Means Clustering dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, p. 522, Jun. 2022, doi: 10.30865/json.v3i4.4242.

[15] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *Jurnal Ilmiah SISFOTENIKA*, vol. 1, no. 1, pp. 12–20, Mei 2011, doi: 10.30700/jst.v1i1.12.

[16] B. Joseph and S. Abraham, "Identifying Slow Learners in an E-Learning Environment Using K-Means Clustering Approach," *Knowledge Management and E-Learning*, vol. 15, no. 4, pp. 539–553, Dec. 2023, doi: 10.34105/j.kmel.2023.15.031.

[17] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *Jurnal Informasi dan Teknologi*, vol. 3, no. 3, pp. 103–108, Sep. 2021, doi: 10.37034/jidt.v3i3.130.

[18] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK dalam Ujian Nasional," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 82–89, Aug. 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[19] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies-Bouldin, Elbow, dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *Jurnal Sains, Teknologi dan Industri*, vol. 21, no. 1, pp. 77–85, Des. 2023, [Online]. Available: https://journal.lppmunindra.ac.id/index.php/string/article/view/16388

[20] M. A. Irwansyah et al., "Analisis Faktor Putus Sekolah dan Kondisi Ruang Kelas Menggunakan Metode Data Mining," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 11, no. 1, pp. 12–21, Apr. 2025, doi: 10.26418/jepin.v11i1.78213.

[21] 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 Komputasi*, vol. 22, no. 2, pp. 115–124, Jun. 2023, doi: 10.33772/jik.v22i2.214.

[22] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *JUKI: Jurnal Komputer dan Informatika*, vol. 6, no. 1, pp. 45–53, Apr. 2024, doi: 10.53842/juki.v6i1.214.

[23] D. Syaripudin and A. R. Raharja, "Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa untuk Menentukan Kelompok Penerima Beasiswa," *Jurnal Teknik dan Ilmu Komputer*, vol. 12, no. 2, pp. 78–85, Jul. 2023, doi: 10.34010/jtik.v12i2.985.

[24] P. S. Hasugian and J. R. Sagala, "Penerapan Data Mining untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 3, pp. 262–268, Des. 2022, [Online]. Available: https://djournals.com/klik/article/view/642

[25] S. Sulistina, P. E. P. Utomo, and B. F. Hutabarat, "Klasterisasi Wilayah Rawan Kriminalitas di Kota Jambi (2022-2024) Menggunakan Algoritma K-Means," *INSTEK: Jurnal Sains dan Teknologi*, vol. 10, no. 2, pp. 110–118, Des. 2025, [Online]. Available: https://journal.uin-alauddin.ac.id/index.php/instek/article/view/42105

[26] Hendrik, Kusrini, and Kusnawi, "Optimasi Penentuan Sentroid Awal pada K-Means untuk Meningkatkan Hasil Evaluasi Davies-Bouldin Index," *Jurnal Teknoinfo*, vol. 18, no. 1, pp. 55–63, Jan. 2024, doi: 10.33365/jti.v18i1.2981.

[27] M. R. Syahkur, D. Hartama, and S. Solikhun, "Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita," *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[28] V. Naik and V. Kamat, "Using Clustering Techniques to Understand Student Involvement in an Online Environment," *International Journal of Information and Education Technology*, vol. 15, no. 5, pp. 1024–1044, May 2025, doi: 10.18178/ijiet.2025.15.5.2308.

[29] R. Hanun, "Perancangan Sistem Aplikasi Penilaian Kedisiplinan Siswa Menggunakan Metode K-Means Elbow Berbasis Website," *Saturnus: Jurnal Teknologi dan Sistem Informasi*, vol. 3, no. 4, pp. 15–29, Sep. 2025, doi: 10.61132/saturnus.v3i4.1045.