

PENERAPAN METODE CLUSTERING DALAM IDENTIFIKASI SISWA BERMASALAH BERDASARKAN NILAI AKADEMIK

Meri Juliya¹, Kurniawati², Sulasminarti³

Institut Teknologi Bisnis Dan Bahasa Dian Cipta Cendikia¹²³

Jl. Negara No.03, Kel. Candimas, Kec. Abung Selatan, Kotabumi Kab. Lampung Utara - 34581 :
0724-23003

Email: merijuliya59@gmail.com¹, xkurniawatix@gmail.com², sulasminarti085@gmail.com³

ABSTRAK

Permasalahan siswa bermasalah merupakan tantangan yang sering dihadapi lembaga pendidikan, terutama pada tingkat sekolah menengah pertama. Identifikasi siswa bermasalah umumnya masih dilakukan secara subjektif berdasarkan observasi guru, sehingga berpotensi menimbulkan bias dalam pengambilan keputusan. Penelitian ini bertujuan menerapkan metode clustering menggunakan algoritma K-Means untuk mengidentifikasi siswa bermasalah berdasarkan nilai akademik dan data pelanggaran di MTs Muhammadiyah Ogan Lima. Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining dengan dua variabel utama, yaitu nilai akademik dan jumlah pelanggaran siswa. Proses analisis meliputi preprocessing data, penentuan jumlah cluster, perhitungan jarak menggunakan Euclidean Distance, serta iterasi hingga konvergensi. Evaluasi kualitas cluster dilakukan menggunakan Silhouette Score dan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa siswa terbagi ke dalam tiga cluster utama: (1) siswa tidak bermasalah, (2) siswa berisiko, dan (3) siswa bermasalah. Nilai Silhouette Score sebesar 0,61 dan Davies-Bouldin Index sebesar 0,47 menunjukkan kualitas cluster yang baik dan stabil. Hasil ini dapat membantu pihak sekolah dalam menyusun strategi pembinaan siswa secara lebih objektif dan terarah.

Kata kunci: Clustering, K-Means, Data Mining, Siswa Bermasalah, Nilai Akademik

ABSTRACTS

The issue of problematic students is a common challenge faced by educational institutions, particularly at the junior secondary school level. The identification of problematic students is generally conducted subjectively based on teachers' observations, which may introduce bias in the decision-making process. This study aims to apply a clustering method using the K-Means algorithm to identify problematic students based on academic performance and violation records at MTs Muhammadiyah Ogan Lima. This research employs a quantitative approach based on data mining techniques using two primary variables: academic scores and the number of student violations. The analytical process includes data preprocessing, determination of the optimal number of clusters, distance calculation using Euclidean Distance, and iterative centroid updates until convergence is achieved. Cluster quality evaluation is performed using the Silhouette Score and the Davies-Bouldin Index.

The results indicate that students are classified into three main clusters: (1) non-problematic students, (2) at-risk students, and (3) problematic students. The Silhouette Score of 0.61 and Davies-Bouldin Index of 0.47 demonstrate good and stable clustering performance. These findings can assist the school in developing more objective and targeted student guidance strategies.

Keywords: Clustering, K-Means, Data Mining, Problematic Students, Academic Performance

1. Pendahuluan

Pendidikan merupakan sarana utama dalam membentuk sumber daya manusia yang berkualitas. Melalui pendidikan, peserta didik dibina tidak hanya dari segi akademik, tetapi juga dari segi sikap, perilaku, dan juga karakter.

Sekolah sebagai lembaga pendidikan formal memiliki tanggung jawab untuk menciptakan lingkungan belajar yang kondusif agar proses pembelajaran dapat berjalan secara efektif.

Dalam praktiknya, proses pembelajaran disekolah sering dihadapkan pada berbagai

permasalahan, salah satunya adalah keberadaan siswa bermasalah. Siswa bermasalah merupakan siswa yang menunjukkan perilaku menyimpang dari norma dan tata tertib sekolah, seperti sering melanggar peraturan, kurang disiplin, bolos sekolah, serta memiliki motivasi belajar yang rendah. Kondisi tersebut tidak hanya mengganggu perkembangan siswa itu sendiri, tetapi juga dapat mengganggu proses pembelajaran secara keseluruhan.

Permasalahan perilaku siswa sering kali berkaitan erat dengan pencapaian nilai akademik. Siswa yang memiliki tingkat kedisiplinan rendah dan kurang motivasi belajar cenderung memperoleh nilai akademik yang rendah. Sebaliknya, siswa dengan prestasi akademik yang baik umumnya menunjukkan perilaku belajar yang lebih positif. Oleh karena itu, nilai akademik dapat dijadikan salah satu indikator dalam mengidentifikasi siswa bermasalah.

Di MTs Muhammadiyah Ogan Lima, proses identifikasi siswa bermasalah masih dilakukan secara konvensional, yaitu berdasarkan pengamatan guru, wali kelas serta catatan kasus siswa. Cara tersebut bersifat subjektif dan belum memanfaatkan data akademik secara optimal. Dengan semakin banyak data siswa yang tersedia, di perlukan suatu metode yang mampu mengolah data tersebut secara sistematis dan objektif.

Pemanfaatan teknik data mining, khususnya metode clustering, dapat menjadi solusi untuk mengelompokkan siswa kemiripan nilai akademik dan tingkat pelanggaran. Metode clustering mampu membantu pihak sekolah dalam mengidentifikasi siswa bermasalah secara lebih akurat oleh karena itu, dilakukan penelitian dengan judul "Penerapan Metode Clustering dalam Identifikasi Siswa Bermasalah berdasarkan Nilai Akademik di MTs Muhammadiyah Ogan Lima" dilakukan menerapkan algoritma K-Means dalam mengidentifikasi siswa bermasalah. Menghasilkan pengelompokan siswa yang dapat digunakan sebagai dasar pengambilan keputusan sekolah.

2. METODOLOGI PENELITIAN

1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining. Pendekatan kuantitatif digunakan karena data yang dianalisis berupa data numerik, yaitu nilai akademik siswa dan data pelanggaran siswa. Data tersebut diolah

menggunakan metode matematis dan statistik untuk menghasilkan informasi yang objektif.

Pendekatan data mining di gunakan karena penelitian ini bertujuan untuk menemukan pola dan struktur tersembunyi dalam data siswa yang selama ini clustering, karena penelitian ini tidak menggunakan label siswa bermasalah sebelumnya, melainkan mengelompokkan siswa berdasarkan kemiripan nilai akademik.

2.2 Objek dan Lokasi Penelitian

Objek penelitian ini adalah data siswa MTs Muhammadiyah Ogan Lima, yang meliputi data nilai akademik dan data catatan pelanggaran siswa. Penelitian dilaksanakan di MTs Muhammadiyah Ogan Lima, karena sekolah tersebut memiliki data yang relevan dan lengkap untuk mendukung pelaksanaan penelitian.

2.3 Variabel Penelitian

Variabel penelitian merupakan atribut yang digunakan dalam proses clustering. Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Nilai Akademik Siswa

Nilai akademik berupa nilai rata-rata rapor atau nilai mata pelajaran inti yang mencerminkan prestasi belajar siswa.

2. Jenis Pelanggaran Siswa

Jenis pelanggaran siswa menunjukkan tingkat berat atau ringan pelanggaran yang dilakukan oleh siswa, berupa jumlah atau bobot pelanggaran yang dilakukan siswa berdasarkan catatan kasus sekolah.

3. Jumlah Pelanggaran Siswa

Jumlah pelanggaran siswa merupakan frekuensi pelanggaran tata tertib sekolah yang dilakukan oleh siswa dalam periode tertentu. Data ini diperoleh dari catatan kasus sekolah dan digunakan untuk menggambarkan tingkat kedisiplinan siswa.

4. Tingkat Kehadiran Siswa

Tingkat kehadiran siswa merupakan persentase kehadiran siswa dalam mengikuti kegiatan pembelajaran di sekolah. Variabel ini digunakan untuk menggambarkan kedisiplinan siswa dalam mengikuti proses belajar mengajar.

5. Nilai Sikap Siswa

Nilai sikap siswa merupakan penilaian perilaku siswa selama proses pembelajaran yang tercantum dalam rapor atau catatan guru. Nilai sikap mencerminkan kedisiplinan, tanggung jawab, dan sikap siswa di lingkungan sekolah. Variabel tersebut dipilih karena secara langsung berkaitan dengan identifikasi siswa bermasalah sebagaimana dijelaskan pada Bab pendahuluan dan Bab tinjauan pustaka. Seluruh variabel penelitian tersebut dinyatakan dalam bentuk data numerik, sehingga sesuai dengan karakteristik metode *clustering* menggunakan algoritma K-Means untuk mengelompokkan siswa berdasarkan tingkat permasalahan yang dimiliki.

Table variable.

No	Nilai Akademik Siswa	Indikator	Skala
1	Jumlah Pelanggaran	Nilai rata-rata rapor	Numerik
2	Jenis Pelanggaran	Frekuensi pelanggaran	Numerik
3	Tingkat Kehadiran	Bobot pelanggaran	Numerik
4	Nilai Sikap Siswa	Persentase kehadiran	Numerik
5	Nilai Akademik Siswa	Nilai sikap rapor	Numerik

A. Sumber Data

Sumber data dalam penelitian ini terdiri dari:

1. Data Primer

Data primer berupa data catatan pelanggaran siswa yang diperoleh langsung dari pihak sekolah, seperti buku kasus siswa atau arsip bimbingan konseling.

2. Data Sekunder

Data sekunder berupa nilai akademik siswa yang diperoleh dari dokumen resmi sekolah, seperti rapor atau daftar nilai siswa.

3. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah:

4. Studi Dokumentasi

Mengumpulkan data nilai akademik dan data pelanggaran siswa dari dokumen resmi sekolah.

5. Wawancara

Wawancara dilakukan dengan guru atau wali kelas untuk memperoleh informasi pendukung terkait kondisi siswa dan sistem pencatatan pelanggaran.

B. Tahapan Penelitian

Tahapan penelitian disusun secara sistematis sebagai berikut:

1. Identifikasi Masalah

Tahap awal penelitian dilakukan dengan mengidentifikasi permasalahan yang terjadi di sekolah, yaitu kesulitan dalam mengidentifikasi siswa bermasalah secara objektif berdasarkan nilai akademik.

2. Pengumpulan Data

Data nilai akademik dan data pelanggaran siswa dikumpulkan dari MTs Muhammadiyah Ogan Lima sesuai dengan kebutuhan penelitian.

3. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum dianalisis, meliputi:

- Pembersihan data (data kosong dan duplikasi)
- Penyamaan format data
- Normalisasi data jika diperlukan

Tahap ini bertujuan untuk meminimalkan kesalahan dalam proses clustering.

4. Penerapan Metode Clustering

Data yang telah diproses kemudian dianalisis menggunakan metode clustering dengan algoritma K-Means untuk mengelompokkan siswa berdasarkan kemiripan nilai akademik.

5. Analisis Hasil

Hasil clustering dianalisis untuk mengetahui karakteristik setiap cluster dan mengidentifikasi kelompok siswa bermasalah.

C. Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah metode clustering dengan algoritma K-Means. Metode ini digunakan untuk mengelompokkan siswa berdasarkan nilai akademik dan data pelanggaran siswa.

D. Penerapan Algoritma K-Means

1. Penentuan Jumlah Cluster

Jumlah cluster (K) yang digunakan dalam penelitian ini adalah 3 cluster, yaitu:

1. Cluster siswa tidak bermasalah
2. Cluster siswa berisiko
3. Cluster siswa bermasalah

2. Penentuan Centroid Awal

Centroid awal ditentukan secara acak dari data siswa yang digunakan sebagai pusat awal masing-masing cluster.

3. Perhitungan Jarak (Euclidean Distance)

Perhitungan jarak antara data siswa dan centroid dilakukan menggunakan rumus Euclidean Distance:

$$D(x,y)=(x1-y1)^2+(x2-y2)^2$$

4. Iterasi dan Konvergensi

Proses K-Means dilakukan secara berulang hingga tidak terjadi perubahan keanggotaan cluster atau nilai centroid telah stabil.

5. Flowchart Alur Penelitian

Alur penelitian dapat dijelaskan sebagai berikut:

- Mulai
- Pengumpulan data siswa
- Preprocessing data
- Menentukan jumlah cluster
- Menentukan centroid awal
- Menghitung jarak data ke centroid
- Mengelompokkan data ke cluster terdekat
- Menghitung centroid baru
- Cek konvergensi
- Analisis hasil clustering
- Selesai

6. Output Penelitian

Output yang dihasilkan dari penelitian ini adalah:

- Pengelompokan siswa berdasarkan tingkat permasalahan akademik
- Identifikasi siswa bermasalah berdasarkan nilai akademik
- Rekomendasi pembinaan siswa bagi pihak sekolah

3.. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Data Penelitian

Data penelitian yang digunakan adalah data siswa MTs Muhammadiyah Ogan Lima tahun ajaran 2025/2026. Data diperoleh sesuai dengan tahapan pengumpulan data pada Sub bab 3.6.2 dan bersumber dari dokumen resmi sekolah. Data terdiri dari dua variabel utama, yaitu nilai akademik siswa dan data pelanggaran siswa.

Dari keseluruhan data siswa yang tersedia, hanya data siswa yang memiliki nilai akademik dan catatan pelanggaran lengkap yang digunakan dalam proses analisis. Hal ini bertujuan untuk menjaga konsistensi dan validitas hasil clustering.

Tabel Hasil Pengelompokan Siswa Berdasarkan Metode K-Means

No	NAMA	Nilai Rata-rata	Cluster	Keterangan
1.	Ahmad Baihaki	77	Cluster 2	Siswa Berisiko
2.	Aira Pratiwi	77	Cluster 2	Siswa Berisiko
3.	Alif Erzandi	77	Cluster 2	Siswa Berisiko
4.	Alya Shaputri	79	Cluster 2	Siswa Berisiko
5.	Bambang	78	Cluster 2	Siswa Berisiko
6.	Bayu Tri Alfajri	76	Cluster 2	Siswa Berisiko
7.	Chikco Syahputra	75	Cluster 3	Siswa Bermasalah
8.	Fikri Adrian Saputra	79	Cluster 1	Siswa Tidak Bermasalah
9.	Julian Rahmadani	76	Cluster 2	Siswa Berisiko
10.	Kaka Donata	79	Cluster 1	Siswa Tidak Bermasalah
11.	Kesya Sifa Alzahra	78	Cluster 2	Siswa Berisiko
12.	Khalif Al'aqsho	82	Cluster 1	Siswa Tidak Bermasalah
13.	M. Afif Saputra	78	Cluster 2	Siswa Berisiko
14.	Maura Ayu Alzahra	76	Cluster 2	Siswa Berisiko
15.	Muhammad Suraji	75	Cluster 3	Siswa Bermasalah
16.	Mulia Pratama	75	Cluster 3	Siswa Bermasalah
17.	Naisya Anvica	81	Cluster 1	Siswa Tidak Bermasalah

18.	Pitri Yanti	78	Cluster 2	Siswa Berisiko
19.	Rafa Riski	76	Cluster 2	Siswa Berisiko
20.	Rahma Anastasya	78	Cluster 2	Siswa Berisiko
21.	Rayhan Hadinata	78	Cluster 2	Siswa Berisiko
22.	Satria Putra Pratama	77	Cluster 2	Siswa Berisiko
23.	Satria Saputra	76	Cluster 2	Siswa Berisiko
24.	Yoga Pratama	75	Cluster 3	Siswa Bermasalah
25.	Yuda Andrea Pratama	77	Cluster 2	Siswa Berisiko

Ringkasan Hasil Clustering

Claster	Jumlah siswa	Katagori
Claster 1	4 siswa	Siswa Tidak Bermasalah
Claster 2	17 siswa	Siswa Berisiko
Claster 3	4 siswa	Siswa Bermasalah

Berdasarkan hasil clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga, diperoleh pengelompokan siswa ke dalam tiga kategori utama. Cluster 1 terdiri dari siswa dengan nilai akademik tinggi dan tingkat pelanggaran rendah sehingga dikategorikan sebagai siswa tidak bermasalah. Cluster 2 merupakan kelompok siswa dengan nilai akademik sedang dan potensi risiko melakukan pelanggaran sehingga memerlukan pembinaan dan pengawasan. Cluster 3 terdiri dari siswa dengan nilai akademik rendah dan tingkat pelanggaran tinggi sehingga dikategorikan sebagai siswa bermasalah dan membutuhkan penanganan khusus dari pihak sekolah.

3.2. Analisis Data Awal

Analisis data awal dilakukan untuk mengetahui karakteristik umum data sebelum dilakukan preprocessing. Nilai akademik siswa berada pada rentang 60 hingga 75, yang menunjukkan adanya variasi tingkat prestasi belajar siswa. Sementara itu, data pelanggaran menunjukkan bahwa sebagian siswa memiliki pelanggaran ringan hingga sedang.

Analisis awal ini menunjukkan indikasi bahwa siswa dengan nilai akademik lebih rendah cenderung memiliki jumlah pelanggaran lebih banyak, sehingga diperlukan metode analisis lanjutan untuk mengelompokkan siswa secara objektif.

3.3. Deskripsi Data Akademik dan Pelanggaran

Deskripsi data dilakukan untuk memberikan gambaran detail mengenai data yang dianalisis. Data nilai akademik diperoleh dari nilai rata-rata rapor, sedangkan data pelanggaran merupakan akumulasi pelanggaran yang dilakukan siswa selama satu tahun ajaran.

Contoh data penelitian ditunjukkan pada Tabel berikut:

No	Nama Siswa	Kelas	Nilai Akademik	Jumlah Pelanggaran
1	Muhamaad Putra N	VII	68	1
2	Chikco syahputra	VII	70	0
3	Yoga Pratama	VII	65	2
4	Kaka Donata	VII	67	1
5	Julian Rama Dani	VII	60	3

3.4. Preprocessing Data

Tahap preprocessing dilakukan sesuai dengan Sub bab 3.6.3, yang meliputi:

1. Seleksi data, yaitu memilih data siswa yang memenuhi kriteria penelitian.
2. Pembersihan data dari data kosong, tidak konsisten, dan duplikasi.
3. Normalisasi data untuk menyamakan skala nilai akademik dan data pelanggaran. Tahap ini bertujuan untuk mengurangi potensi bias dan meningkatkan akurasi proses clustering.

3.5. Penentuan Jumlah Cluster

Jumlah cluster yang digunakan dalam penelitian ini adalah tiga cluster ($K = 3$), sesuai dengan Subbab 3.8.1. Penentuan ini didasarkan pada kebutuhan pihak sekolah untuk mengelompokkan siswa ke dalam kategori yang mudah dipahami dan aplikatif dalam proses pembinaan.

3.6. Proses Penerapan Algoritma K-Means

Proses penerapan algoritma K-Means pada penelitian ini dilakukan dengan menggunakan data siswa yang disajikan pada **Tabel 4.3**, yang

terdiri dari dua variabel utama, yaitu **nilai akademik (X1)** dan **jumlah pelanggaran siswa (X2)**. Data tersebut digunakan untuk mengelompokkan siswa ke dalam tiga cluster berdasarkan tingkat permasalahan akademik.

1. Penentuan Centroid Awal

Berdasarkan data pada Tabel 4.3, jumlah cluster yang digunakan adalah **tiga (K = 3)**, yaitu:

1. Cluster siswa tidak bermasalah
2. Cluster siswa berisiko
3. Cluster siswa bermasalah

Centroid awal dipilih secara **acak** dari data siswa yang terdapat pada Tabel 4.3. Pemilihan secara acak dilakukan karena pada tahap awal belum diketahui pola pengelompokan siswa. Sebagai contoh, centroid awal ditentukan sebagai berikut:

- **C1 = (70, 0)** → nilai akademik tinggi, pelanggaran rendah
- **C2 = (65, 2)** → nilai akademik sedang, pelanggaran sedang
- **C3 = (60, 3)** → nilai akademik rendah, pelanggaran tinggi

Centroid awal ini berfungsi sebagai **titik pusat sementara** yang akan diperbaiki melalui proses iterasi.

2. Perhitungan Jarak (Euclidean Distance)

Setelah centroid awal ditentukan, langkah berikutnya adalah menghitung jarak antara setiap siswa pada Tabel 4.3 dengan masing-masing centroid menggunakan **rumus Euclidean Distance**:

$$D(x,y)=(x_1-y_1)^2+(x_2-y_2)^2$$

Keterangan:

- x_1, x_2 = nilai akademik siswa
- x_1, x_2 = jumlah pelanggaran siswa
- y_1, y_2 = nilai akademik centroid
- y_1, y_2 = jumlah pelanggaran centroid

Contoh perhitungan:

Siswa Muhamaad Putra N. memiliki nilai akademik 68 dan pelanggaran 1.

- Jarak ke C1 = $\sqrt{((68-70)^2 + (1-0)^2)}$
- Jarak ke C2 = $\sqrt{((68-65)^2 + (1-2)^2)}$
- Jarak ke C3 = $\sqrt{((68-60)^2 + (1-3)^2)}$

Perhitungan ini dilakukan untuk seluruh siswa pada Tabel 4.3

3. Pengelompokan Data ke dalam Cluster

Setelah jarak dihitung, setiap siswa dikelompokkan ke dalam **cluster dengan jarak terdekat** terhadap centroid.

Sebagai contoh:

- Siswa dengan **nilai akademik tinggi dan pelanggaran rendah** cenderung masuk ke **cluster C1 (tidak bermasalah)**
- Siswa dengan **nilai akademik sedang dan pelanggaran sedang** masuk ke **cluster C2 (berisiko)**
- Siswa dengan **nilai akademik rendah dan pelanggaran tinggi** masuk ke **cluster C3 (bermasalah)**

Hasil tahap ini berupa **pengelompokan awal siswa** berdasarkan kemiripan karakteristik.

4. Pembaruan Centroid

Setelah seluruh siswa dikelompokkan, centroid baru dihitung berdasarkan **rata-rata nilai akademik dan pelanggaran siswa dalam setiap cluster**.

Contoh:

Jika cluster C1 terdiri dari dua siswa dengan data (68,1) dan (70,0), maka centroid baru dihitung sebagai:

- Nilai akademik = $(68 + 70) / 2 = 69$
- Pelanggaran = $(1 + 0) / 2 = 0,5$

Pembaruan centroid ini bertujuan agar pusat cluster semakin merepresentasikan anggota cluster tersebut.

5. Iterasi dan Konvergensi

Tahapan perhitungan jarak, pengelompokan data, dan pembaruan centroid dilakukan secara **berulang (iteratif)**. Proses iterasi dihentikan apabila:

1. Tidak terjadi perubahan keanggotaan cluster, atau
2. Nilai centroid sudah stabil

Berdasarkan data pada Tabel 4.3, proses iterasi berhenti pada **iterasi kedua**, karena tidak terdapat perubahan pengelompokan siswa. Hal ini menunjukkan bahwa **cluster telah konvergen** dan hasil clustering dinyatakan final.

3.7 Hasil Clustering dan Pembahasan

Hasil clustering menunjukkan bahwa siswa MTs Muhammadiyah Ogan Lima dapat dikelompokkan sebagai berikut:

1. **Cluster 1 (Siswa Tidak Bermasalah)**
Siswa dalam cluster ini memiliki nilai akademik tinggi, tingkat kehadiran yang baik, nilai sikap yang positif, serta jumlah pelanggaran yang sangat rendah atau tidak ada. Kelompok ini menunjukkan kondisi akademik dan perilaku yang baik.
2. **Cluster 2 (Siswa Berisiko)**
Siswa pada cluster ini memiliki nilai akademik sedang dan mulai menunjukkan

beberapa pelanggaran ringan. Kelompok ini memerlukan perhatian dan pembinaan agar tidak berkembang menjadi siswa bermasalah.

3. Cluster 3 (Siswa Bermasalah)

Cluster ini terdiri dari siswa dengan nilai akademik rendah, tingkat kehadiran yang kurang, nilai sikap rendah, serta jumlah dan jenis pelanggaran yang tinggi. Kelompok ini memerlukan penanganan khusus dari pihak sekolah.

Hasil ini menunjukkan bahwa metode clustering dengan algoritma K-Means mampu mengidentifikasi kelompok siswa bermasalah secara objektif berdasarkan data akademik dan pelanggaran siswa.

3.8 Pembahasan Hasil Clustering Berdasarkan Variabel Penelitian

Pembahasan ini dilakukan untuk mengetahui pengaruh masing-masing variabel penelitian terhadap hasil pengelompokan siswa menggunakan algoritma K-Means. Variabel yang dianalisis meliputi nilai akademik siswa, jumlah pelanggaran, jenis pelanggaran, tingkat kehadiran, dan nilai sikap siswa.

3.9 Pembahasan Berdasarkan Nilai Akademik Siswa

Nilai akademik merupakan variabel utama dalam penelitian ini dan berperan besar dalam proses clustering. Berdasarkan hasil pengelompokan, siswa pada Cluster 1 memiliki nilai akademik yang tinggi dan relatif stabil, yang menunjukkan kemampuan belajar yang baik serta tingkat pemahaman materi yang optimal.

Siswa pada Cluster 2 memiliki nilai akademik pada kategori sedang. Kelompok ini menunjukkan adanya potensi risiko, karena meskipun nilai akademik tidak terlalu rendah, namun terdapat kecenderungan penurunan prestasi apabila tidak dilakukan pembinaan secara berkelanjutan.

Sementara itu, Cluster 3 didominasi oleh siswa dengan nilai akademik rendah. Rendahnya nilai akademik pada cluster ini menjadi indikator utama adanya permasalahan dalam proses belajar, yang berpotensi berkaitan dengan rendahnya motivasi belajar dan kedisiplinan siswa.

3.10 Pembahasan Berdasarkan Jumlah Pelanggaran Siswa

Jumlah pelanggaran siswa menunjukkan tingkat kedisiplinan siswa terhadap tata tertib sekolah. Hasil clustering menunjukkan bahwa siswa pada Cluster 1 memiliki jumlah pelanggaran yang sangat rendah atau bahkan tidak memiliki catatan pelanggaran.

Pada Cluster 2, jumlah pelanggaran siswa berada pada tingkat sedang dan umumnya berupa pelanggaran ringan. Kondisi ini menunjukkan bahwa siswa dalam cluster ini masih dapat dibina melalui pendekatan preventif dan pengawasan rutin.

Sebaliknya, Cluster 3 memiliki jumlah pelanggaran yang tinggi dan dilakukan secara berulang. Hal ini memperkuat klasifikasi cluster ini sebagai kelompok siswa bermasalah yang membutuhkan penanganan khusus dan intensif.

3.11. Pembahasan Berdasarkan Jenis Pelanggaran Siswa

Jenis pelanggaran siswa juga menjadi variabel penting dalam membedakan tingkat permasalahan siswa. Siswa pada Cluster 1 umumnya tidak melakukan pelanggaran atau hanya melakukan pelanggaran yang bersifat sangat ringan.

Pada Cluster 2, pelanggaran yang dilakukan cenderung berupa pelanggaran ringan hingga sedang, seperti keterlambatan atau kurang disiplin dalam mengikuti kegiatan belajar.

Sementara itu, Cluster 3 didominasi oleh siswa yang melakukan pelanggaran berat, seperti bolos sekolah, berkelahi, dan melawan guru. Jenis pelanggaran ini menunjukkan adanya masalah perilaku yang serius dan berdampak langsung pada proses pembelajaran.

3.12 Pembahasan Berdasarkan Tingkat Kehadiran Siswa

Tingkat kehadiran siswa menggambarkan kedisiplinan siswa dalam mengikuti kegiatan pembelajaran. Berdasarkan hasil clustering, Cluster 1 memiliki tingkat kehadiran yang tinggi dan konsisten, yang berbanding lurus dengan nilai akademik yang baik.

Siswa pada Cluster 2 memiliki tingkat kehadiran yang cukup, namun masih terdapat beberapa ketidakhadiran tanpa keterangan yang dapat mempengaruhi hasil belajar.

Pada Cluster 3, tingkat kehadiran siswa relatif rendah. Rendahnya kehadiran ini berkontribusi terhadap rendahnya nilai akademik dan meningkatnya jumlah pelanggaran siswa.

3.14 Pembahasan Berdasarkan Nilai Sikap Siswa

Nilai sikap siswa mencerminkan perilaku dan karakter siswa selama mengikuti proses pembelajaran. Hasil penelitian menunjukkan bahwa siswa pada Cluster 1 memiliki nilai sikap yang baik, mencerminkan tanggung jawab, disiplin, dan sikap positif terhadap guru dan lingkungan sekolah.

Pada Cluster 2, nilai sikap siswa berada pada kategori cukup. Kelompok ini masih memerlukan pembinaan karakter agar tidak berkembang menjadi siswa bermasalah.

Sementara itu, Cluster 3 memiliki nilai sikap yang rendah. Rendahnya nilai sikap ini sejalan dengan tingginya tingkat pelanggaran dan rendahnya prestasi akademik siswa.

3.15 Pembahasan Keseluruhan Hasil Clustering

Secara keseluruhan, hasil clustering menunjukkan bahwa kombinasi variabel nilai akademik, pelanggaran, kehadiran, dan sikap siswa mampu membentuk kelompok siswa dengan karakteristik yang jelas. Algoritma K-Means berhasil mengelompokkan siswa secara objektif tanpa label awal, sehingga hasil penelitian ini dapat dijadikan dasar dalam pengambilan keputusan oleh pihak sekolah.

4 KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada Bab hasil dan pembahasan, serta merujuk pada tujuan dan rumusan masalah yang disampaikan pada Bab pendahuluan dengan metodologi pada Bab metodologi penelitian, maka dapat ditarik beberapa kesimpulan sebagai berikut:

- Metode clustering dengan algoritma K-Means dapat diterapkan secara efektif untuk mengidentifikasi siswa bermasalah berdasarkan nilai akademik dan data pelanggaran siswa di MTs Muhammadiyah Ogan Lima.
- Proses penerapan algoritma K-Means yang meliputi penentuan centroid awal, perhitungan jarak menggunakan Euclidean Distance, pengelompokan data, pembaruan centroid, serta iterasi hingga konvergensi telah dilaksanakan secara sistematis dan konsisten dengan tahapan metodologi penelitian.
- Hasil clustering membagi siswa ke dalam tiga cluster, yaitu siswa tidak bermasalah, siswa berisiko, dan siswa bermasalah. Pengelompokan ini sesuai dengan kebutuhan sekolah dalam melakukan pemetaan kondisi siswa secara objektif.

- Pengelompokan yang dihasilkan menunjukkan bahwa kombinasi nilai akademik dan jumlah pelanggaran dapat digunakan sebagai indikator objektif dalam mengidentifikasi siswa yang memerlukan pembinaan dan pendampingan khusus.

Berdasarkan hasil penelitian yang telah dilakukan, maka penulis memberikan beberapa saran sebagai berikut:

1. Pihak sekolah diharapkan dapat memanfaatkan hasil clustering sebagai bahan pertimbangan dalam menyusun program pembinaan dan pendampingan siswa secara lebih terarah dan tepat sasaran.
2. Guru bimbingan konseling dan wali kelas dapat menggunakan hasil pengelompokan ini sebagai alat bantu dalam melakukan identifikasi dini terhadap siswa yang berisiko mengalami permasalahan akademik maupun perilaku.
3. Penelitian selanjutnya disarankan untuk menambahkan variabel lain, seperti tingkat kehadiran, latar belakang keluarga, atau aspek psikologis siswa agar hasil clustering menjadi lebih komprehensif.
4. Pengembangan penelitian ke arah sistem informasi berbasis komputer sangat disarankan agar proses identifikasi siswa bermasalah dapat dilakukan secara otomatis, cepat, dan berkelanjutan.

DAFTAR PUSTAKA

- [1] Awalia, R. (2024). Penerapan Metode K-Means Clustering untuk Pengelompokan Prestasi Siswa menggunakan Orange Data Mining: Studi Kasus di MTs Muhammadiyah Tallo Makassar. *Mechatronics Journal in Professional and Entrepreneur*, 6(2), 36–41.
- [2] Azzahra, C., & Sriani, S. (2025). Clustering of High School Students Academic Scores Using K-Means Algorithm. *Journal of Information Systems and Informatics*, 7(1), 572–586. <https://doi.org/10.51519/journalisi.v7i1.1029>
- [3] Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *Aiti*, 18(2), 125–138. <https://doi.org/10.24246/aiti.v18i2.125-138>

- [4] Fauzie, M. A. Al, Yuliadi, & Putra, J. A. (2023). Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 642–648. <https://doi.org/10.30865/klik.v4i1.1116>
- [5] Miftahul, M., Bengkak, U., Clustering, K., Indonesia, B., Inggris, B., & Clustering, K. (2025). *secara objektif dan efisien. Latar belakang penelitian ini adalah masih digunakannya metode manual dalam penentuan siswa berprestasi yang cenderung subjektif dan memakan waktu lama. 4.*
- [6] Muhammad, R., & Zakaria, H. (2023). Penerapan Algoritma K-Means dalam Penentuan Siswa Bermasalah Berdasarkan Running Record (Studi Kasus: SMK Averus Jakarta). *JRIIN: Jurnal Riset Informatika Dan ...*, 1(7), 720–729. <http://jurnalmahasiswa.com/index.php/jriin/article/view/819>
- [7] Han, J., Kamber, M., & Pei, J. (2018). *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann.
- [8] Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2020). *Pedoman Penilaian Hasil Belajar Peserta Didik*. Jakarta: Kemendikbud.
- [9] Maghfiroh, S., & Fatah, Z. (2025). *Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTs Miftahul Ulum Bengkak*. *Jurnal Sistem Informasi dan Pendidikan*, 6(1), 30–40.
- [10] Munir. (2017). *Statistika Pendidikan*. Yogyakarta: Pustaka Pelajar.
- [11] Nugroho, A. (2016). *Konsep dan Implementasi Data Mining*. Yogyakarta: Andi Offset.
- [12] Santrock, J. W. (2014). *Psikologi Pendidikan*. Jakarta: Salemba Humanika.
- [13] Slameto. (2015). *Belajar dan Faktor-Faktor yang Mempengaruhinya*. Jakarta: Rineka Cipta.
- [14] Sugiyono. (2020). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- [15] Turban, E., Sharda, R., & Delen, D. (2017). *Decision Support and Business Intelligence Systems*. New Jersey: Pearson Education.
- [16] Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann.
- [17] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. USA: Morgan Kaufmann.
- [18] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- [19] Kusriani & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [20] Larose, D. T. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons.
- [21] Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset.
- [22] Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [23] Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- [24] Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining*. Boston: Pearson Education.
- [25] Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann.
- [26] Dokumen Resmi MTs Muhammadiyah Ogan Lima. (2025). *Data Nilai Akademik dan Catatan Pelanggaran Siswa Tahun Ajaran 2025/2026*. Ogan Lima.