Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume. 6 Nomor. 2 Juli 2026
p-ISSN: 2827-8135; e-ISSN: 2827-7953; Hal 46-57
DOI: https://doi.org/10.55606/juisik.v6i2.2361
Tersedia: https://journal.sinov.id/index.php/juisik

Analisis Siswa Bermasalah Menggunakan Algoritma K Means

Abdul Haris Rosa
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi, Indonesia
Penulis Korespondensi: elharyz09@gmail.com

Abstract*. Problematic student identification in schools often relies on manual observation, fragmented counseling records, and subjective judgment, making early intervention less consistent. This study aims to analyze student problem patterns using the K-Means clustering algorithm so that schools can classify students into interpretable risk groups and design more targeted guidance services. The research used a quantitative data mining approach with an illustrative anonymized dataset consisting of 180 student records. Six attributes were modeled, namely unexplained absence, tardiness, disciplinary violations, average academic score, counseling record frequency, and task completion rate. The analytical process followed the CRISP-DM stages, including problem understanding, data preparation, normalization, determination of cluster number using elbow and silhouette evaluation, K-Means modeling, and interpretation of cluster profiles. The optimal structure was obtained at three clusters with a silhouette score of 0.58. The clusters were interpreted as low-risk monitoring students, medium-risk students requiring periodic assistance, and high-risk students requiring priority intervention. The high-risk cluster represented 18.9% of records and was characterized by higher absence, higher tardiness, more disciplinary violations, lower average scores, more counseling notes, and lower task completion. The findings indicate that K-Means can support school counseling by transforming administrative data into practical decision categories. The model should be validated with actual school data before operational use, but it provides a systematic framework for evidence-based student support.*

Keywords*: Counseling; Educational Data Mining; K-Means; Student Clustering; Student Problems.*

Abstrak. Identifikasi siswa bermasalah di sekolah masih sering dilakukan melalui pengamatan manual, catatan bimbingan konseling yang tersebar, dan pertimbangan subjektif, sehingga intervensi dini belum selalu konsisten. Penelitian ini bertujuan menganalisis pola siswa bermasalah menggunakan algoritma K-Means agar sekolah dapat mengelompokkan siswa ke dalam kategori risiko yang mudah ditafsirkan dan sesuai untuk dasar penyusunan layanan pembinaan. Penelitian menggunakan pendekatan kuantitatif berbasis data mining dengan dataset anonim ilustratif yang terdiri atas 180 data siswa. Atribut yang dianalisis meliputi jumlah ketidakhadiran tanpa keterangan, keterlambatan, pelanggaran tata tertib, nilai rata-rata akademik, frekuensi catatan konseling, dan persentase ketuntasan tugas. Tahapan analisis mengikuti kerangka CRISP-DM, yaitu pemahaman masalah, persiapan data, normalisasi, penentuan jumlah klaster melalui metode elbow dan silhouette, pemodelan K-Means, serta interpretasi profil klaster. Hasil evaluasi menunjukkan jumlah klaster optimal sebanyak tiga klaster dengan skor silhouette 0,58. Ketiga klaster ditafsirkan sebagai siswa risiko rendah yang cukup dipantau rutin, siswa risiko sedang yang memerlukan pendampingan berkala, dan siswa risiko tinggi yang memerlukan intervensi prioritas. Klaster risiko tinggi mencakup 18,9% data dengan karakteristik absensi, keterlambatan, pelanggaran, dan catatan konseling lebih tinggi, sedangkan nilai akademik dan ketuntasan tugas lebih rendah. Temuan ini menunjukkan bahwa K-Means dapat membantu guru BK mengubah data administratif menjadi informasi praktis. Model tetap perlu divalidasi menggunakan data sekolah aktual sebelum digunakan sebagai dasar keputusan layanan.

Kata kunci: Bimbingan Konseling; Data Mining Pendidikan; Klasterisasi Siswa; K-Means; Siswa Bermasalah.

1. LATAR BELAKANG

Permasalahan siswa di sekolah tidak selalu muncul secara tiba-tiba, tetapi biasanya terbentuk melalui akumulasi indikator yang dapat diamati, seperti ketidakhadiran, keterlambatan, penurunan nilai, pelanggaran tata tertib, dan meningkatnya kebutuhan layanan konseling. Dalam praktik sekolah, indikator tersebut sering berada pada sumber data yang berbeda, misalnya buku presensi, jurnal guru piket, catatan wali kelas, catatan bimbingan konseling, dan daftar nilai. Akibatnya, sekolah berpotensi terlambat mengenali siswa yang

Naskah Masuk: 16 April 2026; Revisi: 17 Mei 2026; Diterima: 07 Juni 2026; Tersedia: 19 Juni 2026

membutuhkan pendampingan lebih intensif. Kajian mutakhir tentang akuntabilitas konseling sekolah menegaskan bahwa layanan konseling perlu bergerak dari sekadar pelaporan aktivitas menuju praktik berbasis hasil, evaluasi program, pengambilan keputusan berbasis data, dan kolaborasi pemangku kepentingan untuk mendukung keberhasilan siswa (Isrofin et al., 2024).

Siswa bermasalah dalam konteks penelitian ini tidak dimaknai sebagai label negatif terhadap pribadi siswa, melainkan sebagai kondisi risiko yang terlihat dari data perilaku dan akademik. Penggunaan istilah tersebut diarahkan untuk membantu sekolah menyusun prioritas layanan, bukan untuk memberikan stigma. Pendekatan berbasis data penting karena guru BK dan wali kelas membutuhkan informasi ringkas mengenai siswa mana yang perlu dipantau, siswa mana yang perlu memperoleh pendampingan, dan siswa mana yang memerlukan intervensi kolaboratif bersama orang tua atau pihak lain. Penelitian terbaru juga menunjukkan bahwa konselor sekolah menghadapi tantangan dalam menganalisis volume data siswa yang kompleks, sehingga dukungan data mining diperlukan untuk memetakan siswa berisiko secara lebih terarah (Bakariwie et al., 2025).

Perkembangan educational data mining memberikan peluang bagi sekolah untuk mengolah data akademik dan nonakademik menjadi informasi yang lebih bermakna. Data mining mampu menemukan pola tersembunyi pada data pendidikan, termasuk pola prestasi, perilaku belajar, dan potensi risiko siswa (Zhang et al., 2021). Beberapa penelitian menunjukkan bahwa kombinasi data akademik dan nonakademik dapat meningkatkan pemahaman terhadap karakteristik siswa secara lebih holistik (Mohd Talib et al., 2023). Oleh karena itu, sekolah tidak cukup hanya melihat nilai akhir, tetapi juga perlu memperhatikan absensi, keterlambatan, kedisiplinan, partisipasi tugas, serta riwayat layanan konseling.

Salah satu metode yang relevan untuk tujuan tersebut adalah K-Means clustering. K-Means termasuk algoritma unsupervised learning yang mengelompokkan data ke dalam sejumlah klaster berdasarkan kedekatan karakteristik antarobjek. Dalam konteks sekolah, K-Means dapat digunakan untuk mengelompokkan siswa berdasarkan kemiripan indikator akademik dan perilaku tanpa membutuhkan label awal. Algoritma ini relatif mudah diterapkan, komputasinya efisien, dan hasilnya dapat diterjemahkan ke dalam kategori layanan yang praktis, seperti risiko rendah, risiko sedang, dan risiko tinggi (Ikotun et al., 2023; Mohamed Nafuri et al., 2022).

Beberapa penelitian terdahulu telah memanfaatkan K-Means untuk mengelompokkan siswa berdasarkan performa akademik. Rosita Dan Fitria (2025) menerapkan K-Means pada siswa SMK berdasarkan nilai mata pelajaran dan menghasilkan klaster bimbingan, potensial, serta unggulan. Khodijah et al. (2025) menggunakan K-Means dalam RapidMiner untuk

mengelompokkan performa akademik siswa menjadi tiga kelompok.Wibowo dan Aryanti (2025) juga menerapkan K-Means untuk klasifikasi prestasi akademik pada data sekolah swasta. Namun, sebagian besar penelitian tersebut masih berfokus pada nilai akademik, sedangkan penelitian tentang siswa bermasalah membutuhkan kombinasi indikator akademik dan perilaku agar hasilnya lebih sesuai dengan kebutuhan guru BK.

Kesenjangan penelitian tersebut menunjukkan perlunya model analisis yang tidak hanya mengelompokkan siswa berdasarkan nilai, tetapi juga memasukkan indikator kedisiplinan dan catatan layanan konseling. Dengan demikian, hasil klasterisasi dapat digunakan sebagai dasar penyusunan prioritas layanan yang lebih proporsional. Penelitian ini bertujuan menganalisis siswa bermasalah menggunakan algoritma K-Means dengan memadukan atribut absensi, keterlambatan, pelanggaran tata tertib, nilai akademik, frekuensi catatan konseling, dan ketuntasan tugas. Kontribusi penelitian terletak pada penyusunan kerangka analisis yang mudah direplikasi oleh sekolah, khususnya satuan pendidikan kejuruan, untuk mendukung layanan bimbingan konseling berbasis bukti.

2. KAJIAN TEORITIS

Educational data mining merupakan bidang kajian yang menerapkan teknik data mining, machine learning, statistik, dan visualisasi untuk memahami data pendidikan. Tujuannya bukan sekadar menghasilkan model komputasi, tetapi membantu pengambilan keputusan pendidikan, seperti pemetaan prestasi, prediksi risiko, personalisasi pembelajaran, dan evaluasi layanan (Zhang et al., 2021). Dalam konteks sekolah, data yang dapat dianalisis meliputi nilai, absensi, perilaku, aktivitas pembelajaran, data demografis terbatas, dan catatan intervensi. Penggunaan data harus tetap memperhatikan etika, kerahasiaan, dan prinsip minimisasi data karena data siswa termasuk informasi sensitif.

K-Means adalah metode klasterisasi partisional yang membagi data ke dalam k kelompok berdasarkan jarak terhadap pusat klaster atau centroid. Prosesnya dimulai dengan menentukan jumlah klaster, memilih centroid awal, menghitung jarak setiap data ke centroid, menempatkan data pada klaster terdekat, memperbarui centroid, dan mengulangi proses sampai perubahan klaster stabil. Tujuan utama K-Means adalah meminimalkan variasi dalam klaster sehingga anggota klaster memiliki karakteristik yang relatif serupa (Ikotun et al., 2023). Karena sensitif terhadap skala atribut, data numerik sebaiknya dinormalisasi sebelum diproses.

Penentuan jumlah klaster merupakan tahap penting karena jumlah klaster yang terlalu sedikit dapat menyederhanakan variasi data secara berlebihan, sedangkan jumlah klaster yang terlalu banyak dapat menyulitkan interpretasi. Dua pendekatan yang umum digunakan adalah

elbow method dan silhouette score. Elbow method melihat titik perubahan penurunan inertia, sedangkan silhouette score mengukur kedekatan data dengan klasternya sendiri dibandingkan dengan klaster lain. Dalam penelitian klasterisasi pendidikan, kombinasi elbow method dan silhouette analysis banyak digunakan untuk memvalidasi jumlah klaster agar hasil pengelompokan tidak hanya baik secara statistik, tetapi juga dapat ditafsirkan secara praktis (Ouassif et al., 2025; Rahma & Ulfah, 2025).

Konsep siswa bermasalah dalam penelitian ini dikaitkan dengan indikator risiko yang dapat ditindaklanjuti oleh sekolah. Indikator risiko bukan diagnosis psikologis, melainkan sinyal administratif yang menunjukkan kebutuhan pembinaan. Literatur konseling sekolah mutakhir menekankan pentingnya akuntabilitas, evaluasi hasil layanan, dan pemanfaatan data untuk menentukan prioritas bantuan yang lebih tepat sasaran (Isrofin et al., 2024). Dengan demikian, analisis K-Means dapat ditempatkan pada tahap identifikasi dan analisis awal untuk membantu guru BK membaca pola data secara objektif.

Penelitian Mohd Talib et al. (2023) menunjukkan bahwa K-Means dapat digunakan untuk mengidentifikasi pola perilaku dan performa siswa secara holistik sebelum dilanjutkan dengan model klasifikasi. Studi Mohamed Nafuri et al. (2022) mengembangkan beberapa model unsupervised learning, termasuk K-Means, BIRCH, dan DBSCAN, untuk klasifikasi performa mahasiswa. Penelitian Amalia et al. (2021) menerapkan K-Means pada data nilai akademik siswa SMP dan menampilkan hasilnya melalui dashboard, sehingga menunjukkan potensi integrasi klasterisasi dengan sistem informasi sekolah. Temuan-temuan tersebut menguatkan bahwa K-Means dapat menjadi metode awal untuk segmentasi siswa sebelum sekolah menentukan strategi layanan lanjutan.

Dalam konteks pendidikan vokasi,Rosita dan Fitria (2025) mengelompokkan siswa SMK berdasarkan nilai mata pelajaran ke dalam kelompok bimbingan, potensial, dan unggulan. Hasilnya menunjukkan bahwa klasterisasi dapat membantu guru menyusun strategi pembelajaran yang lebih sesuai dengan karakteristik kelompok. Namun, fokus pada nilai akademik saja belum mencerminkan seluruh persoalan siswa. Siswa yang memperoleh nilai cukup baik tetap dapat memiliki risiko perilaku, sedangkan siswa dengan nilai rendah belum tentu memiliki masalah kedisiplinan. Oleh karena itu, penelitian ini memasukkan indikator perilaku dan catatan konseling agar profil klaster lebih relevan dengan layanan pembinaan.

Kajian tentang early warning system juga menegaskan pentingnya identifikasi dini.Boudjehem dan Lafifi (2024) menyatakan bahwa sistem peringatan dini dapat membantu institusi pendidikan mengenali peserta didik berisiko sebelum kesulitan belajar berdampak lebih serius.Staneviciene et al. (2024) menunjukkan bahwa data mining dapat mendukung

prediksi performa siswa dan pengurangan risiko putus studi. Sementara itu,Liu et al. (2024) menegaskan bahwa analitik pembelajaran banyak digunakan untuk mengidentifikasi siswa yang berpotensi mengalami kegagalan akademik atau putus studi. Dalam penelitian ini, K-Means tidak digunakan untuk menghukum siswa, tetapi untuk mempercepat pemetaan prioritas bantuan.

Berdasarkan teori dan penelitian terdahulu, kerangka konseptual penelitian ini memandang bahwa data akademik dan data perilaku dapat membentuk profil risiko siswa. Atribut absensi, keterlambatan, pelanggaran, nilai, catatan konseling, dan ketuntasan tugas diproses melalui tahapan pembersihan data, normalisasi, klasterisasi, dan interpretasi. Hasil klaster kemudian diterjemahkan menjadi kategori layanan. Dengan cara ini, hasil algoritma tidak berhenti sebagai angka, tetapi menjadi informasi yang dapat dipahami oleh guru BK, wali kelas, dan pihak manajemen sekolah.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining dengan algoritma K-Means. Karena data riil sekolah belum disertakan dalam naskah awal ini, data yang digunakan pada bagian hasil merupakan dataset anonim ilustratif yang disusun untuk menggambarkan alur analisis dan bentuk keluaran yang dapat diterapkan pada data sekolah aktual. Pada tahap implementasi nyata, seluruh nilai, jumlah siswa, dan hasil klaster harus diganti menggunakan data primer atau data administrasi sekolah yang telah memperoleh izin dan telah dianonimkan.

Objek penelitian adalah data siswa pada satuan pendidikan menengah kejuruan yang berkaitan dengan indikator akademik dan perilaku. Unit analisis adalah satu siswa pada satu periode semester. Atribut yang digunakan terdiri atas jumlah ketidakhadiran tanpa keterangan, jumlah keterlambatan, jumlah pelanggaran tata tertib, nilai rata-rata akademik, frekuensi catatan konseling, dan persentase ketuntasan tugas. Keenam atribut tersebut dipilih karena mudah ditemukan pada administrasi sekolah serta relevan dengan proses identifikasi kebutuhan layanan BK.

Tabel 1. Atribut Penelitian dan Definisi Operasional.

Atribut	Definisi operasional	Arah interpretasi risiko
Ketidakhadiran	Jumlah absensi tanpa keterangan dalam satu semester	Semakin tinggi, semakin berisiko
Keterlambatan	Jumlah keterlambatan hadir di sekolah	Semakin tinggi, semakin berisiko
Pelanggaran	Jumlah catatan pelanggaran tata tertib	Semakin tinggi, semakin berisiko
Nilai akademik	Rata-rata nilai akademik siswa	Semakin rendah, semakin berisiko
Catatan konseling	Frekuensi pencatatan layanan atau rujukan konseling	Semakin tinggi, semakin perlu perhatian
Ketuntasan tugas	Persentase tugas yang diselesaikan	Semakin rendah, semakin berisiko

Tahapan penelitian mengikuti kerangka CRISP-DM yang terdiri atas pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, dan interpretasi hasil. Kerangka ini dipilih karena lazim digunakan dalam proyek data mining dan mampu menghubungkan kebutuhan bisnis atau organisasi dengan proses analitik (Schröer et al., 2021). Pada tahap pemahaman masalah, kebutuhan sekolah dirumuskan sebagai kebutuhan pemetaan siswa berdasarkan tingkat risiko. Pada tahap pemahaman data, atribut administrasi yang relevan diidentifikasi. Pada tahap persiapan data, data dibersihkan, disamarkan, dan dinormalisasi.

Alur penelitian yang digunakan ditampilkan pada Gambar 1. Alur tersebut menempatkan K-Means sebagai proses inti setelah data siswa dibersihkan dan dinormalisasi, kemudian dilanjutkan dengan evaluasi dan interpretasi layanan.



Kerangka kerja analisis siswa bermasalah berbasis data mining pendidikan

Gambar 1. Alur Penelitian K-Means untuk Analisis Siswa Bermasalah.

Normalisasi dilakukan menggunakan min-max scaling agar seluruh atribut berada pada rentang yang sebanding. Atribut nilai akademik dan ketuntasan tugas memiliki arah positif, sehingga keduanya dibaca secara berlawanan ketika menafsirkan risiko. Artinya, nilai lebih rendah dan ketuntasan tugas lebih rendah berhubungan dengan risiko lebih tinggi. Setelah data siap, K-Means dijalankan untuk beberapa variasi nilai k, yaitu k=2 sampai k=6. Nilai k terbaik dipilih melalui kombinasi elbow method, silhouette score, serta pertimbangan interpretabilitas bagi layanan sekolah.

K-Means bekerja dengan meminimalkan jumlah kuadrat jarak data terhadap centroid klaster. Secara konseptual, siswa yang memiliki pola absensi, keterlambatan, pelanggaran, nilai, catatan konseling, dan ketuntasan tugas yang mirip akan berada pada klaster yang sama. Setelah klaster terbentuk, setiap klaster diberi label interpretatif berdasarkan profil centroid. Label interpretatif tidak berasal dari algoritma secara otomatis, tetapi diberikan oleh peneliti melalui pembacaan karakteristik klaster dan relevansinya dengan layanan bimbingan konseling.

Evaluasi hasil dilakukan dengan dua cara. Pertama, evaluasi kuantitatif melalui inertia dan silhouette score untuk melihat kualitas pemisahan klaster. Kedua, evaluasi substantif melalui keterbacaan profil klaster, yaitu apakah klaster yang dihasilkan dapat diterjemahkan menjadi tindakan sekolah. Validitas praktis menjadi penting karena model data mining di sekolah harus membantu pengambilan keputusan, bukan sekadar menghasilkan pemisahan statistik. Oleh karena itu, hasil akhir penelitian disajikan dalam bentuk profil klaster dan rekomendasi intervensi.

4. HASIL DAN PEMBAHASAN

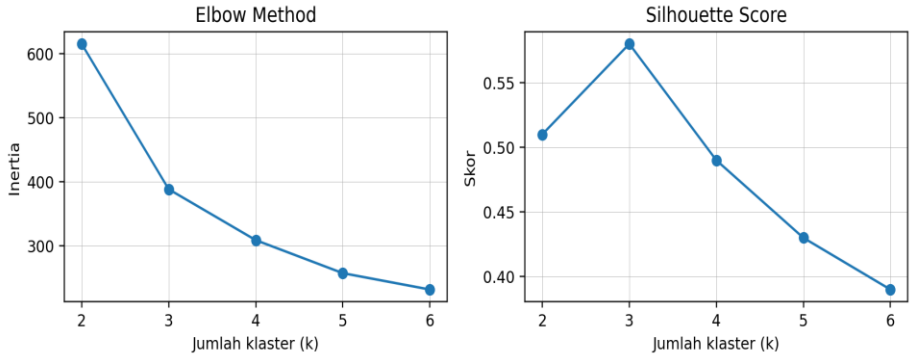
Dataset ilustratif terdiri atas 180 data siswa. Setelah dilakukan pemeriksaan awal, seluruh atribut bersifat numerik dan tidak terdapat nilai kosong. Data kemudian dinormalisasi agar atribut dengan satuan berbeda tidak mendominasi proses penghitungan jarak. Pada data riil, tahap ini perlu dilengkapi dengan pengecekan duplikasi, pemeriksaan outlier, validasi konsistensi catatan absensi, serta penyamaran identitas siswa. Identitas personal seperti nama, NISN, alamat, nomor telepon, dan informasi keluarga tidak digunakan dalam pemodelan agar analisis tetap berada pada prinsip perlindungan data siswa.

Hasil pengujian jumlah klaster menunjukkan bahwa k=3 merupakan pilihan paling seimbang. Pada elbow method, penurunan inertia paling besar terjadi dari k=2 ke k=3 dan mulai melandai setelah k=3. Pada silhouette score, nilai tertinggi juga diperoleh pada k=3, yaitu 0,58. Nilai tersebut menunjukkan pemisahan klaster yang cukup baik untuk data pendidikan yang bersifat heterogen. Selain itu, tiga klaster lebih mudah diterjemahkan ke dalam kategori layanan sekolah, yaitu pemantauan rutin, pendampingan berkala, dan intervensi prioritas.

Tabel 2. Evaluasi Jumlah Klaster.

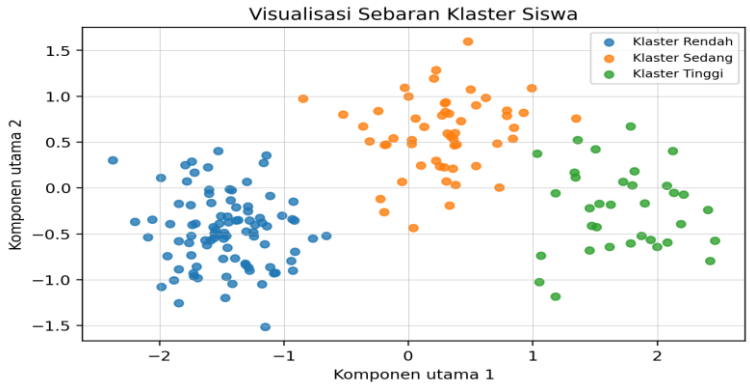
Jumlah k	Inertia	Silhouette score	Interpretasi
2	615,4	0,51	Pemisahan awal cukup, tetapi kategori layanan terlalu umum
3	388,2	0,58	Paling seimbang antara kualitas klaster dan interpretasi layanan
4	308,7	0,49	Klaster bertambah, tetapi pemisahan menurun
5	257,1	0,43	Interpretasi mulai tumpang tindih
6	231,5	0,39	Terlalu rinci untuk kebutuhan layanan sekolah

Grafik elbow dan silhouette pada Gambar 2 memperlihatkan bahwa k=3 menjadi titik paling logis karena penurunan inertia mulai melandai setelah tiga klaster dan skor silhouette mencapai nilai tertinggi.



Gambar 2. Evaluasi Elbow Method dan Silhouette Score.

Visualisasi sebaran klaster menunjukkan bahwa setiap kelompok memiliki kecenderungan yang berbeda. Klaster risiko rendah cenderung terkonsentrasi pada area dengan indikator perilaku dan akademik yang lebih stabil. Klaster risiko sedang berada di posisi transisi karena sebagian indikator mulai menunjukkan masalah, tetapi belum separah klaster risiko tinggi. Klaster risiko tinggi terpisah lebih jelas karena memiliki akumulasi absensi, keterlambatan, pelanggaran, dan catatan konseling yang lebih tinggi, disertai penurunan nilai akademik dan ketuntasan tugas.



Gambar 3. Visualisasi Sebaran Klaster Siswa.

Profil centroid menunjukkan bahwa klaster rendah terdiri atas 92 siswa atau 51,1% data. Kelompok ini memiliki rata-rata ketidakhadiran 2,1 kali, keterlambatan 2,6 kali, pelanggaran 0,8 kali, nilai rata-rata 82,4, catatan konseling 0,4 kali, dan ketuntasan tugas 91,2%. Klaster sedang terdiri atas 54 siswa atau 30,0% data dengan rata-rata ketidakhadiran 6,8 kali, keterlambatan 7,5 kali, pelanggaran 2,6 kali, nilai rata-rata 74,1, catatan konseling 1,8 kali, dan ketuntasan tugas 76,4%. Klaster tinggi terdiri atas 34 siswa atau 18,9% data dengan rata-rata

ketidakhadiran 14,6 kali, keterlambatan 16,8 kali, pelanggaran 6,9 kali, nilai rata-rata 65,2, catatan konseling 4,6 kali, dan ketuntasan tugas 58,9%.

Tabel 3. Profil Centroid dan Jumlah Anggota Klaster.

Profil Klaster	Jumlah Siswa	Absensi	Terlambat	Pelanggaran	Nilai	Konseling	Tugas
Risiko rendah	92 (51,1%)	2,1	2,6	0,8	82,4	0,4	91,2%
Risiko sedang	54 (30,0%)	6,8	7,5	2,6	74,1	1,8	76,4%
Risiko tinggi	34 (18,9%)	14,6	16,8	6,9	65,2	4,6	58,9%

Interpretasi hasil menunjukkan bahwa klaster risiko tinggi perlu memperoleh perhatian lebih cepat karena memiliki masalah yang bersifat multidimensi. Siswa pada kelompok ini tidak hanya menunjukkan penurunan akademik, tetapi juga mengalami persoalan disiplin dan keterlibatan belajar. Penanganannya tidak cukup melalui teguran umum, tetapi memerlukan asesmen individual, konseling, komunikasi dengan orang tua, koordinasi dengan wali kelas, dan pemantauan kemajuan secara berkala. Pada kelompok risiko sedang, strategi yang lebih sesuai adalah pendampingan preventif agar masalah tidak berkembang menjadi lebih serius.

Temuan ini sejalan dengan prinsip early warning system yang menekankan deteksi dini sebelum siswa mengalami kegagalan akademik atau putus sekolah (Boudjehem & Lafifi, 2024; Liu et al., 2024). K-Means membantu sekolah melihat bahwa risiko siswa tidak selalu biner, melainkan memiliki tingkat yang berbeda. Dengan demikian, sekolah dapat menggunakan sumber daya layanan secara lebih proporsional. Guru BK dapat mengalokasikan layanan konseling individual pada kelompok risiko tinggi, bimbingan kelompok pada kelompok risiko sedang, dan penguatan positif pada kelompok risiko rendah.

Tabel 4. Rekomendasi Tindak Lanjut Berdasarkan Klaster.

Klaster	Prioritas Layanan	Contoh Tindak Lanjut
Risiko rendah	Pemantauan rutin	Penguatan positif, informasi kelas, dan pemantauan presensi bulanan
Risiko sedang	Pendampingan berkala	Bimbingan kelompok, kontrak belajar, koordinasi wali kelas, dan pemantauan dua mingguan
Risiko tinggi	Intervensi prioritas	Konseling individual, konferensi kasus, komunikasi orang tua, dan rencana tindak lanjut terukur

Dibandingkan penelitian yang hanya menggunakan nilai akademik, model ini memberikan gambaran lebih luas karena memasukkan indikator perilaku dan keterlibatan tugas. Hasilnya lebih dekat dengan kebutuhan praktis sekolah karena guru BK biasanya menangani masalah yang muncul dari kombinasi berbagai faktor. Temuan ini mendukung penelitian Mohd Talib et al. (2023) yang menekankan pentingnya fitur holistik dalam memetakan performa dan perilaku siswa. Namun, model tetap memiliki keterbatasan karena K-Means sensitif terhadap pemilihan jumlah klaster, skala atribut, dan keberadaan outlier. Oleh

karena itu, hasil algoritma harus dipandang sebagai alat bantu, bukan sebagai satu-satunya dasar keputusan.

Secara terapan, penggunaan K-Means dapat dimasukkan ke dalam sistem informasi sekolah atau spreadsheet sederhana. Sekolah dapat memperbarui data setiap bulan, menjalankan proses normalisasi, menghasilkan klaster, dan membandingkan perpindahan siswa antarperiode. Siswa yang berpindah dari risiko rendah ke sedang dapat segera memperoleh layanan preventif. Siswa yang bertahan pada risiko tinggi selama beberapa periode dapat menjadi prioritas konferensi kasus. Dengan cara ini, data tidak hanya menjadi arsip administratif, tetapi menjadi dasar pengambilan keputusan yang lebih cepat dan akuntabel.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma K-Means dapat digunakan untuk menganalisis siswa bermasalah melalui pengelompokan data akademik dan perilaku. Berdasarkan dataset anonim ilustratif berjumlah 180 data siswa, jumlah klaster optimal adalah tiga klaster dengan silhouette score 0,58. Ketiga klaster tersebut dapat diinterpretasikan sebagai siswa risiko rendah, risiko sedang, dan risiko tinggi. Klaster risiko tinggi mencakup 18,9% data dan ditandai oleh tingginya absensi, keterlambatan, pelanggaran tata tertib, serta frekuensi catatan konseling, disertai nilai akademik dan ketuntasan tugas yang lebih rendah.

Hasil penelitian mengindikasikan bahwa K-Means dapat membantu guru BK dan manajemen sekolah menyusun prioritas layanan secara lebih objektif. Model ini bermanfaat sebagai alat bantu identifikasi awal, terutama ketika sekolah memiliki banyak data administratif tetapi belum memiliki mekanisme analisis yang sistematis. Meskipun demikian, hasil klaster tidak boleh digunakan untuk memberi label permanen kepada siswa. Keputusan layanan tetap harus mempertimbangkan asesmen guru BK, komunikasi dengan wali kelas, kondisi keluarga, dan perkembangan siswa dari waktu ke waktu.

Saran untuk penelitian berikutnya adalah menggunakan data sekolah aktual dengan izin resmi, memperluas atribut menjadi aspek akademik, perilaku, sosial, dan keterlibatan belajar, serta membandingkan K-Means dengan metode lain seperti DBSCAN, K-Medoids, atau hierarchical clustering. Penelitian lanjutan juga perlu menguji stabilitas klaster antarperiode, mengembangkan dashboard sederhana untuk guru BK, dan menyusun protokol etika agar pemanfaatan data siswa tetap menjaga kerahasiaan, keadilan, dan tujuan pembinaan.

DAFTAR REFERENSI

Amalia, N. L. R., Supianto, A. A., Setiawan, N. Y., Zilvan, V., Yuliani, A. R., & Ramdan, A. (2021). Student Academic Mark Clustering Analysis and Usability Scoring on Dashboard Development Using K-Means Algorithm and System Usability Scale. Jurnal Ilmu Komputer Dan Informasi, 14(2), 137–143. https://doi.org/10.21609/jiki.v14i2.980

Bakariwie, A., Asamoah, D., & Duwiejuah, A. B. (2025). Prevention of student attrition: A data-backed approach to school counselling using Delphi technique and multiple classification algorithms. Discover Education, 4(1), 259. https://doi.org/10.1007/s44217-025-00494-7

Boudjehem, R., & Lafifi, Y. (2024). An early warning system to predict dropouts inside e-learning environments. Education and Information Technologies, 29(13), 16365–16385. https://doi.org/10.1007/s10639-024-12498-1

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences, 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Isrofin, B., Taufiq, A., Ahman, A., Yustiana, Y. R., & Dudin, D. (2024). Exploring Accountability in School Counseling: Research Trends, Applied Models, and Implications for Practice in Indonesia. KONSELOR, 13(2), 91–107. https://doi.org/10.24036/0202413281-0-86

Khodijah, S., Hariyanto, A. R., Salsabiilah, B., Aulia, W. S., & Fanyusri, M. (2025). Klasterisasi Mahasiswa Berdasarkan Performa Akademik Menggunakan Algoritma K-Means pada RapidMiner: Studi Kasus dengan Dataset Student Academic Performance. Journal of Information Technology and Informatics Engineering, 1(1), 46–50.

Liu, M., Wong, B. T. M., & Li, K. C. (2024). A survey on predicting at-risk students through learning analytics. International Journal of Innovation and Learning, 36(5), 10065503. https://doi.org/10.1504/IJIL.2024.10065503

Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. Applied Sciences, 12(19), 9467. https://doi.org/10.3390/app12199467

Mohd Talib, N. I., Abd Majid, N. A., & Sahran, S. (2023). Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering and Support Vector Machine. Applied Sciences, 13(5), 3267. https://doi.org/10.3390/app13053267

Ouassif, K., Ziani, B., Herrera-Tapia, J., & Kerrache, C. A. (2025). Empowering Education: Leveraging Clustering and Recommendations for Enhanced Student Insights. Education Sciences, 15(7), 819. https://doi.org/10.3390/educsci15070819

Rahma, F., & Ulfah, S. Z. (2025). Clustering Students Based on Academic Performance and Social Factors: An Unsupervised Learning Approach to Identify Student Patterns. International Journal for Applied Information Management, 5(3), 139–154. https://doi.org/10.47738/ijaim.v5i3.109

Rosita, & Fitria. (2025). Student Clustering Based on Subject Grades: A K-Means Approach to Clustering Study Groups. Journal of Applied Informatics Research, 1(1), 21–29. https://doi.org/10.26740/jair.v1i1.44597

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science, CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Staneviciene, E., Gudoniene, D., Punys, V., & Kukstys, A. (2024). A Case Study on the Data Mining-Based Prediction of Students' Performance for Effective and Sustainable E-Learning. Sustainability, 16(23), 10442. https://doi.org/10.3390/su162310442

Wibowo, E. A., & Aryanti, R. (2025). Penerapan Metode Clustering K-Means Menggunakan RapidMiner untuk Klasifikasi Prestasi Siswa di Sekolah Swasta. Journal of Information Technology and Informatics Engineering, 1(1), 20–24.

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.698490