ANALISIS RISIKO AKADEMIK SISWA DENGAN METODE K-MEANS (STUDI KASUS : SMPN 10 PALEMBANG)

Raihana Nabilatulrahmah, Ayu Triana, Tarisha Ispahan, Adeliani, Suci Fitriani, Amelia, Ken Ditha Tania, Ahmad Rifai
Sistem Informasi, Universitas Sriwijaya
Jalan Raya Palembang - Prabumulih Km. 32 Indralaya,
Kabupaten Ogan Ilir, Sumatera Selatan-Indonesia
09031182227121@student.unsri.ac.id

ABSTRAK

Pendidikan memainkan peranan penting dalam meningkatkan kualitas sumber daya manusia, terutama di kalangan siswa Sekolah Menengah Pertama (SMP). Namun, tingkat pemahaman dan capaian akademik setiap siswa berbeda, yang dipengaruhi oleh beberapa faktor seperti kehadiran, partisipasi kelas, dan keterlibatan dalam tugas. Salah satu tantangan utama dalam dunia pendidikan adalah mengidentifikasi siswa yang berisiko mengalami kesulitan akademik agar dapat diberikan intervensi yang tepat. Penelitian ini dilakukan di SMPN 10 Palembang yang belum memiliki sistem berbasis data mining untuk mengelompokkan siswa berdasarkan tingkat risiko akademik. Penelitian ini bertujuan menganalisis faktor-faktor yang memengaruhi risiko akademik siswa kelas VII menggunakan metode K-Means Clustering. Hasil penelitian menunjukkan kelompok siswa yang terbagi ke dalam 3 cluster: tidak berisiko, berisiko sedang, dan berisiko tinggi, berdasarkan data tingkat kehadiran, nilai akademik, dan partisipasi ekstrakurikuler. Temuan ini dapat menjadi dasar bagi sekolah untuk merancang intervensi yang lebih efektif guna meningkatkan performa akademik dan keterlibatan siswa di sekolah.

Kata kunci: *Data Mining, K-Means Clustering, Risiko Akademik, Analisis Data, Educational Data Mining, Learning Analytics.*

1. PENDAHULUAN

Pendidikan berperan penting dalam membangun sumber daya manusia yang berkualitas. Namun, setiap siswa memiliki tingkat pemahaman dan capaian akademik yang berbeda, yang dipengaruhi oleh berbagai faktor seperti kehadiran, partisipasi kelas, dan keterlibatan dalam tugas. Salah satu tantangan utama dalam dunia pendidikan adalah mengidentifikasi siswa yang berisiko mengalami kesulitan akademik agar dapat diberikan intervensi yang tepat. Tanpa identifikasi dini, siswa berisiko tinggi dapat mengalami kesulitan dalam pembelajaran, yang berdampak pada hasil akademik mereka. Oleh karena itu, diperlukan metode yang efektif untuk menganalisis faktor risiko akademik siswa guna mendukung strategi pembelajaran yang lebih adaptif.

Seiring dengan perkembangan teknologi, penerapan data mining dalam dunia pendidikan semakin relevan dengan kemampuannya dalam mengolah data akademik secara sistematis dan menemukan pola tersembunyi yang berpengaruh terhadap prestasi belajar siswa [1]. Konsep dasar dan teknik dalam data mining itu sendiri dapat diterapkan dalam berbagai bidang, salah satunya adalah pendidikan untuk memberikan pemahaman yang mendalam tentang analisis data akademik [2]. Salah satu metode yang sering digunakan adalah K-Means Clustering, yang mampu mengelompokkan siswa berdasarkan kemiripan karakteristik akademik mereka, seperti nilai, kehadiran, dan partisipasi kelas. Dengan pendekatan ini, sekolah dapat mengidentifikasi kelompok siswa yang berisiko mengalami kesulitan akademik dan merancang strategi pembelajaran yang lebih efektif.

Pada SMPN 10 Palembang, belum tersedia sistem berbasis data mining untuk mengelompokkan siswa berdasarkan tingkat risiko akademik mereka. Oleh karena itu, penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi risiko akademik siswa kelas VII di SMPN 10 Palembang serta menerapkan metode K-Means Clustering untuk mengelompokkan mereka berdasarkan tingkat risiko akademik. Penelitian ini juga mengevaluasi efektivitas model yang dikembangkan dalam membantu sekolah mengidentifikasi siswa yang memerlukan intervensi akademik lebih lanjut. Melalui penelitian ini, diharapkan sekolah dapat menerapkan strategi pembelajaran yang lebih efektif guna meningkatkan kualitas pendidikan secara keseluruhan.

Artikel ini disusun dalam beberapa bagian utama. Pendahuluan menjelaskan latar belakang, permasalahan, dan tujuan penelitian. Literature Review mengulas penelitian terdahulu yang relevan dan membandingkan kontribusi penelitian ini. Metode Penelitian menjelaskan tahapan dalam penerapan metode K-Means Clustering. Hasil dan Pembahasan menampilkan hasil analisis clustering dan evaluasi model. Kesimpulan dan Saran merangkum temuan penelitian dan rekomendasi untuk penelitian selanjutnya.

2. TINJAUAN PUSTAKA

Berbagai penelitian sebelumnya telah menerapkan data mining dalam pendidikan, tetapi sebagian besar masih berfokus pada prediksi nilai

akademik tanpa mempertimbangkan faktor risiko akademik secara menyeluruh. Misalnya, penelitian oleh Sholeh et al. [4] hanya menggunakan regresi linier untuk memprediksi nilai ujian akhir tanpa mempertimbangkan faktor kehadiran dan partisipasi siswa. Sementara itu, penelitian oleh Sari et al. [5] lebih menitikberatkan pada analisis akademik tanpa mempertimbangkan faktor lain yang berpengaruh terhadap pembelajaran. Hal ini menunjukkan perlunya pendekatan yang lebih komprehensif dalam mengidentifikasi siswa berisiko akademik.

Beberapa penelitian telah menerapkan metode K-Means Clustering dalam bidang pendidikan untuk menganalisis dan mengelompokkan siswa berdasarkan performa akademik mereka. Salah satu studi mengombinasikan K-Means dengan Naïve Bayes untuk meningkatkan akurasi prediksi performa akademik siswa, menghasilkan model yang lebih akurat dibandingkan metode tunggal [6]. Studi lain menggunakan K-Means untuk mengelompokkan siswa di tingkat sekolah menengah berdasarkan nilai akademik mereka, yang kemudian dimanfaatkan sebagai dasar dalam pengambilan keputusan bimbingan akademik [7]. Selain itu, metode ini juga diterapkan dalam analisis tren pembelajaran siswa dari waktu ke waktu, memungkinkan sekolah untuk mengidentifikasi pola belajar dan menyesuaikan strategi pembelajaran yang lebih efektif [8].

Dalam penelitian lain, K-Means Clustering digunakan untuk menentukan kelompok siswa yang layak menerima beasiswa berdasarkan data nilai akademik. Hasilnya menunjukkan bahwa metode ini mampu mengelompokkan siswa secara objektif dan membantu institusi pendidikan dalam menyeleksi penerima bantuan akademik [9]. Pendekatan serupa juga diterapkan untuk mengelompokkan siswa berdasarkan tingkat performa, dari yang rendah hingga tinggi, sehingga memungkinkan intervensi akademik yang lebih tepat sasaran [10]. Saputra dan Nataliani mengonfirmasi bahwa K-Means dapat digunakan untuk mengelompokkan siswa berprestasi dengan tingkat akurasi yang cukup tinggi. Penelitian lainnya menemukan bahwa metode ini tidak hanya efektif dalam pengelompokan siswa, tetapi juga dapat dioptimalkan untuk menentukan strategi pembelajaran yang lebih adaptif [11].

Selain K-Means Clustering, berbagai metode lain telah digunakan dalam Educational Data Mining untuk memahami dan meningkatkan kualitas pembelajaran. Salah satu penelitian menerapkan Fuzzy C-Means Clustering untuk menganalisis faktor yang menyebabkan siswa sering membolos. Penelitian lain membandingkan K-Means dengan Decision Tree dalam pengelompokan siswa berprestasi, di mana kombinasi kedua metode ini menghasilkan klasifikasi yang lebih akurat [12].

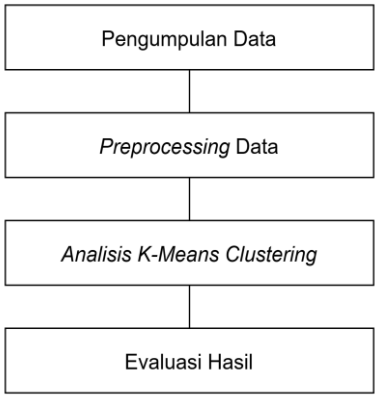
Beberapa studi juga menyoroti bagaimana Educational Data Mining dan Learning Analytics dapat digunakan untuk mengidentifikasi pola pembelajaran siswa serta memprediksi performa akademik mereka. Salah satu pendekatan yang digunakan adalah Regresi Linear, yang diterapkan untuk memprediksi nilai ujian berdasarkan tren akademik sebelumnya. Sementara itu, Azis [14] membandingkan berbagai algoritma machine learning dalam memprediksi performa akademik mahasiswa, menunjukkan bahwa pemilihan metode yang tepat dapat meningkatkan akurasi analisis akademik.

Kharis dan Zili [3] juga menunjukkan bahwa penerapan Learning Analytics dalam data pendidikan dapat mengidentifikasi pola belajar siswa serta mengembangkan sistem rekomendasi berbasis data mining menggunakan Weka dan Python. Secara keseluruhan, berbagai metode dalam EDM telah diterapkan untuk meningkatkan pemahaman terhadap pola pembelajaran siswa, mengoptimalkan strategi akademik, serta mendukung pengambilan keputusan yang lebih berbasis data di lingkungan pendidikan.

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan siswa berdasarkan tingkat risiko akademik mereka, dengan tahapan sebagai berikut:



Gambar 1. Tahapan penelitian

Metode ini telah banyak digunakan dalam berbagai penelitian serupa, seperti yang ditunjukkan dalam studi sebelumnya [13 [15], yang menunjukkan efektivitas K-Means dalam mengelompokkan data akademik secara objektif.

3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data siswa kelas VII di SMPN 10 Palembang. Observasi dilakukan dengan menganalisis dataset yang mencakup catatan akademik, kehadiran, dan partisipasi siswa. Data ini digunakan untuk memahami konteks akademik dan faktor-faktor yang dapat mempengaruhi risiko akademik siswa. Melalui pemahaman ini, penelitian dapat dirancang dengan lebih efektif dan relevan.

3.3. Preprocessing Data

Sebelum dilakukan proses clustering, data yang diperoleh akan melewati tahap preprocessing untuk memastikan kualitasnya. Proses preprocessing meliputi pembersihan data dengan menghapus data yang tidak relevan dan menangani nilai yang hilang (missing values). Selain itu, data juga dinormalisasi agar semua atribut memiliki skala yang seragam sehingga dapat meningkatkan akurasi hasil clustering.

3.4. Analisis K-Means Clustering

Pada tahap analisis K-Means Clustering, dilakukan pengelompokan siswa berdasarkan karakteristik yang relevan dengan risiko akademik. Hasil clustering ini dianalisis untuk mengidentifikasi kelompok siswa yang berisiko tinggi, sedang, dan rendah dalam hal kesulitan akademik. Setelah analisis selesai, evaluasi dilakukan untuk menilai efektivitas model dalam mengidentifikasi risiko akademik melalui analisis statistik deskriptif dan visualisasi data, yang membantu memahami pola dan tren yang muncul. Berdasarkan hasil analisis tersebut, rekomendasi tindakan dapat diberikan kepada pihak sekolah untuk merancang intervensi yang tepat bagi siswa yang berisiko, dengan harapan dapat meningkatkan performa akademik mereka secara keseluruhan.

3.5. Evaluasi Hasil

Tahap evaluasi hasil dilakukan untuk menilai efektivitas model dalam mengidentifikasi risiko akademik. Evaluasi ini mencakup analisis statistik deskriptif untuk memahami karakteristik masing-masing cluster, serta visualisasi data untuk memudahkan pemahaman pola dan tren yang muncul dari hasil clustering. Melalui penggunaan grafik atau diagram, peneliti dapat menggambarkan hasil secara lebih jelas. Berdasarkan analisis ini, rekomendasi tindakan dapat diberikan kepada pihak sekolah untuk merancang intervensi yang tepat bagi siswa yang berisiko, sehingga diharapkan dapat meningkatkan performa akademik mereka secara keseluruhan.

4. HASIL DAN PEMBAHASAN

4.1. Data *Cleaning*

Pada tahap data cleaning, dilakukan pemrosesan awal untuk memastikan data yang digunakan dalam analisis telah memenuhi standar kualitas yang diperlukan. Menurut Yağcı, "Data cleaning bertujuan untuk menghilangkan noise dan ketidakakuratan dalam data yang dapat mengganggu hasil analisis, terutama dalam konteks data pendidikan" [18]. Langkah-langkah utama dalam tahap ini meliputi penghapusan atribut yang tidak relevan, penanganan data yang tidak lengkap, serta pemeriksaan konsistensi data. Salah satu atribut yang dihilangkan dalam proses ini adalah Nama, karena atribut tersebut tidak memiliki kontribusi signifikan dalam proses analisis clustering. Keputusan untuk menghapus atribut ini bertujuan agar dataset lebih sederhana dan analisis lebih fokus pada atribut yang memiliki pengaruh terhadap hasil clustering. Selain itu, atribut yang mengandung data kosong atau tidak lengkap dianalisis lebih lanjut untuk memastikan bahwa tidak ada kehilangan informasi yang dapat mempengaruhi hasil akhir. Metode yang digunakan dalam proses pembersihan data ini melibatkan analisis deskriptif terhadap setiap atribut, termasuk pemeriksaan nilai ekstrem atau outlier yang dapat mengganggu pola clustering. Dengan melakukan data cleaning yang sistematis, dataset menjadi lebih siap untuk dianalisis secara optimal.

4.2. Data *Transformation*

Data yang telah dibersihkan kemudian mengalami proses transformasi agar sesuai dengan kebutuhan algoritma clustering. Salah satu transformasi utama adalah inisialisasi atau konversi data nominal ke dalam bentuk numerik agar dapat diproses oleh algoritma K-Means. Selain itu, normalisasi data sangat diperlukan, terutama dalam algoritma K-Means, yang menurut Masangu et al. "sangat sensitif terhadap skala data, di mana atribut dengan nilai yang lebih besar dapat mendominasi proses clustering jika normalisasi tidak dilakukan terlebih dahulu" [16]. Sebagai contoh, atribut Keaktifan Ekstrakurikuler dikonversi dari bentuk kualitatif menjadi numerik dengan menggunakan skema berikut:

Tabel 1. Keaktifan ekstrakurikuler

Keterangan	Jumlah	Inisiasi
Tidak aktif	18	1
Kurang aktif	9	2
Cukup aktif	6	3
Aktif	1	4

Dari tabel 1 ini memungkinkan algoritma K-Means untuk memahami tingkat keaktifan siswa dalam kegiatan ekstrakurikuler dalam bentuk angka, sehingga dapat digunakan sebagai faktor dalam proses clustering. Selain itu, normalisasi data juga dilakukan untuk memastikan bahwa semua atribut memiliki skala yang seragam. Hal ini penting karena algoritma K-Means sensitif terhadap skala data,]sehingga atribut dengan rentang nilai yang lebih besar dapat mendominasi hasil clustering jika tidak dinormalisasi terlebih dahulu.

4.3. Menghitung Nilai DBI

Davies-Bouldin Index (DBI) adalah salah satu metrik evaluasi yang digunakan untuk mengukur kualitas hasil clustering pada data. Metrik ini menilai sejauh mana cluster yang dihasilkan memiliki homogenitas internal yang tinggi (kesamaan antar data dalam cluster) dan pemisahan yang jelas antara cluster yang berbeda. Nilai DBI yang lebih kecil menunjukkan bahwa clustering lebih optimal, dengan data dalam setiap cluster yang lebih seragam dan jarak antar cluster yang lebih besar.

Perhitungan DBI didasarkan pada dua komponen utama, pertama, Intra-cluster similarity yang mengukur rata-rata jarak setiap data dalam cluster terhadap centroid cluster tersebut. Semakin kecil nilai ini, semakin homogen data dalam cluster. Kedua, Inter-cluster distance yang mengukur jarak antar centroid dari cluster yang berbeda, mencerminkan sejauh mana cluster tersebut terpisah satu sama lain. Semakin besar jarak antar centroid, semakin jelas pemisahan antar cluster. Nilai DBI dihitung dengan mengambil rasio antara intra-cluster similarity dan inter-cluster distance untuk setiap pasangan cluster, kemudian rata-rata rasio tersebut digunakan sebagai nilai akhir DBI. Semakin kecil nilai DBI, semakin baik hasil clustering, karena menunjukkan keseragaman internal yang tinggi dalam cluster dan pemisahan yang jelas antar cluster.

Dalam RapidMiner, proses perhitungan DBI dapat dilakukan setelah melakukan clustering menggunakan algoritma seperti K-Means. Dengan menggunakan operator Cluster Evaluation dan memilih Davies-Bouldin Index sebagai metrik, kita dapat mengukur kualitas clustering untuk berbagai jumlah cluster yang diuji. Nilai DBI yang terendah akan menunjukkan jumlah cluster yang paling optimal, memberikan dasar yang obyektif untuk menentukan jumlah cluster terbaik dalam analisis data.

Gambar 2. Tahapan alur proses rapidminer

4.4. Pengambilan Data (Retrieve Data Mining)

Tahapan pertama dalam proses ini adalah mengambil dataset menggunakan operator Retrieve Data Mining. Chen et al. menjelaskan bahwa pemilihan data yang tepat merupakan kunci untuk mencapai hasil yang akurat dalam analisis clustering, karena data yang relevan dapat memperbaiki kinerja model analisis yang digunakan [19]. Operator ini berfungsi untuk membaca dataset yang telah disimpan dalam Local Repository RapidMiner. Dataset yang digunakan dalam analisis ini mengandung beberapa atribut penting, seperti: Nilai Rata-Rata Akademik (Semua Mata Pelajaran) – berupa nilai rata-rata dari seluruh mata pelajaran yang diikuti siswa; Total Ketidakhadiran (Hari) – jumlah hari siswa tidak hadir dalam kegiatan akademik; Jumlah Ekstrakurikuler (Integer) – berapa banyak kegiatan ekstrakurikuler yang diikuti siswa; dan Keaktifan Ekstrakurikuler (Integer) – tingkat keaktifan siswa dalam ekstrakurikuler. Data inilah yang akan digunakan sebagai dasar untuk membentuk kelompok atau cluster berdasarkan kemiripan karakteristik.

4.5. Duplikasi Dataset (Multiply)

Setelah data diambil, langkah selanjutnya adalah menduplikasi dataset menggunakan operator Multiply. Operator ini berfungsi untuk menciptakan beberapa salinan dataset yang sama sehingga dapat digunakan dalam beberapa percobaan yang berbeda secara paralel. Pada gambar, dataset yang sama diduplikasi menjadi lima jalur pemrosesan yang berbeda. Setiap jalur akan digunakan untuk proses clustering dengan jumlah cluster yang berbeda, yaitu: Clustering k=2 (mengelompokkan data menjadi 2 cluster); Clustering k=3 (mengelompokkan data menjadi 3 cluster); Clustering k=4 (mengelompokkan data menjadi 4 cluster); Clustering k=5 (mengelompokkan data menjadi 5 cluster); dan Clustering k=6 (mengelompokkan data menjadi 6 cluster). Pendekatan ini dilakukan untuk membandingkan hasil clustering dengan jumlah cluster yang berbeda guna menemukan jumlah optimal.

4.6. Proses Clustering dengan K-Means

Setiap jalur pemrosesan menggunakan operator Clustering (K-Means) untuk mengelompokkan data berdasarkan kemiripan karakteristik antar siswa. Operator Clustering (K-Means) akan melakukan iterasi untuk menemukan pusat cluster yang optimal berdasarkan data yang diberikan. Data akan dibagi menjadi beberapa kelompok yang memiliki karakteristik serupa.

4.7. Evaluasi Performa Model Clustering

Setelah clustering dilakukan, masing-masing hasil clustering dikirim ke operator Performance untuk mengukur efektivitas pengelompokan. Evaluasi ini dilakukan untuk mengetahui sejauh mana setiap jumlah cluster (k=2, k=3, k=4, k=5, dan k=6) mampu mengelompokkan siswa dengan baik. Evaluasi performa clustering bertujuan untuk menentukan jumlah cluster terbaik yang paling sesuai dengan pola data. Davies-Bouldin Index (DBI) digunakan untuk mengevaluasi kualitas hasil clustering. Jedidi et al. menjelaskan bahwa penggunaan metrik evaluasi seperti DBI (Davies-Bouldin Index) sangat berguna untuk menilai sejauh mana klaster yang dihasilkan dapat terpisah dengan baik dan memiliki kohesi yang tinggi [17]. Dengan DBI, dapat dipastikan bahwa model menghasilkan klaster yang jelas dan tidak tumpang tindih. Selain itu, Indahyanti et al. juga menambahkan bahwa evaluasi dengan metrik seperti DBI memberikan gambaran objektif tentang kualitas hasil clustering, yang penting untuk memastikan bahwa model memberikan hasil yang valid dan berguna [20]. Di RapidMiner, evaluasi DBI dapat dilakukan dengan menggunakan operator Cluster Evaluation dan memilih Davies-Bouldin Index sebagai

metrik evaluasi. DBI akan dihitung untuk setiap hasil clustering dengan jumlah cluster yang berbeda.

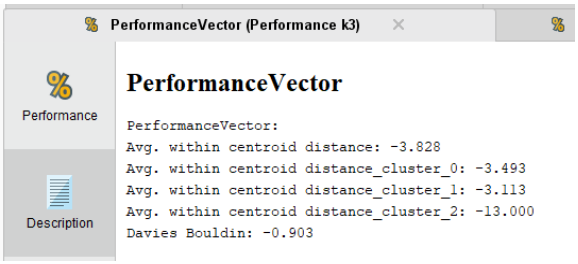
4.8. Pemilihan Cluster Optimal

Setelah evaluasi dilakukan untuk setiap jumlah cluster (k=2, k=3, k=4, k=5, k=6), nilai DBI yang dihasilkan akan dibandingkan. Model akan memilih jumlah cluster yang memiliki nilai DBI terendah sebagai jumlah cluster yang paling optimal. Pada tahap ini, dataset siswa kelas VII di SMPN 10 Palembang yang telah melalui proses data cleaning dan transformasi digunakan untuk menghitung DBI. Dataset tersebut mencakup atribut penting seperti rata-rata nilai akademik, tingkat kehadiran, jumlah ekstrakurikuler, dan tingkat keaktifan ekstrakurikuler. Perhitungan DBI dilakukan untuk berbagai skenario jumlah cluster (k) untuk mengevaluasi kualitas masing-masing konfigurasi dengan hasil sebagai berikut:

Tabel 2. *DBI value*

Cluster	DBI Value
2	-0.739
3	-0.903
4	-0.874
5	-0.804
6	-0.800

Berdasarkan hasil perhitungan DBI seperti pada tabel 2, jumlah cluster dengan nilai DBI terkecil adalah 3, yaitu sebesar -0.903. Cluster ini menunjukkan konfigurasi terbaik karena memiliki homogenitas internal yang tinggi dan jarak antar-cluster yang signifikan, menjadikannya pilihan optimal untuk proses clustering.

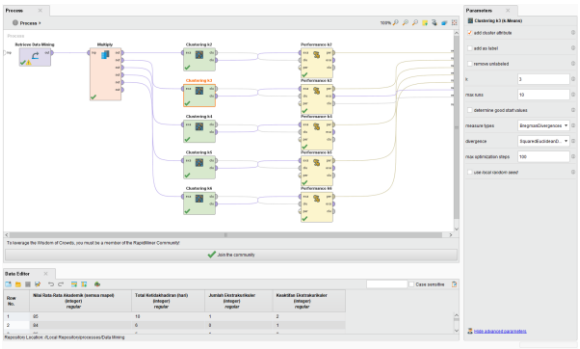


Gambar 3. *Description performance vector*

4.9. Analisa Data dengan Algoritma K-Means

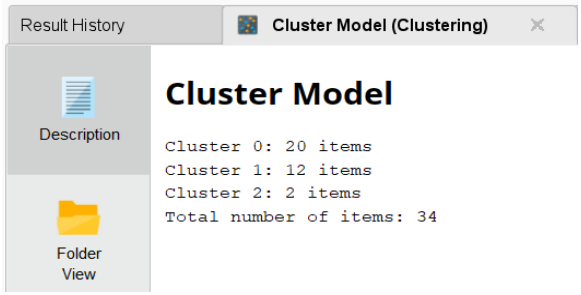
Setelah proses evaluasi selesai dan nilai DBI untuk setiap jumlah cluster dihitung, jumlah cluster yang memiliki nilai DBI terendah, yaitu k=3, dipilih sebagai jumlah cluster yang paling optimal. Untuk memverifikasi hasil clustering dengan k=3, kita dapat melihat distribusi data dalam tiga cluster yang terbentuk dan menganalisis keseragaman data dalam setiap cluster.

Pada gambar 4, di jalur Clustering k=3, pengaturan parameter ditampilkan pada panel sebelah kanan dengan parameter utama berupa jumlah cluster yang ditentukan sebanyak 3 dan max runs sebanyak 10. Penggunaan 10 pada parameter Max Runs bertujuan untuk meminimalkan variasi hasil clustering, sehingga menghasilkan model yang lebih stabil. Proses ini memastikan hasil clustering yang konsisten dan optimal. Untuk melihat hasilnya, kita dapat menyambungkan bagian exa (example) dan clu (cluster) dari operator performance ke titik res (result).



Gambar 4. Melihat clustering untuk k=3 di RapidMiner

Operator Clustering (K-Means) akan melakukan iterasi untuk menemukan pusat cluster yang optimal berdasarkan data yang diberikan kemudian data akan dibagi menjadi beberapa kelompok yang memiliki karakteristik serupa. Berikut hasil pengujian dataset siswa kelas VII di SMPN 10 Palembang yang berjumlah 34 orang:



Gambar 5. Description hasil clustering

4.10. Hasil Clustering dan Interpretasi

Setelah menerapkan algoritma K-Means, dataset yang terdiri dari 34 siswa berhasil dikelompokkan ke dalam tiga cluster yang mencerminkan tingkat risiko akademik mereka. Hasil clustering dapat dilihat dalam tabel berikut:

Tabel 3. Kategori Risiko Akademik

Keterangan	Jumlah	Inisiasi
Tidak berisiko	12	1
Berisiko sedang	20	2
Berisiko tinggi	2	3

Dari tabel 3 dapat diketahui bahwa berdasarkan analisis cluster yang dihasilkan oleh RapidMiner, terdapat 3 cluster yang menggambarkan karakteristik siswa berdasarkan nilai akademik, ketidakhadiran, dan keterlibatan dalam ekstrakurikuler. Siswa dalam Cluster 1 dikategorikan sebagai Tidak Berisiko. Hal ini dikarenakan mereka memiliki tingkat kehadiran yang tinggi dengan absensi antara 0 hingga 3 hari, nilai

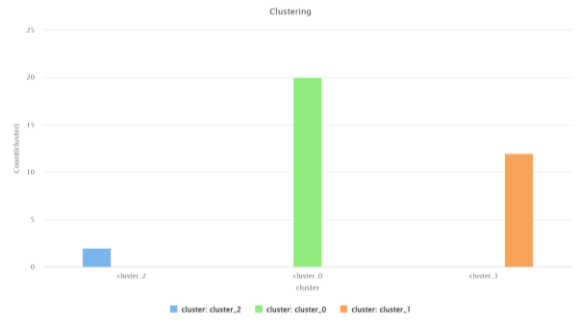
akademik yang cukup baik dalam rentang 85–89, serta tingkat keaktifan ekstrakurikuler yang bisa dikatakan lebih tinggi dibandingkan cluster lainnya. Siswa dalam kelompok ini menunjukkan keseimbangan antara prestasi akademik dan partisipasi dalam kegiatan di luar kelas, yang mencerminkan kedisiplinan dan keterlibatan yang baik dalam lingkungan sekolah.

Sementara itu, Cluster 0 dikategorikan sebagai Berisiko Sedang. Siswa dalam kelompok ini memiliki absensi yang lebih tinggi, berkisar antara 4 hingga 8 hari, dengan nilai akademik yang tetap cukup baik dalam rentang 84–87. Namun, mereka cenderung kurang aktif dalam kegiatan ekstrakurikuler atau bahkan tidak mengikutinya sama sekali. Meskipun performa akademik mereka masih terjaga, ketidakterlibatan dalam kegiatan tambahan dapat menjadi indikasi kurangnya motivasi atau keterlibatan sosial di lingkungan sekolah, yang berpotensi mempengaruhi perkembangan mereka di masa depan. Siswa dalam kategori ini perlu didorong untuk lebih aktif dalam kegiatan sekolah guna meningkatkan keterlibatan mereka.

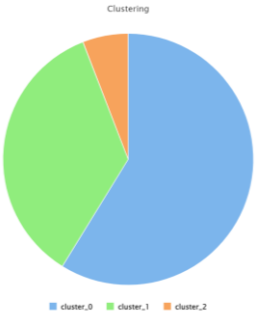
Adapun Cluster 2 itu masuk dalam kategori Berisiko Tinggi, karena siswa dalam kelompok ini memiliki jumlah absensi yang jauh lebih tinggi, yaitu antara 10 hingga 17 hari. Selain itu, nilai akademik mereka cenderung lebih rendah dibandingkan kelompok lainnya, dengan rata-rata 84–85, serta partisipasi ekstrakurikuler yang bervariasi tetapi tidak menonjol. Tingkat ketidakhadiran yang tinggi dapat menjadi indikasi masalah kedisiplinan, kesulitan dalam belajar, atau kurangnya keterlibatan dalam lingkungan sekolah. Oleh karena itu, siswa dalam kelompok ini memerlukan perhatian lebih agar dapat meningkatkan kehadiran, keterlibatan, dan performa akademik mereka.

4.11. Visualisasi Hasil Clustering

Untuk memahami distribusi siswa dalam setiap cluster, digunakan grafik dalam bentuk bar chart dan pie chart yang menggambarkan proporsi jumlah siswa dalam masing-masing kategori. Visualisasi ini membantu dalam mengidentifikasi pola yang muncul dalam dataset serta memberikan gambaran yang lebih jelas kepada pemangku kepentingan, seperti guru dan staf sekolah, mengenai kondisi akademik siswa di SMPN 10 Palembang.



Gambar 6. Bar Chart Distribusi Data



Gambar 7. Pie Chart Distribusi Data

4.12. Implikasi dan Rekomendasi

Berdasarkan hasil clustering, beberapa rekomendasi dapat diberikan untuk membantu siswa dalam masing-masing kategori. Untuk siswa yang termasuk dalam kategori Tidak Berisiko, mereka sebaiknya didorong untuk terus mempertahankan prestasi akademik mereka dan berpartisipasi dalam kegiatan ekstrakurikuler. Selain itu, pemberian penghargaan bagi siswa yang tetap aktif dalam kegiatan ekstrakurikuler dapat menjadi motivasi tambahan agar mereka terus berpartisipasi dalam lingkungan sekolah secara aktif.

Sementara itu, bagi siswa dalam kategori Berisiko Sedang, diperlukan program yang dapat meningkatkan motivasi mereka untuk lebih aktif dalam kegiatan ekstrakurikuler. Program ini dapat berupa workshop, seminar, atau kegiatan lain yang mampu meningkatkan keterlibatan mereka. Selain itu, komunikasi antara guru dan siswa juga perlu ditingkatkan agar hambatan yang dihadapi siswa dalam keterlibatan akademik maupun ekstrakurikuler dapat lebih dipahami dan diatasi dengan strategi yang tepat.

Adapun untuk siswa yang masuk dalam kategori Berisiko Tinggi, mereka membutuhkan perhatian khusus dalam bentuk bimbingan akademik dan konseling yang dapat membantu mereka mengatasi kesulitan belajar. Pendekatan yang lebih personal dapat digunakan untuk memahami faktor-faktor yang menyebabkan rendahnya tingkat kehadiran dan pencapaian akademik mereka. Selain itu, keterlibatan orang tua juga menjadi faktor penting dalam meningkatkan kehadiran dan keterlibatan anak dalam kegiatan sekolah. Orang tua perlu dilibatkan dalam diskusi dan solusi yang ditawarkan oleh pihak sekolah agar dapat memberikan dukungan penuh kepada anak mereka dalam meningkatkan prestasi akademik.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa metode K-Means Clustering adalah pendekatan yang efektif untuk mengelompokkan siswa berdasarkan tingkat risiko akademik mereka. Dengan mengelompokkan siswa ke dalam tiga kategori utama—Tidak Berisiko, Berisiko Sedang, dan Berisiko Tinggi—metode ini memberikan wawasan yang mendalam tentang pola risiko akademik yang ada dalam populasi siswa. Hasil pengelompokan ini membantu pihak sekolah untuk

memahami kondisi siswa secara lebih komprehensif, sehingga memungkinkan perencanaan strategi yang lebih tepat untuk menangani kebutuhan individu maupun kelompok. Salah satu temuan penting dari penelitian ini adalah bahwa siswa dalam cluster Berisiko Tinggi memerlukan perhatian khusus. Data menunjukkan bahwa tingkat absensi yang tinggi, nilai akademik yang rendah, dan partisipasi ekstrakurikuler yang minim menjadi faktor utama yang memengaruhi kinerja mereka. Untuk membantu siswa dalam kategori ini, penelitian merekomendasikan pengembangan program dukungan berbasis data, seperti bimbingan akademik, pendampingan khusus, serta program motivasi. Program-program ini bertujuan untuk meningkatkan pemahaman akademik, mengurangi ketidakhadiran, dan mendorong keterlibatan dalam kegiatan ekstrakurikuler, sehingga siswa lebih termotivasi dan aktif dalam lingkungan sekolah. Selain itu, penelitian ini juga menyarankan pengembangan strategi berbasis teknologi untuk memantau performa siswa secara berkelanjutan. Dengan memanfaatkan hasil clustering ini, sekolah dapat membangun sistem pendukung pengambilan keputusan berbasis data yang memungkinkan intervensi dilakukan lebih cepat dan efektif. Misalnya, siswa dengan Risiko Tinggi dapat segera diidentifikasi dan diberikan bimbingan personalisasi sebelum mereka mengalami penurunan performa yang lebih signifikan. Ke depannya, penelitian ini merekomendasikan penggunaan algoritma lain, seperti Decision Tree atau Fuzzy C-Means, untuk mengevaluasi dan membandingkan akurasi serta efektivitas model yang berbeda. Algoritma tersebut dapat memberikan wawasan tambahan atau hasil yang lebih adaptif terhadap karakteristik data siswa yang berbeda. Pendekatan ini dapat memperkaya hasil penelitian dan memberikan pandangan yang lebih luas dalam menerapkan data mining untuk mendukung peningkatan kualitas pendidikan. Secara keseluruhan, penelitian ini menekankan pentingnya pendekatan berbasis data dalam memahami kebutuhan siswa dan merancang intervensi yang berdampak positif. Dengan kombinasi metode analitik yang tepat, sekolah dapat menciptakan lingkungan belajar yang lebih inklusif dan mendukung keberhasilan siswa secara menyeluruh. Penelitian ini membuka peluang untuk pengembangan lanjutan di bidang edukasi berbasis teknologi dan data mining.

.

DAFTAR PUSTAKA

[1] A. Hudawi, K. Anam, and M. Rahman, "Penerapan Data Mining untuk Menemukan Pola Asosiasi Aktivitas Belajar dan Prestasi Santri Menggunakan Algoritma Apriori," TRILOGI: Jurnal Ilmu Teknologi, Kesehatan, dan Humaniora, vol. 5, no. 4, pp. 653–662, Dec. 2024, doi: 10.33650/trilogi.v5i4.9919.

[2] R. Harman, "Computer Based Information System Journal KATA KUNCI," CBIS JOURNAL, vol. 12, no. 01, 2024, [Online]. Available: http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis

[3] S. A. A. Kharis and A. H. A. Zili, "Learning Analytics dan Educational Data Mining pada Data Pendidikan," JURNAL RISET PEMBELAJARAN MATEMATIKA SEKOLAH, vol. 6, no. 1, pp. 12–20, Mar. 2022, doi: 10.21009/jrpms.061.02.

[4] M. Sholeh, E. Kumalasari Nurnawati, and U. Lestari, "Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner," 2023. [Online]. Available: https://archive.ics.uci.edu/ml/datasheets.php.

[5] D. Permata Sari, F. Sembiring, D. Putri sulisdianto, and Y. Jentiner, "Jurnal Restikom : Riset Teknik Informatika dan Komputer IMPLEMENTASI ALGORITMA FUZZY C-MEANS UNTUK MEMPREDIKSI FAKTOR SISWA MEMBOLOS (STUDI KASUS : FAKTOR SISWA MEMBOLOS DI SMPN 1 PARAKANSALAK)," vol. 2, no. 1, pp. 1–7, 2020, [Online]. Available: https://restikom.nusaputra.ac.id

[6] D. Ayu, N. Wulandari, R. Annisa, ; Lestari Yusuf, and T. Prihatin, "AN EDUCATIONAL DATA MINING FOR STUDENT ACADEMIC PREDICTION USING K-MEANS CLUSTERING AND NAÏVE BAYES CLASSIFIER." [Online]. Available: www.bsi.ac.id

[7] R. Pahlevi Kurniawan, "3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta," vol. 2, no. 2, 2023.

[8] 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.

[9] D. Syaripudin, A. Rachmat Raharja, S. Informasi, F. Kesehatan dan Teknik, and U. Bandung, "PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA DATA NILAI SISWA UNTUK MENENTUKAN KELOMPOK PENERIMA BEASISWA".

[10] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," Journal of Artificial Intelligence and Technology Information (JAITI), vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.

[11] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," Journal of Information Systems and Informatics, vol. 3, no.

3, 2021, [Online]. Available: http://journal-isi.org/index.php/isi

[12] P. S. Hasugian and J. Raphita Sagala, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means," Media Online, vol. 3, no. 3, pp. 262–268, 2022, [Online]. Available: https://djournals.com/klik

[13] E. Muningsih, "KOMBINASI METODE K-MEANS DAN DECISION TREE DENGAN PERBANDINGAN KRITERIA DAN SPLIT DATA," 2022.

[14] A. Rahman Azis, "Analisis Komparasi Algoritma Machine Learning dalam Prediksi Performa Akademik Mahasiswa: Literature Review," Jurnal Ilmu Komputer dan Informatika (JIKI), vol. 4, no. 2, pp. 143–150, 2024, doi: 10.54082/jiki.212.

[15] M. Triandini, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia," Jurnal Informasi dan Teknologi, pp. 167–173, Sep. 2021, doi: 10.37034/jidt.v3i3.120.

[16] L. Masangu, A. Jadhav, and R. Ajoodha, "Predicting student academic performance using data mining techniques," Advances in Science, Technology and Engineering Systems, vol. 6, no. 1, pp. 153–163, 2021, doi: 10.25046/aj060117.

[17] Y. Jedidi et al., "Predicting students' academic performance and modeling using data mining techniques," Mathematical Modeling and Computing, vol. 11, no. 3, pp. 814–825, 2024, doi: 10.23939/mmc2024.03.814.

[18] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," Smart Learning Environments, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40561-022-00192-z.

[19] Y. Chen, J. Sun, J. Wang, L. Zhao, X. Song, and L. Zhai, "Machine Learning-Driven Student Performance Prediction for Enhancing Tiered Instruction," Feb. 2025, [Online]. Available: http://arxiv.org/abs/2502.03143

[20] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Educational Data Mining on Student Academic Performance Prediction: A Survey Educational Data Mining Pada Prediksi Kinerja Akademik Mahasiswa: Sebuah Survey," 2022. [Online]. Available: https://ieeexplore.ieee.org/