

Sistem Cerdas Untuk Meningkatkan Prestasi Siswa Dengan Analisis Rapor Menggunakan Algoritma *K-Means Clustering* Berbasis Web

Muhammad Zaki Zikrillah Baharuddin Nasution^{1*}, Triase²

^{1,2}Program Studi Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Indonesia

Email: ¹zakicoool456@gmail.com, ²triase@uinsu.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 18 November 2025

Direvisi, 12 Desember 2025

Diiterima, 23 Desember 2025

Kata Kunci:

K-Means Clustering,
Sistem Cerdas ,
Prestasi Akademik,
Data Mining,
Davies-Bouldin Index

ABSTRAK

Abstract- SMA Negeri 1 Aek Songsongan still faces obstacles in managing and analyzing academic data, where historical report card data has not been optimally utilized to support targeted learning decisions. This study aims to develop a solution in the form of a web-based intelligent system to analyze student achievement patterns and generate more targeted learning recommendations. The system is designed using the *K-Means Clustering* algorithm through a *Knowledge Discovery in Databases* approach that includes data collection, preprocessing, clustering, cluster quality evaluation using the *Davies-Bouldin Index*, and automatic recommendation generation. The research data consisted of 1,384 report card scores from three academic years, and after preprocessing, 1,075 valid data with eight subject attributes were obtained. The results showed the formation of three clusters with a *Davies-Bouldin Index* value of 0.9678, indicating excellent separation quality. The first cluster describes a group of high-achieving students, the second cluster shows students with moderate to good performance, and the third cluster reflects students with uneven score patterns. The system generates different recommendations for each group in the form of acceleration programs, regular tutoring, and intensive remedial programs. These findings indicate that the use of clustering techniques can support schools in designing more effective and data-driven learning strategies.

Abstrak- SMA Negeri 1 Aek Songsongan masih menghadapi kendala dalam pengelolaan dan analisis data akademik, di mana data historis rapor belum dimanfaatkan secara optimal untuk mendukung keputusan pembelajaran yang tepat sasaran. Penelitian ini bertujuan mengembangkan solusi berupa sistem cerdas berbasis web untuk menganalisis pola prestasi siswa dan menghasilkan rekomendasi pembelajaran yang lebih terarah. Sistem dirancang menggunakan algoritma *K-Means Clustering* melalui pendekatan *Knowledge Discovery in Databases* yang meliputi pengumpulan data, *preprocessing*, proses *clustering*, evaluasi kualitas kluster menggunakan *Davies-Bouldin Index*, serta penyusunan rekomendasi otomatis. Data penelitian terdiri dari 1.384 nilai rapor dari tiga tahun ajaran dan setelah proses *preprocessing* menghasilkan 1.075 data valid dengan delapan atribut mata pelajaran. Hasil penelitian menunjukkan terbentuknya tiga kluster dengan nilai *Davies-Bouldin Index* sebesar 0,9678 yang mengindikasikan kualitas pemisahan yang sangat baik. *Cluster* pertama menggambarkan kelompok siswa berprestasi tinggi, *cluster* kedua menunjukkan siswa dengan performa sedang-baik, dan *cluster* ketiga mencerminkan siswa dengan pola nilai yang tidak merata. Sistem menghasilkan rekomendasi berbeda untuk setiap kelompok berupa program akselerasi, bimbingan reguler, dan remedial intensif. Temuan ini menunjukkan bahwa pemanfaatan teknik *clustering* dapat mendukung sekolah dalam merancang strategi pembelajaran yang lebih efektif dan berbasis data.

Copyright © 2025 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Nama Penulis Korespondensi

Program Studi Teknik Informatika,
Politeknik Purbaya Tegal

Jl. Pancakarya No.1, Kalimati Kaje Talang Tegal, Indonesia

Email: penulis.korespondensi@email.ac.id

1. Pendahuluan

Kemajuan pesat dalam ilmu pengetahuan dan teknologi telah membawa perubahan signifikan

dalam berbagai aspek kehidupan, termasuk sektor pendidikan [1]. Salah satu inovasi yang berkembang pesat adalah penerapan sistem cerdas (*intelligent*

system), yang mampu meningkatkan efisiensi administrasi sekaligus memperkuat proses analisis data untuk pengambilan keputusan yang lebih akurat [2],[3]. Di tengah meningkatnya volume data akademik, institusi pendidikan dituntut untuk mampu mengolah data tersebut menjadi informasi bermakna [4]. Data mining hadir sebagai solusi dengan kemampuan untuk mengekstraksi pola tersembunyi dari kumpulan data besar dan kompleks [5], sehingga dapat menghasilkan wawasan yang bernilai [6]. Pada ranah pendidikan, penggunaan data mining khususnya dalam *Educational Data Mining* (EDM) kini menjadi tren global karena mampu membantu sekolah memprediksi risiko kegagalan belajar, mengidentifikasi gaya belajar, hingga personalisasi pembelajaran berbasis data [7].

SMA Negeri 1 Aek Songsongan yang berlokasi di Kecamatan Aek Songsongan, Kabupaten Asahan, Sumatera Utara, merupakan salah satu sekolah yang menghasilkan data akademik dalam jumlah besar setiap tahun. Namun, data rapor yang tersimpan belum dimanfaatkan secara optimal sebagai dasar pengambilan keputusan. Analisis manual terhadap nilai-nilai tersebut membutuhkan waktu yang lama, rentan kesalahan, dan sulit memberikan gambaran menyeluruh mengenai perkembangan prestasi siswa [8]. Kondisi ini menyebabkan sekolah kesulitan mengidentifikasi siswa yang mengalami peningkatan, penurunan, atau stagnasi dalam pembelajaran [9]. Padahal, pemetaan prestasi yang tepat sangat dibutuhkan untuk memberikan intervensi pembelajaran yang cepat dan efektif.

SMA Negeri 1 Aek Songsongan yang berlokasi di Kecamatan Aek Songsongan, Kabupaten Asahan, Sumatera Utara, merupakan salah satu sekolah yang menghasilkan data akademik dalam jumlah besar setiap tahun. Namun, data rapor yang tersimpan belum dimanfaatkan secara optimal sebagai dasar pengambilan keputusan. Analisis manual terhadap nilai-nilai tersebut membutuhkan waktu yang lama, rentan kesalahan, dan sulit memberikan gambaran menyeluruh mengenai perkembangan prestasi siswa [8]. Kondisi ini menyebabkan sekolah kesulitan mengidentifikasi siswa yang mengalami peningkatan, penurunan, atau stagnasi dalam pembelajaran [9]. Padahal, pemetaan prestasi yang tepat sangat dibutuhkan untuk memberikan intervensi pembelajaran yang cepat dan efektif.

Untuk menjawab kebutuhan tersebut, penelitian ini mengusulkan pengembangan sistem cerdas berbasis web yang menganalisis nilai rapor secara otomatis dengan memanfaatkan algoritma *K-Means Clustering* [10]. Algoritma ini bekerja dengan mengelompokkan siswa berdasarkan kemiripan pola nilai mereka melalui proses pembentukan pusat klaster dan penempatan data berdasarkan jarak terdekat. Hasil pengelompokan memberikan gambaran yang jelas mengenai

distribusi prestasi siswa, sehingga guru dapat merancang intervensi pembelajaran yang lebih tepat sasaran. Pemanfaatan platform web juga sejalan dengan tren terkini transformasi digital di sekolah, yang menekankan aksesibilitas data secara *real-time*, transparansi progres belajar, serta mendukung ekosistem pembelajaran adaptif [11].

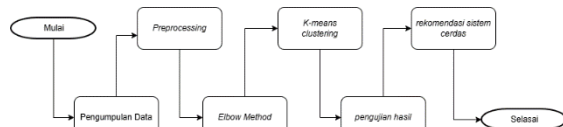
Berbagai penelitian sebelumnya telah menerapkan algoritma *K-Means Clustering* dalam proses pengelompokan data akademik siswa untuk mendukung analisis prestasi dan pengambilan keputusan. Penelitian oleh Dewi [9] menguji hasil pengolahan data nilai rapor siswa semester I–IV menggunakan metode *K-Means* berbantuan *RapidMiner*, dan menghasilkan tiga kelompok, yaitu Sangat Berprestasi (3 siswa), Berprestasi (10 siswa), dan Kurang Berprestasi (12 siswa). Selanjutnya, Shefia [12] melakukan penelitian mengenai pengelompokan prestasi siswa di SD Namorambe menggunakan *K-Means Clustering*, yang menghasilkan sebuah aplikasi berbasis desktop untuk mempermudah pihak sekolah dalam mengolah data dan mengelompokkan siswa. Penelitian oleh Prayudha [7] menerapkan algoritma *K-Means* pada data nilai Matematika, Bahasa Inggris, dan Kimia dengan evaluasi Davies–Bouldin Index, di mana diperoleh dua klaster optimal dengan nilai DBI sebesar 0,519, serta menemukan bahwa atribut Kimia memiliki pengaruh paling besar dalam proses pengelompokan. Sementara itu, Penda Sudarto [13] menerapkan *K-Means* berbasis *RapidMiner* untuk mengelompokkan nilai akademik siswa menjadi tiga klaster, yaitu Unggul (2 siswa), Sedang (4 siswa), dan Rendah (4 siswa). Adapun Nurul Alifia [11] menggunakan algoritma *K-Means* clustering sebagai sistem peringatan dini untuk kegagalan siswa dalam mata kuliah bahasa Indonesia, yang menghasilkan identifikasi tingkat risiko secara terkelompok dengan persentase keberhasilan klasifikasi yang berbeda pada tiap kelas.

Penerapan sistem cerdas berbasis web yang menggunakan algoritma *K-Means Clustering* memungkinkan analisis data rapor siswa di SMA Negeri 1 Aek Songsongan secara otomatis dan terstruktur. Dengan demikian, guru dan pihak sekolah dapat memberikan intervensi dan rekomendasi pembelajaran yang lebih tepat sasaran sesuai dengan kebutuhan masing-masing kelompok siswa [10]. Selain itu, platform berbasis web yang dikembangkan akan memungkinkan akses data yang mudah dan transparan bagi guru, orang tua, dan siswa, sehingga kolaborasi dalam upaya peningkatan prestasi dapat berjalan lebih efektif.

2. Metodologi Penelitian

Penelitian ini menggabungkan teknik penelitian kuantitatif dan kualitatif dengan

menggunakan metode *Mixed Method Research*. Pada metode kualitatif, penelitian ini dilakukan untuk memahami faktor-faktor yang memengaruhi hasil klasterisasi nilai siswa. Data diperoleh melalui observasi difokuskan pada pengecekan kesesuaian format rapor, proses pencatatan nilai. Wawancara juga dilakukan dengan guru serta pihak sekolah mengklarifikasi nilai yang tampak tidak wajar, perubahan kebijakan penilaian, atau kondisi khusus siswa yang dapat memengaruhi hasil analisis. Hasil dari metode ini digunakan untuk menjelaskan dan memperkuat temuan kuantitatif, sehingga hasil klasterisasi tidak hanya berupa angka, tetapi juga memiliki makna yang sesuai dengan kondisi nyata di sekolah. Data dalam penelitian ini terbatas pada nilai rapor siswa dari tiga tahun ajaran di SMA Negeri 1 Aek Songsongan. Adapun data yang digunakan dalam sistem hanya berfokus pada nilai pengetahuan umum dan nilai keterampilan. Metodologi yang diterapkan mengikuti tahapan KDD (*Knowledge Discovery Databases*), yaitu pemilihan data, prapemrosesan, penambangan data dengan *K-Means*, evaluasi pola, dan interpretasi hasil[14]. Alat bantu seperti framework *laravel* dan *flask* serta *Google Colaboratory* juga di gunakan untuk mengembangkan sistem dan mengolah data.



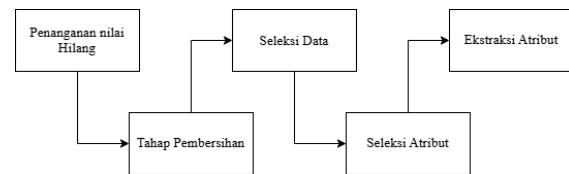
Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data set yang digunakan dalam studi ini diperoleh nilai rapor siswa dari tiga tahun ajaran, yaitu 2020/2021, 2021/2022, dan 2022/2023 di SMA Negeri 1 Aek Songsongan. Data set tersebut terdiri atas nilai pengetahuan dan nilai keterampilan, serta informasi lainnya yang relevan, termasuk NIS, dan Nama siswa. Proses pengambilan data dilakukan dengan mengajukan permohonan resmi kepada bagian Ke- Siswaan di SMA Negeri 1 Aek Songsongan, dengan prosedur yang telah ditetapkan untuk memastikan bahwa data yang digunakan dalam penelitian ini memenuhi standar kualitas dan kelengkapan yang diperlukan. Dataset ini akan digunakan untuk melakukan analisis menggunakan algoritma K-Means, yang bertujuan untuk mengelompokkan siswa berdasarkan data nilai akademik guna mengidentifikasi pola prestasi siswa. Atribut target yang akan klaster adalah hasil belajar siswa, yang dikategorikan ke dalam 3 kluster: Rendah, sedang, sangat baik.

2.2 Preprocessing Data

Preprocessing data dilakukan untuk memastikan bahwa dataset siap untuk dianalisis menggunakan algoritma K-means Clustering [15].



Gambar 2. Alur Preprocessing Data

- Penanganan nilai yang hilang (*Missing Value Handling*): Menghapus data yang hilang agar tidak terjadi Gangguan pada proses data mining.
- Tahap Pembersihan (*Cleaning*): data nilai siswa yang berjumlah 1384 dilihat data yang memiliki atribut data yang salah/typo serta adanya duplikat data dalam data tersebut maka akan dihilangkan.
- Seleksi Data (*Data Selection*): data siswa dari tiga tahun ajaran diseleksi lagi dengan mengambil beberapa data yang akan menjadi kriteria dalam penilaian siswa SMA Negeri 1 Aek Songsongan dalam hal ini ialah nilai pengetahuan umum, nilai keterampilan.
- Seleksi Atribut (*Feature Selection*): Tahap ini bertujuan memilih fitur paling informatif agar hasil clustering bermakna dan mudah dipahami. Data mencakup delapan atribut: Matematika, Bahasa Indonesia, Bahasa Inggris, PKN, Fisika, Kimia, Biologi, dan Keterampilan. Atribut ini dipilih berdasarkan masukan guru serta hasil *Exploratory Data Analysis* (EDA) yang menunjukkan distribusi dan variasi data yang representatif terhadap prestasi akademik siswa.
- Ekstraksi atribut (*Feature Extraction*): nilai rapor yang awalnya terdiri dari beberapa mata pelajaran (atribut) digabungkan menjadi nilai rata-rata.

2.3 Elbow Method

Metode *Elbow* digunakan untuk menentukan jumlah cluster optimal dengan mengamati titik “siku” pada grafik hasil pengelompokan. Grafik ini memperlihatkan penurunan nilai kesalahan (*error*) seiring bertambahnya jumlah cluster. Tujuannya adalah memilih nilai *k* yang cukup kecil namun tetap menghasilkan *within-cluster sum of squares* (WSS) rendah[16]. Dengan algoritma *K-Means*, metode ini menunjukkan hubungan antara jumlah cluster dan tingkat kesalahan. Nilai *k* terbaik ditentukan pada titik ketika penurunan error mulai melambat atau

membentuk siku pada grafik. Evaluasi dilakukan menggunakan *Sum of Squared Error* (SSE), yang mengukur sejauh mana data menyimpang dari pusat clusternya [15]. Secara umum, metode Elbow berfungsi melalui tahapan proses sebagai berikut:

$$SSE = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \varphi_k)^2 \quad (1)$$

Keterangan:

1. $C_k C_k$ = cluster yang terbentuk
2. k = banyak cluster
3. $X_i X_i$ = data x pada fitur ke-i
4. $\varphi_k \varphi_k$ = rata-rata K cluster pada nilai k (k=1,2,3,...,K)

2.4 K-means Clustering

Algoritma K-Means adalah salah satu teknik clustering yang berfungsi untuk mengelompokkan data berdasarkan tingkat kemiripan atributnya. Data dengan karakteristik yang mirip akan dikelompokkan ke dalam satu cluster, sedangkan data dengan karakteristik yang berbeda akan ditempatkan pada cluster yang lain [17]. Secara umum, algoritma *K-Means clustering* bekerja melalui proses berikut:

1. Tentukan nilai k, yang mewakili jumlah cluster yang akan dibuat.
2. Menginisialisasi k pusat kluster (*centroid*) secara acak dari data yang tersedia.
3. Menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus *Euclidean Distance*, untuk menentukan kedekatan antara data dengan centroid tertentu.

$$d(a_i, b_j) = \sqrt{\sum (a_i + b_j)^2} \quad (2)$$

dengan keterangan:

a_i : data kriteria,

b_j : centroid pada ke-j

4. Mengelompokkan setiap data ke dalam cluster berdasarkan jarak terdekat dengan centroid.
5. Memperbarui posisi centroid dengan menghitung nilai rata-rata dari seluruh data yang berada dalam cluster yang sama

Menggunakan rumus:

$$b_j(t+1) = \frac{1}{NS_j} \sum_{j \in S_j} a_j \quad (3)$$

di mana:

$b_j(t+1)$: centroid baru pada iterasi ke (t + 1)

NS_j : banyak data pada cluster S

6. langkah 2 hingga 5 diulangi hingga tidak terjadi perubahan keanggotaan pada masing-masing cluster.

2.5 Pengujian Hasil

Metode *Davies-Bouldin Index* (DBI) digunakan untuk mengevaluasi kualitas hasil dari model *clustering*. DBI menilai hasil pengelompokan berdasarkan dua aspek utama, yaitu kohesi dan

separasi. Nilai kohesi menggambarkan tingkat kedekatan setiap data terhadap titik pusat (*centroid*) dalam satu cluster, sedangkan nilai separasi menunjukkan jarak antar centroid dari masing-masing cluster yang terbentuk.[18] *Davies Bouldin index* dihitung menggunakan rumus:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{s_i + s_j}{d_{ji}} \right) \quad (4)$$

Penjelasan:

1. k: Jumlah klaster.
2. S_i : Kohesi (*scatter*) dari cluster i
3. S_j : Kohesi (*scatter*) dari cluster j
4. D_{ji} : Separasi (jarak) antara centroid cluster i dan centroid cluster j

2.6 Rekomendasi Sistem Cerdas

Tahapan ini merupakan langkah akhir dalam proses penelitian, di mana hasil dari metode *K-Means Clustering* dimanfaatkan untuk menghasilkan rekomendasi secara cerdas dan terarah berdasarkan karakteristik setiap cluster yang terbentuk. Sistem menganalisis pola serta kecenderungan data pada masing-masing kelompok untuk memberikan rekomendasi yang relevan sesuai dengan hasil pengelompokan. Dengan demikian, sistem tidak hanya menampilkan hasil klasterisasi atau pengelompokan data, tetapi juga memberikan informasi analitis dan aplikatif guna mendukung proses pengambilan keputusan.

Pada tahap ini digunakan pendekatan *Rule-Based Recommendation System* sebagai dasar dalam menghasilkan rekomendasi otomatis. Sistem berbasis aturan (*rule-based*) bekerja dengan cara menetapkan seperangkat aturan logika (*if-then rules*) yang menghubungkan karakteristik cluster dengan saran atau tindakan tertentu.

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Penelitian ini menggunakan data nilai rapor siswa dari SMA Negeri 1 Aek Songsongan yang diperoleh dari bagian Ke- Siswa sekolah. Total data yang dikumpulkan sebanyak 1.384 data nilai rapor siswa dari tiga tahun ajaran, yaitu tahun ajaran 2020/2021, 2021/2022, dan 2022/2023. Data yang dikumpulkan merupakan data mentah (*raw data*) yang belum melalui proses *preprocessing*, sehingga masih berpotensi mengandung missing values, data yang duplikat, maupun anomali lainnya. Data yang telah dikumpulkan ini selanjutnya akan melalui tahap *preprocessing* untuk membersihkan dan mempersiapkan data sebelum dilakukan proses klasterisasi. Rincian tersebut disajikan dalam Tabel 1.

Table 1. Dataset

No	Nama	PKN	B.Indo	MM	keterampilan	B.Ing	...	Ekonomi
1	Ade Tri Mega utami	92.6	87.8	91.1	88	86	...	86.8
2	Agam Tanriansyah	89	90.3	87.6	86.8	86.7	...	85.3
3	Ahmad Kholid Setiawan	88.3	85.6	85.6	85.3	84.4	...	88
4	Airil Adril	92.6	87.8	91.1	88	86	...	85.7
...
1384	Frahma Nasution	89	90.3	87.6	86.8	86.7	...	85.3

3.2 Preprocessing Data

a) Missing Value Handling

Dataset awal berjumlah 1.384 data dengan 15 atribut. Ditemukan nilai kosong pada beberapa mata pelajaran, mulai dari 9 hingga 54 nilai kosong, terutama pada Penjas. Nilai kosong tidak diimputasi secara langsung karena sebagian besar teratasi otomatis setelah penghapusan data duplikat pada tahap berikutnya.

b) Data Cleaning

Tahap ini menghapus 520 data duplikat berdasarkan kombinasi NIS dan nama siswa yang melibatkan 211 siswa unik. Duplikasi muncul karena siswa tercatat dua kali dengan nilai berbeda. Penghapusan dilakukan dengan metode *keep='first'* dan menghapus baris bernama kosong atau tidak valid. Setelah pembersihan, data berkurang menjadi 1.075 baris, atau sekitar 22,3% dari data awal dihapus.

c) Data Selection

Dari 15 atribut awal, hanya 10 atribut yang dipertahankan, terdiri dari dua identitas (NIS, nama) dan delapan nilai mata pelajaran inti: PKN, Bahasa Indonesia, Matematika, Bahasa Inggris, Fisika, Kimia, Biologi, dan Keterampilan. Atribut lain seperti Penjas, Geografi, dan Agama dihapus karena tidak relevan dengan fokus penelitian pada prestasi akademik inti.

d) Feature Selection

Delapan mata pelajaran inti dipilih sebagai fitur utama yang paling relevan untuk mengukur prestasi akademik siswa.

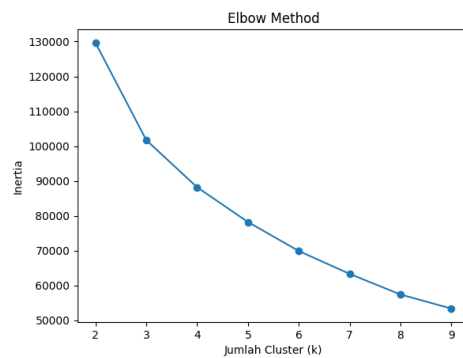
e) Feature Extraction

Fitur baru berupa rata-rata nilai delapan mata pelajaran akademik dibuat untuk menggambarkan performa akademik keseluruhan siswa. Hasil akhir preprocessing menghasilkan dataset bersih dengan 1.075 data dan 11 atribut (2 identitas, 8 nilai, dan 1 rata-rata), disimpan dalam file `data_preprocessed.xlsx` dan siap digunakan untuk analisis lebih lanjut.

Table 2. Hasil Preprocessing Data

No	Nama	PKN	B.Indo	MM	B.Ing	Fisika	Kimia	Biologi	Keterampilan
1	Ade Tri Mega Utami	92.6	87.8	91.1	86	86.3	86.2	87.6	88
2	Agam Tanriansyah	89	90.3	87.6	86.7	85.3	86.8	87.3	86.8
3	Ahmad Kholid Setiawan	88.3	85.6	85.6	84.4	85.3	84.5	86.5	85.3
4	Airil Adril	92.6	87.8	91.1	86	86.3	86.1	87.7	88
...
1075	Najam Nurdianti	90.1	91.7	88	86.2	89.8	66.3	87	88.5

3.3 Elbow Method



Gambar 3. Grafik Elbow

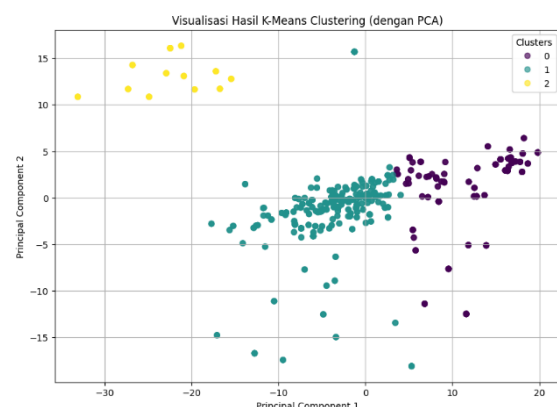
Berdasarkan *Elbow Method*, jumlah cluster optimal yang diperoleh adalah 3 cluster dengan nilai *inertia* sebesar 101719.01 dan algoritma konvergen pada iterasi ke-6. Grafik *Elbow Method* menampilkan hubungan antara jumlah cluster (sumbu X) dan nilai *inertia* (sumbu Y). Titik di mana grafik mulai melandai menandakan jumlah cluster optimal, yang direkomendasikan untuk digunakan dalam proses K-Means clustering.

3.4 K-means Clustering

Algoritma *K-Means Clustering* diterapkan pada data akademik sebanyak 1.075 siswa dengan delapan fitur, yaitu: Pendidikan Kewarganegaraan (PKN), Bahasa Indonesia, Matematika, Bahasa Inggris, Fisika, Kimia, Biologi, dan Keterampilan. Hasil pengujian dengan metode *Elbow* menunjukkan bahwa jumlah cluster optimal adalah tiga ($k = 3$). Setelah proses klusterisasi, terbentuk tiga kelompok siswa dengan proporsi dan karakteristik yang berbeda sebagaimana ditunjukkan pada Tabel 3 berikut.

Table 3. Hasil Data Setiap Cluster

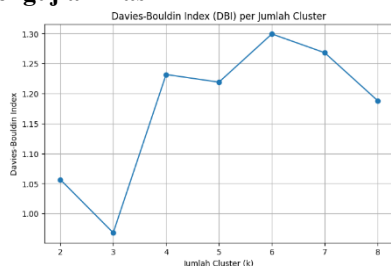
Cluster	Jumlah Siswa	Persentase
Cluster 0	317	29,5%
Cluster 1	706	65,7%
Cluster 2	52	4,8%



Gambar 4. Visualisasi distribusi hasil K-Means Clustering

Untuk memperjelas perbedaan antar cluster, distribusi data siswa divisualisasikan menggunakan *scatter plot* (diagram sebar). Cluster 0 terkonsentrasi pada area nilai tinggi di kedua sumbu yang terdiri dari 317 siswa (29,5%) yang merupakan kelompok berprestasi tinggi dengan nilai centroid berkisar antara 86,63-91,28 di semua mata pelajaran. Cluster 1 berada di area tengah dengan persebaran cukup merata yang merupakan kelompok terbesar dengan 706 siswa (65,7%) yang memiliki performa sedang-baik dengan nilai centroid berkisar 82,45-86,10. Sementara itu, Cluster 2 tersebar tidak teratur dengan perbedaan mencolok antara nilai tinggi dan rendah dengan 52 siswa (4,8%) yang menunjukkan pola nilai tidak merata, di mana nilai Fisika cukup tinggi (88,19) namun nilai Matematika (72,08) dan Kimia (66,16) sangat rendah, mengindikasikan adanya kesulitan belajar spesifik pada mata pelajaran eksakta tertentu.

3.5 Pengujian Hasil



Gambar 5. Davies-bouldin Index (DBI) per Jumlah Cluster

Evaluasi kualitas *clustering* menggunakan *Davies-Bouldin Index* (DBI) menghasilkan nilai 0,9678 yang menunjukkan kualitas pemisahan cluster sangat baik ($DBI < 1,0$). Nilai DBI yang mendekati nol ini mengindikasikan bahwa antar cluster memiliki pemisahan yang jelas dengan overlap minimal, sehingga setiap cluster memiliki karakteristik yang terdefinisi dengan baik [19]. Hasil ini memvalidasi bahwa algoritma K-Means berhasil mengelompokkan siswa berdasarkan pola prestasi akademik mereka secara efektif.

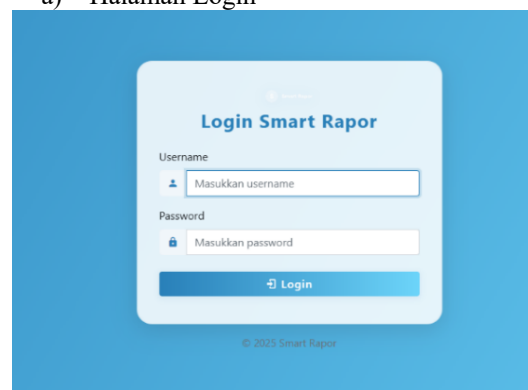
3.6 Rekomendasi sistem Cerdas

Berdasarkan hasil clustering, sistem cerdas mengimplementasikan mekanisme *rule-based recommendation* yang memetakan setiap cluster dengan strategi intervensi pendidikan spesifik. Untuk Cluster 0 dengan 317 siswa berprestasi tinggi, sistem merekomendasikan program akselerasi dan persiapan kompetisi akademik untuk memaksimalkan potensi siswa excellent. Pada Cluster 1 dengan 706 siswa yang memiliki performa sedang-baik, sistem menghasilkan rekomendasi berupa program bimbingan belajar reguler dan pemberian motivasi berkelanjutan untuk mendorong peningkatan prestasi ke level lebih tinggi.

Untuk Cluster 2 dengan 52 siswa yang menunjukkan pola nilai tidak merata, sistem memberikan rekomendasi komprehensif berupa program remedial intensif khususnya untuk Matematika (72,08) dan Kimia (66,16), konseling akademik, penerapan metode pembelajaran diferensiasi, serta pelibatan orang tua dalam monitoring pembelajaran. Anomali pada cluster ini, di mana nilai Fisika tinggi (88,19) namun nilai Matematika dan Kimia rendah, mengindikasikan kesulitan belajar spesifik yang memerlukan pendekatan *multi-stakeholder* dan *personalized learning*.

3.7 Implementasi Sistem

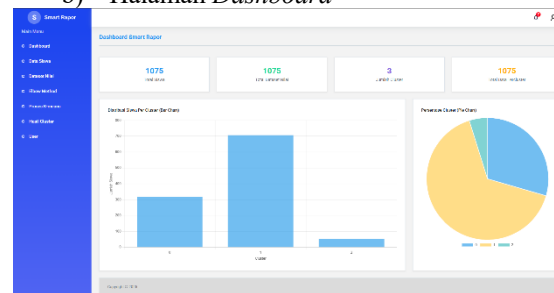
a) Halaman Login



Gambar 6. Halaman Login

Halaman Login pada sistem Smart Rapor berfungsi sebagai gerbang utama bagi pengguna untuk mengakses sistem sesuai rolnya. Admin, siswa, dan orang tua harus memasukkan username serta password terdaftar. Setelah login, pengguna diarahkan ke dashboard data. Jika login gagal, sistem menampilkan notifikasi kesalahan untuk mencegah akses tidak sah.

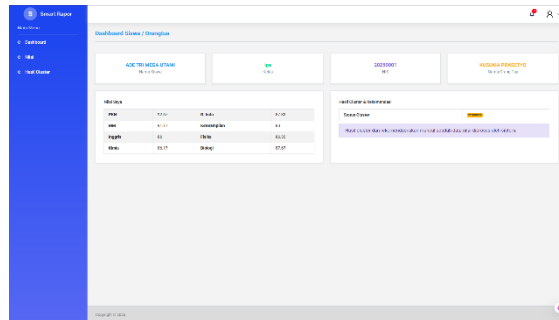
b) Halaman Dashboard



Gambar 7. Halaman Dashboard Admin

Halaman *Dashboard* Admin pada sistem Smart Rapor menampilkan ringkasan data berupa total siswa, total dataset nilai, jumlah cluster, dan total data yang telah terkelompok. Informasi ini dilengkapi dengan grafik batang yang menunjukkan distribusi siswa pada setiap cluster serta diagram lingkaran yang menggambarkan persentase masing-

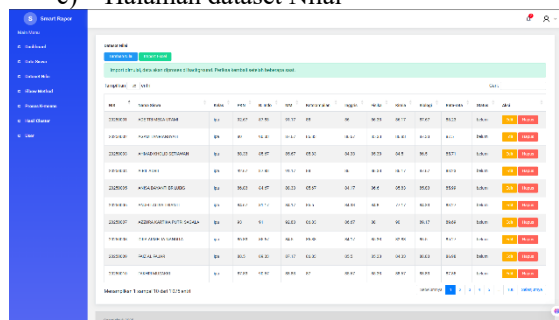
masing cluster, sehingga memudahkan admin dalam memantau hasil analisis nilai siswa.



Gambar 8. Halaman Dashboard Siswa/Orang Tua

Sementara itu, Dashboard Siswa/Orang Tua menampilkan data identitas seperti nama siswa, kelas, NIS, dan nama orang tua. Selain itu, terdapat tabel nilai siswa dari berbagai mata pelajaran yang digunakan sebagai dasar proses *clustering*. Bagian terakhir menampilkan status cluster dan rekomendasi hasil analisis sistem, sehingga siswa dan orang tua dapat mengetahui posisi kinerja akademik serta arahan pengembangan belajar berdasarkan hasil pengelompokan tersebut.

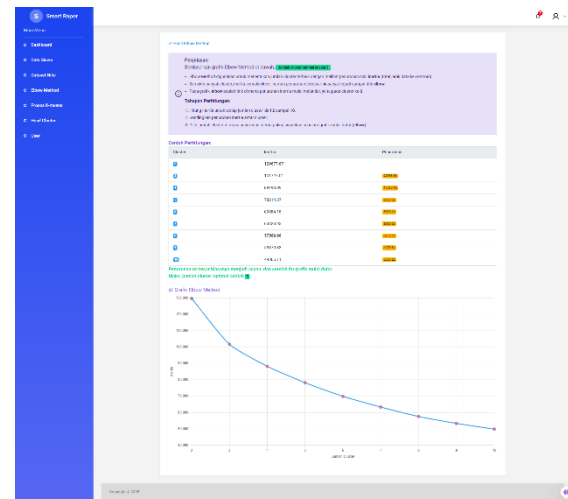
c) Halaman dataset Nilai



Gambar 9. Halaman Dataset Nilai

Halaman Dataset pada sistem Smart Rapor berfungsi untuk mengelola data nilai siswa yang akan dianalisis. Admin dapat menambahkan data dengan dua cara, yaitu melalui *import file* berformat .xls atau .xlsx, serta input manual langsung di sistem. Fitur import memudahkan pengunggahan data dalam jumlah besar, sedangkan input manual digunakan untuk menambah atau memperbaiki data secara individu. Hal ini membuat pengelolaan dataset lebih cepat, fleksibel, dan efisien sebelum dilakukan proses *clustering*.

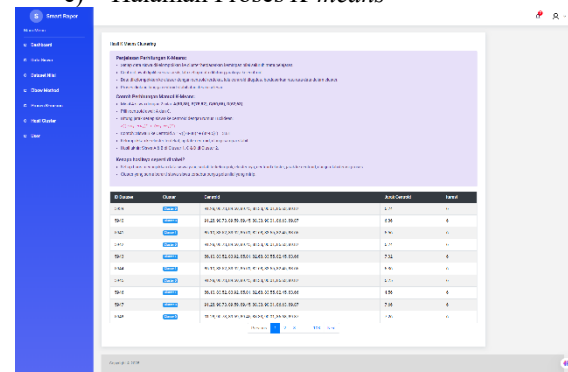
d) Halaman Elbow



Gambar 10. Halaman Elbow Method

Halaman *Elbow Method* pada sistem *Smart Rapor* berfungsi untuk menentukan jumlah cluster yang optimal dalam proses *clustering* data nilai siswa. Pada bagian tahapan perhitungan, sistem menghitung nilai *inertia* untuk setiap jumlah cluster, biasanya dari 2 hingga 10. Selanjutnya, sistem menampilkan hasil penurunan *inertia* antar cluster dalam bentuk tabel agar pengguna dapat melihat perbandingan penurunan yang paling signifikan.

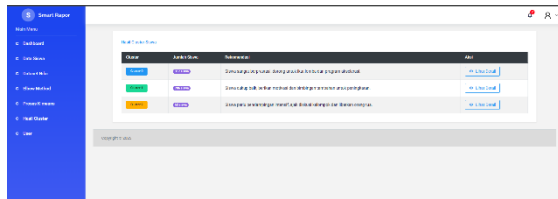
e) Halaman Proses K-means



Gambar 11. Halaman Proses K-means

Halaman Proses K-Means pada sistem Smart Rapor berfungsi untuk menjalankan proses pengelompokan (*clustering*) data nilai siswa berdasarkan kemiripan nilai dari seluruh mata pelajaran. Pada halaman ini, pengguna dapat menentukan sendiri jumlah cluster yang ingin digunakan, sesuai dengan hasil rekomendasi dari Elbow Method atau kebutuhan analisis tertentu.

f) Halaman Hasil Cluster

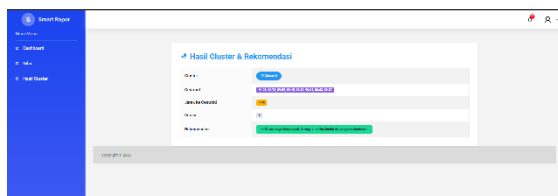


Gambar 12. Halaman Hasil Cluster Admin

Halaman Hasil *Clustering* pada admin sistem Smart Rapor menampilkan ringkasan hasil pengelompokan siswa berdasarkan analisis metode K-Means. Pada halaman ini terdapat informasi berupa jumlah cluster, jumlah siswa di setiap cluster, serta rekomendasi yang dihasilkan untuk masing-masing kelompok. Setiap cluster merepresentasikan tingkat pencapaian siswa yang berbeda, antara lain:

1. Cluster 0 berisi siswa berprestasi tinggi dengan rekomendasi untuk mengikuti lomba atau program akselerasi.
2. Cluster 1 berisi siswa dengan performa cukup baik yang disarankan untuk terus termotivasi dan mengikuti bimbingan tambahan.
3. Cluster 2 berisi siswa yang membutuhkan pendampingan intensif serta dukungan dari guru dan orang tua.

Adapun juga, halaman ini dilengkapi dengan fitur “Lihat Detail” yang menampilkan daftar siswa dalam setiap cluster serta fitur cetak (*print*) untuk admin mencetak laporan hasil clustering sebagai dokumen pendukung atau arsip sekolah.



Gambar 13. Halaman Hasil Cluster & Rekomendasi Siswa

Sementara itu, pada halaman hasil cluster siswa di tampilkan juga informasi tentang pengelompokan siswa secara individu serta memberikan rekomendasi deskripsi otomatis berdasarkan karakteristik Cluster. Informasi ini sangat bermanfaat untuk memahami karakteristik siswa dengan cepat.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian, penerapan sistem cerdas berbasis web menggunakan algoritma *K-Means Clustering* terbukti efektif dalam mengatasi permasalahan pengelolaan dan analisis data rapor siswa di SMA Negeri 1 Aek Songsongan. Sistem mampu mengolah 1.075 data valid dari tiga tahun ajaran dan menghasilkan tiga kelompok siswa dengan karakteristik prestasi yang berbeda. Evaluasi

menggunakan *Davies-Bouldin Index* (DBI) menghasilkan nilai 0,9678 yang menunjukkan kualitas pemisahan cluster yang sangat baik, sehingga memvalidasi efektivitas algoritma dalam mengidentifikasi pola prestasi siswa secara objektif dan terukur.

Penelitian ini Berkontribusi pada penerapan pendekatan *data-driven* dalam pengambilan keputusan pendidikan melalui integrasi metode clustering, evaluasi indeks cluster, dan sistem rekomendasi berbasis rule-based. Hasil ini menunjukkan bahwa pemanfaatan teknik data mining tidak hanya mampu mengungkap pola prestasi yang tidak terlihat secara manual, tetapi juga dapat diimplementasikan dalam bentuk sistem cerdas berbasis web yang adaptif dan mudah digunakan. Secara lebih luas, penelitian ini memperkuat pengembangan ilmu pengetahuan di bidang *Educational Data Mining* (EDM) dan memberikan model implementatif bagi sekolah dalam menerapkan teknologi analitik untuk peningkatan mutu pembelajaran.

Meskipun demikian, penelitian ini memiliki keterbatasan pada cakupan data yang hanya mencakup tiga tahun ajaran serta belum mempertimbangkan variabel non-akademik seperti kehadiran, minat belajar, atau faktor sosial ekonomi. Penelitian selanjutnya direkomendasikan untuk memperluas variabel analisis, membandingkan performa *K-Means* dengan algoritma clustering lain seperti *DBSCAN* atau *hierarchical clustering*, serta mengintegrasikan model *machine learning* prediktif untuk pengembangan *early warning system* sebagai inovasi lanjutan dalam sistem cerdas pendidikan.

Ucapan Terima kasih

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan bantuan dalam pelaksanaan penelitian ini, khususnya kepada Ibu Triase, S.T., M.Kom. atas bimbingan serta arahannya. Penghargaan juga ditujukan kepada editor jurnal atas kesempatan yang diberikan untuk publikasi. Diharapkan hasil penelitian ini dapat memberikan kontribusi dan manfaat bagi para pembaca.

Daftar Pustaka

- [1] M. Imbalo Zaki Hasibuan and T. Triase, “IMPLEMENTASI SISTEM DATABASE NoSQL SECARA REALTIME MENGGUNAKAN FIREBASE REALTIME DATABASE PADA APLIKASI OURTICLE,” *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 2, no. 1, pp. 1–24, 2022, doi: 10.54443/sibatik.v2i1.489.
- [2] D. Amalia, P. Lubis, R. A. Putri, and A. M. Harahap, “PENERAPAN DATA MINING

- UNTUK CLUSTERING KELAYAKAN PENERIMA BPNT MENGGUNAKAN ALGORITMA K-MEANS BERBASIS WEB,” vol. 4307, no. August, pp. 1254–1260, 2024.
- [3] Muttaqin, Yuswardi, A. Maulidinnawati, A. Parewe, I. F. Ashari, and M. Munsarif, “Pengantar Sistem Cerdas,” *Yayasan Kita Menulis*, no. March, pp. 1–222, 2023.
- [4] Y. Suryana and S. F. Nasrullah, “RANCANG BANGUN SISTEM INFORMASI AKADEMIK DI SMA 2 KUNINGAN BERBASIS WEBSITE,” vol. 8, no. 5, pp. 10212–10222, 2024.
- [5] M. Risqi Ananda, N. Sandra, E. Fadhila, A. Rahma, and N. Nurbaiti, “Data Mining dalam Perusahaan PT Indofood Lubuk Pakam,” *Com. Commun. Inf. Technol. J.*, vol. 2, no. 1, pp. 108–119, 2023, doi: 10.47467/commit.v2i1.124.
- [6] S. Sudriyanto, A. Khairi, and A. S. Hikam, “Penerapan Algoritma K-Means Untuk Clustering Santri Pra-Sejahtera Di Yayasan Bantuan Sosial (Ybs) Az-Zainiyyah Pondok Pesantren Nurul Jadid,” *NJCA (Nusantara J. Comput. Its Appl.)*, vol. 8, no. 1, p. 22, 2023, doi: 10.36564/njca.v8i1.234.
- [7] B. A. Prayudha *et al.*, “ALGORITMA K-MEANS UNTUK MENINGKATKAN MODEL KLASTERISASI DATA SISWA SMK SAMUDRA NUSANTARA KABUPATEN CIREBON,” vol. 9, no. 1, pp. 1314–1321, 2025.
- [8] R. Handayani, B. K. Apriani, and M. Mustari, “Pemanfaatan Rapor Pendidikan dalam Perencanaan Berbasis Data untuk Meningkatkan Mutu Sekolah di SDN 44 Ampenan,” vol. 10, pp. 336–342, 2025.
- [9] S. Dewi, S. Defit, and Y. Yuhandri, “Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means,” *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [10] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [11] A. Nurul Alifia, A. Fahrudi Setiawan, and D. Rudhistiar, “Penerapan Algoritma K-Means Clustering Dalam Peringatan Dini Resiko Kegagalan Siswa Pada Mata Pelajaran Bahasa Indonesia,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1174–1181, 2024, doi: 10.36040/jati.v8i2.9075.
- [12] S. N. Br Sembiring, H. Winata, and S. Kusnasari, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [13] J. R. S. Penda Sudarto Hasugian, “Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 3, pp. 262–268, 2022, [Online]. Available: <https://djournal.com/klik>
- [14] D. O. Dacwanda and Y. Nataliani, “Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan,” *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [15] S. Prastiwa and A. Wibowo, “ALGORITMA K-MEANS UNTUK MENGELOMPOKKAN K-MEANS ALGORITHM TO GROUP VILLAGES IN BANTEN PROVINCE BASED ON THE AREA AND NUMBER OF RECIPIENTS OF BULOG,” vol. 3, no. September, pp. 289–295, 2024.
- [16] N. A. Maori, “METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING,” vol. 14, no. 2, pp. 277–287, 2023.
- [17] D. Apriandi, R. M. Sari, and M. I. Sarif, “Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means,” *J. Minfo Polgan*, vol. 13, no. 1, pp. 1117–1129, 2024, doi: 10.33395/jmp.v13i1.13959.
- [18] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [19] D. A. Tarigan, “Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification,” vol. 4, no. 1, pp. 545–552, 2023, doi: 10.30865/klik.v4i1.964.