
Application of the K-Means Algorithm for Student Clustering as a Basis for Learning Strategies at SMK Negeri 4 Tanah Grogot

Fityan Noor¹⁾, Wahyuni²⁾, dan Rizki Galang Rahmadani³⁾

^{1,2,3} Teknik Informatika, STMIK Widya Cipta Dharma

^{1,2,3} Jl. M. Yamin No. 25, Samarinda, Kalimantan Timur 75123

E-mail: fityannoor2004@gmail.com¹, wahyuni@wicida.ac.id², galang@wicida.ac.id³

ABSTRACT

Academic score data at SMK Negeri 4 Tanah Grogot have mainly been used for learning reports and basic evaluation, while patterns of similarity among students have not been analysed systematically. This study applies the K-Means algorithm to group students according to academic-score similarity and uses the results as supporting information for learning strategies. The research followed the Cross Industry Standard Process for Data Mining (CRISP-DM), comprising Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. The final dataset consisted of 34 Grade XI Hospitality students in the 2024/2025 academic year. Nineteen raw attributes were cleaned, religious-subject columns were integrated, the data were transformed into pivot form, and 10 academic features were selected. Min-Max Normalization was applied before K-Means processing. Candidate values of K from 2 to 17 were identified through inertia changes in the Elbow Method, and the final K was chosen using the highest Silhouette Score among the candidates. The final model produced five clusters, an inertia of 13.742869, a Silhouette Score of 0.199861, and convergence in five iterations. Cluster membership was 8, 10, 5, 3, and 8 students. The results were deployed in a web-based system for managing academic data, running clustering, and displaying cluster profiles. Because the Silhouette Score indicates no substantial cluster structure, the results must not be used as a single decision basis and require teacher review alongside non-academic information.

Keywords: K-Means, Student Clustering, CRISP-DM, Silhouette Score, Learning Strategy

Penerapan Algoritma K-Means untuk Pengelompokan Siswa sebagai Dasar Strategi Pembelajaran pada SMK Negeri 4 Tanah Grogot

ABSTRAK

Data nilai akademik di SMK Negeri 4 Tanah Grogot selama ini lebih banyak digunakan untuk pelaporan hasil belajar dan evaluasi sederhana, sedangkan pola kemiripan nilai antarsiswa belum dianalisis secara sistematis. Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan siswa berdasarkan kemiripan nilai akademik dan menggunakan hasilnya sebagai informasi pendukung penyusunan strategi pembelajaran. Penelitian menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM), meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Dataset akhir terdiri atas 34 siswa kelas XI Program Keahlian Perhotelan Tahun Ajaran 2024/2025. Sebanyak 19 atribut mentah dibersihkan, kolom mata pelajaran agama digabungkan, data ditransformasikan ke bentuk *pivot*, dan dipilih 10 fitur nilai akademik. *Min-Max Normalization* diterapkan sebelum proses K-Means. Kandidat nilai K dari 2 sampai 17 diperoleh berdasarkan perubahan *inertia* melalui *Elbow Method*, kemudian nilai K akhir dipilih menggunakan *Silhouette Score* tertinggi pada kandidat tersebut. Model akhir menghasilkan lima *cluster*, *inertia* sebesar 13,742869, *Silhouette Score* sebesar 0,199861, dan mencapai konvergensi dalam lima iterasi. Distribusi anggota *cluster* adalah 8, 10, 5, 3, dan 8 siswa. Hasil penelitian diterapkan dalam sistem berbasis *website* untuk mengelola data akademik, menjalankan *clustering*, serta menampilkan profil *cluster*. Karena nilai *Silhouette Score* menunjukkan belum terdapat struktur *cluster* yang substantif, hasil pengelompokan tidak boleh digunakan sebagai dasar keputusan tunggal dan tetap memerlukan peninjauan guru bersama informasi nonakademik siswa.

Kata Kunci: K-Means, Pengelompokan Siswa, CRISP-DM, *Silhouette Score*, Strategi Pembelajaran

1. PENDAHULUAN

SMK Negeri 4 Tanah Grogot mengelola nilai akademik dari berbagai mata pelajaran sebagai bagian dari proses pelaporan dan evaluasi hasil belajar. Data tersebut memiliki potensi untuk memberikan gambaran mengenai pola kemiripan akademik antarsiswa, tetapi pemanfaatannya masih terbatas pada penyajian nilai dan evaluasi sederhana. Kondisi ini menyebabkan sekolah belum memperoleh informasi agregat yang dapat

membantu memahami kelompok siswa dengan karakteristik nilai yang relatif serupa.

Data penelitian berasal dari leger nilai kelas XI Program Keahlian Perhotelan Tahun Ajaran 2024/2025 semester genap. Data mentah memuat identitas siswa, nilai mata pelajaran, dan ketidakhadiran. Setelah melalui pembersihan, penggabungan kolom mata pelajaran agama, transformasi, dan seleksi fitur, data akademik dapat disusun menjadi matriks numerik yang sesuai untuk proses pengelompokan. Pengolahan tersebut diperlukan karena perbedaan format, nilai kosong, tanda

hubung pada mata pelajaran agama, data duplikat, dan atribut identitas dapat memengaruhi konsistensi analisis.

Pengelompokan secara manual berpotensi dipengaruhi penilaian subjektif dan menjadi semakin sulit ketika jumlah data dan atribut bertambah. Teknik *data mining* dapat digunakan untuk menemukan pola pada data pendidikan, sedangkan *clustering* mengelompokkan objek agar anggota dalam kelompok yang sama memiliki kemiripan lebih tinggi daripada objek pada kelompok lain (Han dkk., 2022). K-Means merupakan metode *clustering* berbasis partisi yang menempatkan data pada *centroid* terdekat dan memperbarui *centroid* secara iteratif sampai kondisi konvergen tercapai (Syakur dkk., 2021). Karakteristik tersebut sesuai dengan data nilai yang bersifat numerik.

Penelitian terdahulu menunjukkan bahwa K-Means telah digunakan untuk mengelompokkan prestasi siswa. Awallia (2024) membagi 21 siswa menjadi tiga kategori berdasarkan nilai rapor, tetapi belum menggunakan evaluasi kualitas *cluster* secara kuantitatif dan belum menghasilkan sistem berbasis *website*. Salsa dkk. (2025) mengolah 282 data siswa dan memperoleh tiga *cluster* dengan Davies-Bouldin Index, tetapi salah satu *cluster* hanya berisi satu siswa dan proses masih bergantung pada perangkat lunak khusus. Nana dkk. (2025) menghasilkan dua hingga tiga kategori akademik, namun jumlah *cluster* tidak konsisten dan tidak tersedia antarmuka yang dapat digunakan langsung oleh sekolah.

Penelitian Ayu dkk. (2023) mengelompokkan siswa berdasarkan nilai akademik dan profil pelajar Pancasila menggunakan MATLAB, sedangkan Sindrawati dkk. (2024) menggunakan nilai siswa untuk pengelompokan penerima beasiswa. Rangkaian penelitian tersebut memperlihatkan manfaat *clustering* pada data pendidikan, tetapi juga menunjukkan ruang pengembangan berupa penentuan jumlah *cluster* yang tidak ditetapkan secara tetap, penyediaan sistem yang dapat digunakan sekolah, dan pembatasan interpretasi agar nomor *cluster* tidak berubah menjadi label kemampuan permanen.

Kebaruan penelitian ini terletak pada penggunaan kombinasi *Elbow Method* dan *Silhouette Score* untuk memperoleh jumlah *cluster* berdasarkan karakteristik data, penerapan seluruh proses dalam sistem berbasis *website*, serta penyajian hasil sebagai analisis agregat yang harus ditinjau guru. *Elbow Method* digunakan untuk memperoleh kandidat berdasarkan perubahan *inertia*, sedangkan *Silhouette Score* memilih kandidat dengan kualitas internal terbaik. Hasil *cluster* tetap dinamai *Cluster 1* sampai *Cluster 5* tanpa label tinggi, sedang, atau rendah.

Tujuan penelitian adalah menerapkan algoritma K-Means untuk mengelompokkan siswa berdasarkan kemiripan nilai akademik sebagai dasar informasi penyusunan strategi pembelajaran di SMK Negeri 4 Tanah Grogot. Kontribusi penelitian meliputi *pipeline* pengolahan data dari leger sampai hasil *clustering*, profil rata-rata setiap *cluster*, arah strategi pembelajaran yang

bersifat rekomendatif, dan sistem berbasis *website* yang membatasi akses Administrator dan Guru sesuai kewenangannya.

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

1. Cakupan permasalahan meliputi pengolahan leger nilai akademik siswa kelas XI Perhotelan, penyiapan data numerik, penentuan jumlah *cluster*, penerapan K-Means, evaluasi internal hasil *clustering*, interpretasi profil *cluster*, serta penerapan hasil pada sistem berbasis *website*.
2. Batasan penelitian adalah penggunaan 34 siswa dari satu kelas pada Tahun Ajaran 2024/2025 semester genap, 10 fitur nilai mata pelajaran, algoritma K-Means tanpa perbandingan metode lain, dan evaluasi utama menggunakan *Silhouette Score*. Data ketidakhadiran, sikap, keterampilan, keaktifan, minat, kondisi psikologis, serta prestasi nonakademik tidak digunakan dalam perhitungan jarak.
3. Rencana hasil yang diperoleh adalah kelompok siswa berdasarkan kemiripan nilai akademik, distribusi anggota dan profil rata-rata setiap *cluster*, arah pembelajaran yang dapat ditinjau guru, serta sistem yang mengelola data siswa, mata pelajaran, nilai, proses *clustering*, dan hasil evaluasi. Nomor *cluster* tidak digunakan sebagai peringkat dan hasil *clustering* tidak menjadi dasar keputusan tunggal.

3. BAHAN DAN METODE

3.1 Algoritma K-Means dan Euclidean Distance

K-Means mengelompokkan data numerik ke dalam K *cluster* berdasarkan jarak terhadap *centroid*. Proses dimulai dengan menentukan *centroid* awal, menghitung jarak setiap data ke seluruh *centroid*, menempatkan data pada *centroid* terdekat, menghitung *centroid* baru dari rata-rata anggota, kemudian mengulangi proses sampai tidak terjadi perubahan keanggotaan atau kondisi konvergen tercapai. K-Means dipilih karena data penelitian terdiri atas nilai mata pelajaran yang dapat direpresentasikan sebagai vektor numerik (Syakur dkk., 2021).

Jarak antara data siswa x dan *centroid* y dihitung menggunakan Euclidean Distance sebagaimana Persamaan (1) (Han dkk., 2022). Semakin kecil jarak yang diperoleh, semakin dekat karakteristik nilai siswa terhadap *centroid* yang dibandingkan. Dalam penelitian ini p menyatakan jumlah fitur mata pelajaran, sedangkan x_j dan y_j masing-masing menyatakan nilai fitur ke- j pada data siswa dan *centroid*.

$$d(x,y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (1)$$

Pada Persamaan (1), $d(x,y)$ adalah jarak antara data x dan *centroid* y , sedangkan p adalah jumlah fitur mata pelajaran.

3.2 Normalisasi dan Evaluasi Cluster

Nilai akademik berada pada skala 0 sampai 100, tetapi setiap mata pelajaran memiliki nilai minimum, maksimum, dan penyebaran yang berbeda. *Min-Max*

Normalization digunakan untuk mengubah setiap fitur ke rentang 0 sampai 1 agar kontribusi fitur dalam perhitungan jarak lebih seimbang (García dkk., 2023). Rumus normalisasi ditunjukkan pada Persamaan (2), dengan x sebagai nilai asli, $x_{\min}$ sebagai nilai minimum, $x_{\max}$ sebagai nilai maksimum, dan x' sebagai nilai hasil normalisasi.

Kualitas hasil *clustering* dievaluasi menggunakan *Silhouette Score*. Nilai $a(i)$ menyatakan rata-rata jarak data ke- i terhadap anggota *cluster* yang sama, sedangkan $b(i)$ menyatakan rata-rata jarak minimum data tersebut terhadap *cluster* lain. Nilai *Silhouette* setiap data dihitung dengan Persamaan (3), kemudian dirata-ratakan untuk seluruh data. Nilai mendekati 1 menunjukkan penempatan yang baik, nilai mendekati 0 menunjukkan tumpang tindih antarkelompok, dan nilai negatif menunjukkan kemungkinan penempatan yang kurang sesuai (Zhang & Li, 2023). Kriteria penelitian menyatakan nilai di bawah 0,25 belum menunjukkan struktur *cluster* yang substantif.

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (2)$$

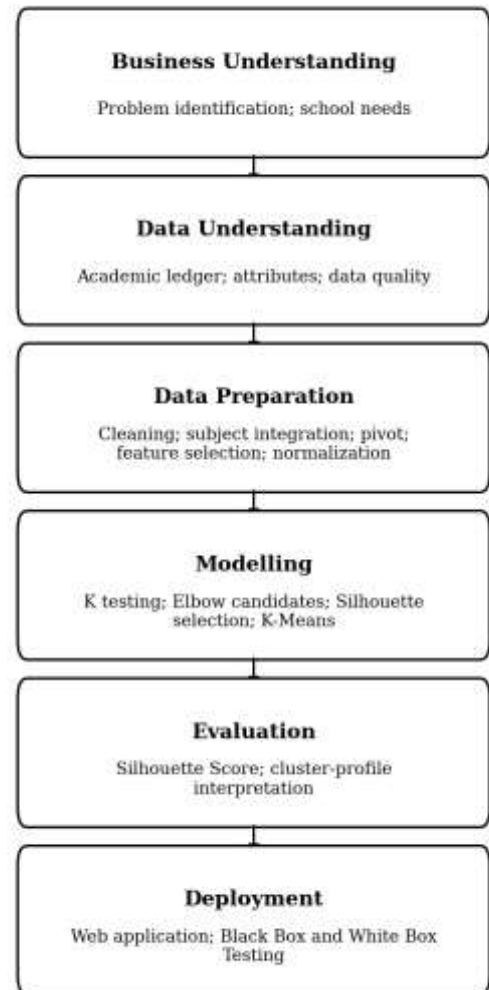
$$s(i) = [b(i) - a(i)] / \max\{a(i), b(i)\} \quad (3)$$

Persamaan (2) mengubah nilai ke rentang 0 sampai 1. Persamaan (3) membandingkan kohesi data pada *cluster* asal dengan separasinya terhadap *cluster* terdekat.

3.3 Tahapan CRISP-DM

Metode penelitian menggunakan CRISP-DM karena menyediakan tahapan *data mining* yang terstruktur, terdokumentasi, dan dapat ditinjau kembali apabila ditemukan ketidaksesuaian pada data atau hasil pemodelan (Hidayati dkk., 2021; Salsabila dkk., 2023). Enam tahap yang digunakan adalah *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Alur penelitian diringkas pada Gambar 1.

Business Understanding digunakan untuk memahami masalah pemanfaatan nilai akademik dan kebutuhan sekolah. *Data Understanding* mengidentifikasi jumlah data, atribut, tipe, rentang nilai, kelengkapan, dan duplikasi. *Data Preparation* mencakup pembersihan, penggabungan mata pelajaran agama, transformasi *pivot*, seleksi fitur, dan normalisasi. *Modelling* menguji beberapa nilai K dan menjalankan K -Means. *Evaluation* menilai kualitas *cluster* menggunakan *Silhouette Score* dan menafsirkan profil rata-rata. *Deployment* menerapkan proses pada sistem *website* serta melakukan *Black Box Testing* dan *White Box Testing*.



Gambar 1. Bagan Alur Penelitian

Figure 1. Research Workflow

Sumber: Hasil pengolahan peneliti (2026).

3.4 Sumber Data dan Persiapan Data

Pengumpulan data dilakukan melalui studi pustaka, studi dokumen, dan wawancara dengan Wakil Kepala Sekolah Bidang Kurikulum. Studi dokumen menghasilkan leger nilai akademik siswa, daftar mata pelajaran, identitas siswa, data ketidakhadiran, dan informasi tahun ajaran. Wawancara digunakan untuk memahami pengelolaan nilai, proses evaluasi, serta kebutuhan sistem pengelompokan.

Dataset mentah memiliki 19 atribut. Proses pembersihan menghapus spasi berlebih, menyeragamkan format kelas, menyaring header, memvalidasi nilai pada rentang 0 sampai 100, dan mencegah duplikasi. Tanda hubung pada kolom agama tidak diperlakukan sebagai nilai nol; beberapa kolom agama digabungkan menjadi satu fitur Pendidikan Agama dan Budi Pekerti. Data kemudian diubah ke bentuk *pivot* agar satu baris mewakili satu siswa dan satu kolom mewakili satu mata pelajaran.

Atribut ID, NISN, nama, kelas, dan ketidakhadiran dipertahankan sebagai informasi, tetapi dikeluarkan dari perhitungan jarak. Sepuluh fitur yang digunakan adalah Pendidikan Agama dan Budi Pekerti, Pendidikan Pancasila, Bahasa Indonesia, Pendidikan Jasmani Olahraga dan Kesehatan, Sejarah, Muatan Lokal Bahasa Daerah, Matematika, Bahasa Inggris, Perhotelan, serta Projek Kreatif dan Kewirausahaan.

Tabel 1. Fitur yang Digunakan dalam Clustering

Table 1. Features Used for Clustering

Atribut / Attribute	Status
Pendidikan Agama dan Budi Pekerti	Ya / Yes
Pendidikan Pancasila	Ya / Yes
Bahasa Indonesia	Ya / Yes
Pendidikan Jasmani, Olahraga, dan Kesehatan	Ya / Yes
Sejarah	Ya / Yes
Muatan Lokal Bahasa Daerah	Ya / Yes
Matematika (Umum)	Ya / Yes
Bahasa Inggris	Ya / Yes
Perhotelan	Ya / Yes
Projek Kreatif dan Kewirausahaan	Ya / Yes

3.5 Penentuan Jumlah Cluster dan Pemodelan

Pengujian jumlah *cluster* dilakukan pada $K=2$ sampai $K=17$. Untuk setiap K , sistem menjalankan K-Means dan mencatat *inertia* serta *Silhouette Score*. Selisih penurunan *inertia* antarnilai K dihitung, kemudian rata-rata penurunan digunakan untuk memperoleh kandidat *Elbow Method*. Nilai K dengan penurunan sekurangnya sama dengan rata-rata penurunan ditetapkan sebagai kandidat. Dari kandidat tersebut, sistem memilih nilai K dengan *Silhouette Score* tertinggi untuk proses *clustering* akhir.

Setelah K terpilih, K-Means dijalankan terhadap data hasil normalisasi. Setiap siswa ditempatkan pada *centroid* terdekat, *centroid* diperbarui berdasarkan rata-rata anggota, dan perhitungan diulang sampai konvergen. Sistem menyimpan nomor *cluster*, jumlah anggota, *inertia*, *Silhouette Score*, jumlah iterasi, dan profil nilai untuk ditampilkan pada halaman hasil.

3.6 Implementasi dan Pengujian Sistem

Deployment dilakukan pada aplikasi berbasis *website* dengan *Apache* sebagai *web server*, *PHP* untuk antarmuka dan proses aplikasi, *MySQL* untuk penyimpanan data, serta *Python* untuk persiapan data, K-Means, perhitungan *inertia*, dan *Silhouette Score*. Administrator dapat mengelola data siswa, mata pelajaran, nilai, pengguna, impor leger, dan proses *clustering*. Guru hanya dapat melihat dashboard dan hasil pengelompokan.

Black Box Testing memeriksa keluaran fungsi berdasarkan masukan pengguna, sedangkan *White Box Testing* memverifikasi alur logika penentuan jumlah *cluster* dan pemilihan kandidat. Pengujian numerik juga membandingkan normalisasi, Euclidean Distance, penempatan *cluster*, pembaruan *centroid*, *inertia*, dan *Silhouette Score* dengan perhitungan manual pada data uji berukuran kecil (Zen dkk., 2024).

4. PEMBAHASAN

4.1 Hasil Persiapan Data

Proses persiapan menghasilkan matriks akhir 34×10 atau 340 nilai numerik. Penyatuan kolom agama menghilangkan perbedaan representasi antarjenis pendidikan agama, sedangkan transformasi *pivot* memastikan setiap siswa memiliki susunan fitur yang konsisten. Atribut identitas dan ketidakhadiran tidak digunakan pada perhitungan K-Means agar jarak hanya mencerminkan kemiripan nilai akademik.

Normalisasi tetap digunakan meskipun nilai *Silhouette* pada data normalisasi tidak dinyatakan lebih tinggi daripada data asli. Tujuan utamanya adalah menyeragamkan rentang fitur dan mencegah mata pelajaran dengan penyebaran lebih besar mendominasi Euclidean Distance. Oleh karena itu, perbedaan nilai *inertia* antara data asli dan data normalisasi tidak dibandingkan langsung karena keduanya berada pada skala berbeda.

4.2 Hasil Pemodelan

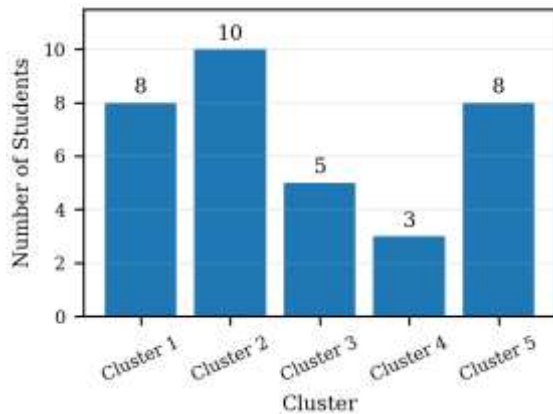
Pengujian $K=2$ sampai $K=17$ menghasilkan lima *cluster* sebagai model akhir. K-Means mencapai kondisi konvergen dalam lima iterasi dengan *inertia* sebesar 13,742869. *Silhouette Score* model akhir adalah 0,199861. Ringkasan parameter dan hasil pemodelan ditampilkan pada Tabel 2.

Distribusi anggota *cluster* adalah 8 siswa pada *Cluster* 1, 10 siswa pada *Cluster* 2, 5 siswa pada *Cluster* 3, 3 siswa pada *Cluster* 4, dan 8 siswa pada *Cluster* 5. Distribusi tersebut ditampilkan pada Gambar 2. Perbedaan jumlah anggota menunjukkan variasi pola nilai akademik, tetapi belum dapat dimaknai sebagai tingkatan kemampuan karena nomor *cluster* hanya berfungsi sebagai identitas kelompok.

Tabel 2. Ringkasan Hasil Pemodelan

Table 2. Summary of Modelling Results

Parameter	Hasil / Result
Jumlah siswa / Number of students	34
Atribut mentah / Raw attributes	19
Fitur akhir / Final features	10
Rentang K / K range	2–17
K terpilih / Selected K	5
<i>Inertia</i>	13,742869
<i>Silhouette Score</i>	0,199861
Iterasi / Iterations	5



Gambar 2. Distribusi Anggota Cluster

Figure 2. Cluster Membership Distribution

Sumber: Hasil pengolahan peneliti (2026).

4.3 Profil Cluster dan Arah Pembelajaran

Profil *centroid* dikembalikan ke skala nilai 0 sampai 100 agar mudah dibaca guru. Cluster 1 memiliki rata-rata keseluruhan 84,09 dan relatif memerlukan penguatan Matematika serta Sejarah. Cluster 2 memiliki rata-rata 81,53 dan memerlukan perhatian pada Matematika serta Bahasa Indonesia. Cluster 3 memiliki rata-rata 82,34 dan relatif memerlukan penguatan Matematika serta Sejarah. Cluster 4 memiliki rata-rata tertinggi, yaitu 86,73, sehingga dapat diarahkan pada pengayaan. Cluster 5 memiliki rata-rata 84,03 dan relatif memerlukan penguatan Bahasa Indonesia serta Bahasa Inggris.

Arah pembelajaran diturunkan dari profil rata-rata, bukan dari label kemampuan tetap. Kelompok yang memiliki rata-rata relatif lebih rendah pada mata pelajaran tertentu dapat memperoleh remedial, latihan bertahap, pembelajaran kontekstual, pendampingan terstruktur, atau tutor sebaya. Kelompok dengan penguasaan relatif lebih baik dapat memperoleh pengayaan dan tugas proyek dengan tingkat kesulitan lebih tinggi. Pendekatan tersebut sejalan dengan gagasan bahwa interpretasi *cluster* perlu memperhatikan karakteristik dan kebutuhan belajar, bukan sekadar nomor kelompok (Marantika dkk., 2023).

Tabel 3 merangkum jumlah anggota, rata-rata keseluruhan, fokus penguatan, dan arah pembelajaran. Setiap rekomendasi harus ditinjau berdasarkan kondisi nyata di kelas. Guru dapat menggabungkan hasil *clustering* dengan observasi, hasil asesmen, keterampilan, sikap, kehadiran, dan keaktifan sebelum menentukan tindakan pembelajaran.

Tabel 3. Profil dan Arah Pembelajaran Setiap Cluster

Table 3. Cluster Profiles and Learning Directions

Cluster	n	Rata-rata / Mean	Arah Pembelajaran / Learning Direction
1	8	84,09	Penguatan Matematika dan Sejarah; latihan bertahap dan pembelajaran kontekstual.
2	10	81,53	Penguatan Matematika dan Bahasa Indonesia; pendampingan, remedial, dan latihan berulang.

Cluster	n	Rata-rata / Mean	Arah Pembelajaran / Learning Direction
3	5	82,34	Penguatan Matematika dan Sejarah; remedial terarah dan tutor sebaya.
4	3	86,73	Pengayaan dan tugas proyek lebih menantang, dengan peninjauan kebutuhan individu.
5	8	84,03	Penguatan Bahasa Indonesia dan Bahasa Inggris melalui literasi dan presentasi.

4.4 Evaluasi Kualitas Cluster

Silhouette Score sebesar 0,199861 berada di bawah 0,25. Berdasarkan kriteria yang digunakan dalam penelitian, nilai tersebut menunjukkan belum terdapat struktur *cluster* yang bermakna atau substantif. Kedekatan anggota dalam *cluster* belum sepenuhnya terpisah dari *cluster* lain sehingga masih terdapat tumpang tindih antarkelompok.

Temuan ini penting karena kualitas internal yang lemah membatasi cara hasil digunakan. Sistem dapat membantu merangkum pola nilai, tetapi tidak dapat menyatakan bahwa seluruh siswa dalam satu *cluster* memiliki kebutuhan yang sama. Hasil pengelompokan hanya menjadi informasi awal dan tidak layak digunakan sebagai dasar keputusan tunggal. Pelaporan nilai evaluasi secara terbuka juga mencegah interpretasi berlebihan terhadap hasil K-Means.

4.5 Implementasi Sistem dan Hasil Pengujian

Sistem *website* menyediakan halaman login, dashboard, pengelolaan data siswa, pengelolaan mata pelajaran, input nilai, impor leger, proses *clustering*, hasil *cluster*, dan pengelolaan pengguna. Halaman proses *clustering* memungkinkan Administrator memilih data yang akan diolah, menjalankan normalisasi dan K-Means, serta menghitung *Silhouette Score*. Halaman hasil menampilkan distribusi anggota, profil rata-rata mata pelajaran, nilai evaluasi, dan daftar siswa pada setiap *cluster*.

Pengujian fungsional mencakup login benar dan salah, tambah-ubah-hapus data siswa, mata pelajaran, dan nilai, impor file valid dan tidak valid, proses *clustering*, perhitungan *Silhouette Score*, serta penyajian distribusi *cluster*. Seluruh skenario yang dilaporkan memberikan hasil berhasil. Ringkasan fungsi utama disajikan pada Tabel 4.

White Box Testing menghasilkan empat jalur independen pada logika penentuan jumlah *cluster*. Kasus uji mencakup data kurang dari tiga siswa, pengujian K pada data yang memenuhi syarat, perulangan perhitungan *inertia* dan *Silhouette Score*, serta pemilihan kandidat *Elbow Method*. Hasil pengujian menunjukkan alur yang diuji berjalan sesuai kondisi yang diharapkan, tetapi keberhasilan pengujian program tidak sama dengan kualitas *cluster*; kualitas tetap ditentukan melalui evaluasi *Silhouette Score*.

Tabel 4. Ringkasan Pengujian Fungsional
Table 4. Summary of Functional Testing

Fungsi / Function	Hasil / Result
Login	Berhasil / Passed
Data siswa	Berhasil / Passed
Mata pelajaran	Berhasil / Passed
Input nilai	Berhasil / Passed
Import leger	Berhasil / Passed
Proses <i>clustering</i>	Berhasil / Passed
Hasil <i>clustering</i>	Berhasil / Passed

4.6 Pembahasan Kontribusi dan Keterbatasan

Dibandingkan penelitian terdahulu yang menetapkan dua atau tiga kategori secara langsung, penelitian ini menguji rentang K dan memisahkan proses pencarian kandidat melalui *Elbow Method* dari proses pemilihan melalui *Silhouette Score*. Pendekatan ini membuat jumlah *cluster* mengikuti karakteristik data, meskipun kualitas akhir tetap rendah. Implementasi pada *website* juga mengurangi ketergantungan pengguna terhadap perangkat lunak analitik khusus.

Kontribusi lain adalah penggunaan profil *cluster* sebagai analisis agregat tanpa menjadikan nomor *cluster* sebagai label prestasi. Pendekatan ini lebih berhati-hati daripada kategorisasi tinggi, sedang, dan rendah karena hasil K-Means bergantung pada data, skala, inisialisasi, dan parameter. Dengan demikian, guru tetap menjadi pihak yang menilai kesesuaian arah pembelajaran.

Keterbatasan penelitian meliputi jumlah data yang hanya berasal dari satu kelas, penggunaan 10 fitur nilai akademik tanpa atribut nonakademik, evaluasi internal pada satu *dataset*, belum adanya pengujian kestabilan keanggotaan pada beberapa inisialisasi, belum adanya perbandingan dengan metode *clustering* lain, dan belum dilakukannya pengujian efektivitas strategi pembelajaran secara langsung. Sistem juga belum menyediakan penonaktifan akun dan pencatatan log aktivitas secara lengkap.

5. KESIMPULAN

Penerapan K-Means terhadap data nilai akademik 34 siswa kelas XI Perhotelan menghasilkan matriks 34×10 setelah proses pembersihan, penggabungan mata pelajaran agama, transformasi *pivot*, seleksi fitur, dan normalisasi. Pengujian K=2 sampai K=17 menghasilkan lima *cluster*, *inertia* 13,742869, dan konvergensi dalam lima iterasi. Distribusi anggota adalah 8, 10, 5, 3, dan 8 siswa, dengan rata-rata keseluruhan masing-masing 84,09; 81,53; 82,34; 86,73; dan 84,03. Profil tersebut dapat digunakan untuk mengidentifikasi mata pelajaran yang relatif memerlukan penguatan atau pengayaan, tetapi nomor *cluster* tidak menunjukkan urutan kemampuan. Nilai *Silhouette Score* sebesar 0,199861 menunjukkan belum terdapat struktur *cluster* yang substantif dan masih terdapat tumpang tindih antarkelompok. Oleh karena itu, hasil *clustering* hanya berfungsi sebagai informasi pendukung dan harus dikombinasikan dengan penilaian guru terhadap aspek akademik maupun nonakademik. Sistem berbasis *website*

berhasil menerapkan pengelolaan data, proses K-Means, penyajian distribusi anggota, profil *cluster*, dan pembatasan akses berdasarkan peran pengguna.

6. SARAN

1. *Dataset* utama perlu diverifikasi bersama sekolah dan diberi identitas versi agar jumlah siswa, atribut, nilai kosong, duplikasi, serta tanda hubung pada mata pelajaran agama konsisten dan dapat direproduksi.
2. Seluruh *pipeline* perlu mencatat parameter seperti *random state*, *n init*, rentang K, dan versi pustaka. Kestabilan *cluster* perlu diuji melalui beberapa inisialisasi atau resampling dan dilengkapi Davies-Bouldin Index serta Calinski-Harabasz Index.
3. Penelitian selanjutnya dapat membandingkan K-Means dengan Fuzzy C-Means, Hierarchical *Clustering*, atau DBSCAN setelah konsistensi data dipastikan, melibatkan lebih banyak kelas dan tahun ajaran, serta menguji efektivitas strategi pembelajaran secara langsung.
4. Penggunaan data tambahan seperti kehadiran, keterampilan, sikap, dan prestasi nonakademik harus memiliki kriteria jelas dan persetujuan sekolah. Keamanan sistem perlu ditingkatkan melalui password hash, pembatasan sesi, status akun, log aktivitas, pencadangan, masa simpan data, dan penyamaran identitas siswa untuk publikasi.

7. REFERENSI

- Awalia, R. (2024). Penerapan metode K-Means *clustering* untuk pengelompokan prestasi siswa menggunakan Orange *Data Mining*: Studi kasus di MTs Muhammadiyah Tallo Makassar. *Mechatronics Journal in Professional and Entrepreneur*.
- Ayu, S., Novriyenni, & Imeldawaty, G. (2023). Pengelompokan data siswa berdasarkan profil pelajar Pancasila menggunakan metode *clustering*. *Indonesian Journal of Education and Computer Science*, 2(2), 107–120.
- García, S., Luengo, J., & Herrera, F. (2023). Data preprocessing in *data mining* (2nd ed.). Springer.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Hidayati, N., Suntoro, J., & Setiaji, G. G. (2021). Perbandingan algoritma klasifikasi untuk prediksi cacat software dengan pendekatan CRISP-DM. *Jurnal Sains dan Informatika*, 7(2), 117–126. <https://doi.org/10.34128/jsi.v7i2.313>
- Marantika, J. N. R., Tomasouw, B. P., & Wenno, I. H. (2023). Pengelompokan siswa dalam pembelajaran berdiferensiasi berdasarkan karakteristik dan kebutuhan belajar. *Jurnal Pendidikan dan Pembelajaran*.
- Nana, S., Nining, R., & Annisa, A. S. (2025). Optimalisasi prestasi akademik siswa melalui pengelompokan indeks prestasi dengan K-Means *clustering*. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 4(2), 198–207.
- Narayana Sakti Aji, Fauzan Natsir, & Siti Istianah. (2023). Penentuan penjualan barang berdasarkan pengelompokan produk dengan K-Means *clustering*

-
- metode CRISP-DM pada CV Sembako Dina. *Journal Zetroem*, 5(2), 119–126.
<https://doi.org/10.36526/ztr.v5i2.3041>
- Salsabila, F., Fitrianti, I., Umaidah, Y., & Heryana, N. (2023). Penerapan metode CRISP-DM untuk analisis pengelompokan pekerja informal di Provinsi Jawa Barat dengan algoritma K-Means. *Jurnal Statistika dan Informatika*, 28(2).
- Salsa, N. A., Fuad, N. H., & Antonius, Y. K. (2025). Penerapan *data mining* dalam penilaian kinerja akademik siswa/i SMP YPI Pulogadung dengan metode K-Means *clustering*. *Jurnal Ilmiah Informatika (JIF)*.
- Sindrawati, Dodi, S., & Agung, R. R. (2024). Penerapan algoritma K-Means *clustering* pada data nilai siswa untuk menentukan kelompok penerima beasiswa. *Jurnal Sistem Informasi dan Informatika*.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2021). Integration K-Means *clustering* method and *elbow method* for identification of the best customer profile *cluster*. *IOP Conference Series: Materials Science and Engineering*.
- Zen, dkk. (2024). Pengujian perangkat lunak berbasis metode *black box testing* pada sistem informasi akademik. *Jurnal Rekayasa Sistem dan Teknologi Informasi*.
- Zhang, L., & Li, M. (2023). *Cluster validation methods: Silhouette and beyond*. Springer.