

Segmentasi Siswa Berdasarkan Karakteristik Akademik Dan Non-Akademik Menggunakan Algoritma *K-Means Clustering* Untuk Strategi Pembelajaran Personal (Studi Kasus: SMP Muhammadiyah 03 Medan)

Sylvia Giovani^{1*}, Andi Zullhery

¹Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info	ABSTRAK
Article history: Received Oktober 3, 2026 Revised Oktober 4, 2026 Accepted Oktober 4, 2026 Kata Kunci: Data Mining, K-Means Clustering, Segmentasi Siswa, Strategi Pembelajaran Personal, Elbow Method Keywords: Data Mining, K-Means Clustering, Student Segmentation, Personalized Learning Strategies, Elbow Method	Perbedaan karakteristik akademik dan non-akademik siswa menyebabkan guru memerlukan strategi pembelajaran sesuai kebutuhan setiap kelompok. Penelitian ini bertujuan membangun sistem segmentasi siswa berbasis web menggunakan algoritma K-Means Clustering sebagai dasar penyusunan strategi pembelajaran personal. Data yang digunakan merupakan data siswa kelas VII SMP Muhammadiyah 03 Medan yang meliputi nilai Matematika, Bahasa Indonesia, IPA, dan tingkat partisipasi ekstrakurikuler. Tahapan penelitian meliputi validasi data, konversi nilai ekstrakurikuler ke bentuk numerik, normalisasi menggunakan metode Z-Score, penentuan jumlah cluster menggunakan Elbow Method, serta segmentasi menggunakan algoritma K-Means Clustering. Sistem dikembangkan menggunakan Python dan Streamlit, dilengkapi visualisasi hasil clustering, interpretasi cluster, serta rekomendasi strategi pembelajaran. Hasil penelitian menunjukkan jumlah cluster optimal adalah tiga. Setiap cluster memiliki karakteristik akademik dan non-akademik berbeda sehingga sistem dapat memberikan rekomendasi strategi pembelajaran sesuai kelompok siswa. Pengujian Black Box Testing menunjukkan seluruh fungsi sistem berjalan sesuai harapan. Sistem yang dibangun mampu membantu guru melakukan segmentasi siswa secara sistematis dan mendukung penyusunan strategi pembelajaran yang lebih tepat sesuai karakteristik siswa. ABSTRACT <i>Differences in students' academic and non-academic characteristics require teachers to adopt instructional strategies tailored to each student group. This study develops a web-based student segmentation system using K-Means Clustering as a foundation for personalized learning strategies. Data consist of records from 7th-grade students at SMP Muhammadiyah 03 Medan, covering Mathematics, Indonesian, Natural Sciences, and extracurricular participation. Research stages include data validation, conversion of extracurricular participation into numerical values, Z-Score normalization, determination of the number of clusters through the Elbow Method, and K-Means segmentation. The system uses Python and Streamlit to visualize clustering results, interpret clusters, and recommend learning strategies. Results indicate three optimal clusters with distinct academic and non-academic characteristics, enabling appropriate recommendations for each group. Black Box Testing confirms that all system functions operate as intended. The developed system assists teachers in systematically segmenting students and supports more precise learning strategies aligned with student characteristics and the needs of each group.</i> <i>This is an open access article under the CC BY license</i>



Corresponding Author:

Sylvia Giovani
Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara,
Medan, Indonesia
Email: sylviaiovani030@gmail.com

1. PENDAHULUAN

Pendidikan global saat ini berupaya untuk menciptakan pembelajaran yang semakin individual, fleksibel, dan didukung oleh data. Di Indonesia, Kurikulum Merdeka juga mendorong pembelajaran yang dipersonalisasi sebagai strategi untuk memaksimalkan potensi siswa. Meskipun transformasi pendidikan semakin meningkat, banyak sekolah di Indonesia masih belum mampu menerapkan personalisasi secara tepat. Sementara generasi saat ini mendorong pembelajaran yang relevan, adaptif, dan berfokus pada kebutuhan individu.

Kurangnya personalisasi dalam strategi pembelajaran mengakibatkan beberapa siswa tidak mendapatkan dukungan yang sesuai. Ketidakakuratan pengelompokan siswa berdampak pada efektivitas pembelajaran dan distribusi kelompok yang kurang optimal [1]. Pengelompokan siswa berdasarkan karakteristik mereka menjadi penting untuk mendukung pembelajaran personal [2]. Dengan membagi siswa ke dalam kelompok berdasarkan karakteristik yang relatif seragam, baik akademik maupun non-akademik, sekolah dapat mengembangkan strategi intervensi, bimbingan, dan pengajaran yang lebih sesuai.

Sekolah menghadapi tantangan dalam menilai kinerja siswa saat menggunakan metode non-personalisasi yang mengabaikan keragaman siswa [3]. Siswa di SMP Muhammadiyah 03 Medan memiliki karakteristik akademik dan non-akademik yang beragam. Kondisi tersebut menunjukkan bahwa sekolah memerlukan pendekatan pembelajaran yang mampu mengakomodasi perbedaan karakteristik serta kebutuhan setiap siswa secara menyeluruh. Pengelompokan secara manual membutuhkan waktu dan dapat menghasilkan penilaian yang berbeda-beda sehingga diperlukan pengolahan data yang terstruktur.

K-Means telah digunakan dalam pengelompokan siswa berdasarkan profil nilai untuk membantu penentuan jurusan perguruan tinggi [4]. Pembelajaran personalisasi memungkinkan proses pembelajaran disesuaikan dengan kemampuan dan kebutuhan setiap siswa [5], sedangkan pembelajaran adaptif memberikan kesempatan kepada siswa untuk mengulang materi yang belum dikuasai atau melanjutkan setelah memahami materi sebelumnya [6]. Differentiated Instruction menyesuaikan konten, proses, produk, dan lingkungan belajar berdasarkan kesiapan, minat, serta profil belajar siswa [7].

Clustering membagi data ke dalam kelompok berdasarkan kesamaan karakteristik tanpa membutuhkan label kelas yang telah ditentukan sebelumnya [8]. Prinsipnya adalah membentuk kelompok dengan kemiripan tinggi di dalam kelompok dan kemiripan lebih rendah terhadap objek pada kelompok lain [9]. K-Means menentukan pusat kelompok dan menempatkan data berdasarkan jarak terdekat terhadap centroid; penerapannya juga telah digunakan pada data calon siswa dan penentuan siswa kelas unggulan [10], [11].

Kesenjangan yang menjadi perhatian penelitian ini adalah pemanfaatan kombinasi karakteristik akademik dan non-akademik siswa tingkat sekolah menengah pertama sebagai dasar penyusunan strategi pembelajaran personal. Data nilai Matematika, Bahasa Indonesia, IPA, dan ekstrakurikuler belum dimanfaatkan melalui pengolahan terstruktur untuk memperoleh kelompok siswa dengan karakteristik relatif serupa. Penelitian ini menerapkan algoritma K-Means pada data siswa SMP

Muhammadiyah 03 Medan untuk mengidentifikasi kelompok dan profil kebutuhan belajar, serta membangun sistem berbasis web yang menyajikan hasil segmentasi dan rekomendasi strategi pembelajaran sesuai karakteristik masing-masing kelompok.

2. METODE

2.1 Data dan Variabel Penelitian

Data yang digunakan merupakan data siswa kelas VII SMP Muhammadiyah 03 Medan yang bersumber dari arsip dokumen sekolah. Variabel akademik meliputi nilai Matematika, Bahasa Indonesia, dan IPA, sedangkan variabel non-akademik direpresentasikan melalui nilai ekstrakurikuler. Dataset diunggah dalam format Microsoft Excel (.xlsx), dibaca menggunakan Pandas, dan disimpan dalam DataFrame untuk diproses. Atribut nomor, nama siswa, dan kelas hanya berfungsi sebagai identitas atau informasi pendukung dan tidak digunakan dalam perhitungan clustering. Struktur variabel yang digunakan disajikan pada Tabel 1.

Tabel 1. Variabel untuk segmentasi siswa

Variabel	Representasi	Peran
Matematika	Nilai numerik	Karakteristik akademik
Bahasa Indonesia	Nilai numerik	Karakteristik akademik
IPA	Nilai numerik	Karakteristik akademik
Ekstrakurikuler	Kategori A, B, C	Karakteristik non-akademik

Penelitian hanya menggunakan algoritma K-Means dan tidak membandingkan kinerjanya dengan algoritma clustering lainnya. Keluaran penelitian dibatasi pada prototipe sistem web berupa dashboard analisis yang menampilkan hasil klasterisasi, profil siswa, dan strategi pembelajaran. Sistem diakses oleh pengguna atau guru untuk membantu penyusunan strategi yang sesuai dengan karakteristik kelompok siswa.

2.2 Validasi dan Pra-pemrosesan Data

Validasi dataset merupakan tahap pemeriksaan terhadap file yang diunggah untuk memastikan bahwa data memenuhi kebutuhan sistem sebelum dilakukan proses segmentasi. Tahap ini bertujuan untuk mengurangi kesalahan akibat format file, struktur kolom, atau data yang tidak sesuai. Pemeriksaan dataset yang dilakukan sistem meliputi:

Pemeriksaan meliputi format Microsoft Excel, nama dan kelengkapan kolom KELAS, Matematika, B.Indonesia, IPA, dan Ekstrakurikuler, serta kesesuaian tipe data numerik dan kategorikal. Jika kolom yang dibutuhkan tidak tersedia, sistem memberikan notifikasi dan proses segmentasi tidak dapat dilanjutkan. Baris yang memiliki nilai kosong pada atribut clustering tidak diikutsertakan dalam segmentasi; sistem menampilkan jumlah data yang tidak diproses.

Nilai ekstrakurikuler berbentuk kategori A, B, dan C dikonversi menjadi numerik agar dapat diproses menggunakan K-Means. Konversi mengikuti rentang nilai yang ditetapkan sekolah, sebagaimana ditunjukkan pada Tabel 2. Setelah atribut dipilih dan data kategorikal dikonversi, sistem membentuk dataset akhir yang siap digunakan pada tahap standardisasi Z-Score.

Tabel 2. Konversi kategori ekstrakurikuler

Kategori	Rentang nilai sekolah	Nilai numerik
A	90–100	3
B	80–89	2
C	70–79	1

2.3 Standardisasi Z-Score

Standardisasi dilakukan untuk menyetarakan skala setiap atribut sebelum perhitungan jarak Euclidean. Tanpa penyetaraan skala, variabel dengan rentang lebih besar dapat memberikan pengaruh yang lebih kuat terhadap hasil pengelompokan [12]. Nilai pada suatu variabel ditransformasikan sehingga mempunyai rata-rata 0 dan simpangan baku 1. Proses tersebut diterapkan pada empat atribut penelitian. Perhitungan rata-rata, simpangan baku, dan Z-Score dinyatakan pada Persamaan (1)–(3).

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (2)$$

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

Pada persamaan tersebut, μ merupakan rata-rata atribut, σ merupakan simpangan baku, x_i merupakan nilai observasi ke- i , n merupakan jumlah data, dan z merupakan hasil standardisasi. Sistem menggunakan StandardScaler pada pustaka Scikit-learn. Hasil standardisasi menjadi masukan bagi penentuan jumlah cluster dan proses segmentasi.

2.4 Penentuan Jumlah Cluster dan K-Means

Jumlah cluster ditentukan menggunakan Elbow Method melalui perhitungan WCSS (Within-Cluster Sum of Squares) untuk beberapa alternatif nilai k [13]. Sistem menguji $k = 1$ hingga $k = 10$, kemudian menampilkan grafik hubungan jumlah cluster dengan inertia atau WCSS. Titik siku digunakan sebagai pertimbangan untuk menetapkan jumlah cluster ketika penurunan WCSS mulai melandai.

Setelah jumlah cluster ditetapkan, centroid awal ditentukan menggunakan K-Means++ pada pustaka Scikit-learn. Setiap data ditempatkan pada cluster dengan centroid terdekat menggunakan Euclidean Distance, kemudian centroid diperbarui berdasarkan rata-rata nilai anggota cluster. Perhitungan jarak dan pembaruan centroid diulang sampai posisi centroid stabil atau memenuhi kondisi penghentian [14]. Tujuan K-Means adalah meminimalkan SSE sebagaimana Persamaan (4), sedangkan jarak dan pembaruan centroid dinyatakan pada Persamaan (5) dan (6).

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} |x_j - \mu_i|^2 \quad (4)$$

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (5)$$

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (6)$$

J merupakan total error, k merupakan jumlah cluster, C_i merupakan kelompok ke- i , x_j merupakan data anggota kelompok, dan μ_i merupakan centroid kelompok tersebut. Pada jarak Euclidean, x_i dan c_i menunjukkan nilai atribut data dan centroid. Centroid baru dihitung dari seluruh anggota cluster, dengan n_k sebagai jumlah anggota. Penelitian menggunakan empat atribut setelah standardisasi Z-Score sehingga pengelompokan mempertimbangkan kedekatan karakteristik akademik dan non-akademik.

2.5 Perancangan Sistem dan Pengujian

Unified Modeling Language digunakan sebagai media pemodelan kebutuhan, alur proses, struktur sistem, dan hubungan antarkomponen [15]. Pemodelan mencakup Use Case Diagram untuk interaksi pengguna, Activity Diagram untuk alur aktivitas, Class Diagram untuk struktur data dan komponen, serta Sequence Diagram untuk urutan interaksi dalam proses segmentasi. Sistem dikembangkan menggunakan Python dan Streamlit dengan dukungan Pandas, NumPy, Scikit-learn, Matplotlib, dan OpenPyXL.

Alur sistem dimulai dari pengunggahan data, validasi dan preprocessing, konversi nilai ekstrakurikuler, standardisasi Z-Score, penentuan jumlah cluster, penerapan K-Means, visualisasi, interpretasi, hingga penyajian strategi pembelajaran. Data diproses dalam memori dan hasil segmentasi dapat diunduh dalam format Excel. Fitur pilihan kelas hanya memfilter tampilan data dan tidak mengubah hasil clustering.

Metode pengujian yang digunakan dalam penelitian ini adalah Black Box Testing. Metode ini digunakan untuk menguji fungsi sistem berdasarkan masukan (input) dan keluaran (output) yang dihasilkan tanpa memperhatikan struktur atau kode program di dalam sistem. Pengujian dilakukan dengan memberikan berbagai skenario masukan pada setiap fitur aplikasi untuk mengetahui apakah sistem telah memberikan hasil sesuai dengan yang diharapkan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Validasi dan Implementasi Pengolahan Data

Halaman Segmentasi Siswa mengintegrasikan unggah file Excel, validasi, preprocessing, konversi kategori, dan standardisasi. Hasil validasi pada tampilan sistem menunjukkan 263 siswa sebagai data awal dan 259 siswa sebagai data yang siap diproses. Data dengan nilai kosong pada atribut clustering tidak digunakan dalam segmentasi. Tabel 3 merangkum jumlah data pada hasil implementasi sistem.

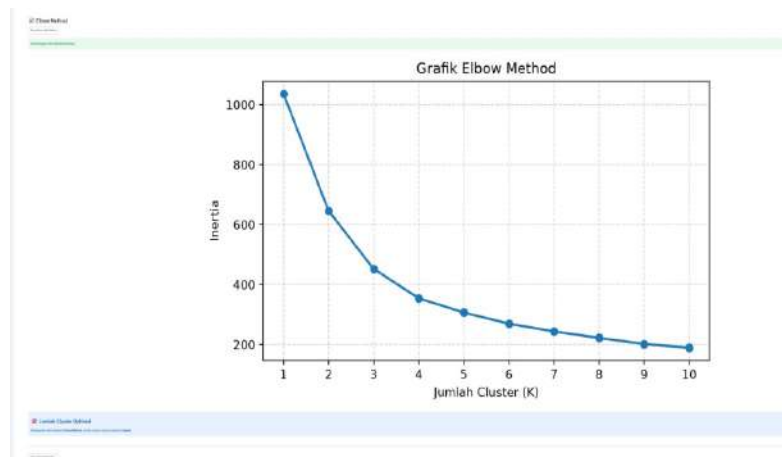
Tabel 3. Jumlah data pada hasil validasi sistem

Tahap	Jumlah siswa
Data awal	263
Data siap diproses	259

Sistem menampilkan preview data yang lolos validasi beserta hasil konversi ekstrakurikuler. Hasil standardisasi Z-Score tersedia untuk diperiksa sebelum proses K-Means. Setelah data siap, pengguna dapat menampilkan grafik Elbow Method dan menjalankan segmentasi. Implementasi ini menyatukan tahapan pengolahan yang diperlukan untuk memetakan siswa secara terstruktur.

3.2 Penentuan Tiga Cluster

Grafik Elbow Method pada Gambar 1 menunjukkan bahwa inertia mengalami penurunan yang cukup besar dari $k = 1$ hingga $k = 3$, kemudian penurunannya mulai melandai. Berdasarkan titik siku tersebut, jumlah cluster yang digunakan adalah $k = 3$. Nilai tersebut menjadi dasar proses K-Means berikutnya; sistem menampilkan informasi jumlah cluster setelah perhitungan Elbow berhasil dilakukan.



Gambar 1. Grafik Elbow Method pada hasil implementasi sistem

Penentuan jumlah cluster berkaitan dengan karakteristik data yang dianalisis. Elbow Method digunakan untuk mengamati perubahan WCSS pada alternatif jumlah kelompok [13]. Setelah tiga cluster ditetapkan, sistem menjalankan K-Means pada data yang telah melalui validasi, preprocessing, dan standardisasi. Hasil segmentasi disajikan bersama nilai centroid, label cluster pada setiap siswa, dan visualisasi kelompok.

3.3 Distribusi Anggota dan Karakteristik Cluster

Hasil segmentasi membentuk Cluster 1 sebanyak 130 siswa, Cluster 2 sebanyak 83 siswa, dan Cluster 3 sebanyak 46 siswa. Sistem menginterpretasikan Cluster 1 sebagai kelompok dengan kemampuan akademik rendah dan partisipasi ekstrakurikuler sedang; Cluster 2 memiliki kemampuan akademik tinggi dan partisipasi ekstrakurikuler sedang; Cluster 3 memiliki kemampuan akademik sedang dan partisipasi ekstrakurikuler tinggi. Distribusi dan profil masing-masing kelompok ditampilkan pada Tabel 4.

Tabel 4. Distribusi dan interpretasi hasil segmentasi

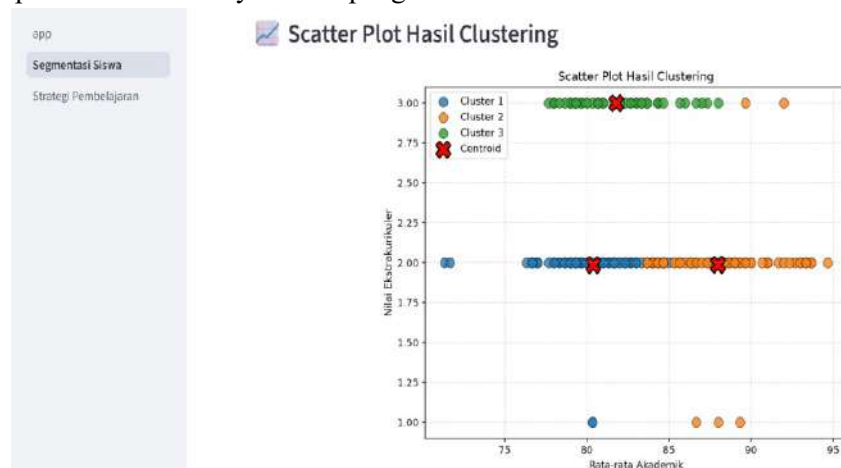
Cluster	Jumlah siswa	Kemampuan akademik	Partisipasi ekstrakurikuler
1	130	Rendah	Sedang
2	83	Tinggi	Sedang
3	46	Sedang	Tinggi

Nilai centroid akhir pada tampilan sistem menunjukkan posisi setiap cluster berdasarkan empat atribut setelah standardisasi. Tabel 5 menyajikan nilai centroid yang digunakan sebagai dasar interpretasi. Cluster 2 mempunyai centroid akademik lebih tinggi dibandingkan kelompok lain, sedangkan Cluster 3 mempunyai nilai centroid ekstrakurikuler yang lebih tinggi.

Tabel 5. Nilai centroid akhir dalam skala Z-Score

Cluster	Matematika	B. Indonesia	IPA	Ekstrakurikuler
1	-0,323	-0,622	-0,6279	-0,4311
2	0,6973	1,0663	1,1254	-0,4232
3	-0,3452	-0,1661	-0,2561	1,9819

Scatter plot pada Gambar 2 digunakan untuk membantu pengguna memahami penyebaran siswa berdasarkan hasil clustering. Setiap titik merepresentasikan seorang siswa, perbedaan warna menunjukkan cluster, dan simbol X menunjukkan posisi centroid. Sistem juga menampilkan distribusi jumlah siswa tiap cluster serta menyediakan pengunduhan hasil dalam format Excel.



Gambar 2. Scatter plot hasil clustering siswa

Penggabungan nilai akademik dan ekstrakurikuler memberikan gambaran profil siswa yang lebih menyeluruh. Hasil pengelompokan dapat digunakan sebagai dasar penentuan strategi pembelajaran yang sesuai dengan karakteristik setiap kelompok. Pemanfaatan K-Means pada penelitian ini berkaitan dengan penelitian pengelompokan siswa berdasarkan gaya belajar [1], profil nilai untuk pemilihan jurusan [4], data calon siswa [10], serta penentuan siswa kelas unggulan [11]. Pada penelitian ini, tindak lanjut segmentasi disajikan dalam sistem web melalui interpretasi kelompok dan rekomendasi strategi pembelajaran personal.

3.4 Rekomendasi Strategi Pembelajaran Personal

Halaman Strategi Pembelajaran menampilkan rekomendasi setelah proses segmentasi berhasil dilakukan. Sebelum segmentasi, sistem memberikan pesan agar pengguna menjalankan proses pengelompokan terlebih dahulu. Setiap rekomendasi memuat karakteristik cluster dan pendekatan pembelajaran yang telah ditentukan dalam penelitian. Pemetaan strategi pada Tabel 6 diambil dari tampilan rekomendasi sistem.

Tabel 6. Strategi pembelajaran sesuai karakteristik cluster

Cluster	Strategi pembelajaran
1	Remedial; tutor sebaya; latihan terstruktur
2	Pengayaan; Project Based Learning; Problem Based Learning
3	Project Based Learning; Cooperative Learning; pengembangan kepemimpinan

Pada Cluster 1, remedial dilakukan melalui pembelajaran ulang, latihan tambahan, atau bimbingan khusus bagi siswa yang belum mencapai kompetensi. Tutor sebaya menunjuk siswa yang memiliki pemahaman lebih baik untuk membantu temannya memahami materi. Latihan terstruktur disusun secara bertahap untuk memperkuat pemahaman konsep siswa.

Pada Cluster 2, pengayaan memberikan materi, tugas, atau aktivitas yang lebih menantang kepada siswa yang sudah menguasai kompetensi dasar. Project Based Learning memberikan proyek yang

berkaitan dengan materi pembelajaran sehingga siswa menghasilkan produk atau karya melalui perencanaan, pelaksanaan, dan presentasi. Problem Based Learning menyajikan masalah nyata untuk mendorong berpikir kritis, analisis, serta penemuan solusi secara mandiri maupun kelompok.

Pada Cluster 3, Project Based Learning digunakan bersama Cooperative Learning dan pengembangan kepemimpinan. Guru mengorganisasi siswa dalam kelompok kecil agar bekerja sama, berdiskusi, dan saling membantu mencapai tujuan pembelajaran. Pengembangan kepemimpinan dilakukan dengan memberi kesempatan menjadi ketua kelompok, koordinator proyek, atau pemimpin kegiatan pembelajaran untuk melatih kepemimpinan dan tanggung jawab.

Hubungan antara profil cluster dan strategi tersebut mendukung pembelajaran yang disesuaikan dengan kesiapan dan kebutuhan siswa [5]–[7]. Rekomendasi yang dihasilkan merupakan bahan pertimbangan bagi guru dalam menyusun pembelajaran. Pengembangan lebih lanjut dapat mempertimbangkan hasil evaluasi belajar siswa dan masukan guru agar rekomendasi menjadi lebih adaptif sesuai karakteristik dan kebutuhan kelompok.

3.5 Hasil Pengujian Fungsional

Pengujian Black Box Testing dilakukan pada fungsi aplikasi mulai dari membuka sistem hingga pengunduhan hasil dan penyajian strategi pembelajaran. Tabel 7 menyajikan hasil 15 skenario yang dilaporkan dalam penelitian. Semua skenario memberikan keluaran sesuai fungsi yang diharapkan.

Tabel 7. Hasil pengujian Black Box Testing

No.	Fitur yang diuji	Hasil pengujian	Status
1	Membuka aplikasi	Halaman utama berhasil ditampilkan	Berhasil
2	Upload file excel	File berhasil diunggah dan data ditampilkan	Berhasil
3	Validasi data	Sistem menampilkan jumlah data awal dan data siap diproses	Berhasil
4	Konversi nilai ekstrakurikuler	Nilai berhasil dikonversi menjadi angka	Berhasil
5	Normalisasi Z-Score	Data normalisasi berhasil ditampilkan	Berhasil
6	Menampilkan Grafik Elbow	Grafik dan informasi cluster optimal berhasil ditampilkan	Berhasil
7	Proses Segmentasi	Segmentasi berhasil dilakukan	Berhasil
8	Menampilkan nilai centroid	Nilai centroid berhasil ditampilkan	Berhasil
9	Menampilkan hasil segmentasi	Data segmentasi berhasil ditampilkan	Berhasil
10	Menampilkan Scatter Plot	Scatter plot berhasil ditampilkan	Berhasil
11	Menampilkan distribusi cluster	Grafik distribusi berhasil ditampilkan	Berhasil
12	Menampilkan interpretasi cluster	Interpretasi cluster berhasil ditampilkan	Berhasil
13	Download hasil segmentasi	File berhasil diunduh	Berhasil
14	Halaman Strategi Pembelajaran (belum segmentasi)	Pesan peringatan berhasil ditampilkan	Berhasil
15	Halaman Strategi Pembelajaran (sesudah segmentasi)	Strategi Pembelajaran berhasil ditampilkan	Berhasil

Berdasarkan hasil pengujian menggunakan metode Black Box Testing, seluruh fitur yang diuji pada aplikasi segmentasi siswa telah berjalan sesuai dengan fungsi yang diharapkan. Setiap skenario pengujian menghasilkan keluaran yang sesuai dengan rancangan sistem. Fitur unggah data, validasi dan preprocessing, konversi nilai ekstrakurikuler, standardisasi Z-Score, penentuan jumlah cluster menggunakan Elbow Method, proses segmentasi menggunakan algoritma K-Means, penyajian centroid, hasil segmentasi, visualisasi, interpretasi cluster, rekomendasi strategi pembelajaran, hingga pengunduhan hasil segmentasi dapat dijalankan dengan baik. Dengan demikian, berdasarkan hasil pengujian fungsional, sistem telah menjalankan seluruh fungsi yang diuji sesuai dengan kebutuhan penelitian.

Penelitian dibatasi pada penggunaan K-Means dan tidak membandingkan hasilnya dengan algoritma clustering lain. Variabel yang digunakan juga terbatas pada tiga nilai mata pelajaran dan nilai ekstrakurikuler. Pengembangan berikutnya dapat mencakup data kelas yang lebih luas, penambahan kehadiran, sikap, kedisiplinan, atau prestasi, serta perbandingan dengan Hierarchical Clustering, DBSCAN, atau Fuzzy C-Means. Sistem dapat diperluas melalui autentikasi, manajemen data siswa, penyimpanan basis data, dan laporan yang lebih lengkap.

4. KESIMPULAN

Penelitian menghasilkan aplikasi segmentasi siswa berbasis web menggunakan Python dan Streamlit yang mengolah karakteristik akademik dan non-akademik melalui validasi, preprocessing, konversi kategori ekstrakurikuler, standardisasi Z-Score, Elbow Method, dan K-Means Clustering. Hasil Elbow menunjukkan tiga cluster, dengan distribusi 130 siswa pada Cluster 1, 83 siswa pada Cluster 2, dan 46 siswa pada Cluster 3. Kelompok tersebut memiliki perbedaan karakteristik akademik dan partisipasi ekstrakurikuler yang menjadi dasar penyajian rekomendasi strategi pembelajaran personal. Sistem menyajikan tabel segmentasi, nilai centroid, visualisasi, interpretasi, dan hasil yang dapat diunduh. Pengujian Black Box Testing menunjukkan seluruh fungsi utama yang diuji berjalan sesuai harapan. Rekomendasi dapat digunakan sebagai bahan pertimbangan bagi guru untuk menentukan strategi pembelajaran sesuai karakteristik masing-masing kelompok.

REFERENSI

- [1] D. Maryono, C. W. Budiyanto, and A. A. Pamungkas, "Implementation of K-Means clustering for optimization of student grouping based on index of learning styles in programming classes," *IJIE (Indonesian Journal of Informatics Education)*, vol. 6, no. 2, pp. 84–89, 2022, doi: 10.20961/ijie.v6i2.68151.
- [2] R. Amelia, S. N. R. Izzah, M. A. Hikmah, and M. Y. A. Bakar, "Memahami gaya belajar siswa: Kunci keberhasilan personalisasi pembelajaran," *Jurnal Ilmiah Nusantara*, vol. 2, no. 1, pp. 287–300, 2025, doi: 10.61722/jinu.v2i1.3276.
- [3] 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, 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [4] M. R. Palevi and Z. Indra, "Implementasi algoritma K-Means clustering dengan pendekatan active learning pada siswa SMA untuk menentukan jurusan ke perguruan tinggi," *Jurnal SAINTIKOM*, vol. 23, no. 1, pp. 26–36, 2024, doi: 10.53513/jis.v23i1.9553.
- [5] Y. Al Fadillah, A. R. Akbar, and Gusmaneli, "Strategi desain pembelajaran adaptif untuk meningkatkan pengalaman belajar di era digital," *Jurnal Pendidikan Sains dan Teknologi Terapan*, vol. 1, no. 4, pp. 354–362, 2024.
- [6] Y. B. Widodo, S. Sibuea, and M. Narji, "Kecerdasan buatan dalam pendidikan: Meningkatkan pembelajaran personalisasi," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 2, pp. 602–615, 2024, doi: 10.37012/jtik.v10i2.2324.

- [7] S. M. Koimah, N. A. Zahra, E. Prasitini, S. K. Sasmita, and N. Sari, "Implementasi pembelajaran berdiferensiasi untuk memenuhi kebutuhan belajar siswa yang beragam," *Jurnal Ilmu Sosial dan Budaya Indonesia*, vol. 2, no. 2, pp. 58–66, 2024, doi: 10.61476/49j96838.
- [8] S. Pitafi, T. Anwar, and Z. Sharif, "A taxonomy of machine learning clustering algorithms, challenges, and future realms," *Applied Sciences*, vol. 13, no. 6, Art. no. 3529, 2023, doi: 10.3390/app13063529.
- [9] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. Ramírez Vargas, E. Bautista Thompson, and I. Ashraf, "A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective," *Symmetry*, vol. 15, no. 9, Art. no. 1679, 2023, doi: 10.3390/sym15091679.
- [10] Asmana, Y. A. Wijaya, and Martanto, "Clustering data calon siswa baru menggunakan metode K-Means di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 552–559, 2022, doi: 10.36040/jati.v6i2.5236.
- [11] A. Sulistiyawati and E. Supriyanto, "Implementasi algoritma K-Means clustering dalam penentuan siswa kelas unggulan," *Jurnal Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [12] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS ONE*, vol. 19, no. 12, Art. no. e0310839, 2024, doi: 10.1371/journal.pone.0310839.
- [13] A. J. Onumanyi, D. N. Molokomme, S. J. Isaac, and A. M. Abu-Mahfouz, "AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset," *Applied Sciences*, vol. 12, no. 15, Art. no. 7515, 2022, doi: 10.3390/app12157515.
- [14] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and H. Jia, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.
- [15] K. Nistrina and L. Sahidah, "Unified Modelling Language (UML) untuk perancangan sistem informasi penerimaan siswa baru di SMK Marga Insan Kamil," *J-SIKA: Jurnal Sistem Informasi Karya Anak Bangsa*, vol. 4, no. 1, pp. 17–23, 2022.