



Optimalisasi Strategi Pembelajaran Siswa Melalui Identifikasi Gaya Belajar Menggunakan Klasterisasi K-Means dan Klasifikasi K Nearest Neighbor

Ilsa Hidayat*, Musli Yanto, Rini Sovia

Fakultas Ilmu Komputer, Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
Jl. Lubuk Begalung, Padang, Indonesia

Email: ^{1,*}ilsahidayat2001@gmail.com, ²musli_yanto@upiptk.ac.id, ³rini_sovia@upiptk.ac.id

Email Penulis Korespondensi: ilsahidayat2001@gmail.com

Submitted: 31/01/2026; Accepted: 30/04/2026; Published: 30/04/2026

Abstrak—Ketepatan dalam menyesuaikan strategi pengajaran dengan karakteristik belajar siswa menjadi penting karena dapat menentukan efektivitas proses pembelajaran. Salah satu faktor kunci dalam peningkatan mutu pembelajaran adalah kesesuaian antara strategi pengajaran guru dengan gaya belajar siswa. Ketidaksesuaian antara kedua aspek tersebut dapat mengurangi efektivitas proses pembelajaran dan berdampak pada rendahnya capaian hasil belajar. Berdasarkan hal tersebut penelitian ini bertujuan untuk mengoptimalkan strategi pembelajaran siswa melalui penerapan model klasterisasi K-Means dan klasifikasi K-Nearest Neighbor. Kinerja Algoritma K-Means mampu melakukan pengelompokan gaya belajar serta pembedaan pelabelan gaya belajar, K-Nearest Neighbor digunakan untuk melakukan klasifikasi terhadap data yang telah diberi label oleh algoritma K-Means. Dataset penelitian ini berjumlah 200 data siswa bersumber dari SMP Negeri 1 Panyabungan dari Hasil 20 butir pertanyaan yang dijawab siswa. Hasil penelitian menunjukkan bahwa kombinasi algoritma K-Means dan K-Nearest Neighbor menghasilkan performa yang baik dengan nilai accuracy sebesar 0,92, precision sebesar 0,92, recall sebesar 0,92, dan F1-score sebesar 0,91. Kontribusi penelitian ini diharapkan dapat memperkaya literatur terkait penerapan model K-Means dan K-Nearest Neighbor dalam optimalisasi strategi pembelajaran, serta membantu guru di SMP Negeri 1 Panyabungan dalam merancang dan menerapkan strategi pembelajaran yang lebih efektif dan sesuai dengan kebutuhan siswa.

Kata Kunci: K-Means; K-Nearest Neighbor; Strategi Pembelajaran; Gaya Belajar; Personalisasi Pembelajaran

Abstract—Accuracy in adjusting teaching strategies to student learning characteristics is important because it can determine the effectiveness of the learning process. One of the key factors in improving the quality of learning is the suitability between teachers' teaching strategies and students' learning styles. The mismatch between the two aspects can reduce the effectiveness of the learning process and have an impact on low learning outcomes. Based on this, this study aims to optimize students' learning strategies through the application of the K-Means clustering model and the K-Nearest Neighbor classification. The performance of the K-Means Algorithm is able to classify learning styles and determine the labeling of learning styles, K-Nearest Neighbor is used to classify data that has been labeled by the K-Means algorithm. This research dataset amounted to 200 student data sourced from SMP Negeri 1 Panyabungan from the results of 20 questions answered by students. The results showed that the combination of the K-Means and K-Nearest Neighbor algorithms produced good performance with an accuracy value of 0.92, precision of 0.92, recall of 0.92, and F1-score of 0.91. The contribution of this research is expected to enrich the literature related to the application of the K-Means and K-Nearest Neighbor models in optimizing learning strategies, as well as assisting teachers at SMP Negeri 1 Panyabungan in designing and implementing learning strategies that are more effective and in accordance with the needs of students.

Keywords: K-Means; K-Nearest Neighbor; Learning Strategy; Learning Style; Personalized Learning

1. PENDAHULUAN

Peningkatan kualitas pendidikan merupakan agenda strategis Kementerian Pendidikan Kebudayaan Riset dan Teknologi pada jenjang pendidikan menengah [1]. Faktor kunci untuk peningkatan mutu pembelajaran adalah kesesuaian antara strategi pengajaran guru dengan gaya belajar siswa [2]. Ketepatan dalam menyesuaikan strategi pengajaran dengan karakteristik belajar siswa menjadi penting untuk efektivitas proses pembelajaran [3]. Identifikasi gaya belajar sejak dini diperlukan agar guru lebih mudah merancang pembelajaran yang lebih tepat sasaran serta memenuhi kebutuhan belajar setiap siswa [4].

Ketidaksesuaian antara kedua aspek tersebut dapat mengurangi efektivitas proses pembelajaran dan berdampak pada rendahnya capaian hasil belajar siswa [5]. Permasalahan terjadi pada setiap siswa memiliki cara belajar yang berbeda-beda, tetapi sebagian besar guru masih menggunakan metode konvensional seperti ceramah dan mencatat. Akibat dari permasalahan tersebut membuat siswa mengalami kesulitan untuk memahami pelajaran. Permasalahan tersebut berdampak terhadap pencapaian pembelajaran siswa, di mana sebagian siswa dapat memahami materi dengan cepat, sementara siswa lain membutuhkan pendekatan cara lain agar dapat menguasai pelajaran secara optimal. saat ini belum tersedia sistem berbasis data yang dapat digunakan untuk mengidentifikasi dan mengklasifikasikan gaya belajar siswa. Gaya belajar sangat penting untuk menyesuaikan metode pembelajaran secara tepat sasaran, sehingga strategi pembelajaran siswa dapat dioptimalkan [6].

Model VARK merupakan kerangka gaya belajar yang paling banyak digunakan. Konsep ini diperkenalkan oleh Neil Fleming pada tahun 2006 dan kemudian dievaluasi oleh University of Florida [7]. Berbagai penelitian menunjukkan bahwa setiap individu memiliki kecenderungan gaya belajar yang berbeda [8]. Pembelajar visual cenderung lebih mudah memahami informasi melalui penggunaan gambar, diagram, atau model yang



menggambarkan data tertulis [9]. Pembelajaran auditori lebih fokus pada penjelasan lisan seperti yang disampaikan langsung oleh pengajar maupun melalui rekaman suara untuk kemudian diolah menjadi catatan, cara belajar yang digunakan biasanya melalui pendekatan ceramah, podcast, atau rekaman audio [10]. Berbeda dengan pembelajar membaca/menulis lebih memilih materi berupa teks, seperti catatan kuliah yang terperinci atau buku ajar, karena mereka cenderung memahami informasi dengan menuliskannya dalam bentuk kata-kata dan teks [11]. Adapun pembelajar kinestetik lebih mengutamakan pengalaman nyata melalui praktik langsung dan aktivitas nyata sebagai teknik utama dalam memahami materi [12].

Keberagaman gaya belajar yang tidak diperhatikan sering kali menjadi faktor penyebab rendahnya pencapaian akademik siswa [13]. Hal ini disebabkan oleh kenyataan bahwa setiap siswa belajar sesuai dengan strategi dan mengandalkan keterampilan mereka [14]. Pendekatan yang dapat digunakan untuk mengatasi permasalahan ini adalah teknik data *mining*, yang berfungsi untuk mengekstraksi informasi dari dataset besar [15]. Data *mining* merupakan suatu proses untuk mengidentifikasi pola, hubungan, atau informasi yang tersembunyi pada sekumpulan data besar, data *mining* juga melibatkan pencarian korelasi atau pola antara ribuan atribut dalam basis data yang besar [16].

Penelitian yang memanfaatkan teknik data *mining* bertujuan untuk menganalisis dan mengoptimalkan strategi pembelajaran berdasarkan data yang diperoleh dari gaya belajar siswa. Salah satu teknik pada data *mining* adalah *clustering*. *Clustering* merupakan proses mengelompokkan data berdasarkan kesamaan fitur atau karakteristik tanpa memerlukan label kategori sebelumnya [17]. *clustering* digunakan untuk mengelompokkan siswa berdasarkan gaya belajar mereka. Algoritma yang sering digunakan dalam *clustering* adalah algoritma *K-Means clustering* [18]. *K-Means* adalah algoritma yang membagi data ke dalam sejumlah cluster berdasarkan kedekatannya dengan pusat cluster (*centroid*) [19]. Proses ini dimulai dengan memilih nilai *K* titik acak sebagai pusat cluster awal. Setiap data kemudian dikelompokkan ke cluster yang memiliki pusat terdekat. Setelah itu, pusat cluster dihitung ulang berdasarkan rata-rata data dalam cluster tersebut, dan proses ini diulang hingga posisi pusat cluster tidak berubah [20]. Dalam konteks pendidikan, algoritma *K-Means* dapat digunakan untuk mengelompokkan siswa ke dalam kelompok cluster sesuai gaya belajar setiap siswa.

Teknik data *mining* juga dapat dilihat pada kinerja metode klasifikasi. Klasifikasi merupakan teknik untuk mengkategorikan data ke dalam kelas tertentu berdasarkan atribut yang dimilikinya. Salah satu algoritma yang banyak digunakan dalam klasifikasi adalah *K-Nearest Neighbors* [21]. Algoritma KNN adalah metode klasifikasi sebuah data berdasarkan mayoritas kelas dari nilai *K* data terdekat yang ada dalam dataset [22]. Pada algoritma KNN, setiap data dalam ruang fitur dihitung jaraknya (biasanya menggunakan *Euclidean distance*) dari data yang akan diklasifikasikan, dan kemudian dipilih nilai *K* tetangga terdekat. Kelas dari data tersebut akan ditentukan berdasarkan kelas mayoritas dari tetangga-tetangga terdekatnya [23]. KNN dapat digunakan untuk memprediksi kelas atau kategori gaya belajar siswa dari hasil cluster yang di peroleh. Berbagai penelitian sebelumnya telah menunjukkan bahwa penggabungan metode *clustering* dan *classification* telah banyak digunakan.

Penelitian yang di lakukan oleh Daulay & Wandri, (2025) berjudul *Integrating K-Means Clustering and K-Nearest Neighbor Classification for Effective Scholarship Recipient Selection* membuktikan bahwa metode *K-Means* dan *K-NN* efektif dalam seleksi penerima Beasiswa KIP Kuliah. Dari 1.257 peserta, *K-Means* mengelompokkan data ke dalam tiga klaster, dan *K-NN* dengan nilai *K* optimal 155 memilih 155 peserta yang layak. Model ini mencapai akurasi 89,72% dengan waktu pemrosesan hanya 16 detik, menunjukkan bahwa kombinasi ini cepat, akurat, dan efisien dalam seleksi beasiswa [25].

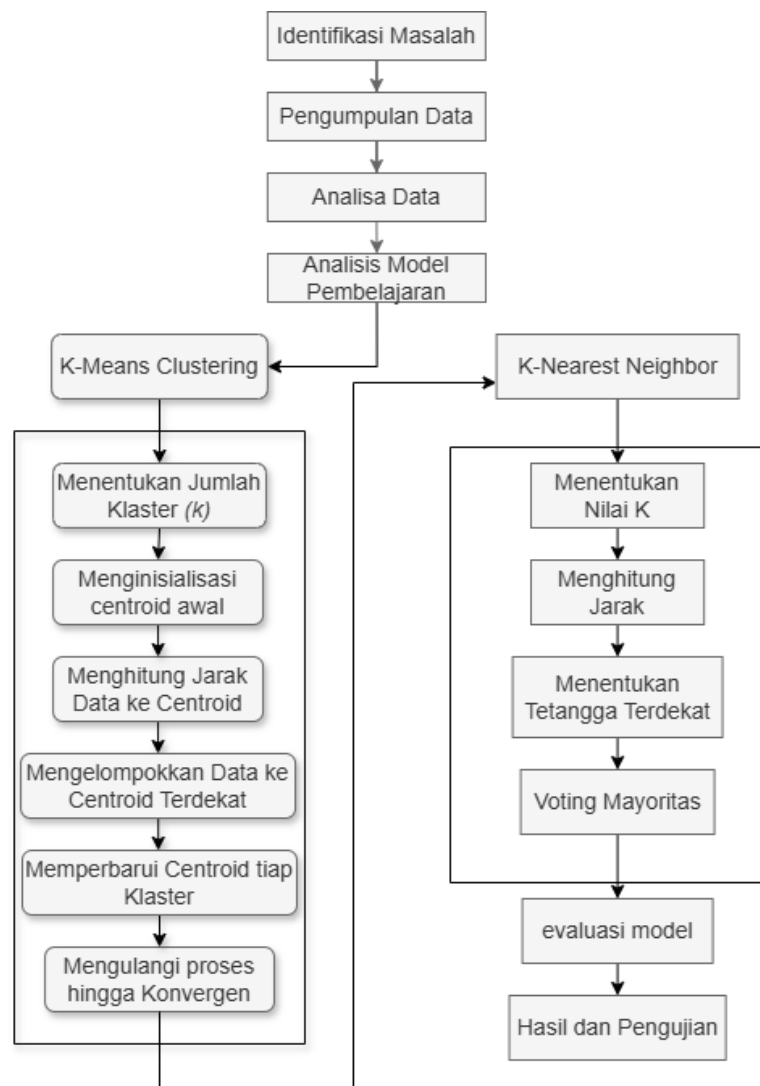
Sementara penelitian berjudul *Identifying learning styles in MOOCs environment through machine learning predictive modeling* menunjukkan hasil *Decision Tree* mencapai akurasi lebih dari 99% dalam memprediksi gaya belajar siswa di platform MOOC, melampaui algoritma lain seperti *Neural Networks*, *Random Forests*, *Naive Bayes*, dan *K-Nearest Neighbors*. Selain itu, *clustering* dengan *K-Means* berhasil membagi siswa ke dalam empat kelompok gaya belajar serupa pada dimensi visual, verbal, aktif, reflektif, sekuensial, dan global. Secara keseluruhan, temuan ini menegaskan bahwa teknik *machine learning* mampu mengidentifikasi gaya belajar siswa secara akurat, sehingga mendukung penerapan pembelajaran yang lebih personal dan efektif [26].

Dari berbagai penelitian sebelumnya, jelas terlihat bahwa algoritma *clustering* dan *classification* secara konsisten memberikan manfaat nyata di berbagai bidang. Hasil tersebut menunjukkan bahwa integrasi data *mining* dengan data yang bersumber dari aktivitas maupun aspek fisiologis memiliki potensi besar dalam membangun sistem yang adaptif, mulai dari optimasi kinerja sistem hingga melakukan penyesuaian strategi pembelajaran digital. *K-Means Clustering* juga dikenal efektif untuk menemukan pola cluster serta membentuk kelompok data berdasarkan tingkat kemiripan karakteristik. Pada konteks pendidikan, algoritma ini berperan penting untuk mengelompokkan siswa sesuai kecenderungan gaya belajarnya. Atas dasar permasalahan yang telah di jabarkan, penelitian ini difokuskan pada Optimalisasi Strategi Pembelajaran Siswa dengan Model Klasterisasi *K-Means* Dan Klasifikasi *K-NN* Di SMP Negeri 1 Panyabungan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan menerapkan algoritma *clustering* dan *K-Nearest Neighbors*. Metodologi penelitian disusun untuk memastikan penelitian ini dapat mencapai tujuan yang telah ditetapkan

melalui pendekatan yang sistematis dan terstruktur. Metodologi penelitian berperan sebagai kerangka kerja yang memfasilitasi pengolahan dan analisis data secara efektif. Berdasarkan kerangka tersebut, alur penelitian disusun dengan mencakup langkah-langkah tertentu yang akan dijelaskan lebih lanjut pada gambar di bawah ini.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan secara teknis bagaimana tahapan penelitian dilakukan, dengan alur kerja yang tersusun rapi dan sistematis, proses penelitian dimulai dari identifikasi masalah. Pada tahap ini, permasalahan yang akan diteliti dianalisis terlebih dahulu melalui pengamatan langsung di lapangan, serta diskusi dan wawancara di SMP Negeri 1 Panyabungan untuk mengumpulkan informasi yang relevan. Proses ini bertujuan untuk mengidentifikasi berbagai persoalan yang ada supaya dapat ditentukan lingkup penelitian secara lebih jelas.

Proses pengumpulan data pada penelitian ini dilakukan melalui penyusunan kuesioner yang terdiri dari 20 butir pernyataan yang telah di validasi oleh pakar di bidang terkait. Kuesioner tersebut disebarakan kepada 200 siswa kelas VII sebagai responden, hasil pengisiannya digunakan untuk mengidentifikasi gaya belajar masing-masing siswa. Instrumen ini berfungsi mengklasifikasikan siswa ke dalam empat kategori gaya belajar, yaitu visual, auditori, kinestetik, dan *read/write*.

Tahap analisis pada data berguna untuk memahami karakteristik data yang telah dikumpulkan, sehingga dapat memberikan gambaran awal mengenai pola maupun informasi penting yang terkandung di dalamnya. Proses ini mencakup pemeriksaan distribusi data, identifikasi nilai ekstrem, serta pencarian hubungan antarvariabel yang relevan dengan tujuan penelitian. Analisa data berfungsi sebagai dasar untuk menentukan metode pengolahan yang tepat, baik melalui teknik *clustering* maupun klasifikasi. Dengan demikian, tahap ini memastikan bahwa data siap digunakan pada proses selanjutnya secara lebih sistematis dan terarah.

2.1 Proses Algoritma K-Means Clustering

Analisis model pembelajaran dilakukan dengan menerapkan algoritma *K-Means Clustering* dan *K-Nearest Neighbor* [27]. Algoritma *K-Means* digunakan untuk melakukan pelabelan data dengan membentuk kelompok



gaya belajar siswa berdasarkan karakteristik tertentu. Selanjutnya algoritma *K-Nearest Neighbor* digunakan untuk melakukan klasifikasi terhadap data yang telah diberi label oleh algoritma *K-Means*. Berikut ini dijabarkan tahapan-tahapan algoritma *K-Means* [28]:

- Menentukan nilai K yang optimal untuk menetapkan jumlah kluster.
- Memilih K titik awal sebagai *centroid* secara acak (Tidak harus berasal dari data pada dataset).
- Menetapkan setiap data ke dalam suatu kluster dengan cara menghitung jarak antara data dan masing-masing *centroid*, kemudian memilih kluster dengan jarak minimum. Perhitungan jarak dilakukan menggunakan *Euclidean distance* Persamaan 1.

$$D_E = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}, \quad i = 1, 2, 3, \dots, n \quad (1)$$

Pada persamaan tersebut, A_i merupakan nilai fitur ke- i dari suatu objek data, B_i menyatakan nilai fitur ke- i dari *centroid*, dan n menunjukkan jumlah fitur atau dimensi data yang digunakan dalam perhitungan jarak *Euclidean*.

- Menghitung ulang variasi *centroid* yang dipilih sebelumnya dan menentukan *centroid* baru untuk setiap kluster berdasarkan rata-rata seluruh data yang termasuk ke dalam kluster tersebut pada persamaan 2.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Pada persamaan tersebut, v_{ij} merupakan nilai rata-rata (*centroid*) untuk elemen ke- i pada variabel ke- j , N_i menyatakan jumlah data dalam kluster ke- i , X_{kj} merupakan nilai data ke- k pada variabel ke- j , dan simbol \sum menyatakan proses penjumlahan seluruh nilai data dari indeks ke-1 hingga N_i .

- Mengulangi langkah 3 untuk menetapkan kembali kluster apabila diperlukan, berdasarkan *centroid* baru yang diperoleh pada langkah d.
- Jika masih terjadi perpindahan anggota kluster, maka kembali ke langkah d. Jika tidak ada perubahan, maka proses selesai dan model siap digunakan [28].

2.2 Proses Algoritma K-Nearest Neighbors

Klasifikasi *K-Nearest Neighbor* dilakukan setelah mendapatkan label gaya belajar dari algoritma *k means*. Data yang telah di label akan diklasifikasikan dengan tahap algoritma *K-Nearest Neighbor*. Berikut ini dijabarkan tahapan-tahapan algoritma *K-Nearest Neighbor* [29]:

- Menentukan Nilai K
Menentukan nilai k penting karena semakin besar nilai k , semakin banyak informasi dari tetangga yang dipertimbangkan. Namun, nilai k yang terlalu besar dapat menyebabkan model menjadi lebih umum dan kurang sensitif terhadap *noise* dalam data. Oleh karena itu, pemilihan k yang tepat sangat mempengaruhi hasil klasifikasi [29].
- Menghitung Jarak
Setelah nilai k ditentukan, langkah selanjutnya menghitung jarak antara data uji (data yang ingin diprediksi) dengan seluruh data pelatihan. Proses ini dilakukan menggunakan metrik jarak untuk mengukur seberapa dekat data uji dengan masing-masing data pelatihan. Salah satu metrik yang paling umum digunakan adalah *Euclidean distance*, yang dihitung dengan Persaman 3 [29].

$$D = \sqrt{\sum_{i=1}^n (x_1 - y_2)^2} \quad (3)$$

D merupakan jarak *Euclidean* antara data uji dan data pelatihan, n menyatakan jumlah atribut atau dimensi data, x_i adalah nilai atribut ke- i dari data uji, y_i merupakan nilai atribut ke- i dari data pelatihan, dan simbol \sum menyatakan proses penjumlahan selisih kuadrat seluruh atribut dari indeks ke-1 hingga n .

- Menentukan Tetangga Terdekat
Setelah menghitung jarak selanjutnya memilih nilai K tetangga terdekat berdasarkan jarak terkecil yang telah dihitung. Pada langkah ini kita memilih K data pelatihan yang memiliki jarak paling kecil dari data uji. Dengan memilih tetangga terdekat, kita memastikan bahwa klasifikasi dilakukan berdasarkan data yang paling relevan [29].
- Voting Mayoritas
Langkah terakhir dalam algoritma *K-NN* adalah melakukan voting mayoritas untuk menentukan kelas dari data testing. Setelah menemukan K tetangga terdekat, kita akan melihat kelas yang dimiliki oleh tetangga-tetangga tersebut dan memilih kelas yang paling banyak muncul. Kelas yang paling sering muncul di antara tetangga terdekat akan menjadi prediksi untuk data uji [29].

Evaluasi model dilakukan setelah tahap klasifikasi selesai yaitu dengan membandingkan data yang *actual* dan data hasil klasifikasi dari algoritma *KNN*. Tahap ini dilakukan dengan mengukur kinerja model melalui indikator tertentu, seperti akurasi, presisi, recall, atau metrik lainnya yang relevan dengan jenis data dan metode yang digunakan. Hasil evaluasi memberikan gambaran objektif mengenai kelebihan maupun keterbatasan model dalam mengolah serta memprediksi data. Dengan demikian, evaluasi model menjadi dasar untuk memastikan kualitas model dan menentukan apakah diperlukan perbaikan atau penyempurnaan lebih lanjut.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk membantu guru di SMP Negeri 1 Panyabungan dalam merancang dan menerapkan strategi pembelajaran yang lebih efektif serta sesuai dengan kebutuhan siswa melalui pemanfaatan data penelitian. Tahapan analisis diawali dengan proses input dataset awal yang terdiri atas 200 siswa, kemudian data tersebut digunakan sebagai dasar dalam proses pengolahan menggunakan algoritma *K-Means* dan *K-Nearest Neighbor*. Pada tahap *K-Means*, data gaya belajar siswa dikelompokkan berdasarkan kemiripan pola nilai melalui beberapa proses, yaitu menentukan jumlah kluster, menginisialisasi *centroid* awal, menghitung jarak setiap data ke *centroid*, mengelompokkan data ke *centroid* terdekat, memperbarui *centroid* pada setiap kluster, serta mengulangi proses tersebut hingga mencapai kondisi konvergen. Setelah hasil pengelompokan diperoleh, tahap berikutnya menggunakan algoritma *K-Nearest Neighbor* dengan menentukan nilai K, menghitung jarak antar data, menentukan tetangga terdekat, melakukan voting mayoritas, serta mengevaluasi model untuk memperoleh hasil pengujian. Melalui tahapan tersebut, penelitian ini diharapkan dapat menghasilkan pemetaan dan klasifikasi gaya belajar siswa yang lebih sistematis sehingga dapat menjadi acuan bagi guru dalam menentukan strategi pembelajaran yang sesuai. Berikut ini disajikan tabel dataset awal yang digunakan dalam penelitian ini.

Tabel 1. Dataset Awal

No	Nama siswa	V	A	R	K
1	Umar Bakri	8	8	3	3
2	Ahmad Yusuf	12	12	14	16
3	Zazkia Ameliana	17	9	7	15
4	Safa Ardiyah Fitri	14	14	17	14
5	Cita Anur	5	6	10	2
...
196	Muhammad Zulhadi	18	16	9	20
197	Sandra dewi	20	2	11	12
198	Muliana	0	1	2	4
199	Yeni	1	3	15	10
200	Anto	20	1	9	11

Berdasarkan Tabel 1, dataset awal penelitian ini memuat data siswa beserta skor kecenderungan gaya belajar pada empat atribut, yaitu V, A, R, dan K. Atribut V menunjukkan gaya belajar visual, A menunjukkan auditori, R menunjukkan membaca/menulis, dan K menunjukkan kinestetik. Setiap siswa memiliki nilai yang berbeda dari keempat atribut tersebut, sehingga dataset ini mampu menggambarkan variasi pola gaya belajar antar siswa. Dengan demikian, Tabel 1 memperlihatkan data mentah sebelum dilakukan proses pengolahan dan tahapan selanjutnya hingga klasifikasi gaya belajar.

3.1 Praprocessing Data

Tahap *praprocessing* data merupakan tahapan untuk menyiapkan data agar layak digunakan pada proses pengolahan dan analisis menggunakan algoritma data *mining*. tahap ini dilakukan untuk penyesuaian terhadap struktur dan atribut data agar sesuai dengan kebutuhan metode yang digunakan. Data awal yang digunakan dalam penelitian ini ditampilkan pada Tabel 2.

Tabel 2. Hasil Processing Data

No	X1	X2	X3	X4
1	8	8	3	3
2	12	12	14	16
3	17	9	7	15
4	14	14	17	14
5	5	6	10	2
...
196	18	16	9	20
197	20	2	11	12
198	0	1	2	4
199	1	3	15	10
200	20	1	9	11

Berdasarkan Tabel 2, hasil *preprocessing* menunjukkan bahwa atribut non-numerik, seperti nama siswa, dihapus dari proses perhitungan karena tidak memiliki pengaruh langsung terhadap analisis menggunakan algoritma *K-Means*. Atribut yang digunakan selanjutnya hanya berupa nilai gaya belajar, yang dinyatakan dalam bentuk variabel X1, X2, X3, dan X4 agar struktur data lebih teratur dan konsisten. Variabel X1, X2, X3, dan X4 masing-masing mewakili atribut V, A, R, dan K. Dengan demikian, dataset telah disederhanakan dan siap



digunakan pada tahap pengelompokan, karena seluruh atribut yang tersisa bersifat numerik dan sesuai untuk analisis kluster.

3.2 Analisis Algoritma K-Means Clustering

Analisis menggunakan algoritma *K-Means* dilakukan untuk membentuk label kluster gaya belajar siswa guna mempermudah proses pada tahap klasifikasi. Tahapan awal pada penerapan algoritma *K-Means* adalah penentuan jumlah kluster dan nilai *K*. untuk penelitian ini nilai *K* ditetapkan sebanyak empat kluster yang optimal untuk merepresentasikan pola gaya belajar siswa. Penentuan nilai *K* juga digunakan sebagai dasar pemilihan centroid awal yang dilakukan secara acak. Pada tahap ini, data ke-161 dipilih sebagai centroid pertama (*C1*), data ke-199 sebagai centroid kedua (*C2*), data ke-5 sebagai centroid ketiga (*C3*), dan data ke-198 sebagai centroid keempat (*C4*).

Setelah *centroid* awal ditentukan, tahap selanjutnya adalah menetapkan setiap data ke dalam kluster tertentu dengan cara menghitung jarak antara data dan masing-masing *centroid*. Data akan dimasukkan ke dalam kluster yang memiliki jarak paling minimum berdasarkan perhitungan jarak *Euclidean*, sebagaimana ditunjukkan pada Persamaan (1). Selanjutnya dilakukan perhitungan ulang *centroid* untuk setiap kluster dengan menentukan *centroid* baru berdasarkan nilai rata-rata seluruh data yang tergabung dalam kluster tersebut menggunakan Persamaan (2). Hasil *centroid* pada iterasi terbaru kemudian dibandingkan dengan *centroid* pada iterasi sebelumnya. Apabila masih terjadi perubahan keanggotaan kluster, maka proses iterasi dilanjutkan dengan perhitungan *centroid* yang baru. Namun, apabila nilai *centroid* telah sama dengan iterasi sebelumnya atau telah mencapai kondisi konvergen, maka proses iterasi dihentikan. Setelah label kluster terbentuk, tahapan selanjutnya adalah proses klasifikasi data. Berikut disajikan hasil pengelompokan kluster dalam bentuk tabel.

Tabel 3. Hasil Pengelompokan Algoritma K-Means

No	Nama Siswa	X1	X2	X3	X4	Cluster	Kategori Gaya Belajar
1	Umar Bakri	8	8	3	3	4	Kinestetik
2	Ahmad Yusuf	12	12	14	16	1	Visual
3	Zazkia Ameliana	17	9	7	15	2	Auditori
4	Safa Ardiyah Fitri	14	14	17	14	1	Visual
5	Cita Anur	5	6	10	2	4	Kinestetik
...
196	Muhammad Zulhadi	18	16	9	20	1	Visual
197	Sandra dewi	20	2	11	12	3	Membaca/Menulis
198	Muliana	0	1	2	4	4	Kinestetik
199	Yeni	1	3	15	10	4	Kinestetik
200	Anto	20	1	9	11	3	Membaca /Menulis

Tabel 3 menunjukkan bahwa proses klusterisasi menggunakan algoritma *K-Means* telah mencapai kondisi konvergen pada iterasi ke-15. Kondisi konvergen ini menunjukkan bahwa pusat kluster sudah stabil dan tidak mengalami perubahan yang berarti pada iterasi berikutnya. Berdasarkan hasil klusterisasi tersebut, terbentuk empat kluster gaya belajar siswa, yaitu kluster 1 sebagai gaya belajar visual, kluster 2 sebagai gaya belajar auditori, kluster 3 sebagai gaya belajar membaca/menulis, dan kluster 4 sebagai gaya belajar kinestetik. Distribusi anggota kluster menunjukkan bahwa gaya belajar visual dan membaca/menulis memiliki jumlah anggota yang sama, yaitu masing-masing sebanyak 68 siswa atau sebesar 34% dari total data. Sementara itu, gaya belajar auditori dan kinestetik masing-masing terdiri atas 32 siswa atau sebesar 16%. Hasil ini menunjukkan bahwa gaya belajar visual dan membaca/menulis menjadi kategori yang paling dominan pada data, sedangkan gaya belajar auditori dan kinestetik memiliki proporsi yang lebih kecil pada data. Label kluster yang dihasilkan dari proses *K-Means* selanjutnya digunakan sebagai label aktual dalam proses pembentukan dan pengujian model klasifikasi menggunakan algoritma *K-Nearest Neighbor*. Dengan demikian, label kluster tersebut berfungsi sebagai data pembandingan terhadap hasil prediksi yang dihasilkan oleh algoritma KNN. Pada tahap pengujian, data ke-140 sampai data ke-200 digunakan untuk membandingkan label aktual hasil dari klusterisasi dengan label prediksi dari algoritma KNN. Perbandingan ini bertujuan untuk mengetahui sejauh mana algoritma KNN mampu mengenali pola gaya belajar siswa berdasarkan hasil klusterisasi sebelumnya .

3.3 Analisis Algoritma K-Nearest Neighbor

Analisis klasifikasi dilakukan menggunakan algoritma *K-Nearest Neighbor* dengan memanfaatkan data yang telah memiliki label kluster dari tahap sebelumnya. Pada tahap awal, data berlabel kluster seperti yang ditampilkan pada Tabel 3 dipersiapkan sebagai dasar proses klasifikasi. Seluruh dataset yang terdiri dari 200 data dibagi menjadi dua bagian. Sebanyak 70% digunakan sebagai data *training* untuk melatih model dari data 1 hingga data 139, sedangkan 30% digunakan sebagai data *testing* untuk menguji kinerja model dari data 140 hingga data 200. Proses pembagian data *training* dan *testing* ini tidak dilakukan secara *random* pada proses penelitian. Tahap berikutnya adalah penentuan nilai *K* yang pada penelitian ini ditetapkan sebesar 4. Nilai *K* ini digunakan untuk menghitung

jarak *Euclidean* antara setiap data *testing* dan seluruh data *training*. Setelah jarak dihitung, dilakukan identifikasi tetangga terdekat untuk menentukan pola gaya belajar siswa. Pada tahap akhir, kelas data ditentukan melalui mekanisme voting mayoritas berdasarkan kelas dari tetangga terdekat. Hasil klasifikasi kemudian ditampilkan pada tabel berikut yang terdiri dari 60 data yang telah di klasifikasikan.

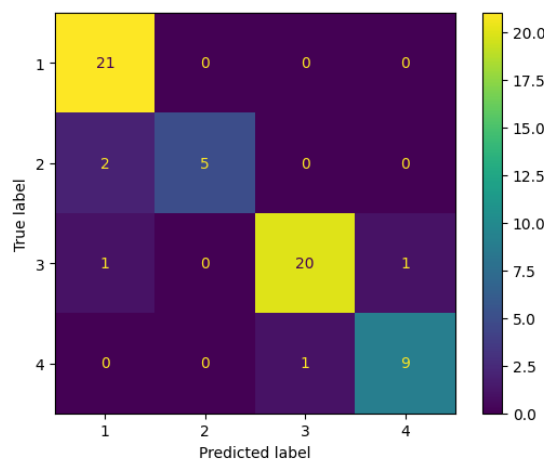
Tabel 4. Hasil Klasifikasi Algoritma K-Nearest Neighbor

No	Nama Siswa	V	A	R	K	Label Gaya belajar	Y
1	Randy Hasan	15	14	18	15	Visual	1
2	Afifah Syahira	16	16	17	5	Membaca/Menulis	3
3	Nia Ramadhani	10	10	8	6	Kinestetik	4
4	Khalifa Amini Lubis	13	13	16	8	Membaca/Menulis	3
5	Eka Sahrinan	12	10	13	10	Membaca/Menulis	3
...
56	Muhammad Zulhadi	18	16	9	20	Visual	1
57	Sandra dewi	20	2	11	12	Membaca/Menulis	3
58	Muliana	0	1	2	4	Kinestetik	4
59	Yeni	1	3	15	10	Membaca/Menulis	3
60	Anto	20	1	9	11	Membaca/Menulis	3

Berdasarkan Tabel 4, hasil klasifikasi algoritma *K-Nearest Neighbor* menunjukkan kategori gaya belajar siswa berdasarkan nilai V, A, R, dan K. Kolom label gaya belajar menampilkan hasil klasifikasi dalam bentuk kategori, sedangkan kolom Y merupakan hasil prediksi model yang dinyatakan dalam kode numerik. Nilai prediksi tersebut selanjutnya dibandingkan dengan label aktual dari data testing. Dalam penelitian ini, data *testing* yang digunakan adalah data hasil clustering pada Tabel 3, tepatnya untuk data nomor 140 sampai 200. Perbandingan antara nilai aktual dan nilai prediksi inilah yang menjadi dasar penyusunan *confusion matrix* untuk mengevaluasi kinerja model KNN.

3.4 Evaluasi Model

Setelah tahap klasifikasi selesai, kinerja model diuji menggunakan *confusion matrix* dengan membandingkan data aktual dan hasil prediksi dari algoritma *K-Nearest Neighbor*. Data siswa pada baris ke-140 hingga ke-200 dalam dataset digunakan sebagai data testing, di mana label cluster pada Tabel 3 berperan sebagai nilai aktual. Nilai aktual ini menjadi acuan untuk mengevaluasi hasil prediksi KNN yang direpresentasikan oleh kode numerik pada kolom Y di Tabel 4. Dengan melakukan komparasi antara label aktual dan prediksi, tingkat akurasi dan kesesuaian klasifikasi yang dihasilkan oleh model dapat diukur secara presisi, berdasarkan akumulasi prediksi yang benar (*true*) dan salah (*false*) untuk masing-masing kategori gaya belajar. Selanjutnya, hasil perbandingan ini digunakan untuk menghitung nilai *accuracy*, *precision*, *recall*, dan *F1-score* indikator performa model sebelum dilakukan interpretasi hasil lebih lanjut. Visualisasi dari *confusion matrix* yang menunjukkan performa klasifikasi model disajikan pada Gambar di bawah ini, sehingga pembaca dapat melihat secara langsung bagaimana KNN mengenali pola gaya belajar siswa berdasarkan label aktual dari *K-Means* dan hasil prediksi model.



Gambar 2. Hasil confusion Matrix

Berdasarkan hasil *confusion matrix* pada gambar 2 model klasifikasi menggunakan algoritma *K-Nearest Neighbor* mampu mengklasifikasikan data gaya belajar siswa dengan baik ke dalam empat kelas. Sebagian besar data pada setiap kelas berhasil diprediksi dengan benar, dengan tingkat kesalahan klasifikasi yang relatif kecil. Nilai akurasi yang diperoleh sebesar 0,92 menunjukkan bahwa model memiliki tingkat ketepatan yang tinggi. Selain itu, nilai *precision* sebesar 0,92, *recall* sebesar 0,92, dan *F1-score* sebesar 0,91 mengindikasikan



keseimbangan yang baik antara kemampuan model dalam melakukan prediksi yang tepat dan kemampuan dalam mengenali data pada setiap kelas. Dengan demikian, model KNN dinilai efektif dan layak digunakan untuk klasifikasi gaya belajar siswa pada penelitian ini.

4. KESIMPULAN

Berdasarkan hasil penelitian penerapan algoritma *K-Means* dan *K-Nearest Neighbor* terbukti efektif dalam mengoptimalkan strategi pembelajaran siswa melalui pengelompokan dan klasifikasi gaya belajar. Algoritma *K-Means* berhasil membagi siswa ke dalam empat klaster, yaitu visual, auditori, membaca/menulis, dan kinestetik. Distribusi klaster menunjukkan bahwa gaya belajar visual dan membaca/menulis masing-masing terdiri atas 68 siswa atau 34% dari total data. Sementara itu, gaya belajar auditori dan kinestetik masing-masing beranggotakan 32 siswa atau 16%. Temuan ini menggambarkan bahwa gaya belajar visual dan membaca/menulis memiliki proporsi paling dominan dalam data penelitian. Selanjutnya, algoritma KNN yang dilakukan tanpa *random* menunjukkan kinerja klasifikasi yang baik dengan akurasi 0,92, *precision* 0,92, *recall* 0,92, dan *F1-score* 0,91. Hasil tersebut mengindikasikan bahwa kombinasi *K-Means* dan KNN layak digunakan sebagai pendekatan analitik untuk merancang strategi pembelajaran yang adaptif dan sesuai dengan karakteristik siswa. Dengan demikian, hasil penelitian ini memberikan dasar empiris bagi pengembangan pembelajaran berbasis data yang responsif terhadap kebutuhan dan preferensi belajar siswa di lingkungan pendidikan. Pada penelitian selanjutnya, pengolahan data dengan *K-Means* dapat ditingkatkan melalui optimalisasi penentuan centroid awal agar proses konvergensi lebih efisien. Inisialisasi centroid berbasis purity dapat dipertimbangkan untuk membentuk klaster awal yang lebih representatif. Selain itu, penggunaan *Felder–Silverman Learning Style Model* (FSLSM) secara komprehensif serta perbandingan beberapa algoritma supervised learning dapat dilakukan untuk meningkatkan akurasi klasifikasi dan prediksi.

REFERENCES

- [1] M. Astuti, F. Ismail, S. Fatimah, W. Puspita, and Herlina, "The Relevance Of The Merdeka Curriculum In Improving The Quality Of Islamic Education In Indonesia," *Int. J. Learn. Teach. Educ. Res.*, vol. 23, no. 6, pp. 56–72, 2024, doi: 10.26803/ijlter.23.6.3
- [2] Y. J. A. Khasawneh, R. Alsarayreh, A. A. Al Ajlouni, H. M. Eyadat, M. N. Ayasrah, and M. A. S. Khasawneh, "An examination of teacher collaboration in professional learning communities and collaborative teaching practices," *J. Educ. e-Learning Res.*, vol. 10, no. 3, pp. 446–459, 2023, doi: 10.20448/jeelr.v10i3.4841
- [3] E. G. Rincon-Flores *et al.*, "Improving the learning-teaching process through adaptive learning strategy," *Smart Learn. Environ.*, vol. 11, no. 1, 2024, doi: 10.1186/s40561-024-00314-9
- [4] S. Chan, S. Maneewan, and R. Koul, "Teacher educators' teaching styles: relation with learning motivation and academic engagement in pre-service teachers," *Teach. High. Educ.*, vol. 28, no. 8, pp. 2044–2065, 2023, doi: 10.1080/13562517.2021.1947226
- [5] I. Miguel, A. David, C. Henar, G. Sanz, A. Bustillo, and I. M. Alonso, "Evaluation of the novelty effect in immersive Virtual Reality learning experiences," *Virtual Real.*, vol. 28, no. 1, pp. 1–23, 2024, doi: 10.1007/s10055-023-00926-5
- [6] M. C. Malacapay, "The Influence of Learning Styles and Attitudes on Academic Performance of College Students in a Flipped Learning Environment," *Int. J. Instr.*, vol. 17, no. 4, pp. 623–644, 2024, doi: 10.29333/iji.2024.17435a
- [7] D. Chinnapun and U. Narkkul, "Enhancing Learning in Medical Biochemistry by Teaching Based on VARK Learning Style for Medical Students," *Adv. Med. Educ. Pract.*, vol. 15, pp. 895–902, 2024, doi: 10.2147/AMEP.S472532
- [8] A. Gayef, A. Çaylan, and S. A. Temiz, "Learning styles of medical students and related factors," *BMC Med. Educ.*, vol. 23, no. 1, pp. 1–11, 2023, doi: 10.1186/s12909-023-04267-4
- [9] S. G. Essa, T. Celik, Human-Hendricks, and N. Emelia, "Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review," *IEEE Access*, vol. 11, no. April, pp. 48392–48409, 2023, doi: 10.1109/ACCESS.2023.3276439
- [10] A. R. Shaidullina *et al.*, "Learning styles in science education at university level: A systematic review," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 19, no. 7, pp. 1–10, 2023, doi: 10.29333/ejmste/13304
- [11] R. R. Iyer and R. Sethuraman, "Role of ehealth literacy, learning styles, and patterns of web-based e-content access for seeking health information among dental university students in Vadodara, India," *ournal Educ. Heal. Promot.*, vol. 1, no. 56, February 2024, pp. 1–6, 2024, doi: 10.4103/jehp.jehp_750_23
- [12] E. El-Saftawy, A. A. A. Latif, A. M. ShamsEldeen, M. A. Alghamdi, A. M. Mahfoz, and B. E. Aboulhoda, "Influence of applying VARK learning styles on enhancing teaching skills: application of learning theories," *BMC Med. Educ.*, vol. 24, no. 1, 2024, doi: 10.1186/s12909-024-05979-x
- [13] H. Taş and M. B. Minaz, "The Effects of Learning Style-Based Differentiated Instructional Activities on Academic Achievement and Learning Retention in the Social Studies Course," *SAGE Open*, vol. 14, no. 2, pp. 1–14, 2024, doi: 10.1177/21582440241249290
- [14] Y. Yousaf, M. Shoaib, M. A. Hassan, and U. Habiba, "An intelligent content provider based on students learning style to increase their engagement level and performance," *Interact. Learn. Environ.*, vol. 31, no. 5, pp. 2737–2750, 2023, doi: 10.1080/10494820.2021.1900875
- [15] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, "A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry (Basel)*, vol. 15, no. 9, pp. 1–44, 2023, doi: 10.3390/sym15091679
- [16] A. Sánchez, C. Vidal-Silva, G. Mancilla, M. Tupac-Yupanqui, and J. M. Rubio, "Sustainable e-Learning by Data



- Mining—Successful Results in a Chilean University,” *Sustain.*, vol. 15, no. 2, pp. 1–16, 2023, doi: 10.3390/su15020895
- [17] S. M. Miraftebadeh, C. G. Colombo, M. Longo, and F. Foiadelli, “K-Means and Alternative Clustering Methods in Modern Power Systems,” *IEEE Access*, vol. 11, no. September, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640
- [18] I. Hidayat, E. Darnila, and Y. Afrillia, “Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma K-Means,” *G-Tech J. Teknol. Terap.*, vol. 7, no. 3, pp. 1218–1226, 2023, doi: 10.33379/gtech.v7i3.2880
- [19] M. Yang, L. Huang, and C. Tang, “K-Means Clustering with Local Distance Privacy,” *Big Data Min. Anal.*, vol. 6, no. 4, pp. 433–442, 2023, doi: 10.26599/BDMA.2022.9020050
- [20] E. Setyaningsih, N. Hidayat, U. Lestari, and A. Septiarini, “Modification of K-Means and K-Mode Algorithms To Enhance the Performance of Clustering Student Learning Styles in the Learning Management System,” *ICIC Express Lett.*, vol. 17, no. 1, pp. 49–59, 2023, doi: 10.24507/icicel.17.01.49
- [21] J. Li, J. Zhang, J. Zhang, and S. Zhang, “Quantum KNN Classification With K Value Selection and Neighbor Selection,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 43, no. 5, pp. 1332–1345, 2024, doi: 10.1109/TCAD.2023.3345251
- [22] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, “Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications,” *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00973-y
- [23] S. V. Razavi-Termeh, A. Sadeghi-Niaraki, S. Razavi, and S. M. Choi, “Enhancing flood-prone area mapping: fine-tuning the K-nearest neighbors (KNN) algorithm for spatial modelling,” *Int. J. Digit. Earth*, vol. 17, no. 1, pp. 1–29, 2024, doi: 10.1080/17538947.2024.2311325
- [24] S. Daulay and R. Wandri, “Integrating K-Means Clustering and K-Nearest Neighbor Classification for Effective Scholarship Recipient Selection,” 2025. doi: 10.32520/stmsi.v14i1.4818
- [25] S. Daulay and R. Wandri, “Integrating K-Means Clustering and K-Nearest Neighbor Classification for Effective Scholarship Recipient Selection,” *Sistemasi*, vol. 14, no. 1, p. 235, 2025, doi: 10.32520/stmsi.v14i1.4818
- [26] M. Jebbari, B. Cherradi, S. Hamida, and A. Raihani, *Identifying learning styles in MOOCs environment through machine learning predictive modeling*, vol. 29, no. 16. Springer US, 2024. doi: 10.1007/s10639-024-12637-8
- [27] G. Vardakas, I. Papakostas, and A. Likas, “Efficient error minimization in kernel k-means clustering,” *Pattern Anal. Appl.*, vol. 28, no. 2, 2025, doi: 10.1007/s10044-025-01463-4
- [28] G. R. Kanagachidambaresan and N. Bharathi, *Learning Algorithms for Internet of Things*, 1st ed. Apress Berkeley, CA, 2024. doi: //doi.org/10.1007/979-8-8688-0530-1
- [29] H. Zhou, *Learn Data Mining Through Excel*, 2nd ed. Apress Berkeley, CA, 2023. doi: 10.1007/978-1-4842-9771-1_10