

Analysis of Student Interest Area Recommendations Using K-Means and the Perceptron Algorithm at MTsN 3 Solok

Hafis Hidayattullah ¹⁾; Rini Sovia ²⁾; Larissa Navia Rani ³⁾

^{1,2,3)} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

Email: ¹⁾ hafis2608@gmail.com

How to Cite :

Hidayattullah, H., Sovia, R., Rani, L.N. (2026). Analysis of Student Interest Area Recommendations Using K-Means and the Perceptron Algorithm at MTsN 3 Solok Jurnal Media Computer Science, 5(4)

ARTICLE HISTORY

Received [03 Agustus 2026]

Revised [25 Oktober 2026]

Accepted [27 Oktober 2026]

KEYWORDS

Wireshark, Windump, Network Security Monitoring.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Penentuan bidang akademik sebelum melanjutkan pendidikan ke Sekolah Menengah Kejuruan (SMK) perlu didukung informasi yang menggambarkan kemampuan siswa. Penelitian ini bertujuan memberikan rekomendasi awal bidang akademik dengan menerapkan K-Means Clustering dan Single Layer Perceptron (SLP). K-Means digunakan untuk membentuk tiga tingkat kemampuan akademik, yaitu Tinggi, Sedang, dan Rendah, berdasarkan sembilan nilai mata pelajaran. Selanjutnya, SLP mengenali kecenderungan siswa pada bidang Sains, Sosial, atau Kreatif berdasarkan nilai rapor dan selisih dua rata-rata bidang tertinggi. Dataset mencakup 1.000 data nilai rapor tahun ajaran 2021/2022 sampai 2024/2025. K-Means menghasilkan Silhouette Score 0,6734 dan konvergen pada iterasi ke-7. Pada 150 data testing, SLP memperoleh accuracy 97,33%, precision 97,36%, recall 97,33%, dan F1-score 97,34%, dengan 146 prediksi benar. Hasil kedua metode digabungkan menjadi rekomendasi awal yang memuat tingkat kemampuan dan kecenderungan bidang sebagai bahan pertimbangan bagi siswa dan sekolah, bukan keputusan mutlak dalam menentukan jurusan. Pemilihan akhir tetap mempertimbangkan minat, bakat, kondisi siswa, dan arahan guru.

ABSTRACT

Choosing an academic field before entering Vocational High School (SMK) should be supported by information reflecting students' abilities. This study aims to provide an initial field recommendation using K-Means Clustering and Single Layer Perceptron (SLP). K-Means forms three levels of ability, namely High, Medium, and Low, based on nine subject grades. SLP identifies tendencies toward Science, Social, or Creative fields using report card grades and the difference between the two highest field averages. The dataset contains 1,000 records from the 2021/2022 to 2024/2025 academic years. K-Means produced a Silhouette Score of 0.6734 and converged at the seventh iteration. On 150 testing records, SLP achieved 97.33% accuracy, 97.36% precision, 97.33% recall, and a 97.34% F1-score, with 146 correct predictions. Both outputs were combined into an initial recommendation containing ability level and field tendency for students and schools, rather than a final decision. Final selection should consider interests, talents, conditions, and teacher guidance.

PENDAHULUAN

Pemilihan bidang akademik ke SMK merupakan keputusan penting karena pilihan tersebut dapat memengaruhi proses belajar dan perkembangan siswa, sementara keputusan yang kurang tepat dapat menurunkan motivasi, memengaruhi prestasi, dan membuat siswa merasa tidak sesuai

dengan bidang yang dipilih sehingga penentuan jurusan perlu dilakukan dengan pertimbangan yang cukup (Tampubolon, 2024). Dalam praktiknya, sebagian siswa masih menentukan jurusan karena mengikuti teman, keinginan orang tua, atau belum memahami kemampuan akademiknya sendiri, padahal kondisi tersebut dapat membuat pilihan pendidikan tidak sesuai dengan kekuatan siswa dan menimbulkan kesulitan ketika menjalani pembelajaran yang lebih khusus pada jenjang berikutnya (Pribadi et al., 2024). Kondisi tersebut menunjukkan perlunya rekomendasi awal yang dapat membantu siswa, guru, dan sekolah membaca kemampuan akademik secara lebih terukur, terutama karena arahan masih banyak dilakukan melalui pembacaan nilai, wawancara singkat, dan pertimbangan guru secara manual, sehingga data nilai perlu diolah menjadi informasi yang lebih jelas sebelum digunakan sebagai bahan pertimbangan dalam memilih bidang akademik menuju SMK (Aravik et al., 2025).

Nilai rapor yang terdiri atas beberapa mata pelajaran tidak selalu mudah dibaca apabila hanya dilihat satu per satu, sehingga machine learning dapat membantu mengolah nilai tersebut untuk menemukan pola yang teratur dan menyajikannya menjadi informasi yang lebih mudah dipahami oleh siswa dan sekolah dalam melakukan bimbingan akademik (Banarsari et al., 2023). Nilai rapor yang terdiri atas beberapa mata pelajaran tidak selalu mudah dibaca apabila hanya dilihat satu per satu, sehingga machine learning dapat membantu mengolah nilai tersebut untuk menemukan pola yang teratur dan menyajikannya menjadi informasi yang lebih mudah dipahami oleh siswa dan sekolah dalam melakukan bimbingan akademik (Maulani et al., 2025). Meskipun nilai dapat membantu menggambarkan kecenderungan akademik, hasil pengolahannya tidak dapat digunakan sebagai satu-satunya dasar dalam menentukan jurusan karena pilihan pendidikan juga perlu mempertimbangkan minat, bakat, motivasi, kondisi siswa, arahan guru, dan faktor lain di luar nilai rapor, sehingga hasil analisis penelitian ini ditempatkan sebagai rekomendasi awal dan bukan keputusan akhir bagi siswa (Alfarizi et al., 2023).

Pengolahan data dalam penelitian ini dilakukan melalui dua tahap utama, yaitu pengelompokan dan klasifikasi, dengan K-Means Clustering digunakan lebih dahulu untuk membagi siswa berdasarkan kemiripan pola nilai akademik karena metode ini bekerja pada data numerik dengan menghitung kedekatan setiap data terhadap pusat kelompok atau centroid sampai terbentuk kelompok yang memiliki karakteristik relatif serupa (Alawi et al., 2023). Hasil cluster kemudian dianalisis berdasarkan rata-rata nilai centroid untuk menentukan tingkat kemampuan akademik siswa menjadi Tinggi, Sedang, dan Rendah, sehingga makna kelompok tidak ditentukan dari nomor cluster, tetapi dari karakteristik nilai yang mewakili setiap kelompok agar hasil pengelompokan lebih mudah dipahami dalam konteks pendidikan (Dewi & Pakereng, 2023; Yudhistira & Andika, 2023). Setelah tingkat kemampuan diperoleh, SLP digunakan untuk mengklasifikasikan kecenderungan bidang akademik siswa ke dalam Sains, Sosial, atau Kreatif melalui proses pembelajaran bobot dan bias berdasarkan data masukan, sehingga K-Means memberikan informasi mengenai tingkat kemampuan sedangkan SLP memberikan arah kecenderungan bidang akademik yang kemudian digabungkan dalam satu hasil rekomendasi (Handayani et al., 2025).

Beberapa penelitian terdahulu menunjukkan bahwa K-Means telah banyak digunakan untuk mengolah data pendidikan, salah satunya penelitian Yudhistira dan Andika yang mengelompokkan data nilai siswa menjadi tiga cluster dan memperoleh Silhouette Score sebesar 0,489, tetapi hasilnya masih berupa kelompok berdasarkan karakteristik data dan belum diarahkan untuk menentukan kecenderungan bidang akademik siswa menuju SMK (Aulia Sitorus, 2025; Permana et al., 2026). Penelitian Prayudha dan rekan juga menggunakan K-Means pada data akademik siswa Sekolah Menengah Kejuruan dan menghasilkan dua cluster dengan nilai Davies-Bouldin Index sebesar 0,519, yang menunjukkan metode tersebut dapat digunakan untuk memetakan kelompok akademik, namun keluarannya masih berfokus pada tingkat kelompok nilai dan belum digabungkan dengan klasifikasi bidang yang lebih spesifik (Hayati et al., 2025; Wedo et al., 2025). Sementara itu, Mufakhrasy menggunakan Perceptron untuk klasifikasi program studi mahasiswa baru berdasarkan nilai akademik dan jurusan asal dengan kesesuaian keluaran 54,28 persen pada 35 data pengujian,

sehingga penggunaan K-Means bersama SLP untuk menghasilkan tingkat kemampuan sekaligus kecenderungan bidang akademik siswa masih cukup terbatas mutlak (Zahara et al., 2023).

Berdasarkan kajian sebelumnya, penelitian ini menggabungkan K-Means Clustering dan SLP dalam rekomendasi awal bidang akademik siswa karena hasil clustering dapat menunjukkan tingkat kemampuan setelah karakteristik kelompok dianalisis, sedangkan klasifikasi dapat menentukan arah kecenderungan bidang berdasarkan pola nilai sehingga kedua keluaran saling melengkapi. Penelitian menggunakan 1.000 data nilai rapor siswa pada sembilan mata pelajaran selama tahun ajaran 2021/2022 sampai 2024/2025, dengan K-Means membentuk kategori Tinggi, Sedang, dan Rendah, sedangkan SLP mengklasifikasikan kecenderungan Sains, Sosial, atau Kreatif, kemudian kedua hasil digabungkan menjadi rekomendasi seperti Tinggi-Sains, Sedang-Sosial, atau Rendah-Kreatif untuk memberikan gambaran kemampuan dan arah bidang dalam satu hasil. Penelitian ini bertujuan menghasilkan rekomendasi awal bidang akademik berbasis nilai rapor serta mengevaluasi kualitas pengelompokan dan kemampuan klasifikasi, sedangkan kontribusi penelitian ini adalah menggabungkan informasi tingkat kemampuan dan kecenderungan bidang akademik dalam satu hasil rekomendasi yang dapat digunakan sebagai bahan pertimbangan awal bagi siswa dan sekolah.

LANDASAN TEORI

Machine learning merupakan pendekatan komputasi yang mempelajari pola dari data untuk menghasilkan pengelompokan, klasifikasi, prediksi, atau rekomendasi (Maulani et al., 2025). Pendekatan supervised learning menggunakan target kelas, sedangkan unsupervised learning mencari struktur data tanpa label awal sehingga keduanya dapat digunakan secara berurutan dalam penelitian ini (Bintoro et al., 2024). Dalam penelitian pendidikan, pendekatan ini dapat digunakan untuk mengolah data nilai menjadi informasi yang lebih terstruktur selama hasil model tetap dipahami sesuai ruang lingkup data yang digunakan (Handayani et al., 2025).

Clustering membentuk kelompok berdasarkan kemiripan data dan dapat menyederhanakan sekumpulan nilai menjadi pola kemampuan yang lebih mudah ditafsirkan (Alawi et al., 2023). K-Means menggunakan centroid sebagai pusat cluster dan menempatkan setiap data pada pusat terdekat sampai proses pengelompokan stabil (Hapsari & Indriyani, 2024). Kualitas hasil juga dipengaruhi pemilihan centroid awal dan perhitungan jarak antar data sehingga tahap inisialisasi perlu diperhatikan (Khan et al., 2024).

Single Layer Perceptron menghubungkan input langsung ke neuron output melalui bobot dan bias sehingga struktur pembelajarannya relatif mudah ditelusuri (Aulia Sitorus, 2025). Perceptron memperbarui parameter ketika prediksi tidak sesuai dengan target dan dapat digunakan sebagai model dasar klasifikasi pada data akademik yang terstruktur (Permana et al., 2026). Kemampuan model tetap dipengaruhi keterpisahan pola antarkelas karena Perceptron membentuk batas keputusan yang bersifat linear (Taghinezhad & Ghasemzadeh, 2024).

Pada klasifikasi multi kelas, nilai keluaran dibandingkan untuk menentukan kelas akhir, kemudian prediksi dievaluasi terhadap target pada data uji (Gori et al., 2024). Dalam penelitian ini, K-Means digunakan untuk tingkat kemampuan dan SLP untuk kecenderungan bidang, lalu kedua hasil digabungkan sebagai rekomendasi awal. Kerangka berpikir tersebut menempatkan evaluasi sebagai tahap untuk menilai kualitas cluster dan klasifikasi sebelum hasil digunakan sebagai informasi pendukung bagi siswa.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk mengolah nilai rapor menjadi dua keluaran, yaitu tingkat kemampuan akademik dan kecenderungan bidang. K-Means digunakan sebagai unsupervised learning, sedangkan SLP digunakan sebagai supervised learning sebelum kedua hasil digabungkan menjadi rekomendasi awal (Bintoro et al., 2024). Data diperiksa, dibagi

menjadi training, validation, dan testing, lalu dinormalisasi sebelum pemodelan; kerangka penelitian disajikan pada Gambar 1.

Kerangka penelitian dimulai dari identifikasi masalah pada proses arahan bidang akademik, dilanjutkan pengumpulan 1.000 data nilai rapor siswa kelas IX, pemeriksaan atribut, pembentukan X1-X9 dan X10, pembagian data, serta normalisasi. K-Means memproses X1-X9 untuk membentuk kategori Tinggi, Sedang, dan Rendah melalui perhitungan jarak dan pembaruan centroid, sedangkan SLP memproses X1-X10 untuk mempelajari kecenderungan Sains, Sosial, atau Kreatif melalui pembaruan bobot dan bias. Keluaran kedua metode kemudian digabungkan menjadi rekomendasi awal dan dievaluasi menggunakan Silhouette Score untuk K-Means serta confusion matrix, accuracy, precision, recall, dan F1-score untuk SLP.

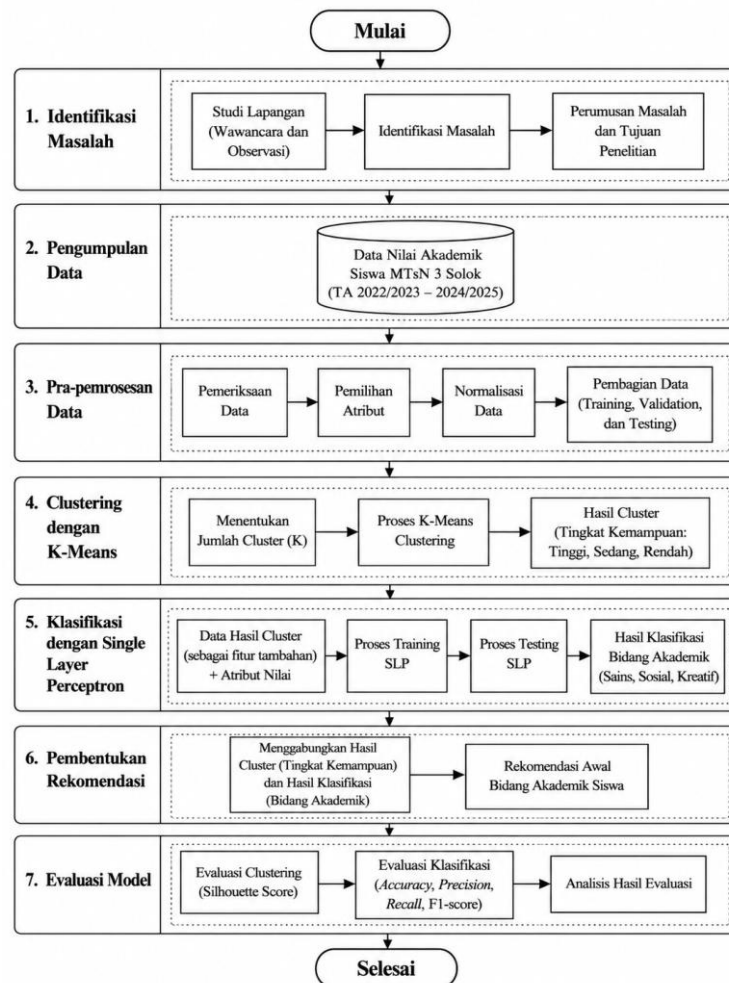
Data penelitian terdiri atas sembilan nilai mata pelajaran yang dipetakan menjadi X1 sampai X9, sedangkan X10 dibentuk dari selisih dua rata-rata bidang akademik tertinggi untuk menunjukkan tingkat dominansi kecenderungan siswa. Atribut identitas tidak digunakan dalam pemodelan karena tidak berhubungan dengan proses perhitungan, sedangkan label target SLP ditentukan berdasarkan rata-rata tertinggi dari kelompok Sains, Sosial, dan Kreatif. Dataset selanjutnya dibagi secara stratified menjadi 70% data training, 15% data validation, dan 15% data testing agar proporsi setiap kelas tetap relatif seimbang pada masing-masing bagian.

Normalisasi dilakukan menggunakan Min-Max pada rentang 0,1 sampai 0,9 agar setiap atribut berada pada skala yang sama sebelum digunakan dalam proses pemodelan. Nilai minimum dan maksimum diperoleh dari data training, kemudian parameter yang sama diterapkan pada data validation dan testing agar informasi dari data pengujian tidak masuk ke dalam proses pembentukan model. Data yang telah dinormalisasi selanjutnya digunakan dalam K-Means dengan jumlah cluster $K=3$ dan jarak Euclidean, sedangkan pembaruan centroid dilakukan secara berulang sampai keanggotaan cluster tidak mengalami perubahan.

Penentuan kategori hasil K-Means tidak mengikuti nomor cluster secara langsung, tetapi dilakukan dengan membandingkan karakteristik centroid dan nilai akademik pada masing-masing kelompok sehingga diperoleh tingkat Tinggi, Sedang, dan Rendah. Pada tahap klasifikasi, SLP memiliki sepuluh input dan tiga neuron output yang mewakili kelas Sains, Sosial, dan Kreatif, dengan bobot serta bias awal bernilai 0 dan learning rate sebesar 0,01. Proses training dilakukan menggunakan data training, sedangkan data validation digunakan untuk memantau perkembangan model dan data testing hanya digunakan pada pengukuran performa akhir.

Hasil prediksi SLP ditentukan berdasarkan nilai output terbesar, kemudian dipadukan dengan tingkat kemampuan yang diperoleh dari K-Means untuk membentuk rekomendasi akhir. Apabila nilai X10 lebih dari 0,03, sistem menampilkan satu bidang utama, sedangkan X10 sebesar 0,00 sampai 0,03 digunakan untuk menampilkan dua bidang yang memiliki nilai rata-rata sangat berdekatan. Aturan tersebut digunakan pada tahap penyajian rekomendasi agar kecenderungan yang hampir seimbang tetap dapat ditampilkan tanpa mengubah hasil evaluasi murni model SLP.

Tahap akhir metode penelitian dilakukan dengan membandingkan hasil pengelompokan dan klasifikasi terhadap tujuan yang telah ditetapkan. Hasil K-Means digunakan untuk menunjukkan tingkat kemampuan akademik, sedangkan hasil SLP menunjukkan kecenderungan bidang yang kemudian disatukan dalam bentuk rekomendasi awal. Seluruh proses tersebut disusun secara berurutan agar setiap tahap pengolahan data dapat ditelusuri dan hasil yang diperoleh tetap sesuai dengan karakteristik data siswa.



Gambar 1 Kerangka Penelitian

K-Means

K-Means mengelompokkan kemampuan akademik dari sembilan nilai mata pelajaran X1-X9 yang telah dinormalisasi. Setiap data ditempatkan pada centroid terdekat dan proses diulang sampai pengelompokan stabil (Hapsari & Indriyani, 2024). Centroid diperbarui dari rata-rata anggota cluster pada setiap iterasi (Handoko et al., 2025).

Tabel 1. Parameter Utama K-Means

Hasil	Keterangan
Input	X1-X9 (sembilan nilai mata pelajaran)
Jumlah cluster	K = 3
Ukuran jarak	Euclidean distance
Kriteria berhenti	Keanggotaan cluster tidak berubah
Evaluasi	Silhouette Score
Input	X1-X9 (sembilan nilai mata pelajaran)
Jumlah cluster	K = 3

Tahapan K-Means meliputi penetapan K=3, penentuan centroid awal, perhitungan jarak Euclidean, penempatan data ke centroid terdekat, pembaruan centroid, dan pengulangan sampai stabil. Inisialisasi centroid dapat memengaruhi jalannya iterasi sehingga hasil akhir tetap diperiksa setelah konvergensi (Khan et al., 2024). Persamaan jarak dan pembaruan centroid ditunjukkan pada Persamaan (1) dan (2).

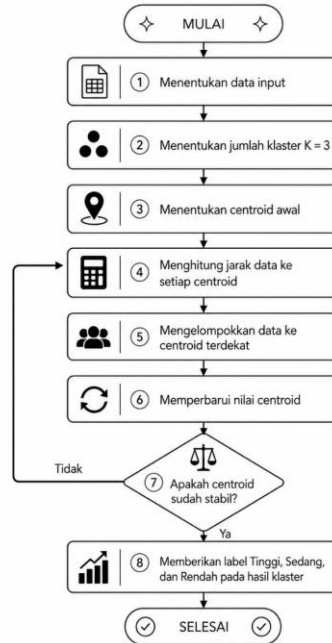
$$d(x, c_j) = \sqrt{\sum_{i=1}^m (x_i - c_{ji})^2} \dots\dots\dots(1)$$

$$c_j = \left(\frac{1}{n_j}\right) \sum_{i=1}^{n_j} x_i \dots\dots\dots(2)$$

Keterangan: x : data siswa,

c_j : centroid cluster ke-j,

m : jumlah atribut, dan n_j jumlah anggota cluster ke-j.

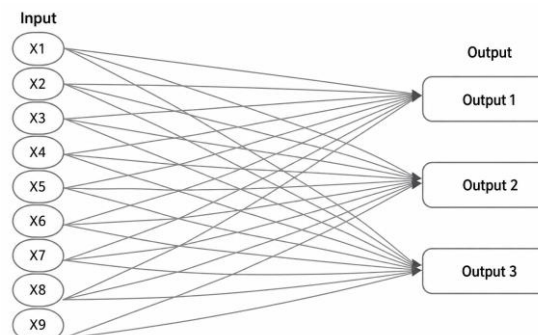


Gambar 2 Alur Proses K-Means

Setelah konvergen, label Tinggi, Sedang, dan Rendah ditetapkan dari karakteristik centroid dan statistik nilai asli, bukan dari nomor cluster. Kualitas cluster dinilai dengan Silhouette Score; nilai mendekati 1 menunjukkan pemisahan yang lebih baik (Sholeh et al., 2023). Nilai mendekati nol atau negatif menunjukkan data berada dekat batas antarkluster (Dien et al., 2024).

Single Layer Perceptron (SLP)

SLP digunakan untuk mengklasifikasikan kecenderungan Sains, Sosial, atau Kreatif. Input terdiri atas X1-X9 dan X10, dengan X10 sebagai selisih dua rata-rata bidang tertinggi; target dibentuk dari rata-rata tertinggi pada kelompok Sains, Sosial, atau Kreatif. SLP menghubungkan input langsung ke neuron output melalui bobot dan bias sehingga prosesnya relatif mudah ditelusuri (Taghinezhad & Ghasemzadeh, 2024).



Gambar 3. Struktur Konseptual Single Layer Perceptron

Gambar 3 menunjukkan hubungan antara variabel input dan tiga neuron output yang mewakili Sains, Sosial, dan Kreatif. Implementasi penelitian menambahkan X10 sebagai variabel dominansi sehingga SLP memproses sepuluh input dalam proses klasifikasi. Nilai keluaran terbesar dipilih sebagai kelas prediksi setelah seluruh input dihitung menggunakan bobot dan bias masing-masing.

Tabel 2. Parameter Pelatihan SLP

Parameter	Nilai	Keterangan
Jumlah input	10	X1-X10
Jumlah output	3	Sains, Sosial, Kreatif
Learning rate	0,01	Laju pembaruan bobot
Bobot dan bias awal	0	Kondisi awal model
Maximum epoch	20.000	Batas pelatihan
Random seed	42	Pengacakan data
Pembagian data	700/150/150	Training/validation/testing

SLP menggunakan 10 input, 3 neuron output, bobot dan bias awal 0, learning rate 0,01, maximum epoch 20.000, dan random seed 42. Skor tiap kelas dihitung dari input, bobot, dan bias, kemudian argmax menentukan kelas dan parameter diperbarui ketika prediksi salah sesuai prinsip pembelajaran Perceptron (Taye, 2023). Prosedur pelatihan menggunakan data training, dipantau melalui data validation, dan selanjutnya diuji pada data testing untuk melihat kemampuan model pada data yang tidak digunakan selama pembentukan bobot (Géron, 2022). Persamaan fungsi skor dan pembaruan bobot serta bias SLP ditunjukkan pada persamaan 3 dan 4.

$$y_k = \sum_{i=1}^{10} w_{(ki)}x_i + b_k; \hat{y} = \operatorname{argmax}_{k(y_k)} \dots\dots\dots(3)$$

$$w_t \leftarrow w_t + \alpha x; w_p \leftarrow w_p - \alpha x; b_t \leftarrow b_t + \alpha; b_p \leftarrow b_p - \alpha \dots\dots\dots(4)$$

Keterangan: y_k : Skor kelas,
 w : bobot,
 x : input
 b : bias
 x : input
 α : learning rate
 t : kelas target
 p : kelas salah prediksi

Confusion Matrix

Confusion matrix membandingkan target dan prediksi pada setiap kelas, kemudian hasilnya digunakan untuk menghitung accuracy, precision, recall, dan F1-score (Gori et al., 2024). Accuracy menunjukkan proporsi prediksi benar, precision menunjukkan ketepatan prediksi suatu kelas, recall menunjukkan kemampuan model menemukan anggota kelas, dan F1-score merangkum keseimbangan antara precision dan recall. Metrik tersebut digunakan bersama pola kesalahan antar kelas agar performa SLP dapat dibaca secara lebih lengkap.

Tabel 3. Dasar Pembacaan Confusion Matrix dan Metrik

Komponen	Makna
TP	Data kelas tertentu diprediksi benar sebagai kelas tersebut
FP	Data kelas lain diprediksi sebagai kelas tersebut
FN	Data kelas tersebut diprediksi sebagai kelas lain

Accuracy	$\text{Accuracy} = (\text{jumlah prediksi benar} / \text{seluruh data uji}) \times 100\%$
Precision	$\text{TP} / (\text{TP} + \text{FP})$
Recall	$\text{TP} / (\text{TP} + \text{FN})$
F1-score	$2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$

Metrik evaluasi dibaca bersama pola kesalahan pada tiap kelas agar performa SLP tidak hanya dinilai dari persentase prediksi benar. Perbandingan nilai antar metrik membantu melihat apakah model memiliki kinerja yang relatif seimbang pada kelas Sains, Sosial, dan Kreatif. Hasil evaluasi selanjutnya digunakan sebagai dasar untuk menilai kelayakan model pada data testing internal penelitian.

HASIL DAN PEMBAHASAN

Hasil penelitian disajikan mengikuti alur pengolahan data, yaitu dataset, analisis variabel, hasil K-Means, hasil SLP, dan evaluasi. Penyajian difokuskan pada hasil akhir yang berkaitan langsung dengan tujuan penelitian agar pembahasan tetap ringkas dan mudah diikuti. Perhitungan berulang selama proses pelatihan tidak ditampilkan seluruhnya karena bagian ini menekankan hasil bersih dari metode yang digunakan.

Dataset Pendidikan

Dataset terdiri atas 1.000 data nilai rapor siswa kelas IX MTsN 3 Solok tahun ajaran 2021/2022-2024/2025. Setiap data memuat sembilan nilai mata pelajaran, sedangkan identitas siswa tidak digunakan dalam pemodelan. Kelengkapan dan konsistensi data diperiksa sebelum normalisasi; contoh data tersaji pada Tabel 4.

Tabel 4 Sampel Data Nilai Rapor

X1	X2	X3	X4	X5	X6	X7	X8	X9
86,20	88,10	89,50	91,00	90,00	86,20	85,40	89,10	89,90
90,40	92,10	90,90	90,40	91,60	89,00	89,90	89,70	89,60
88,50	92,00	92,10	93,50	89,90	90,10	90,30	91,60	90,50
86,90	88,50	89,00	90,90	89,70	86,10	84,30	89,60	88,40
85,80	89,40	90,50	89,20	90,50	90,40	88,00	87,80	89,70
84,90	90,50	91,90	88,50	91,30	90,00	88,00	87,70	90,30

Tabel 4 menunjukkan adanya variasi nilai antar siswa pada sembilan mata pelajaran yang digunakan dalam penelitian. Variasi tersebut menjadi dasar K-Means untuk membaca kemiripan pola kemampuan akademik antar siswa. Data yang sama selanjutnya digunakan dalam pembentukan variabel SLP untuk mempelajari kecenderungan bidang akademik.

Analisis Variabel

Sembilan mata pelajaran dibagi menjadi tiga aspek, yaitu Sains dari Matematika, IPA, dan TIK, Sosial dari IPS, Bahasa Indonesia, dan Bahasa Inggris, serta Kreatif dari Prakarya, Seni Budaya, dan PJOK. Pembagian aspek digunakan untuk menghitung rata-rata bidang yang menjadi dasar pembentukan target kecenderungan akademik. Variabel X10 dihitung dari selisih dua rata-rata aspek tertinggi untuk menunjukkan tingkat dominansi kecenderungan bidang siswa.

Tabel 5. Variabel Penelitian

Variabel	Mata Pelajaran / Atribut	Aspek
X1	Matematika	Sains
X2	IPA	Sains

Variabel	Mata Pelajaran / Atribut	Aspek
X3	TIK	Sains
X4	IPS	Sosial
X5	Bahasa Indonesia	Sosial
X6	Bahasa Inggris	Sosial
X7	Prakarya	Kreatif
X8	Seni Budaya	Kreatif
X9	PJOK	Kreatif
X10	Selisih dua rata-rata bidang tertinggi	Tambahan SLP

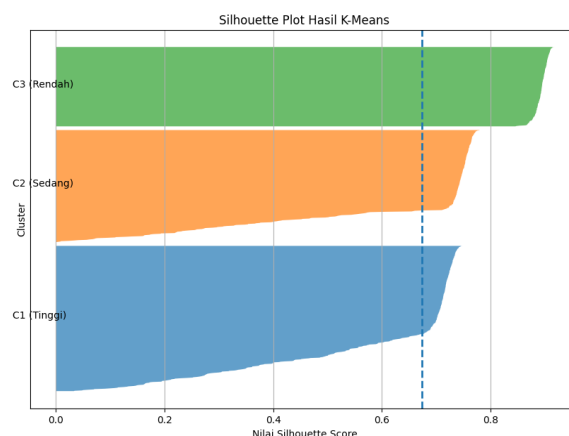
K-Means menggunakan variabel X1 sampai X9 sebagai dasar pengelompokan tingkat kemampuan akademik siswa. SLP menggunakan X1 sampai X10 untuk mempelajari kecenderungan Sains, Sosial, atau Kreatif. Kedua keluaran diproses secara terpisah terlebih dahulu, kemudian digabungkan pada tahap pembentukan rekomendasi akhir.

Hasil K-Means

K-Means mencapai kondisi stabil pada iterasi ke-7 setelah tidak terdapat perubahan yang berarti pada pembentukan cluster. Hasil akhir terdiri atas 431 siswa pada C1 dengan kategori Tinggi dan rata-rata 88,74, 333 siswa pada C2 dengan kategori Sedang dan rata-rata 79,01, serta 236 siswa pada C3 dengan kategori Rendah dan rata-rata 63,80. Komposisi tersebut digunakan sebagai hasil pengelompokan tingkat kemampuan akademik sebelum digabungkan dengan keluaran SLP.

Tabel 6. Hasil Akhir Pengelompokan K-Means

Cluster	Label	Jumlah	Rata-rata Nilai
C1	Tinggi	431	88,74
C2	Sedang	333	79,01
C3	Rendah	236	63,80



Gambar 4. Visualisasi Silhouette Score K-Means

Silhouette Score sebesar 0,6734 menunjukkan struktur tiga cluster memiliki pemisahan yang cukup baik. Gambar 4 memperlihatkan mayoritas nilai silhouette positif, meskipun beberapa data C1 dan C2 berada dekat batas cluster. K=2 menghasilkan 0,7094, tetapi K=3 dipertahankan karena penelitian membutuhkan tiga tingkat kemampuan.

Hasil K-Means Single Layer Perceptron

SLP dilatih dengan 700 data training dan 150 validation, lalu diuji pada 150 data testing. Model berhenti pada epoch 10.585 dengan error training 0, akurasi training 100%, dan validation 98,67%; pada testing, 146 dari 150 data diprediksi benar. Confusion matrix mencatat 70/72 Sains, 36/37 Sosial, dan 40/41 Kreatif terklasifikasi benar.

Tabel 7. Confusion Matrix SLP pada 150 Data Testing

Target / Prediksi	Sains	Sosial	Kreatif
Sains	70	1	1
Sosial	0	36	1
Kreatif	1	0	40

Empat kesalahan klasifikasi muncul pada data yang memiliki skor antarkelas relatif berdekatan. Kondisi tersebut menunjukkan bahwa sebagian data berada dekat dengan batas keputusan yang dibentuk oleh model. Secara keseluruhan, SLP mampu mereplikasi pola target pada data testing internal, tetapi keluarannya tetap dipahami sebagai kecenderungan berbasis nilai.

Hasil Evaluasi

Evaluasi menunjukkan K-Means memperoleh Silhouette Score sebesar 0,6734, sehingga struktur tiga cluster memiliki pemisahan yang cukup baik pada data penelitian. SLP mencapai accuracy 97,33%, precision 97,36%, recall 97,33%, dan F1-score 97,34% pada 150 data testing. Nilai tersebut menunjukkan bahwa fungsi pengelompokan dan klasifikasi telah berjalan cukup baik pada dataset internal yang digunakan.

	Kelas	precision	recall	f1-score	support
0	Sains	98.59	97.22	97.90	72
1	Sosial	97.30	97.30	97.30	37
2	Kreatif	95.24	97.56	96.39	41
3	accuracy	97.33	97.33	97.33	1
4	macro avg	97.04	97.36	97.19	150
5	weighted avg	97.36	97.33	97.34	150

Gambar 5. Classification Report Single Layer Perceptron

Gambar 5 menunjukkan F1-score Sains 97,90%, Sosial 97,30%, Kreatif 96,39%, dan weighted average 97,34%. Rekomendasi akhir menggabungkan tingkat kemampuan K-Means dengan kecenderungan bidang SLP, misalnya Tinggi-Sains atau Sedang-Sosial. $X_{10}=0,00$ menampilkan dua bidang Seimbang, $X_{10} 0,01-0,03$ menampilkan Hampir Seimbang, dan $X_{10}>0,03$ menampilkan satu bidang Dominan. Ambang 0,03 bersifat heuristik dan bukan batas pedagogis baku.

Performa tinggi perlu dibaca hati-hati karena label SLP dibentuk dari rata-rata nilai yang juga berasal dari fitur input, sehingga terdapat risiko target leakage. Uji 5-fold cross-validation tetap sekitar 98%, tetapi X_{10} tidak meningkatkan performa rata-rata dibanding X_1-X_9 . Dengan demikian, model telah teruji baik pada data internal, namun validasi pedagogis tetap memerlukan guru BK, tes minat-bakat, pilihan siswa, dan data jurusan aktual.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil membentuk rekomendasi awal bidang akademik dengan membagi fungsi K-Means dan SLP secara terpisah. K-Means mengelompokkan 1.000 siswa menjadi 431 kategori Tinggi, 333 Sedang, dan 236 Rendah serta menghasilkan Silhouette Score 0,6734, sedangkan SLP pada 150 data testing memperoleh accuracy 97,33%, precision 97,36%, recall 97,33%, dan F1-score 97,34%. Gabungan keduanya menghasilkan informasi tingkat kemampuan dan kecenderungan Sains, Sosial, atau Kreatif yang dapat digunakan sebagai bahan pertimbangan awal, bukan sebagai keputusan mutlak penentuan jurusan SMK.

Saran

Pengembangan berikutnya perlu memvalidasi rekomendasi dengan guru BK, tes minat-bakat, pilihan siswa, dan jurusan SMK yang benar-benar dipilih, serta menggunakan data dari sekolah atau tahun ajaran lain untuk menilai kemampuan generalisasi. Perbandingan dengan metode klasifikasi lain juga diperlukan karena label SLP dibentuk dari nilai yang sama dengan fitur input, sehingga risiko target leakage masih perlu dikendalikan. X10 sebaiknya diperlakukan sebagai informasi dominansi atau fitur eksperimen sampai manfaatnya terhadap performa dan validitas pedagogis dapat dibuktikan.

DAFTAR PUSTAKA

- Alawi, S. J. S., Shaharane, I. N. M., & Jamil, J. M. (2023). *CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS*. 2(1), 41–55.
- Alfarizi, M., Teknologi, A., & Pendidikan, T. (2023). *Evolusi Kecerdasan Buatan dalam Pendidikan : Teknologi , Etika , dan Akselerasi Indonesia* (Vol. 0). <https://doi.org/10.55981/brin.1354.c1061>
- Aravik, H., Marnisah, L., & Nubila, H. (2025). *Sosialisasi Pemilihan Jurusan Kuliah Pada Sekolah SMA di Kota Palembang*. 221–228.
- Aulia Sitorus, N. S. A. (2025). *Klasifikasi Biner Menggunakan Single Layer Perceptron Pada Data Kuesioner Pelamar Asisten Lab Komputer*.
- Banarsari, A., Rizki Nurfadilah, D., & Zainul Akmal, A. (2023). *Pemanfaatan Teknologi Pendidikan Pada Abad 21 Arum*. 6(Snip 2022), 459–464.
- Bintoro, P., Sari, R., Wiharjo, E., & Asari, A. (2024). *Pengantar Machine Learning* (A. Asari (ed.)). PT MAFY MEDIA LITERASI INDONESIA.
- Dewi, S., & Pakereng, M. A. I. (2023). *Implementasi principal component analysis pada k-means untuk klasterisasi tingkat pendidikan penduduk kabupaten semarang*. 8(4), 1186–1195.
- Dien, H. E., Ratsanjani, M. H., Saputra, A. A., Retno, A., & Hayati, T. (2024). *Analisis Klasterisasi Patok Jalan Berbasis Geospasial Menggunakan K-Means dan Evaluasi Davies-Bouldin Analysis of Geospatial-Based Road Milestone Clustering Using K-Means and Davies-Bouldin Evaluation*. 215–224. <https://doi.org/10.56873/jpkm.v9i2.5471>
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd Editio). O'Reilly Media.
- Gori, T., Sunyoto, A., Fatta, H. Al, Yogyakarta, U. A., Korespondensi, P., Selection, C. F., & Siswa, K. A. (2024). *PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA DATA PREPROCESSING AND CLASSIFICATION FOR PREDICTING STUDENT*. 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Handayani, S., Devianto, Y., Niqotaini, Z., & Wijanti, Y. (2025). *Machine Learning di Dunia Pendidikan* (N. Umilizah (ed.)).
- Handoko, K., Astiti, S., Novrian, W., Putra, T. N., & Darmansah, D. (2025). *K-Means Clustering Menggunakan Aplikasi Rapidminer* (E. Efitra & I. K. Sari (eds.)). PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=dQBjEQAQBAJ>

- Hapsari, R. K., & Indriyani, T. (2024). *Data Mining* (Cetakan Pe). PT. Sonpedia Publishing Indonesia.
- Hayati, R., Wijayati, I. W., Nugroho, F. arief, & Wardoyo, T. H. (2025). *ASESMEN PEMBELAJARAN Teori dan Praktik* (A. Cahyo Purnomo (ed.)). Sada Kurnia Pustaka.
- Khan, A. A., Salman, M., Batool, A., Raza, M. S., & Bashir, M. A. (2024). K-Means Centroids Initialization Based on Differentiation Between Instances Attributes. *International Journal of Intelligent Systems*, 2024(1), 7086878.
- Maulani, G., Hasan, F. N., Setiawan, D., Bowo, I. T., Ardhana, V. Y. P., Ramdhani, Y., Inayah, I., Ardiantoro, L., Sugianto, C. A., Chandra, R., Afandi, I. R., Septyandy, M. R., & Safitri, R. (2025). *Machine Learning*. CV. Mega Press Nusantara. <https://books.google.co.id/books?id=RbIPEQAAQBAJ>
- Permana, N. P., Muhammad, I. F., Ilmawan, F., & Sovia, R. (2026). *Prediksi Kelulusan Siswa Berdasarkan Nilai Akademik Menggunakan Algoritma Perceptron*. 6(November 2025).
- Pribadi, A., Maulana, H., & Aditiawan, F. (2024). *Sistem pakar pemilihan jurusan sma (ipa dan ips)*. 8(3), 4070–4077.
- Sholeh, M., Aeni, K., Informatika, P. S., Informatika, P. S., Sains, F., Peradaban, U., & Bouldin, D. (2023). *PERBANDINGAN EVALUASI METODE DAVIES BOULDIN , ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN*. 8(1).
- Taghinezhad, F., & Ghasemzadeh, M. (2024). A New Structure for Perceptron in Categorical Data Classification. *Journal of AI and Data Mining*, 12(3), 409–421.
- Tampubolon, J. (2024). *Strategi Pemilihan Program Studi di Perguruan Tinggi Bidang Ilmu Pendidikan Berdasarkan Peluang Pengembangan Akademik dan Prospek Karir*. 28–35. <https://doi.org/10.52622/jam.v3i2.294>
- Taye, M. M. (2023). *Understanding of Machine Learning with Deep Learning :*
- Wedo, M., Nurcahyo, G. W., & Sovia, R. (2025). *Implementasi Metode Profile Matching dalam Sistem Pendukung Keputusan untuk Seleksi Penerimaan Siswa Baru*. 7(3). <https://doi.org/10.32877/bt.v7i3.2229>
- Yudhistira, A., & Andika, R. (2023). *Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering*. 1(1), 20–28.