

DATA MINING PENDIDIKAN: PREDIKSI GAYA BELAJAR MAHASISWA TEKNIK MENGGUNAKAN MACHINE LEARNING

Sumarlin^{*1}, Dewi Anggraini²

^{1,2}Sekolah Tinggi Informatika dan Komputer Uyelindo Kupang, Kupang
Email: ¹shumarlin@gmail.com, ²thewifoeh@gmail.com

^{*}Penulis Korespondensi

(Naskah masuk: 06 September 2024, diterima untuk diterbitkan: 19 Juni 2025)

Abstrak

Dalam platform online, pembelajar yang berbeda memiliki gaya belajar yang berbeda berdasarkan perilaku belajar. Oleh karena itu, menganalisis perilaku dan mendeteksi gaya belajar mahasiswa adalah penting untuk memberikan rekomendasi sumber daya yang tepat, sehingga meningkatkan hasil belajar mahasiswa. Untuk memprediksi gaya belajar mahasiswa, dihitung dan dibandingkan kinerja algoritma pembelajaran mesin seperti regresi logistik, pohon penentuan, K-Nearest neighbour, support vector machine, neural network, dan Naive Bayes. Dataset terdiri dari seratus mahasiswa teknik yang belajar Arsitektur Komputer selama satu semester. Studi berbasis data seperti ini sangat penting untuk membangun sistem analisis pembelajaran di institusi pendidikan tinggi dan membantu proses pengambilan keputusan. Hasilnya menunjukkan bahwa model yang disarankan mencapai akurasi klasifikasi sebesar 65–78% dengan hanya empat parameter digunakan: nilai akhir, predikat, program studi, dan jenis kelamin. Hasil menunjukkan bahwa algoritma K-Nearest Neighbour memiliki tingkat akurasi 78% tertinggi dibandingkan dengan algoritma machine learning lainnya. Ini menunjukkan bahwa ada korelasi yang signifikan antara data aktual dan data prediksi. Hasilnya menunjukkan bahwa 78% sampel diklasifikasikan dengan benar. Hasil empiris dari penelitian ini memungkinkan pemahaman yang lebih baik tentang proses penggalian data pendidikan perguruan tinggi saat ini. Pemahaman ini dapat digunakan untuk mempertimbangkan faktor-faktor yang perlu dipertimbangkan oleh para mahasiswa teknik saat membuat keputusan tentang proses pembelajaran.

Kata kunci: Gaya Belajar, Pembelajaran Mesin, Penggalian Data.

EDUCATIONAL DATA MINING: PREDICTING OF ENGINEERING STUDENTS' LEARNING STYLES USING MACHINE LEARNING

Abstract

In online platforms, different learners have different learning styles based on learning behavior. Therefore, analyzing behavior and detecting student learning styles is important to provide appropriate resource recommendations, thereby improving student learning outcomes. To predict student learning styles, the performance of machine learning algorithms such as logistic regression, determination trees, K-Nearest neighbors, support vector machines, neural networks, and Naive Bayes are calculated and compared. The dataset consists of one hundred engineering students studying Computer Architecture for one semester. Data-based studies like this are essential for building learning analytics systems in higher education institutions and aiding decision-making processes. The results show that the proposed model achieves a classification accuracy of 65–78% with only four parameters used: final grade, predicate, study program, and gender. The results show that the K-Nearest Neighbor algorithm has the highest accuracy rate of 78% compared to other machine learning algorithms. This shows that there is a significant correlation between the actual data and the predicted data. The results show that 78% of the samples were classified correctly. The empirical results of this research enable a better understanding of the current process of mining higher education education data. This understanding can be used to consider factors that engineering students need to consider when making decisions about the learning process.

Keywords: Learning Style, Machine Learning, Data Mining,

1. PENDAHULUAN

Data Mining (DM) mudah digunakan dengan pengetahuan pemrograman apapun, topik analitik pembelajaran semakin populer di bidang pendidikan (E. Christodoulou, 2019). Metode baru ini terutama berkonsentrasi pada pembuatan cara untuk menangani dan mengeksplorasi sejumlah besar informasi yang diwakili dalam berbagai jenis data. Menurut Zhang Zhong Biao (2023), penerapan penambangan data dalam pendidikan memberikan pemahaman yang lebih baik tentang sinergi antara lingkungan pendidikan dan pembelajaran yang dibuat oleh institusi untuk mendorong proses pembelajaran. WEKA, Rapidminer, dan lainnya merupakan jenis alat penambangan data dibangun dalam bahasa JAVA dan mendukung berbagai tugas penambangan data biasa dan memungkinkan implementasi sejumlah algoritma penambangan dan analisis data terukur (Garg, 2020).

Penggunaan metode data mining tradisional untuk memecahkan masalah pendidikan dikenal sebagai Penggalian data pendidikan (EDM). EDM mencakup data pendidikan seperti informasi mahasiswa, catatan pendidikan, hasil ujian, partisipasi mahasiswa di kelas, dan frekuensi pertanyaan mahasiswa (Fernandes et al., 2019). EDM telah berkembang menjadi alat yang berguna dalam beberapa tahun terakhir untuk memprediksi prestasi akademik, mengidentifikasi pola tersembunyi dalam data pendidikan, dan meningkatkan lingkungan belajar dan mengajar. Dengan bantuan EDM, analisis pembelajaran telah memperoleh dimensi baru (Waheed et al., 2020). EDM memainkan peran yang sangat penting dalam 'menambang' informasi paling tepat tentang perilaku mahasiswa serta mengukur keefektifan proses pembelajaran (Sana et al., 2019).

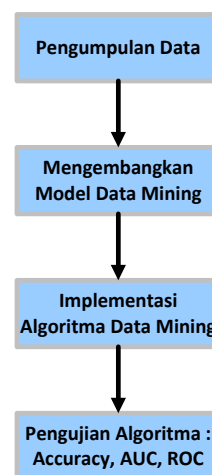
Analisis pembelajaran mencakup berbagai aspek pengumpulan informasi mahasiswa secara bersamaan, memeriksa dan menganalisis lingkungan belajar untuk mendapatkan pemahaman yang lebih baik, dan menunjukkan kinerja mahasiswa. Bidang penelitian yang sedang berkembang adalah penggunaan analisis pembelajaran untuk meningkatkan hasil pendidikan. Meskipun penelitian tentang analisis pembelajaran lebih banyak dilakukan dalam memprediksi kinerja akademik mahasiswa juga merupakan bagian dari analisis pembelajaran (Waheed et al., 2020), yang mencakup klasifikasi gaya belajar mahasiswa (KN Prashanth Kumar, 2024). E-learning adalah gudang data aktivitas proses belajar mahasiswa. Ini dapat digunakan sebagai dasar untuk mengklasifikasikan gaya belajar mahasiswa. Hasil belajar mahasiswa dipengaruhi oleh gaya belajar visual, auditory, read/write, dan kinestetik (VARK) yang diidentifikasi berdasarkan karakteristik mahasiswa (Milan Klement, 2024). Peluang baru untuk meningkatkan pengalaman belajar dan mengoptimalkan interaksi pengguna dengan

platform teknologi muncul sebagai hasil dari data proses pendidikan yang dikumpulkan (Shorfuzzaman et al., 2019). Banyak hal yang telah diperbaiki oleh pemrosesan data pendidikan. Ini termasuk memprediksi perilaku mahasiswa, pembelajaran analitis, dan pendekatan baru kebijakan pendidikan (Capuano & Toti, 2019; Viberg et al., 2018).

Berbagai teknik data mining yang diterapkan peneliti untuk mempelajari trend atau pola tersembunyi dalam data pendidikan dapat memberikan wawasan berharga bagi lembaga pendidikan dalam hal menjadikan proses pembelajaran adaptif dengan kebutuhan mahasiswa. Wawasan ini dapat membantu institusi mencapai tujuan akhir mereka dalam meningkatkan kinerja akademik mahasiswa dalam sistem pembelajaran berbantuan teknologi di dunia modern (Kudratdeep Aulakh, 2023). Performa dari algoritma Logistik Regression, Decision Trees, KNN, SVM, Neural Networks, Naïve Bayes yang merupakan bagian dari algoritma pembelajaran mesin (Mustafa Yağcı, 2022), dihitung dan dibandingkan dalam mengklasifikasi gaya belajar mahasiswa. Oleh karena itu, tujuan utama dari penelitian ini adalah untuk memahami bagaimana EDM dapat diterapkan pada sistem e-learning dengan fokus pada perspektif peningkatan hasil belajar mahasiswa (Azadi dkk. al., 2021, Aqeel dkk., 2022, Su dkk., 2021, Abbas, 2020), dengan mengklasifikasi mahasiswa berdasarkan gaya belajar VARK. Hasilnya, model ini memungkinkan evaluasi dan peningkatan komponen sistem pendidikan untuk memastikan kualitas pendidikan dengan menemukan pola tersembunyi dalam data pendidikan.

2. METODE PENELITIAN

Detail kumpulan data, metode pra-pemrosesan, dan algoritma pembelajaran mesin yang digunakan dalam penelitian ini dibahas dalam bagian ini. Berikut merupakan prosedur dalam penelitian ini.



Gambar 1. Prosedur penelitian

1. Pengumpulan
Data - Langkah awal adalah mengumpulkan data yang akan menjadi dasar bagi proses data mining.
2. Mengembangkan Model Data Mining –
Setelah data terkumpul, model dikembangkan berdasarkan data tersebut. Langkah ini melibatkan pemilihan fitur yang sesuai dan perancangan model yang disesuaikan dengan permasalahan yang akan dipecahkan.
3. Implementasi Algoritma Data Mining
Setelah model dikembangkan, langkah berikutnya adalah mengimplementasikan algoritma data mining yang telah dipilih dan menerapkannya pada dataset.
4. Pengujian Algoritma Accuracy, AUC, ROC, F1, Recall, Presisi.
Langkah terakhir adalah menguji kinerja algoritma yang telah diimplementasikan menggunakan berbagai metrik seperti Accuracy, Area Under the Curve (AUC), dan analisis kurva Receiver Operating Characteristic (ROC) untuk menentukan efektivitas model.

2.1 Dataset

Institusi pendidikan secara rutin menyimpan semua data yang tersedia tentang mahasiswa di media elektronik. Data disimpan dalam database untuk diproses. Data ini bisa bermacam-macam jenis dan volumenya, mulai dari demografi mahasiswa hingga prestasi akademis mereka. Pada penelitian ini data diambil dari sistem e-learning, dimana seluruh catatan mahasiswa disimpan pada sebuah sistem pada kampus STIKOM Uyelindo. Dalam catatan tersebut, program studi, jenis kelamin, konten, nilai akhir, dan predikat dari 100 mahasiswa yang telah mengambil mata kuliah arsitektur komputer tahun akademik 2022-2023 dipilih sebagai dataset. Tabel 1 menunjukkan sebaran mahasiswa menurut satuan akademik. Selain itu.

Table 1. Dataset

Program Studi	Jumlah Mahasiswa
Sistem Informasi	60
Teknik Informatika	40
Total Number of Students	100

Semua dataset yang digunakan bersumber dari database e-learning, variable konten terdapat konten video, audio dan textual yang digunakan mahasiswa pada e-learning, nilai akhir diperoleh dari total nilai tugas dan ujian tengah semester dengan presentasi

penilaian akhir sebesar 35%. Predikat terdiri atas baik, tidak baik, sangat baik dan cukup baik, jika nilai akhir lebih besar 85 maka prediksi sanga baik.

2.2 Identifikasi dan Pengumpulan Data

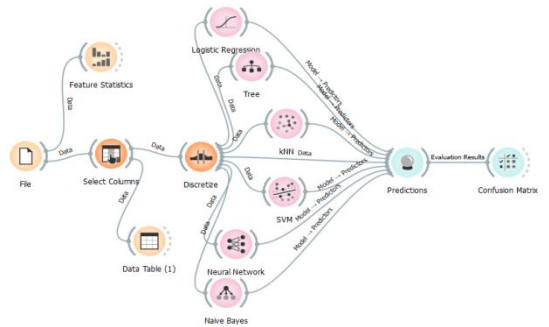
Pada tahap ini, perlu diputuskan dari mana data akan diambil, jenis data apa yang akan digunakan, dan apakah data yang dikumpulkan digunakan sesuai dengan tujuan. Pengurangan jumlah variabel yang digunakan untuk memprediksi hasil tertentu dikenal sebagai pemilihan fitur hasil, mengurangi kompleksitas, meningkatkan efisiensi komputasi algoritma, dan mencegah overfitting.

2.3 Membangun model DM dan implementasi algoritma

Gaya belajar mahasiswa dapat diklasifikasi dengan menggunakan LR, DT, RF, KNN, SVM, NN, dan NB. Klasifikasi dinilai dengan melakukan validasi silang sepuluh kali lipat. Dalam proses DM, ada dua tujuan utama. Tujuan pertama adalah melakukan prediksi dengan menganalisis data database (model prediktif), dan tujuan kedua adalah menjelaskan perilaku. Model prediktif menggunakan data dengan hasil yang diketahui dan kemudian mengklasifikasi nilai hasil untuk kumpulan data dengan hasil yang tidak diketahui. Dalam model deskriptif, pola pada data yang ada didefinisikan untuk mengambil keputusan. Metode pembelajaran mesin secara otomatis membuat model yang mencocokkan data masukan dengan nilai target yang diharapkan ketika masalah pengoptimalan yang diawasi muncul. Di sisi lain, teknik statistik menggunakan model untuk berhasil memprediksi nilai keluaran berdasarkan data masukan yang tersedia.

3. HASIL DAN PEMBAHASAN

Perangkat lunak machine learning Orange digunakan untuk menyelesaikan semua tahap percobaan. Orange adalah alat pemrograman DM yang kuat dan mudah digunakan untuk ilmuwan data ahli dan pemula. Orange menumpuk widget ke dalam alur kerja untuk melakukan analisis data. Pengambilan data, pra-pemrosesan data, visualisasi, pemodelan, dan evaluasi adalah semua fungsi yang dapat dilakukan oleh setiap widget. Serangkaian tindakan atau tindakan yang akan dilakukan pada platform untuk melakukan tugas tertentu disebut jalur kerja. Menggabungkan berbagai bagian dalam alur kerja dapat digunakan untuk membuat bagan analisis data yang komprehensif. Diagram alur kerja yang dirancang digambarkan pada Gambar 2.



Gambar 2. Alur kerja model yang dirancang

Dataset ini mencakup nilai akhir, predikat, jenis kelamin, program studi, dan konten belajar dari seratus mahasiswa yang mengikuti kelas Arsitektur komputer pada tahun akademik 2022–2023. Bagian dari data set ditampilkan dalam Tabel 3. Dalam proses pengembangan model, variabel terikat meliputi konten atau gaya belajar, sedangkan variabel bebas berupa program studi, jenis kelamin, nilai akhir, dan predikat. Model variabel ditunjukkan dalam Tabel 2.

Table 2. Model Variabel		
Variabel Bebas	Variabel Terikat (Target)	Atribut Meta
Nilai Akhir	Konten / Gaya Belajar	Student_ID
Predikat		
Jenis kelamin		
Program Studi		

Nilai akhir dari mata kuliah arsitektur komputer dikategorikan berdasarkan model diskritisasi yang sama lebar setelah model variabel ditentukan. Kriteria yang digunakan untuk mengubah nilai akhir ke dalam format kategorikal ditunjukkan dalam Tabel 3.

Table 3. Kategori Kriteria Nilai Akhir	
Kategori	Kriteria
Very Deficient	Grade < 65
Average	65 < grade <75
Good	75 < grade <85
Excellent	85>grade

Gambar 2 menunjukkan bahwa Nilai aktual sebenarnya ada di kolom Gaya belajar. Nilai yang diprediksi oleh model yang diusulkan ada pada kolom LN, DT, KNN, SVM, NN, dan NB. Misalnya, gambar 6 menunjukkan bahwa gaya belajar aktual SIS11 adalah auditori. Di sisi lain, gaya belajar visual untuk prediksi model SVM, NN, NB, dan LR, dan gaya belajar auditori untuk prediksi model DT dan KNN. Hal yang sama pada SIS12 dimana nilai gaya belajar actual adalah visual sedangkan nilai pada model prediksi DT dan KNN adalah gaya belajar read/write, model prediksi SVM,

NN dan LR adalah Visual sedangkan model prediksi Naïve Bayes adalah Auditori.

No	Type	KNN	SVM	Naive Bayes	Logistic Regression	Learning Style	StudentID	Study Program	Gender	Final Score	Predicate
1	0.75 - Auditori	0.85 - Auditori	0.25 - Visual	0.14 - Visual	0.3 - Visual	Visual	SIS11	Informasi Sys.	Male	78.89	Good
2	0.50 - Read/Write	0.25 - Read/Write	0.25 - Visual	0.38 - Auditori	0.30 - Visual	Visual	SIS12	Informasi Sys.	Male	88.89	Excellent
3	0.25 - Visual	0.45 - Visual	0.25 - Visual	0.38 - Visual	0.28 - Auditori	0.22 - Visual	SIS13	Informasi Sys.	Male	81.25	Excellent
4	1.00 - Auditori	0.85 - Auditori	0.25 - Visual	0.38 - Auditori	0.34 - Visual	Auditory	SIS14	Informasi Sys.	Male	74.13	Excellent
5	0.50 - Visual	0.50 - Visual	0.28 - Visual	0.27 - Visual	0.19 - Visual	Visual	SIS15	Informasi Sys.	Male	75.00	Good
6	0.25 - Auditori	0.45 - Auditori	0.25 - Visual	0.22 - Visual	0.27 - Auditori	0.22 - Visual	SIS16	Informasi Sys.	Female	86.13	Excellent
7	0.50 - Visual	0.25 - Visual	0.25 - Visual	0.22 - Visual	0.27 - Auditori	0.20 - Visual	SIS17	Informasi Sys.	Female	86.75	Excellent
8	0.50 - Auditori	0.25 - Read/Write	0.25 - Visual	0.22 - Visual	0.27 - Auditori	0.22 - Visual	SIS18	Informasi Sys.	Female	86.00	Excellent
9	0.50 - Read/Write	0.25 - Read/Write	0.25 - Visual	0.38 - Visual	0.30 - Auditori	0.20 - Visual	SIS19	Informasi Sys.	Male	88.00	Excellent
10	0.50 - Visual	0.25 - Visual	0.25 - Visual	0.28 - Visual	0.14 - Visual	0.18 - Visual	SIS20	Informasi Sys.	Female	81.25	Good
11	0.50 - Visual	0.50 - Read/Write	0.25 - Visual	0.28 - Visual	0.14 - Visual	0.18 - Visual	SIS21	Informasi Sys.	Female	82.50	Good
12	0.25 - Visual	0.25 - Visual	0.25 - Visual	0.38 - Auditori	0.22 - Visual	Auditory	SIS22	Informasi Sys.	Male	86.00	Excellent
13	0.25 - Auditori	0.45 - Auditori	0.25 - Visual	0.27 - Auditori	0.27 - Visual	Auditory	SIS23	Informasi Sys.	Female	86.00	Excellent
14	0.50 - Auditori	0.25 - Visual	0.25 - Visual	0.22 - Visual	0.27 - Auditori	0.28 - Visual	SIS24	Informasi Sys.	Female	87.00	Excellent
15	0.50 - Visual	0.50 - Visual	0.25 - Visual	0.12 - Visual	0.24 - Visual	0.21 - Visual	SIS25	Informasi Sys.	Male	88.00	Average
16	0.50 - Read/Write	0.50 - Visual	0.28 - Visual	0.10 - Visual	0.14 - Visual	Read/Write	SIS26	Informasi Sys.	Male	74.38	Good
17	0.50 - Visual	0.25 - Visual	0.28 - Visual	0.16 - Visual	0.14 - Visual	Visual	SIS27	Informasi Sys.	Male	77.00	Good
18	0.50 - Visual	0.50 - Read/Write	0.25 - Visual	0.28 - Visual	0.20 - Visual	Kinesthetic	SIS28	Informasi Sys.	Female	82.50	Good
19	0.50 - Visual	0.50 - Visual	0.25 - Visual	0.28 - Visual	0.20 - Visual	Visual	SIS29	Informasi Sys.	Female	83.00	Good

Gambar 3. Model dan keputusan akhir model prediktif

3.1 Evaluasi kinerja model

Performa model dievaluasi dengan matriks konfusi, akurasi klasifikasi (CA), presisi, recall, f-score (F1), dan metrik area under roc curve (AUC).

3.2 Confusion Matrix

Matriks konfusi menunjukkan kondisi kumpulan data saat ini serta jumlah prediksi model yang benar/salah. Matriks konfusi ditampilkan dalam Tabel 4. Performa model dihitung berdasarkan berapa banyak instance yang diklasifikasikan dengan benar dan berapa banyak instance yang diklasifikasikan dengan salah. Jumlah sampel sebenarnya dalam set pengujian ditunjukkan dalam baris, dan estimasi model ditunjukkan dalam kolom.

		Predicate	
		Positive (1)	Negative (0)
Actual	Positive (1)	TP	FP
	Negative (0)	FN	TN

Pada Tabel 3, true positive (TP) dan true negative (TN) menunjukkan jumlah instance yang diklasifikasikan dengan benar. False Positif (FP) menunjukkan jumlah instance yang diprediksi 1 (positif) padahal seharusnya berada di kelas 0 (negatif), dan false negative (FN) menunjukkan jumlah instance yang diprediksi 0 (negatif) padahal seharusnya berada di kelas 1.

Untuk algoritma LR, matriks konfusi 4 × 4 ditunjukkan pada Gambar 3. Diagonal utama menunjukkan persentase kejadian yang diprediksi dengan benar, dan elemen lain dalam matriks menunjukkan persentase kesalahan prediksi.

		Predicted				Σ
		Auditory	Read/Write	Visual	kinesthetic	
Actual	Auditory	4.5 %	13.6 %	81.8 %	0.0 %	22
	Read/Write	0.0 %	20.8 %	70.8 %	8.3 %	24
	Visual	0.0 %	6.8 %	88.6 %	4.5 %	44
	kinesthetic	0.0 %	20.0 %	50.0 %	30.0 %	10
Σ		1	13	79	7	100

Gambar 4. Confusion Matrix of the LR Algorithm

Gambar 3 menunjukkan bahwa 30% mahasiswa yang kategori gaya belajar kinestetik, 88,6% mahasiswa gaya belajar visual, 20,8%

kategori gaya belajar read/write dan 4,5% gaya belajar auditori, diprediksi dengan benar.

		Predicted				Σ
		Auditory	Read/Write	Visual	kinesthetic	
Actual	Auditory	63.6 %	4.5 %	31.8 %	0.0 %	22
	Read/Write	20.8 %	66.7 %	8.3 %	4.2 %	24
	Visual	9.1 %	11.4 %	77.3 %	2.3 %	44
	kinesthetic	10.0 %	30.0 %	40.0 %	20.0 %	10
Σ		24	25	47	4	100

Gambar 5. Confusion Matrix of the DT Algoritma

Gambar 4 menunjukkan bahwa 20% mahasiswa masuk kategori gaya belajar kinestetik, 77,3% mahasiswa gaya belajar visual, 66,7% kategori gaya belajar read/write dan 63,6% gaya belajar auditori, diprediksi dengan benar.

		Predicted				Σ
		Auditory	Read/Write	Visual	kinesthetic	
Actual	Auditory	54.5 %	9.1 %	36.4 %	0.0 %	22
	Read/Write	8.3 %	45.8 %	41.7 %	4.2 %	24
	Visual	9.1 %	9.1 %	77.3 %	4.5 %	44
	kinesthetic	20.0 %	40.0 %	30.0 %	10.0 %	10
Σ		20	21	55	4	100

Gambar 6. Confusion Matrix of the KNN Algoritma

Gambar 5 menunjukkan bahwa 10% mahasiswa masuk kategori gaya belajar kinestetik, 77,3% mahasiswa gaya belajar visual, 45,8% kategori gaya belajar read/write dan 54,5% gaya belajar auditori, diprediksi dengan benar.

		Predicted				Σ
		Auditory	Read/Write	Visual	kinesthetic	
Actual	Auditory	31.8 %	4.5 %	63.6 %	0.0 %	22
	Read/Write	12.5 %	33.3 %	54.2 %	0.0 %	24
	Visual	6.8 %	13.6 %	79.5 %	0.0 %	44
	kinesthetic	30.0 %	40.0 %	30.0 %	0.0 %	10
Σ		16	19	65	0	100

Gambar 7. Confusion Matrix of the NN Algoritma

Gambar 6 menunjukkan bahwa tidak ada mahasiswa yang masuk kategori gaya belajar kinestetik, 79,5% mahasiswa gaya belajar visual, 33,3% kategori gaya belajar read/write dan 31,8% gaya belajar auditori, diprediksi dengan benar.

		Predicted				Σ
		Auditory	Read/Write	Visual	kinesthetic	
Actual	Auditory	31.8 %	0.0 %	54.5 %	13.6 %	22
	Read/Write	20.8 %	4.2 %	50.0 %	25.0 %	24
	Visual	18.2 %	0.0 %	70.5 %	11.4 %	44
	kinesthetic	0.0 %	0.0 %	40.0 %	60.0 %	10
Σ		20	1	59	20	100

Gambar 8. Confusion Matrix of the NB Algoritma

Gambar 7 menunjukkan bahwa 60% mahasiswa yang masuk kategori gaya belajar kinestetik, 70,5% mahasiswa gaya belajar visual, 4,2% kategori gaya belajar read/write dan 31,8% gaya belajar auditori, diprediksi dengan benar.

Akurat klasifikasi: CA adalah rasio yang benar dari prediksi (TP + TN) dengan jumlah instance total (TP + TN + FP + FN). Presisi: Presisi adalah rasio dari jumlah kasus positif yang diklasifikasikan dengan benar dibandingkan dengan jumlah kasus yang diprediksi positif secara keseluruhan. menghasilkan nilai dalam interval [0,1]. Recall: Jumlah instance positif yang diklasifikasikan dengan benar dibandingkan dengan total instance kelas aktualnya positif adalah recall. Tingkat positif sebenarnya adalah istilah lain untuk penarikan kembali. menghasilkan nilai dalam interval [0,1]. Kriteria-F (F1): Ada hubungan berlawanan antara presisi dan perolehan. Oleh karena itu, rata-rata harmonik dari kedua kriteria dihitung untuk mendapatkan hasil yang lebih akurat dan sensitif. Ini disebut kriteria-F.

3.3 Receiver operating characteristics (ROC) curve

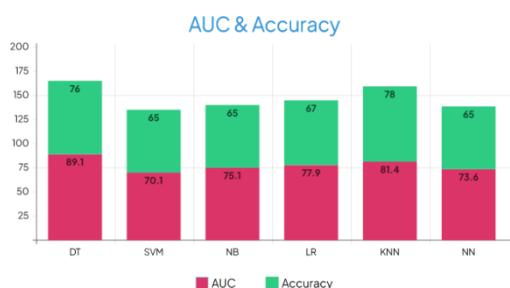
Kurva AUC-ROC digunakan untuk menilai kinerja masalah klasifikasi. Ini adalah metrik yang banyak digunakan untuk mengevaluasi kinerja algoritma pembelajaran mesin, terutama dalam kasus kumpulan data yang tidak seimbang. AUC-ROC juga menunjukkan seberapa baik model memprediksi.

3.4 AUC: Area under the ROC curve

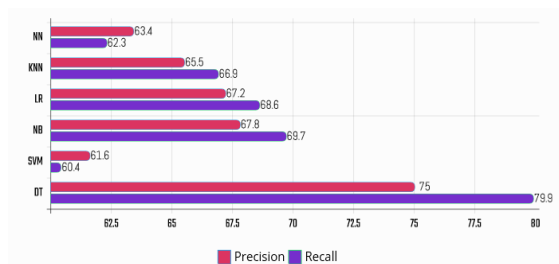
Algoritma pembelajaran mesin lebih baik dalam membedakan kelas tertentu jika area yang dicakupnya lebih besar. Tabel 4 menunjukkan nilai AUC, kriteria klasifikasi (CA), kriteria F-F1, presisi, dan nilai recall model. Nilai AUC ideal adalah 1.

Table 4. Nilai AUC, CA, F1, presisi dan recall model

Model	AUC	Classificati on Accuracy (CA)	F1	Preci sion	Reca ll
Decision Tree	0,891	0,760	0,648	0,750	0,799
Support Vector Machine	0,701	0,650	0,690	0,616	0,604
Naïve Bayes	0,751	0,650	0,697	0,678	0,697
Logistic Regression	0,779	0,670	0,698	0,672	0,686
K-Nearest Neighbour	0,814	0,780	0,660	0,655	0,669
Neural Network	0,736	0,650	0,653	0,634	0,623



Gambar 9. Nilai AUC dan Accuracy Algoritma Machine Learning



Gambar 10. Precision dan Recall Algoritma Machine Learning

AUC algoritma DT, SVM, NB, LR, KNN, dan NN adalah 0.89, 0.70, 0.75, 0.77, 0.81, dan 0.73. Akurasi klasifikasi algoritma DT, SVM, NB, LR, KNN, dan NN juga adalah 0.76, 0.65, 0.65, 0.67, 0.78, dan 0.65. Sebagai contoh, hasil menunjukkan bahwa algoritma DT memiliki tingkat akurasi 76%. Ini menunjukkan bahwa ada korelasi yang signifikan antara data aktual dan data prediksi. Hasilnya menunjukkan bahwa 76% sampel diklasifikasikan dengan benar.

Model baru yang didasarkan pada algoritma pembelajaran mesin diusulkan dalam penelitian ini untuk memprediksi gaya belajar mahasiswa teknik. Data yang digunakan adalah perkuliahan mata kuliah arsitektur komputer. Untuk memprediksi nilai gaya belajar mahasiswa, kinerja algoritma pembelajaran mesin seperti DT, SVM, NB, LR, KNN, dan NN dihitung dan dibandingkan. Dua parameter yang menjadi fokus dalam penelitian ini. Yang pertama adalah prediksi gaya belajar yang didasarkan pada data perkuliahan mahasiswa sebelumnya. Parameter kedua adalah perbandingan indikator kinerja algoritma pembelajaran mesin.

4. KESIMPULAN

Hasilnya menunjukkan bahwa model yang diusulkan mencapai akurasi klasifikasi sebesar 65–78%. Berdasarkan hasil ini, dapat dikatakan bahwa program studi, jenis kelamin, nilai akhir, dan predikat mahasiswa adalah beberapa prediktor yang penting untuk digunakan dalam memprediksi gaya belajar mahasiswa. Algoritma DT, SVM, NB, LR, KNN, dan NN sangat akurat untuk memprediksi gaya belajar mahasiswa.

Hasil penelitian ini akan dibandingkan dengan penelitian lainnya yang menggunakan data akademik mahasiswa, seperti penelitian yang dilakukan oleh (Janka Pecuchova, 2023),

melakukan prediksi awal mahasiswa yang berisiko putus sekolah dari mata kuliah pembelajaran. Kapasitas untuk memprediksi hasil akademik mahasiswa merupakan hal yang berharga bagi institusi pendidikan dimanapun yang bertujuan untuk meningkatkan kinerja dan ketekunan mahasiswa. Berdasarkan prediksi yang dihasilkan, mahasiswa yang diidentifikasi berisiko mengalami retensi atau kinerja akademik dapat diberikan dukungan lebih tepat waktu (Sumyea Helal, 2018). Menerapkan metode pembelajaran mesin untuk memprediksi performa mahasiswa secara keseluruhan dalam kursus. Model prediktif akhir yang disajikan dapat memprediksi dengan akurasi yang cukup tinggi mahasiswa yang berprestasi rendah dan tinggi dalam suatu kursus berdasarkan aktivitas dan data penilaian awal (Nouri, J., 2019). Penelitian yang menggunakan model regresi hybrid untuk mengoptimalkan keakuratan prediksi kinerja akademik mahasiswa, yang diukur sebagai nilai masa depan dalam mata pelajaran yang berbeda, dan pengklasifikasi multi-label yang dioptimalkan untuk memprediksi nilai kualitatif dan pengaruh tersebut. dari berbagai faktor yang berhubungan dengan prestasi mahasiswa yang diperoleh (Abdullah Alshantqi, 2020). model prediksi yang akurat dapat dibangun berdasarkan pola berurutan yang berasal dari data perilaku mahasiswa, yang mampu mengidentifikasi mahasiswa yang berkinerja buruk di awal perkuliahan (Galina Deeva, 2022).

Hasil penelitian ini juga menunjukkan bahwa algoritma DT dan KNN memiliki akurasi klasifikasi tertinggi sedangkan algoritma NN, NB, dan SVM memiliki akurasi klasifikasi terendah. Hasil ini menunjukkan bahwa, ketika algoritma pembelajaran mesin digunakan untuk memprediksi gaya belajar mahasiswa, algoritma DT dan KNN ternyata lebih akurat. Hasil ini dibandingkan dengan penelitian sebelumnya di mana algoritma pembelajaran mesin digunakan untuk memprediksi kinerja akademik berdasarkan berbagai variable. Penelitian yang dilakukan oleh (Delali Kwasi Dake, 2022) Membandingkan kinerja algoritma SVM, DT, perceptron multilayer, dan algoritma random forest. Algoritme pembelajaran mesin diuji akurasi, presisi, perolehan, pengukuran F, dan ROC menggunakan teknik validasi silang komparatif 10 kali lipat dan 5 kali lipat. Selain teknik validasi silang, mekanisme pemilihan fitur chi-kuadrat diterapkan untuk memahami waktu pelatihan dan akurasi algoritma random forest muncul sebagai algoritma dengan kinerja terbaik. Model pembelajaran mesin berdasarkan algoritma random forest (RF), support vector machine (SVM), dan algoritma K-nearest neighbour (KNN) digunakan untuk memprediksi kinerja akademik mahasiswa sekolah dasar. Performa prediksi masing-masing model pembelajaran mesin dievaluasi menggunakan akurasi dan validasi dalam sampel pengujian. Akurasi rata-rata model RF, SVM, dan KNN dalam

memprediksi kinerja akademik mahasiswa sekolah dasar dalam sampel pelatihan ditemukan algoritma RF lebih tinggi daripada akurasi algoritma SVM dan KNN (Keyun Xu , 2023).

Hasil menunjukkan bahwa model yang disarankan dapat memprediksi gaya belajar mahasiswa dengan 76% akurat. Berdasarkan hasil ini, dapat disimpulkan bahwa model ini memiliki kemampuan untuk memprediksi gaya belajar mahasiswa di masa mendatang, memberikan mahasiswa kesempatan untuk meninjau kembali aspek belajar mereka untuk menyesuakannya dengan gaya belajar yang telah ditetapkan untuk mencapai hasil belajar yang lebih baik. Penelitian yang dilakukan oleh Bello Ahmad Muhammad dkk (2024) mengusulkan pendekatan deteksi gaya belajar tambahan untuk pendidikan online dengan teknik penyematan grafik bipartit. Model gaya belajar Felder-Silverman (FSLSM) yang dipilih setiap kali untuk mengidentifikasi dan mengelompokkan pembelajar menggunakan algoritma pengelompokan k-means. Pendekatan yang diusulkan dapat diintegrasikan ke dalam sistem pendidikan yang berbeda.

Penelitian berikutnya mengidentifikasi atribut-atribut baru dan memperkecil atribut-atribut yang diidentifikasi sebelumnya, yang akan membantu mengidentifikasi gaya belajar pelajar. Penulis menerapkan algoritma klasifikasi dan membandingkan keakuratan berbagai algoritma pada kumpulan data. Berbagai pola menarik diamati dalam perilaku pelajar ketika mempelajari berbagai jenis konsep dalam situasi yang berbeda (Fareeha Rasyid, 2021). Penelitian yang mengusulkan arsitektur multi-Step hybrid baru berdasarkan sistem Colony sistem dan jaringan saraf tiruan untuk meningkatkan ketepatan identifikasi gaya pembelajaran. Dua varian berbeda diusulkan dan dievaluasi dengan data dari 75 mahasiswa; Hasilnya menunjukkan nilai presisi tinggi, mengungguli pendekatan otomatis yang ada untuk identifikasi gaya belajar (Jason Bernard, 2022).

DAFTAR PUSTAKA

- ABBAS, J. 2021. The impact of coronavirus (SARS-CoV2) epidemic on individuals mental health: The protective measures of Pakistan in managing and sustaining transmissible disease. *Psychiatra Danubina*. <https://doi.org/10.24869/PSYD.2020.472>
- AHMAD MUHAMMAD, B., QI, C., WU, Z., & KABIR AHMAD, H. 2022. GRL-LS: A learning style detection in online education using graph representation learning. In *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2022.117138>
- AISSAOUI, O. EL, EL MADANI, Y. E. A., OUGHDIR, L., & ALLIOUI, Y. EL. 2019. Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2019.01.012>
- ALSHANQITI, A., & NAMOUN, A. 2020. Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.3036572>
- AQEEL, M., REHNA, T., SHUJA, K. H., & ABBAS, J. 2022. Comparison of Students' Mental Wellbeing, Anxiety, Depression, and Quality of Life During COVID-19's Full and Partial (Smart) Lockdowns: A Follow-Up Study at a 5-Month Interval. *Frontiers in Psychiatry*. <https://doi.org/10.3389/fpsyt.2022.835585>
- ASISH, S. M., KULSHRESHTH, A. K., & BORST, C. W. 2022. Detecting distracted students in educational VR environments using machine learning on eye gaze data. *Computers and Graphics (Pergamon)*, 109, 75–87. <https://doi.org/10.1016/j.cag.2022.10.007>
- AULAKH, K., ROUL, R. K., & KAUSHAL, M. 2023. E-learning enhancement through educational data mining with Covid-19 outbreak period in backdrop: A review. In *International Journal of Educational Development*. <https://doi.org/10.1016/j.ijedudev.2023.102814>
- AZADI, N. A., ZIAPOUR, A., LEBNI, J. Y., IRANDOOST, S. F., ABBAS, J., & CHABOKSAVAR, F. 2021. The effect of education based on health belief model on promoting preventive behaviors of hypertensive disease in staff of the Iran University of Medical Sciences. *Archives of Public Health*. <https://doi.org/10.1186/s13690-021-00594-4>
- AZZI, I., JEGHAL, A., RADOUANE, A., YAHYAOUY, A., & TAIRI, H. 2020. A robust classification to predict learning styles in adaptive E-learning systems. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-019-09956-6>
- BERNARD, J., POPESCU, E., & GRAF, S. 2022. Improving online education through automatic learning style identification using a multi-step architecture with ant colony system and artificial neural networks. *Applied Soft Computing*. <https://doi.org/10.1016/j.asoc.2022.109779>
- BIAO, Z. Z. 2023. Design and realization of data mining simulation and methodological models. *Journal of King Saud University - Science*, 35(10), 102964. <https://doi.org/10.1016/j.jksus.2023.102964>
- BORHANI, K., & WONG, R. T. K. 2023. An

- artificial neural network for exploring the relationship between learning activities and students' performance. *Decision Analytics Journal*, 9(September), 100332. <https://doi.org/10.1016/j.dajour.2023.100332>
- CAPUANO, N., & TOTI, D. 2019. Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2018.03.034>
- CHRISTODOULOU, E., MA, J., COLLINS, G. S., STEYERBERG, E. W., VERBAKEL, J. Y., & VAN CALSTER, B. 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. In *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2019.02.004>
- DAKE, D. K., & BUABENG-ANDOH, C. 2022. Using Machine Learning Techniques to Predict Learner Drop-out Rate in Higher Educational Institutions. *Mobile Information Systems*. <https://doi.org/10.1155/2022/2670562>
- DEEVA, G., DE SMEDT, J., SAINT-PIERRE, C., WEBER, R., & DE WEERDT, J. 2022. Predicting student performance using sequence classification with time-based windows. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2022.118182>
- EL AISSAOUI, O., EL MADANI EL ALAMI, Y., OUGHDIR, L., & EL ALLIOUI, Y. 2018. Integrating web usage mining for an automatic learner profile detection: A learning styles-based approach. *2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018*. <https://doi.org/10.1109/ISACV.2018.8354021>
- EL AISSAOUI, O., OUGHDIR, L., & EL ALLIOUI, Y. 2022. A Literature Review on Student Modeling Purposes. https://doi.org/10.1007/978-3-030-90633-7_64
- FARHANA, S. 2021. Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm. *Procedia Computer Science*, 194, 224–228. <https://doi.org/10.1016/j.procs.2021.10.077>
- FERNANDES, E., HOLANDA, M., VICTORINO, M., BORGES, V., CARVALHO, R., & ERVEN, G. VAN. 2019. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*. <https://doi.org/10.1016/j.jbusres.2018.02.012>
- FERREIRA, L. D., SPADON, G., CARVALHO, A. C., & RODRIGUES, J. F. 2018. A comparative analysis of the automatic modeling of Learning Styles through Machine Learning techniques. *Proceedings - Frontiers in Education Conference, FIE*. <https://doi.org/10.1109/FIE.2018.8659191>
- GARG, A. (2020). Online Education: A Learner's Perspective During COVID-19. *Asia-Pacific Journal of Management Research and Innovation*. <https://doi.org/10.1177/2319510x211013594>
- HELAL, S., LI, J., LIU, L., EBRAHIMIE, E., DAWSON, S., MURRAY, D. J., & LONG, Q. 2018. Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knsys.2018.07.042>
- KLEMENT, M. 2014. How do my Students Study? An Analysis of Students' of Educational Disciplines Favorite Learning Styles According to VARK Classification. *Procedia - Social and Behavioral Sciences*. <https://doi.org/10.1016/j.sbspro.2014.04.326>
- LARA, J. A., LIZCANO, D., MARTÍNEZ, M. A., PAZOS, J., & RIERA, T. 2014. A system for knowledge discovery in e-learning environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA. *Computers and Education*. <https://doi.org/10.1016/j.compedu.2013.10.009>
- MADUSHANI, J. P. S. S., SANDAMAL, R. M. K., MEDDAGE, D. P. P., PASINDU, H. R., & GOMES, P. I. A. 2023. Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers. *Transportation Engineering*, 13(April), 100190. <https://doi.org/10.1016/j.treng.2023.100190>
- MUHAJIR, D., AKBAR, M., BAGASKARA, A., & VINARTI, R. 2021. Improving classification algorithm on education dataset using hyperparameter tuning. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2021.12.171>
- NOURI, J., SAQR, M., & FORS, U. 2019. Predicting performance of students in a flipped classroom using machine learning: Towards automated data-driven formative feedback. *ICSIT 2019 - 10th International Conference on Society and Information Technologies, Proceedings, January 2022*, 79–82.
- PARDAMEAN, B., SUPARYANTO, T., CENGGORO, T. W., SUDIGYO, D., & ANUGRAHANA, A. 2022. AI-Based Learning Style Prediction in Online Learning for Primary Education. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2022.3160177>
- PECUCHOVA, J., & DRLIK, M. 2023. Predicting Students at Risk of Early Dropping Out from Course Using Ensemble Classification

- Methods. *Procedia Computer Science*, 225, 3223–3232.
<https://doi.org/10.1016/j.procs.2023.10.316>
- PRADHAN, D., SAHOO, B., MISRA, B. B., & PADHY, S. 2020. A multiclass SVM classifier with teaching learning based feature subset selection for enzyme subclass classification. *Applied Soft Computing Journal*.
<https://doi.org/10.1016/j.asoc.2020.106664>
- PRASHANTH KUMAR, K. N., HARISH KUMAR, B. T., & BHUVANESH, A. 2024. Spectral clustering algorithm based web mining and quadratic support vector machine for learning style prediction in E-learning platform. *Measurement: Sensors*, 31(June 2023), 100962.
<https://doi.org/10.1016/j.measen.2023.100962>
- RASHEED, F., & WAHID, A. 2021. Learning style detection in E-learning systems using machine learning techniques. *Expert Systems with Applications*.
<https://doi.org/10.1016/j.eswa.2021.114774>
- RONG, L. J. 2022. Distance learning quality assessment of universities based on interval monotonic decision tree algorithm. *Computers and Electrical Engineering*.
<https://doi.org/10.1016/j.compeleceng.2022.108116>
- SHORFUZZAMAN, M., HOSSAIN, M. S., NAZIR, A., MUHAMMAD, G., & ALAMRI, A. 2019. Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Computers in Human Behavior*.
<https://doi.org/10.1016/j.chb.2018.07.002>
- SU, Z., MCDONNELL, D., WEN, J., KOZAK, M., ABBAS, J., ŠEGALO, S., LI, X., AHMAD, J., CHESHMEHZANGI, A., CAI, Y., YANG, L., & XIANG, Y. T. 2021. Mental health consequences of COVID-19 media coverage: the need for effective crisis communication practices. In *Globalization and Health*.
<https://doi.org/10.1186/s12992-020-00654-4>
- TROUSSAS, C., CHRYSAFIADI, K., & VIRVOU, M. 2021. Personalized tutoring through a stereotype student model incorporating a hybrid learning style instrument. *Education and Information Technologies*.
<https://doi.org/10.1007/s10639-020-10366-2>
- TUAHA, S., SIDDIQUI, I. F., & ALI ARAIN, Q. 2019. Analyzing Students' Academic Performance through Educational Data Mining. *3C Tecnología Glosas de Innovación Aplicadas a La Pyme*.
<https://doi.org/10.17993/3ctecno.2019.specialissue2.402-421>
- VIBERG, O., HATAKKA, M., BÄLTER, O., & MAVROUDI, A. 2018. The current landscape of learning analytics in higher education. In *Computers in Human Behavior*.
<https://doi.org/10.1016/j.chb.2018.07.027>
- WAHEED, H., HASSAN, S. U., ALJOHANI, N. R., HARDMAN, J., ALELYANI, S., & NAWAZ, R. 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*.
<https://doi.org/10.1016/j.chb.2019.106189>
- XU, K., & SUN, Z. 2023. Predicting academic performance associated with physical fitness of primary school students using machine learning methods. In *Complementary Therapies in Clinical Practice*.
<https://doi.org/10.1016/j.ctcp.2023.101736>
- YAĞCI, M. 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*.
<https://doi.org/10.1186/s40561-022-00192-z>

Halaman ini sengaja dikosongkan