

**KLASIFIKASI PERFORMA AKADEMIK SISWA
MENGUNAKAN PENDEKATAN *MACHINE LEARNING***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Kadek Aditya Ananta Wisnu Wardana

21.11.4392

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**KLASIFIKASI PERFORMA AKADEMIK SISWA
MENGUNAKAN PENDEKATAN *MACHINE LEARNING***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Kadek Aditya Ananta Wisnu Wardana

21.11.4392

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**Klasifikasi Performa Akademik Siswa Menggunakan Pendekatan
*Machine Learning***

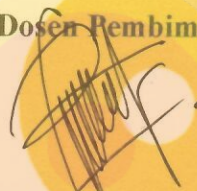
yang disusun dan diajukan oleh

Kadek Aditya Ananta Wisnu Wardana

21.11.4392

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Februari 2025

Dosen Pembimbing,



Robert Marco, S.T., M.T., Ph.D

NIK. 190302228

HALAMAN PENGESAHAN

SKRIPSI

**Klasifikasi Performa Akademik Siswa Menggunakan Pendekatan
*Machine Learning***

yang disusun dan diajukan oleh

Kadek Aditya Ananta Wisnu Wardana

21.11.4392

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Emigawaty, M.Kom
NIK. 190302226

Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Kadek Aditya Ananta Wisnu Wardana
NIM : 21.11.4392

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Teknik *Machine Learning* untuk Klasifikasi Performa Akademik Siswa Menggunakan Dataset Ibrahim Aljarah

Dosen Pembimbing : Robert Marco, S.T., M.T., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Februari 2025

Yang Menyatakan,



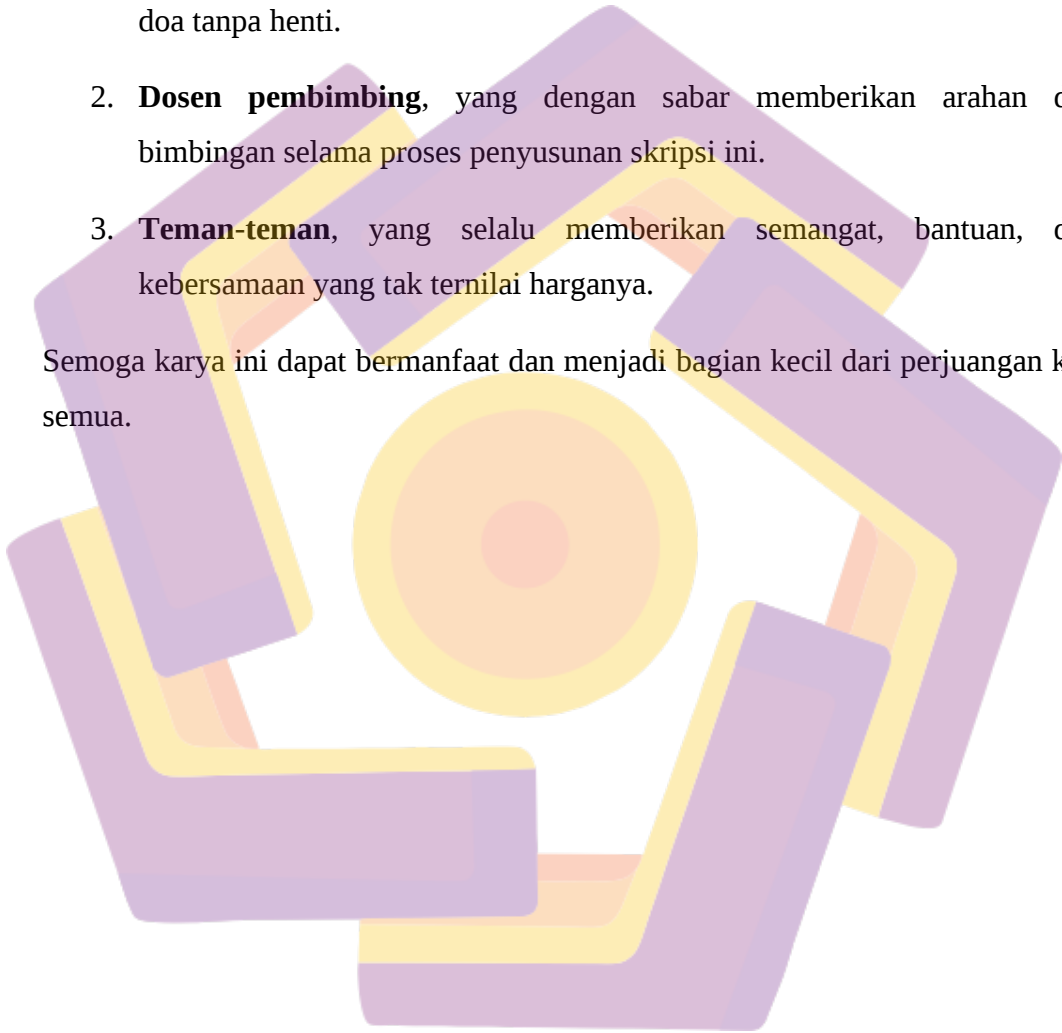
Kadek Aditya Ananta Wisnu Wardana

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan dengan penuh rasa terima kasih dan penghargaan kepada:

1. **Orang tua saya**, yang senantiasa memberikan dukungan, kasih sayang, dan doa tanpa henti.
2. **Dosen pembimbing**, yang dengan sabar memberikan arahan dan bimbingan selama proses penyusunan skripsi ini.
3. **Teman-teman**, yang selalu memberikan semangat, bantuan, dan kebersamaan yang tak ternilai harganya.

Semoga karya ini dapat bermanfaat dan menjadi bagian kecil dari perjuangan kita semua.



KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan judul “Klasifikasi Performa Akademik Siswa Menggunakan Pendekatan Machine Learning” sebagai salah satu syarat untuk memperoleh gelar S.Kom di Informatika, Universitas Amikom Yogyakarta.

Pada kesempatan ini, saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. **Robert Marco, S.T., M.T., Ph.D.**, yang telah memberikan bimbingan, arahan, dan dukungan selama penyusunan skripsi ini.
2. **Keluarga**, yang selalu memberikan dukungan dan doa.
3. Semua pihak yang telah membantu langsung maupun tidak langsung.

Saya menyadari skripsi ini masih memiliki kekurangan, oleh karena itu kritik dan saran sangat saya harapkan untuk perbaikan ke depan. Semoga skripsi ini bermanfaat bagi pengembangan ilmu pengetahuan dan masyarakat.

Yogyakarta, 18 Februari 2025

Penulis

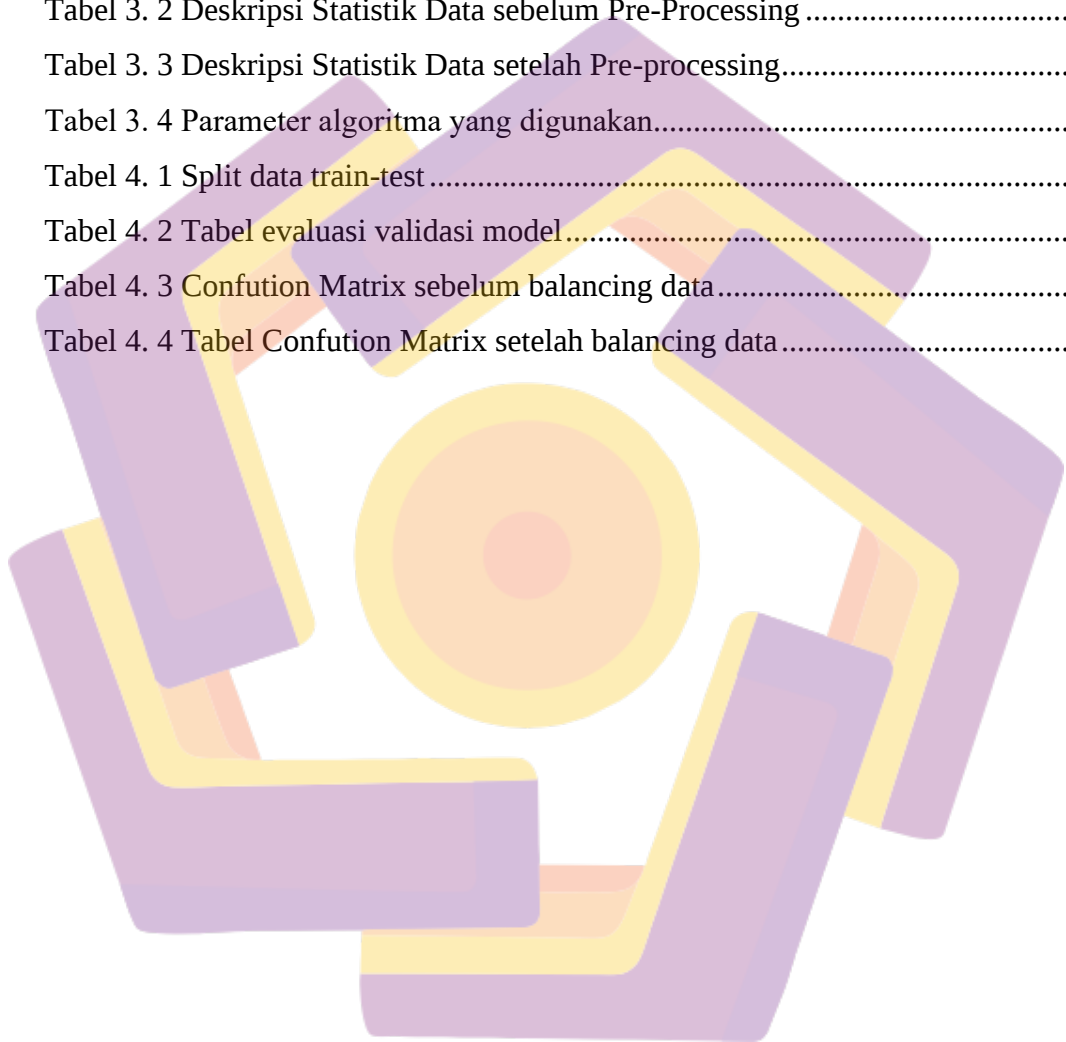
DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	10

2.2.1	Machine Learning	10
2.2.2	Klasifikasi	11
2.2.3	Algoritma Random Forest.....	12
2.2.4	Algoritma Extreme Gradient Boost	14
2.2.5	Algoritma Logistic Regression	18
2.2.6	Algoritma Decision Tree.....	22
2.2.7	Evaluasi.....	23
BAB III METODE PENELITIAN		28
3.1	Alur Penelitian	28
3.2	Alat dan Bahan.....	35
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Proses Pengumpulan Data.....	37
4.2	Exploratory Data Analysis (EDA).....	37
4.3	Preprocessing	40
4.4	Modelling.....	43
4.5	Evaluasi dan Perbandingan Kinerja Algoritma.....	44
BAB V PENUTUP		50
5.1	Kesimpulan	50
5.2	Saran	50
REFERENSI		51
LAMPIRAN.....		56

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Confution Matrix	25
Tabel 3. 1 Deskripsi Dataset	29
Tabel 3. 2 Deskripsi Statistik Data sebelum Pre-Processing	30
Tabel 3. 3 Deskripsi Statistik Data setelah Pre-processing.....	32
Tabel 3. 4 Parameter algoritma yang digunakan.....	33
Tabel 4. 1 Split data train-test	43
Tabel 4. 2 Tabel evaluasi validasi model.....	44
Tabel 4. 3 Confution Matrix sebelum balancing data.....	48
Tabel 4. 4 Tabel Confution Matrix setelah balancing data.....	49

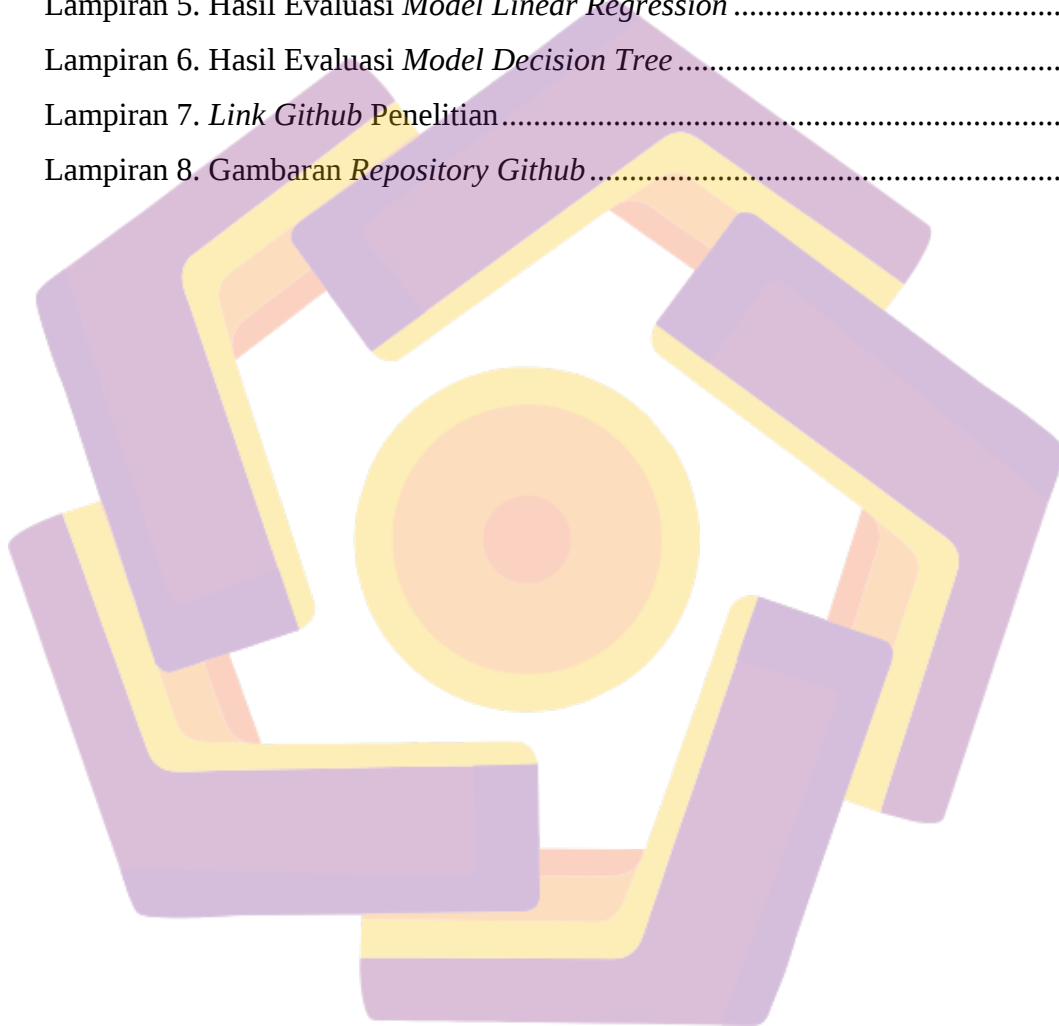


DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	28
Gambar 4. 1 Students Academic Performance Dataset	37
Gambar 4. 2 Informasi Umum Dataset	37
Gambar 4. 3 Ringkasan Statistik Dataset.....	38
Gambar 4. 4 Visualisasi Fitur Numerik	38
Gambar 4. 5 Visualisasi Fitur Kategorikal	40
Gambar 4. 6 Dataset sebelum dilakukan penanganan data duplikat.....	40
Gambar 4. 7 Dataset setelah dilakukan penanganan data duplikat	41
Gambar 4. 8 Dataset sebelum dilakukan label encoder	41
Gambar 4. 9 Dataset setelah dilakukan label encoder	41
Gambar 4. 10 Dataset sebelum dilakukan normalisasi data.....	42
Gambar 4. 11 Dataset setelah dilakukan normalisasi data.....	42
Gambar 4. 12 Fitur Class sebelum dilakukan penanganan ketidakseimbangan data	42
Gambar 4. 13 Fitur Class setelah dilakukan penanganan ketidakseimbangan data	43
Gambar 4. 14 Confution Matrix Random Forest	46
Gambar 4. 15 Confution Matrix Extreme Gradient Boost	46
Gambar 4. 16 Confution Matrix Logistic Regression	46
Gambar 4. 17 Confution Matrix Decision Tree	47

DAFTAR LAMPIRAN

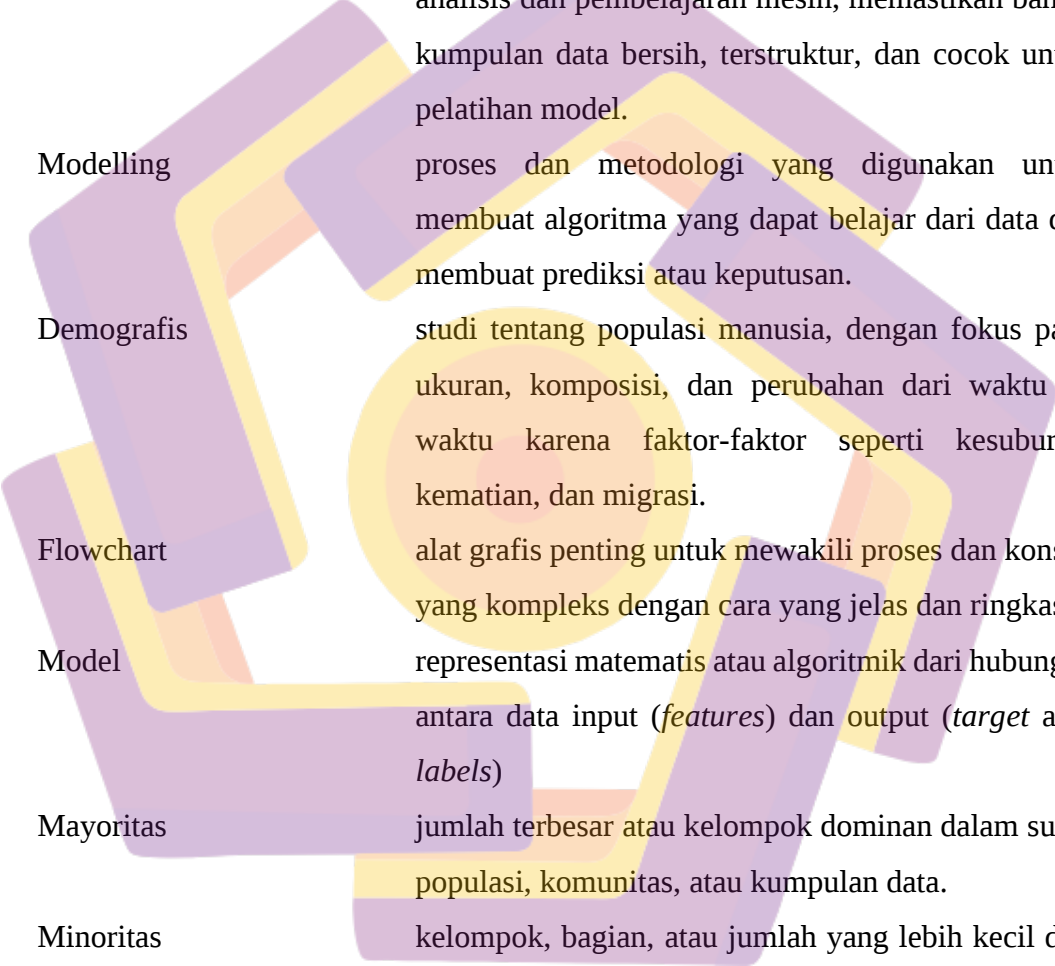
Lampiran 1. Tampilan <i>website kaggle</i>	56
Lampiran 2. <i>Dataset Performance Academic</i>	56
Lampiran 3. Hasil Evaluasi <i>Model Random Forest</i>	56
Lampiran 4. Hasil Evaluasi <i>Model Extreme Gradient Boost</i>	57
Lampiran 5. Hasil Evaluasi <i>Model Linear Regression</i>	57
Lampiran 6. Hasil Evaluasi <i>Model Decision Tree</i>	57
Lampiran 7. <i>Link Github Penelitian</i>	58
Lampiran 8. <i>Gambaran Repository Github</i>	58



DAFTAR LAMBANG DAN SINGKATAN

RF	Random Forest
XGBoost	Extreme Gradient Boost
LR	Logistic Regression
DT	Decision Tree
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
EDA	Exploratory Data Analysis
ROS	Random Over Sampling
CSV	Comma-Separated Values
AUC	Area Under the Curve

DAFTAR ISTILAH



Algoritma	urutan instruksi terstruktur yang dirancang untuk memecahkan masalah tertentu secara efisien dan akurat.
Preprocessing	langkah penting dalam mempersiapkan data untuk analisis dan pembelajaran mesin, memastikan bahwa kumpulan data bersih, terstruktur, dan cocok untuk pelatihan model.
Modelling	proses dan metodologi yang digunakan untuk membuat algoritma yang dapat belajar dari data dan membuat prediksi atau keputusan.
Demografis	studi tentang populasi manusia, dengan fokus pada ukuran, komposisi, dan perubahan dari waktu ke waktu karena faktor-faktor seperti kesuburan, kematian, dan migrasi.
Flowchart	alat grafis penting untuk mewakili proses dan konsep yang kompleks dengan cara yang jelas dan ringkas
Model	representasi matematis atau algoritmik dari hubungan antara data input (<i>features</i>) dan output (<i>target</i> atau <i>labels</i>)
Mayoritas	jumlah terbesar atau kelompok dominan dalam suatu populasi, komunitas, atau kumpulan data.
Minoritas	kelompok, bagian, atau jumlah yang lebih kecil dari keseluruhan dalam suatu populasi, komunitas, atau kumpulan data
Dataset	kumpulan data yang terorganisasi dalam bentuk tertentu dan digunakan untuk berbagai keperluan analisis, pelatihan model, atau pengambilan keputusan

INTISARI

Demografis sebagai salah satu factor utama yang mempengaruhi rendahnya performa akademik siswa dalam pendidikan. Beberapa studi sebelumnya yang telah dilakukan untuk mengatasi performa akademik siswa belum menghasilkan akurasi yang cukup memuaskan dengan masalah utama seperti ketidakseimbangan data dan outlier data. Oleh karena itu, digunakan beberapa teknik *Machine Learning*, yaitu *K-Nearest Neighbors*, *Support Vector Machine*, *Random Forest*, dan *Extreme Gradient Boost* dalam mengklasifikasikan performa akademik siswa berdasarkan dataset Ibrahim Aljarah. Metode penelitian yang digunakan dalam penelitian ini meliputi *preprocessing* data, serta pelatihan dan evaluasi model *Machine Learning*. Dataset diproses dengan teknik normalisasi dan mengatasi ketidakseimbangan data. Model *Machine Learning* yang diuji meliputi *K-Nearest Neighbors*, *Support Vector Machine*, *Random Forest*, dan *Extreme Gradient Boost*. Setiap model dilatih menggunakan teknik *cross-validation* dan dievaluasi berdasarkan metrik akurasi, presisi, recall, f1-score, dan AUC untuk menentukan performa terbaik dalam klasifikasi performa akademik siswa. Hasil penelitian menunjukkan bahwa *random forest* menjadi algoritma dengan nilai akurasi paling tinggi yaitu dengan akurasi sebesar 0,86, recall sebesar 0,86, presisi sebesar 0,88, f1-score sebesar 0,86, dan AUC sebesar 0,9561. Penelitian ini memberikan kontribusi bagi institusi pendidikan dalam meningkatkan proses evaluasi akademik berbasis data. Penelitian ini dapat digunakan untuk dataset lain dan menerapkan algoritma canggih seperti *deep learning*.

Kata kunci: Performa Akademik Siswa, *K-Nearest Neighbors*, *Support Vector Machine*, *Random Forest*, dan *Extreme Gradient Boost*

ABSTRACT

Demographics as one of the main factors affecting students' low academic performance in education. Several previous studies conducted to address students' academic performance have not achieved satisfactory accuracy due to major issues such as data imbalance and outliers. Therefore, several Machine Learning techniques, namely K-Nearest Neighbors, Support Vector Machine, Random Forest, and Extreme Gradient Boost, are used to classify students' academic performance based on the Ibrahim Aljarah dataset. The research methodology used in this study includes data preprocessing, as well as training and evaluation of Machine Learning models. The dataset is processed using normalization techniques and methods to address data imbalance. The Machine Learning models tested include K-Nearest Neighbors, Support Vector Machine, Random Forest, and Extreme Gradient Boost. Each model is trained using cross-validation techniques and evaluated based on accuracy, precision, recall, f1-score, and AUC metrics to determine the best performance in classifying students' academic performance. The results of the study indicate that Random Forest is the algorithm with the highest accuracy, achieving an accuracy of 0.86, recall of 0.86, precision of 0.88, f1-score of 0.86, and an AUC of 0.9561. This study contributes to educational institutions in improving data-driven academic evaluation processes. This research can be applied to other datasets and implement advanced algorithms such as deep learning.

Keyword: *Students Performance Academic, K-Nearest Neighbors, Support Vector Machine, Random Forest, dan Extreme Gradient Boost*