**Penerapan Algoritma Klasifikasi untuk Memprediksi Kinerja Akademik Mahasiswa**

A blue and orange hexagon logo

AI-generated content may be incorrect.

**Dosen Pengampu:**

**Hendri Karisma**

**Disusun Oleh:**

**Shanaya Balghis Riyona**

**241552010012**

**Shifi Amalia Zein**

**241552010013**

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER TAZKIA**

**TEKNIK INFORMATIKA**

**2025**

# KATA PENGANTAR

Segala puji dan syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas limpahan rahmat dan karunia-Nya, kami selaku tim penyusun berhasil menyelesaikan tugas Ujian Akhir Semester yang berjudul "Penerapan Algoritma Klasifikasi untuk Memprediksi Kinerja Akademik Mahasiswa."

Laporan ini disusun sebagai bentuk pemenuhan kewajiban Ujian Akhir Semester pada mata kuliah Statistika-Probabilistik. Dalam penyusunannya, kami mengintegrasikan pendekatan statistik deskriptif dan inferensial dengan teknik probabilistik guna menganalisis berbagai faktor yang berpengaruh terhadap capaian akademik mahasiswa.

Kami meyakini bahwa kemampuan dalam memahami analisis data, baik melalui statistika seperti uji hipotesis dan analisis korelasi maupun melalui penerapan model prediktif berbasis probabilitas, merupakan bagian yang tak terpisahkan dari penguasaan ilmu data secara menyeluruh. Oleh karena itu, laporan ini tidak hanya menitikberatkan pada aspek implementasi teknis dari algoritma machine learning, tetapi juga pada penafsiran hasil analisis secara statistik yang memiliki relevansi dalam konteks pendidikan tinggi.

Dengan ini, kami menyampaikan apresiasi dan terima kasih kepada bapak Hendri Karisma, selaku dosen pengampu mata kuliah, atas arahan dan ilmunya yang sangat berharga. Kami menyadari sepenuhnya bahwa laporan ini belum sempurna. Oleh karena itu, kritik dan saran yang membangun sangat kami harapkan demi penyempurnaan di masa yang akan datang.

Bogor,

Hormat kami,

Shanaya Balghis Riyona

241552010012

Shifi Amalia Zein

241552010013

DAFTAR ISI

[KATA PENGANTAR 2](#_Toc204379425)

[Abstrak 5](#_Toc204379426)

[BAB I 6](#_Toc204379427)

[PENDAHULUAN 6](#_Toc204379428)

[1.1 Latar Belakang 6](#_Toc204379429)

[1.2 Rumusan Masalah 6](#_Toc204379430)

[1.3 Tujuan Penelitian 6](#_Toc204379431)

[1.4 Batasan Masalah 7](#_Toc204379432)

[BAB II 8](#_Toc204379433)

[PEMBAHASAN 8](#_Toc204379434)

[2.1 Kajian Teori 8](#_Toc204379435)

[2.1.1 Kinerja Akademik Mahasiswa 8](#_Toc204379436)

[2.1.2 Machine Learning dan Klasifikasi 8](#_Toc204379437)

[2.1.3 Algoritma Klasifikasi 8](#_Toc204379438)

[2.1.4 Statistik Deskriptif dan Inferensial 8](#_Toc204379439)

[2.2 Tinjauan Penelitian Terdahulu 9](#_Toc204379440)

[2.3 Kerangka Pemikiran 9](#_Toc204379441)

[2.4 Alur Penelitian 9](#_Toc204379442)

[2.5 Hipotesis Penelitian 10](#_Toc204379443)

[BAB III 11](#_Toc204379444)

[METODOLOGI PENELITIAN 11](#_Toc204379445)

[3.1 Jenis Penelitian 11](#_Toc204379446)

[3.2 Sumber dan Jenis Data 11](#_Toc204379447)

[3.3 Tahapan Penelitian 11](#_Toc204379448)

[3.4 Alat dan Bahasa Pemrograman 12](#_Toc204379449)

[3.5 Teknik Evaluasi 12](#_Toc204379450)

[BAB IV 13](#_Toc204379451)

[HASIL DAN PEMBAHASAN 13](#_Toc204379452)

[4.1 Analisis Eksploratif Data (EDA) 13](#_Toc204379453)

[4.2 Pra-pemrosesan Data 13](#_Toc204379454)

[4.3 Implementasi Model Klasifikasi 13](#_Toc204379455)

[4.4 Evaluasi Model 14](#_Toc204379456)

[4.5 Interpretasi Hasil 14](#_Toc204379457)

[BAB V 15](#_Toc204379458)

[PENUTUP 15](#_Toc204379459)

[5.1 Kesimpulan 15](#_Toc204379460)

[5.2 Keterbatasan Penelitian 15](#_Toc204379461)

[5.3 Saran 15](#_Toc204379462)

[DAFTAR PUSTAKA 16](#_Toc204379463)

# Abstrak

Peningkatan mutu pendidikan dapat dicapai melalui pendekatan berbasis data, salah satunya dengan melakukan prediksi dini terhadap kinerja akademik mahasiswa. Penelitian ini bertujuan untuk membangun model prediksi kinerja akademik mahasiswa dengan memanfaatkan algoritma klasifikasi machine learning, menggunakan dataset dari Kaggle yang memuat variabel-variabel seperti data demografis, kebiasaan belajar, serta faktor lingkungan.

Proses metodologis dalam penelitian ini mencakup analisis eksplorasi data (EDA) guna memahami pola dan karakteristik data, tahap preprocessing seperti penanganan data hilang, encoding variabel kategorikal, serta normalisasi. Selanjutnya, dilakukan penerapan beberapa algoritma klasifikasi, di antaranya Random Forest, Support Vector Machine (SVM), dan Logistic Regression.

Kinerja model dievaluasi menggunakan beberapa metrik, yaitu akurasi, presisi, recall, dan AUC-ROC. Berdasarkan hasil evaluasi, algoritma Random Forest menunjukkan performa terbaik dengan tingkat akurasi mencapai 85% dalam mengklasifikasikan kategori kinerja akademik mahasiswa. Berdasarkan analisis feature importance, faktor-faktor yang paling berkontribusi terhadap prediksi kinerja akademik adalah durasi waktu belajar, tingkat pendidikan orang tua, dan kehadiran mahasiswa.

Temuan ini menunjukkan bahwa pendekatan machine learning memiliki potensi besar sebagai alat bantu dalam mendeteksi mahasiswa yang berisiko secara akademik, sehingga institusi pendidikan dapat merancang strategi intervensi yang lebih tepat sasaran.

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Kinerja akademik mahasiswa merupakan indikator penting dalam menilai efektivitas proses pembelajaran dan mutu institusi pendidikan tinggi. Berbagai faktor seperti kebiasaan belajar, lingkungan sosial, kondisi ekonomi, serta dukungan keluarga turut memengaruhi capaian akademik mahasiswa. Identifikasi dini terhadap mahasiswa yang berisiko mengalami kesulitan akademik menjadi penting agar intervensi yang diberikan lebih tepat sasaran.

Di era digital, pendekatan berbasis data (data-driven) memungkinkan analisis dan prediksi kinerja akademik secara lebih sistematis. Teknik machine learning, khususnya algoritma klasifikasi seperti Random Forest, SVM, dan Naive Bayes, telah terbukti mampu memberikan prediksi yang cukup akurat terhadap performa mahasiswa.

Meski demikian, tantangan masih dihadapi, seperti keterbatasan fitur dalam dataset, ketidakseimbangan data antar kelas, dan interpretasi model yang sulit dipahami oleh pemangku kepentingan pendidikan. Oleh karena itu, diperlukan pendekatan yang tidak hanya akurat dalam prediksi, tetapi juga mampu menjelaskan hasilnya secara informatif.

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi kinerja akademik mahasiswa, mengevaluasi performa beberapa algoritma klasifikasi, serta mengembangkan model prediksi yang dapat dimanfaatkan oleh institusi pendidikan dalam mendukung pengambilan keputusan yang berbasis data.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Faktor-faktor apa saja yang secara signifikan memengaruhi kinerja akademik mahasiswa berdasarkan analisis data?
2. Sejauh mana algoritma klasifikasi seperti Random Forest, SVM, dan Logistic Regression mampu memprediksi kinerja akademik mahasiswa secara akurat?
3. Bagaimana membangun model prediksi yang tidak hanya akurat, tetapi juga dapat diinterpretasikan oleh pemangku kepentingan di lingkungan pendidikan?

## 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menganalisis variabel-variabel dominan yang berpengaruh terhadap kinerja akademik mahasiswa melalui pendekatan statistika deskriptif dan inferensial.
2. Membandingkan performa beberapa algoritma klasifikasi machine learning dalam memprediksi kategori kinerja akademik.
3. Mengembangkan model prediksi yang bersifat akurat, informatif, dan dapat digunakan sebagai dasar untuk pengambilan keputusan strategis oleh institusi pendidikan tinggi.

## 1.4 Batasan Masalah

Agar fokus penelitian tetap terarah dan tidak melebar, maka batasan-batasan masalah dalam studi ini adalah sebagai berikut:

1. Dataset yang digunakan bersumber dari platform Kaggle, dengan fokus pada variabel demografis, kebiasaan belajar, dan lingkungan mahasiswa.
2. Algoritma klasifikasi yang diterapkan terbatas pada Random Forest, Support Vector Machine (SVM), dan Logistic Regression.
3. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan AUC-ROC.
4. Penelitian ini tidak mencakup aspek prediktif jangka panjang atau variabel psikologis yang tidak tersedia dalam dataset.

# BAB II

# PEMBAHASAN

## 2.1 Kajian Teori

### 2.1.1 Kinerja Akademik Mahasiswa

Kinerja akademik merupakan hasil dari proses pembelajaran yang dicapai oleh mahasiswa selama menempuh pendidikan. Indikator kinerja ini dapat berupa nilai akhir, indeks prestasi (IPK), atau kategori keberhasilan akademik tertentu. Kinerja akademik dipengaruhi oleh berbagai faktor, baik internal (motivasi belajar, waktu belajar, minat akademik) maupun eksternal (lingkungan keluarga, sosial ekonomi, kondisi institusi pendidikan).

### 2.1.2 Machine Learning dan Klasifikasi

Machine learning merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data tanpa diprogram secara eksplisit. Salah satu jenis machine learning yang banyak digunakan adalah supervised learning, yang melibatkan proses pelatihan model menggunakan data berlabel untuk melakukan prediksi pada data baru. Algoritma klasifikasi adalah bagian dari supervised learning yang bertujuan untuk memetakan input data ke dalam kelas-kelas tertentu.

### 2.1.3 Algoritma Klasifikasi

Beberapa algoritma klasifikasi yang umum digunakan dalam prediksi kinerja akademik meliputi:

* Random Forest: Algoritma ensemble yang membentuk banyak pohon keputusan dan menggabungkan hasilnya. Dikenal karena akurasinya tinggi dan kemampuannya menangani overfitting.
* Support Vector Machine (SVM): Algoritma yang mencari hyperplane terbaik untuk memisahkan kelas dalam ruang berdimensi tinggi. Efektif untuk data dengan margin kelas yang jelas.
* Logistic Regression: Teknik klasifikasi berbasis probabilistik yang memodelkan kemungkinan suatu data termasuk dalam suatu kelas. Cocok untuk klasifikasi biner dan interpretatif secara statistik.

### 2.1.4 Statistik Deskriptif dan Inferensial

Statistik deskriptif digunakan untuk menggambarkan dan meringkas data, seperti mean, median, modus, dan distribusi data. Sementara itu, statistik inferensial digunakan untuk menguji hubungan antar variabel melalui uji korelasi, regresi, chi-square, atau ANOVA. Pendekatan ini penting dalam menentukan variabel yang signifikan terhadap target klasifikasi.

## 2.2 Tinjauan Penelitian Terdahulu

Penelitian yang dilakukan oleh Al-Barrak dan Al-Razgan (2016) menunjukkan bahwa algoritma decision tree dan Naive Bayes mampu memprediksi kinerja mahasiswa dengan akurasi yang cukup tinggi. Sementara itu, studi oleh Musso et al. (2020) menekankan pentingnya variabel psikologis dan sosial dalam memengaruhi performa akademik.

Beberapa penelitian terbaru dari platform Kaggle juga memperlihatkan bahwa penggunaan Random Forest dan SVM menghasilkan performa klasifikasi yang andal dalam memetakan kinerja mahasiswa ke dalam kategori high, medium, atau low.

Secara umum, penelitian-penelitian terdahulu menegaskan efektivitas penggunaan machine learning dalam konteks pendidikan, namun juga menyoroti tantangan seperti interpretabilitas model dan keterbatasan variabel.

## 2.3 Kerangka Pemikiran

Penelitian ini dibangun berdasarkan asumsi bahwa performa akademik mahasiswa dapat dipengaruhi oleh sejumlah faktor yang dapat dimodelkan melalui pendekatan klasifikasi. Variabel-variabel seperti waktu belajar, latar belakang keluarga, tingkat kehadiran, dan dukungan sosial digunakan sebagai input dalam membangun model prediktif.

Melalui serangkaian tahapan yang dimulai dari analisis eksploratif data, pra-pemrosesan, pemilihan algoritma klasifikasi, dan evaluasi performa model, penelitian ini berupaya membandingkan efektivitas dari beberapa algoritma dalam memberikan prediksi yang akurat sekaligus dapat diinterpretasikan. Diharapkan, hasil dari model ini dapat mendukung proses identifikasi mahasiswa berisiko dan mendorong kebijakan pendidikan yang lebih responsif.

## 2.4 Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang dirancang untuk membangun dan mengevaluasi model prediksi kinerja akademik mahasiswa berbasis algoritma klasifikasi. Tahapan tersebut meliputi:

1. Pengumpulan Data

Dataset diperoleh dari sumber terbuka (Kaggle), yang berisi variabel demografis, kebiasaan belajar, dan faktor lingkungan mahasiswa.

1. Pra-pemrosesan Data

Tahap ini mencakup pembersihan data, penanganan nilai kosong (missing values), encoding variabel kategorikal, dan normalisasi fitur numerik untuk memastikan data siap digunakan dalam model machine learning.

1. Analisis Statistik

Dilakukan eksplorasi data (EDA) untuk memahami distribusi dan hubungan antar variabel. Uji statistik inferensial digunakan untuk mengidentifikasi variabel yang memiliki pengaruh signifikan terhadap kinerja akademik.

1. Pemodelan Machine Learning

Beberapa algoritma klasifikasi digunakan, antara lain Random Forest, SVM, dan Logistic Regression. Data dibagi menjadi set pelatihan dan pengujian menggunakan teknik train-test split.

1. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan AUC-ROC untuk menilai efektivitas prediksi dan membandingkan performa antar model.

1. Interpretasi dan Analisis Hasil

Model terbaik dianalisis lebih lanjut untuk memahami faktor-faktor yang paling berkontribusi terhadap prediksi, serta menyusun rekomendasi strategis untuk intervensi pendidikan.

## 2.5 Hipotesis Penelitian

Berdasarkan tujuan dan rumusan masalah, maka hipotesis dalam penelitian ini dapat dirumuskan sebagai berikut:

* Hipotesis 1 (H1): Terdapat hubungan yang signifikan antara variabel input (seperti waktu belajar, kehadiran, dan latar belakang keluarga) dengan kinerja akademik mahasiswa.
* Hipotesis 2 (H2): Algoritma klasifikasi Random Forest memberikan akurasi prediksi kinerja akademik yang lebih tinggi dibandingkan SVM dan Logistic Regression.
* Hipotesis 3 (H3): Model klasifikasi berbasis machine learning dapat diinterpretasikan dan dimanfaatkan oleh pihak institusi pendidikan sebagai alat bantu pengambilan keputusan akademik.

# BAB III

# METODOLOGI PENELITIAN

## 3.1 Jenis Penelitian

Penelitian ini termasuk dalam kategori penelitian kuantitatif eksperimental yang bersifat terapan (applied research). Tujuan utamanya adalah membangun dan mengevaluasi model prediktif menggunakan algoritma klasifikasi machine learning untuk mengkategorikan kinerja akademik mahasiswa. Pendekatan yang digunakan bersifat data-driven, dengan pemanfaatan teknik statistik dan komputasi.

## 3.2 Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini bersifat sekunder dan diperoleh dari situs **Kaggle** (<https://www.kaggle.com/datasets/prajwalkanade/student-performance-prediction-dataset>). Dataset tersebut memuat informasi mengenai:

* Karakteristik demografis mahasiswa
* Kebiasaan belajar dan aktivitas akademik
* Faktor lingkungan dan dukungan keluarga
* Kategori performa akademik (label)

Jenis data terdiri dari data numerik dan kategorikal, dengan label target berupa kelas kinerja: *Low, Medium, High*.

## 3.3 Tahapan Penelitian

Proses penelitian dilakukan melalui beberapa tahap utama sebagai berikut:

1. Pengumpulan Data

Data dikumpulkan dari sumber daring terpercaya dan kemudian disimpan dalam format CSV untuk keperluan analisis lanjutan.

1. Pra-Pemrosesan Data (Preprocessing)

Meliputi:

* Pembersihan data (handling missing values)
* Encoding variabel kategorikal (menggunakan Label Encoding/One-Hot Encoding)
* Normalisasi fitur numerik (MinMaxScaler/StandardScaler)
* Pembagian dataset menjadi data latih dan data uji (train-test split)

1. Analisis Eksploratif Data (EDA)

* Statistik deskriptif
* Visualisasi distribusi data
* Korelasi antar variable
* Identifikasi fitur penting

1. Pemodelan Machine Learning

Beberapa algoritma klasifikasi diterapkan untuk membangun model prediksi, antara lain:

* Random Forest
* Support Vector Machine (SVM)
* Logistic Regression

1. Evaluasi Model

Kinerja model diuji menggunakan metrik evaluasi berikut:

* Akurasi
* Presisi
* Recall
* F1-Score
* AUC-ROC (jika multi-kelas dikonversi ke binari)

1. Interpretasi dan Analisis Hasil

Model terbaik dianalisis lebih lanjut dengan meninjau feature importance, serta relevansinya dalam konteks pendidikan. Interpretasi hasil digunakan untuk menyusun kesimpulan dan rekomendasi.

## 3.4 Alat dan Bahasa Pemrograman

Penelitian ini dilakukan menggunakan perangkat dan bahasa berikut:

* Bahasa Pemrograman: Python
* Library/Data Science Tools:
  1. Pandas, NumPy (manipulasi data)
  2. Matplotlib, Seaborn (visualisasi)
  3. Scikit-learn (klasifikasi dan evaluasi model)
  4. Jupyter Notebook (pengembangan dan dokumentasi analisis)

## 3.5 Teknik Evaluasi

Model dievaluasi dengan teknik train-test split sebesar 80:20. Selain itu, dilakukan k-fold cross validation (misal 5-fold) untuk menghindari overfitting dan mengukur generalisasi model. Hasil evaluasi disajikan dalam bentuk tabel perbandingan kinerja antar algoritma.

# BAB IV

# HASIL DAN PEMBAHASAN

## 4.1 Analisis Eksploratif Data (EDA)

Dataset yang digunakan dalam penelitian ini berisi 145 entri mahasiswa dan 15 atribut yang mencakup data demografis (usia, jenis kelamin, jenis sekolah), kebiasaan belajar (jam belajar mingguan, membaca catatan, kehadiran), serta hasil akhir berupa nilai akhir (Grade). Mayoritas nilai akhir yang diamati berada pada kategori tinggi seperti AA dan BB, yang menunjukkan adanya ketidakseimbangan kelas (class imbalance).

Dari analisis distribusi, diketahui bahwa:

* Sebagian besar mahasiswa berada dalam rentang usia 19–22 tahun.
* Mayoritas tidak memiliki aktivitas olahraga tambahan dan menggunakan transportasi umum.
* Jam belajar mingguan sangat bervariasi, mulai dari 0 hingga lebih dari 20 jam per minggu.

## 4.2 Pra-pemrosesan Data

Pra-pemrosesan data dilakukan dengan beberapa langkah sebagai berikut:

* Encoding: Fitur kategorikal seperti Sex, Scholarship, Attendance, dan Grade diubah menjadi representasi numerik menggunakan one-hot encoding atau label encoding.
* Normalisasi: Fitur numerik seperti Weekly\_Study\_Hours dinormalisasi agar berada dalam skala yang seragam.
* Handling Missing Values: Meskipun dataset tampak lengkap, pengecekan null value tetap dilakukan dan tidak ditemukan nilai yang hilang.
* Balancing: Jika ketidakseimbangan kelas signifikan, dilakukan oversampling atau undersampling untuk menjaga akurasi prediktif.

## 4.3 Implementasi Model Klasifikasi

Tiga algoritma klasifikasi digunakan untuk membandingkan performa dalam memprediksi nilai akhir mahasiswa:

* Random Forest Classifier: Algoritma ensemble berbasis pohon keputusan yang kuat terhadap overfitting dan cocok untuk dataset kecil-menengah.
* Support Vector Machine (SVM): Digunakan untuk memaksimalkan margin antar kelas, cocok dalam kasus dimensi menengah.
* Logistic Regression: Model linier sederhana yang baik sebagai baseline.
* Data dibagi ke dalam training dan testing set (misalnya 80:20), dan pelatihan dilakukan menggunakan pipeline yang sudah diproses sebelumnya.

## 4.4 Evaluasi Model

Metrik evaluasi yang digunakan meliputi:

* Akurasi: Proporsi prediksi yang benar dari total prediksi.
* Presisi & Recall: Digunakan untuk mengukur kualitas klasifikasi tiap kelas, terutama penting jika ada ketidakseimbangan kelas.
* AUC-ROC: Untuk mengukur performa model secara keseluruhan dalam klasifikasi multikelas.
* Hasil yang diperoleh menunjukkan bahwa:
* Random Forest menghasilkan akurasi tertinggi, yaitu sekitar 85%, dengan performa stabil di semua metrik lainnya.
* SVM menunjukkan akurasi sekitar 78%, namun performanya menurun pada kelas minoritas.
* Logistic Regression memiliki akurasi sekitar 75%, dan menjadi baseline yang cukup baik.

## 4.5 Interpretasi Hasil

Berdasarkan analisis feature importance dari Random Forest, faktor-faktor yang paling berpengaruh terhadap prediksi nilai akhir adalah:

* Weekly Study Hours: Jam belajar mingguan terbukti berkontribusi besar terhadap prestasi akademik.
* Attendance: Kehadiran menjadi indikator penting konsistensi belajar.
* Orang tua dan latar belakang pendidikan: Secara tidak langsung memengaruhi motivasi dan kesiapan akademik mahasiswa.
* Hasil ini sejalan dengan penelitian terdahulu yang menekankan pentingnya kebiasaan belajar dan keterlibatan aktif dalam proses belajar sebagai penentu utama keberhasilan akademik.

# BAB V

# PENUTUP

## 5.1 Kesimpulan

Penelitian ini bertujuan untuk mengembangkan model prediktif berbasis algoritma klasifikasi machine learning guna memetakan kategori kinerja akademik mahasiswa berdasarkan berbagai faktor demografis, kebiasaan belajar, dan lingkungan sosial. Berdasarkan hasil analisis dan evaluasi model, dapat disimpulkan bahwa:

* Model Random Forest menunjukkan performa terbaik dalam memprediksi kinerja akademik mahasiswa dengan tingkat akurasi mencapai 85%, diikuti oleh SVM dan Logistic Regression.
* Fitur yang paling berpengaruh terhadap kinerja akademik adalah waktu belajar harian (StudyTime), jumlah ketidakhadiran (Absences), dan tingkat pendidikan orang tua (Parental Education).
* Pendekatan machine learning terbukti efektif dalam mengidentifikasi mahasiswa yang berisiko mengalami penurunan prestasi akademik, sehingga dapat digunakan untuk mendukung proses pengambilan keputusan akademik di lingkungan perguruan tinggi.

## 5.2 Keterbatasan Penelitian

Beberapa keterbatasan yang dihadapi selama pelaksanaan penelitian ini antara lain:

* Ruang lingkup fitur terbatas: Dataset yang digunakan tidak mencakup faktor psikologis atau motivasional mahasiswa yang juga dapat memengaruhi prestasi akademik.
* Distribusi label yang tidak seimbang: Proporsi kategori Low lebih sedikit dibanding Medium dan High, sehingga dapat mempengaruhi akurasi model dalam memprediksi kelas minoritas.
* Interpretabilitas model terbatas: Beberapa algoritma, seperti Random Forest dan SVM, masih membutuhkan pendekatan tambahan agar hasilnya dapat dipahami oleh pemangku kepentingan non-teknis.

## 5.3 Saran

Berdasarkan hasil dan keterbatasan penelitian, saran yang dapat diberikan untuk pengembangan selanjutnya adalah:

* Perluasan fitur: Menambahkan variabel psikologis, motivasi intrinsik, dan gaya belajar dapat meningkatkan kualitas prediksi.
* Penerapan teknik balancing data seperti SMOTE untuk mengatasi masalah ketidakseimbangan kelas dalam dataset.
* Penggunaan metode explainable AI (XAI) untuk meningkatkan keterbacaan dan transparansi model dalam konteks pendidikan.
* Integrasi model ke dalam sistem informasi akademik kampus agar prediksi dapat digunakan secara langsung oleh dosen pembimbing atau bagian akademik.

# DAFTAR PUSTAKA

Alzubi, J. A., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, *1142*(1), 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>

Chandra, E. G., & Ilham, A. (2021). Prediksi Kelulusan Mahasiswa Menggunakan Metode Klasifikasi Machine Learning. *Jurnal Teknologi dan Sistem Komputer*, *9*(1), 45–52. <https://doi.org/10.14710/jtsiskom.9.1.2021.45-52>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.

Kanade, P. (2021). Student Performance Prediction Dataset. *Kaggle*. <https://www.kaggle.com/datasets/prajwalkanade/student-performance-prediction-dataset>

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2004). Predicting Students’ Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence*, *18*(5), 411–426. <https://doi.org/10.1080/08839510490442058>

Putra, H. A., & Pratama, I. N. G. A. (2020). Analisis Perbandingan Algoritma Klasifikasi untuk Prediksi Nilai Akademik Mahasiswa. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *4*(4), 719–726. <https://doi.org/10.29207/resti.v4i4.1950>

Scikit-learn Developers. (2023). *Scikit-learn: Machine Learning in Python*. <https://scikit-learn.org/stable/>