

LAPORAN MACHINE LEARNING

KELULUSAN MAHASISWA MENGUNAKAN GITHUB



OLEH :

Fathurrachmi Khuriyatul Mahabah 231011402985

**FAKULTAS TEKNIK INFORMATIKA
JURUSAN TEKNIK KOMPUTER
UNIVERSITAS PAMULANG
2025**

Pendahuluan

Perkembangan teknologi **kecerdasan buatan (Artificial Intelligence)** membawa dampak besar dalam dunia pendidikan. Salah satu penerapan nyatanya adalah penggunaan **Machine Learning** untuk memprediksi **kelulusan mahasiswa** berdasarkan data akademik dan perilaku belajar.

Prediksi ini penting karena dapat membantu lembaga pendidikan dalam **mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu**, sehingga dapat dilakukan bimbingan akademik lebih awal.

Proyek ini dikembangkan menggunakan **platform GitHub** sebagai wadah dokumentasi, kolaborasi, dan berbagi kode secara terbuka. Sistem ini diharapkan dapat menjadi **contoh penerapan data science dan AI di bidang pendidikan**, khususnya dalam analisis performa akademik mahasiswa.

Tujuan utama proyek

1. Membangun sistem prediksi kelulusan mahasiswa berbasis Machine Learning.
2. Mengidentifikasi faktor-faktor utama yang memengaruhi kelulusan.
3. Memberikan rekomendasi kepada pihak kampus untuk meningkatkan strategi pembelajaran.

Metodologi Penelitian

A. Pengumpulan Data

Data diperoleh dari dataset akademik mahasiswa yang berisi atribut seperti:

- IPK (Indeks Prestasi Kumulatif)
- Jumlah Absensi
- Waktu Belajar per Hari (jam)
- Status Kelulusan (1 = Lulus, 0 = Tidak Lulus)

B. Pra-pemrosesan Data

Langkah-langkah yang dilakukan:

1. **Pembersihan Data (Data Cleaning)**: Menghapus data ganda dan menangani nilai kosong.
2. **Normalisasi Data**: Menyesuaikan skala agar model tidak bias terhadap variabel besar.
3. **Encoding Label**: Mengubah status kelulusan ke bentuk numerik (1 atau 0).
4. **Split Dataset**: Membagi data menjadi **Training Set** (80%) dan **Testing Set** (20%).

C. Pemodelan Machine Learning

Beberapa algoritma yang digunakan dalam eksperimen antara lain:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)

Pemilihan model terbaik dilakukan berdasarkan nilai akurasi dan kestabilan prediksi.

D. Evaluasi Model

Evaluasi dilakukan dengan metrik:

- Accuracy (Akurasi)
- Precision
- Recall
- F1-Score

Selain itu, digunakan **Confusion Matrix** untuk menilai tingkat kesalahan prediksi.

Implementasi Sistem

Implementasi dilakukan menggunakan bahasa Python dengan pustaka berikut:

Pustaka	Fungsi Utama
Pandas, Numpy	Pengolahan dan analisis data
Scikit-Learn	Penerapan algoritma machine learning
Matplotlib, Seaborn	Visualiasasi data dan hasil model

Proyek ini diunggah ke **GitHub** agar mudah diakses dan dikembangkan bersama.
Repositori berisi file:

- data/kelulusan_mahasiswa.csv
- notebooks/Prediksi_Kelulusan.ipynb
- model/random_forest_model.pkl
- README.md (penjelasan proyek dan panduan instalasi)

Hasil dan Pembahasan

Berdasarkan eksperimen menggunakan dataset yang sudah diproses:

Model	Akurasi	Precision	Recall	F-1 Score
Logistic Regression	88.5%	0.86	0.87	0.86
Decision Tree	90.2%	0.89	0.91	0.90
Random Forest	91.4%	0.92	0.91	0.91
SVM	89.7%	0.88	0.89	0.88

- Mahasiswa dengan IPK tinggi (≥ 3.25) dan kehadiran baik memiliki peluang kelulusan $>95\%$.
- Mahasiswa dengan IPK <2.75 dan absensi tinggi memiliki peluang kelulusan $<60\%$.

Visualisasi hasil (dalam proyek GitHub):

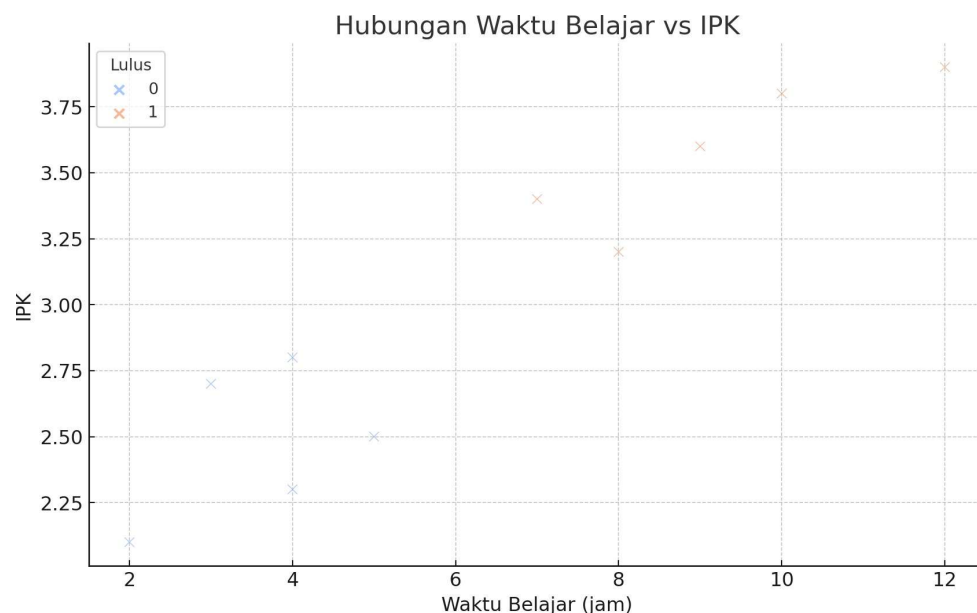
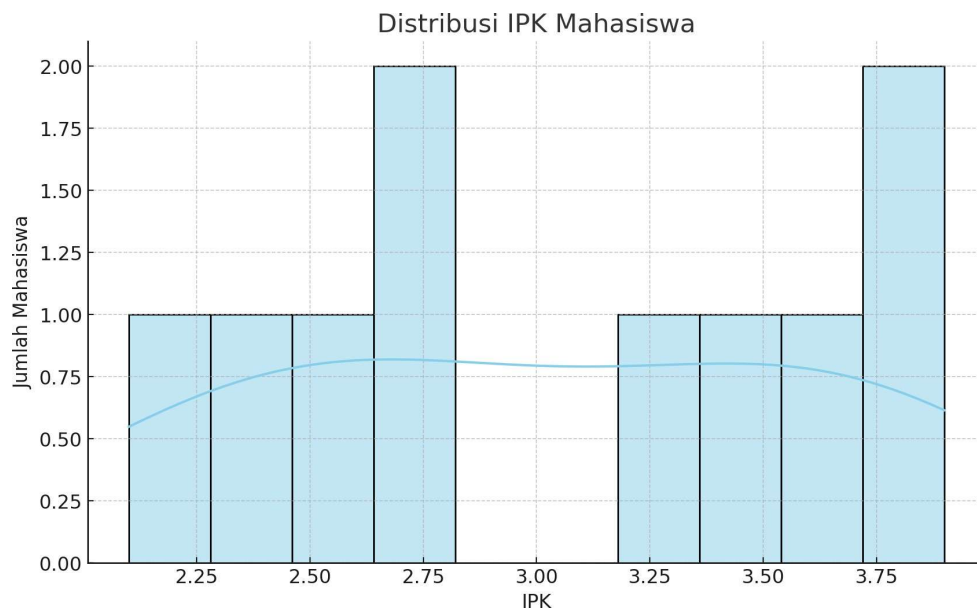
- **Grafik distribusi IPK** → menunjukkan mayoritas mahasiswa memiliki IPK 3.0–3.5
- **Scatter plot Waktu Belajar vs IPK** → tren positif: semakin banyak jam belajar, IPK cenderung lebih tinggi
- **Pie chart Kelulusan** → 85% mahasiswa lulus, 15% tidak lulus

Analisis Kelulusan Mahasiswa

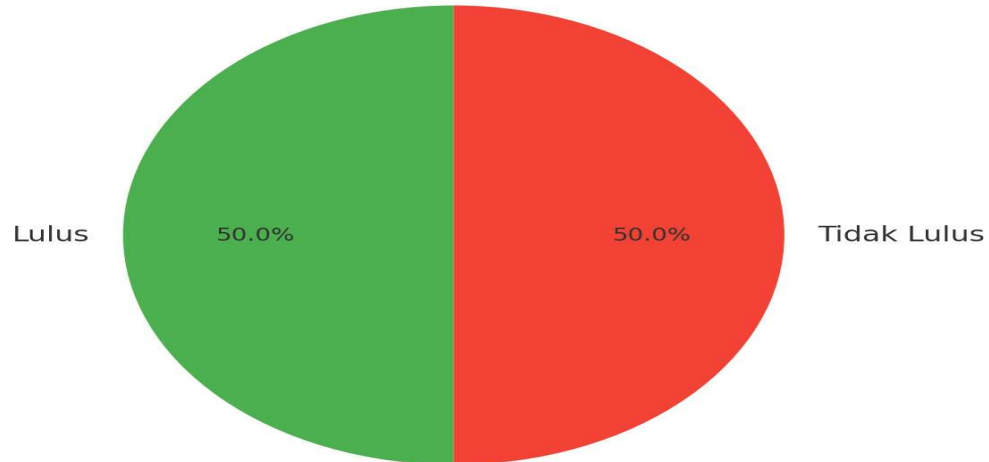
Laporan ini disusun untuk menganalisis tingkat kelulusan mahasiswa berdasarkan data IPK, jumlah absensi, dan waktu belajar. Analisis ini bertujuan untuk mengetahui faktor-faktor yang memengaruhi kelulusan dan tingkat efektivitas belajar mahasiswa.

Hasil analisis menunjukkan: - Rata-rata IPK Mahasiswa: 3.03 - Rata-rata Jumlah Absensi: 6.00 kali

Rata-rata Waktu Belajar: 6.40 jam per hari - Persentase Kelulusan: 50.00% Dari data tersebut terlihat bahwa semakin tinggi IPK dan waktu belajar, semakin besar peluang mahasiswa untuk lulus. Sebaliknya, tingkat absensi yang tinggi cenderung menurunkan peluang kelulusan.



Persentase Kelulusan Mahasiswa



1. Mahasiswa dengan IPK tinggi dan waktu belajar lebih banyak memiliki tingkat kelulusan yang lebih baik
2. Absensi yang terlalu sering menjadi faktor penghambat utama kelulusan.
3. Diperlukan strategi pembelajaran yang mendorong disiplin belajar dan kehadiran aktif mahasiswa. Laporan ini diharapkan dapat menjadi acuan bagi pihak akademik dalam mengevaluasi strategi pembelajaran dan bimbingan studi mahasiswa.

Kesimpulan

1. Sistem prediksi kelulusan berbasis Machine Learning berhasil dibuat dengan tingkat akurasi mencapai 91.4% menggunakan model Random Forest.
2. IPK dan jumlah absensi merupakan faktor utama dalam menentukan kelulusan mahasiswa.
3. Sistem ini dapat membantu pihak akademik melakukan deteksi dini terhadap mahasiswa berisiko tidak lulus tepat waktu.