

LAPORAN TUGAS PERTEMUAN 4

MATA KULIAH MACHINE LEARNING

Nama: Galih Naufal Faturrohman

NIM: 231011402731

Program Studi Informatika

Universitas Pamulang

Tahun Akademik 2025

Laporan Analisis Data Kelulusan Mahasiswa

Laporan ini dibuat sebagai bagian dari Tugas Pertemuan 4 pada mata kuliah **Machine Learning**. Analisis ini bertujuan untuk memahami faktor-faktor yang berpengaruh terhadap kelulusan mahasiswa. Proses dilakukan dengan menggunakan bahasa pemrograman Python serta pustaka **pandas**, **seaborn**, dan **scikit-learn**.

Langkah-langkah Analisis

- Membaca dataset kelulusan_mahasiswa.csv menggunakan pandas.

- Melakukan pengecekan nilai kosong (missing values) dan duplikasi data.
- Visualisasi data menggunakan boxplot, histogram, scatterplot, dan heatmap untuk memahami distribusi IPK dan korelasi antar variabel.
- Menambahkan dua fitur baru: Rasio_Absensi dan IPK_x_Study.
- Menyimpan dataset hasil pemrosesan ke dalam file processed_kelulusan.csv.
- Melakukan pembagian data menjadi data latih, validasi, dan uji dengan perbandingan 70:15:15.

Pembagian Dataset

Dataset	Persentase	Tujuan
Training	70%	Melatih model
Validation	15%	Menyesuaikan parameter model
Testing	15%	Menguji performa model

Kesimpulan

Dari hasil analisis ini dapat disimpulkan bahwa faktor-faktor seperti **IPK**, **jumlah absensi**, dan **waktu belajar** memiliki pengaruh signifikan terhadap tingkat kelulusan mahasiswa. Dataset yang telah diproses dapat digunakan lebih lanjut untuk membangun model machine learning seperti **Decision Tree**, **Random Forest**, atau **Logistic Regression** guna memprediksi kelulusan mahasiswa di masa depan.