

# Laporan Pertemuan 5 – Machine Learning

## *Analisis Model Logistic Regression dan Random Forest*

**Nama:** Galih Naufal Faturrohman

**NIM:** 231011402731

**Program Studi:** Teknik Informatika – Universitas Pamulang

## 1. Pendahuluan

Analisis terhadap faktor-faktor yang memengaruhi kelulusan mahasiswa merupakan hal penting dalam bidang pendidikan tinggi.

Dengan menggunakan pendekatan *Machine Learning*, data akademik seperti IPK, jumlah absensi, dan waktu belajar dapat dimanfaatkan untuk membangun model prediktif yang membantu memetakan kemungkinan kelulusan mahasiswa.

Pada tugas ini, dua algoritma *Machine Learning* yaitu **Logistic Regression** dan **Random Forest** digunakan untuk mengklasifikasikan status kelulusan mahasiswa berdasarkan dataset *processed\_kelulusan.csv*.

## 2. Tujuan

Analisis ini bertujuan untuk:

- 1 Melakukan pra-pemrosesan dan pembagian dataset.
- 2 Membangun model *Logistic Regression* sebagai baseline.
- 3 Mengembangkan model *Random Forest* dengan tuning parameter untuk meningkatkan akurasi.
- 4 Mengevaluasi performa model menggunakan metrik **F1-score**, **ROC-AUC**, dan **confusion matrix**.

## 3. Metode Analisis

Data dibagi menjadi tiga bagian, yaitu data latih (70%), validasi (15%), dan uji (15%). Proses preprocessing dilakukan menggunakan *SimpleImputer* (median) untuk menangani nilai kosong dan *StandardScaler* untuk normalisasi fitur numerik.

Dua model diterapkan:

- **Logistic Regression** (baseline) dengan *class\_weight = balanced*
- **Random Forest Classifier** dengan 300 estimator dan *hyperparameter tuning* menggunakan **GridSearchCV**

Pencarian parameter optimal dilakukan untuk *max\_depth* dan *min\_samples\_split* dengan skema validasi silang (*StratifiedKfold*).

## 4. Hasil dan Pembahasan

Model **Logistic Regression** memberikan performa awal yang cukup baik namun terbatas dalam menangani hubungan non-linear antar variabel.

Setelah menggunakan **Random Forest**, performa meningkat secara signifikan dengan nilai **F1-macro** lebih tinggi pada data validasi.

Proses *hyperparameter tuning* menghasilkan konfigurasi model terbaik dengan kedalaman pohon yang moderat dan jumlah minimal sampel split yang optimal.

Evaluasi model menunjukkan bahwa:

- Nilai **F1-score (test)** meningkat dibanding baseline.
- *Confusion matrix* menunjukkan klasifikasi lebih tepat untuk kelas “Lulus”.
- Nilai **ROC-AUC** tinggi, menandakan kemampuan diskriminasi model terhadap dua kelas cukup baik.

Hasil ini mengindikasikan bahwa **Random Forest** lebih efektif dalam menangani data dengan pola kompleks dibandingkan model linear seperti Logistic Regression.

## 5. Kesimpulan

- 1 Model *Machine Learning* dapat diterapkan untuk memprediksi kelulusan mahasiswa berdasarkan data akademik dan perilaku belajar.
- 2 **Random Forest** memberikan hasil prediksi yang lebih baik daripada **Logistic Regression** karena mampu menangani hubungan non-linear dan variasi fitur.
- 3 Proses *hyperparameter tuning* meningkatkan performa model dengan signifikan.
- 4 Evaluasi menggunakan metrik F1-score dan ROC-AUC membuktikan model memiliki kemampuan klasifikasi yang cukup stabil.

5 Untuk penelitian selanjutnya, disarankan memperluas dataset dan menambahkan variabel baru agar model dapat melakukan generalisasi dengan lebih baik.

**Catatan:**

Seluruh proses analisis, mulai dari pembacaan data, pra-pemrosesan, pelatihan model, hingga evaluasi, dilakukan menggunakan bahasa pemrograman Python pada file `main5.py`.