# LAPORAN PRAKTIKUM MACHINE LEARNING

Pertemuan 6 — Random Forest untuk Klasifikasi

Pada pertemuan ini, fokus pembahasan diarahkan kepada penerapan algoritma Random Forest sebagai salah satu metode klasifikasi pada kasus prediksi kelulusan mahasiswa. Model ini dipilih karena memiliki performa yang baik pada data berskala kecil hingga besar, serta dapat bekerja efektif pada data dengan fitur numerik seperti IPK, jumlah absensi, dan waktu belajar.

Random Forest merupakan algoritma yang terdiri dari sekumpulan pohon keputusan (decision trees) yang digabungkan menggunakan teknik bootstrap aggregation (bagging). Gagasan di balik metode ini adalah menghasilkan prediksi yang lebih stabil dan akurat daripada hanya menggunakan satu pohon keputusan saja. Dengan menggabungkan banyak model sederhana, Random Forest mengurangi risiko overfitting dan meningkatkan generalisasi.

Dataset yang digunakan pada praktikum ini berasal dari hasil pemrosesan data sebelumnya (Pertemuan 4 & 5), di mana fitur penting seperti IPK, Rasio Absensi, serta IPK\_x\_Study telah disiapkan. Proses pembagian data dilakukan dengan stratifikasi, memastikan proporsi kelas lulus (1) dan tidak lulus (0) tetap seimbang antara train dan test set.

Setelah data siap, model Random Forest dibangun menggunakan parameter dasar seperti jumlah estimator sebanyak 300 pohon, serta penggunaan class\_weight='balanced' untuk mengatasi ketidakseimbangan kelas jika muncul. Setelah proses pelatihan, model kemudian dievaluasi menggunakan metrik F1-score untuk melihat keseimbangan antara precision dan recall.

Selain itu dilakukan validasi silang (cross-validation) menggunakan metode StratifiedKFold agar evaluasi lebih konsisten dan tidak bias terhadap satu pembagian data saja. Proses tuning parameter (GridSearchCV) juga dilakukan untuk menemukan konfigurasi pohon terbaik seperti max\_depth dan min\_samples\_split.

Dari hasil evaluasi, model Random Forest menunjukkan performa yang unggul dibanding baseline Logistic Regression pada validation set. Setelah pemilihan model terbaik berdasarkan metrik F1-skor dan hasil tuning, model final kemudian diuji menggunakan test set untuk memvalidasi performanya secara objektif dan menghindari overfitting.

Confusion matrix dan ROC-AUC juga digunakan sebagai tambahan analisis untuk melihat seberapa baik model membedakan kelas prediksi. ROC-AUC memberikan gambaran kemampuan model dalam memprediksi probabilitas kelas benar.

Kesimpulannya, Random Forest menjadi pilihan model yang baik dalam kasus ini karena stabil, tidak mudah overfitting, dan mampu memberikan akurasi yang memadai pada dataset kecil. Implementasi ini memenuhi kaidah pipeline machine learning sesuai prinsip evaluasi ilmiah.

## Kesimpulan

Penerapan Random Forest dalam klasifikasi kelulusan mahasiswa pada pertemuan ini berhasil memberikan hasil yang cukup baik. Model dapat mengeneralisasi data dengan optimal serta memberikan nilai F1-score yang kompetitif. Dengan demikian, Random Forest dapat digunakan sebagai alternatif model yang kuat untuk menyelesaikan permasalahan klasifikasi serupa.