# Laporan Pertemuan 5 - Modeling Machine Learning

## Pendahuluan

Pada pertemuan ini, dilakukan proses pemodelan machine learning dengan tujuan untuk memprediksi kelulusan mahasiswa berdasarkan beberapa variabel seperti IPK, jumlah absensi, waktu belajar, serta fitur turunan yang telah dibuat pada pertemuan sebelumnya. Fokus utama pada tahap ini meliputi pemilihan model, pembagian data, training model, evaluasi performa, serta justifikasi pemilihan model akhir.

## Pembagian Dataset

Dataset yang digunakan berasal dari output hasil feature engineering pada pertemuan sebelumnya. Data dibagi menjadi dua bagian, yaitu 70% untuk data pelatihan (training) dan 30% untuk data pengujian (testing). Pembagian ini dilakukan menggunakan fungsi train\_test\_split dari library sklearn dengan metode stratified agar perbandingan kelas pada label tetap seimbang antara train dan test.

## Pemilihan Model

Model yang dipilih pada tahap ini adalah Logistic Regression. Alasan pemilihan model ini adalah karena model tersebut sangat sesuai untuk kasus klasifikasi biner seperti kelulusan mahasiswa (0 = Tidak Lulus, 1 = Lulus). Selain itu, Logistic Regression memiliki interpretabilitas yang baik serta dapat memberikan insight terhadap hubungan antara fitur dan label.

## Training Model

Model dilatih menggunakan data training yang telah disiapkan sebelumnya. Selama pelatihan, parameter default dari Logistic Regression digunakan karena data relatif kecil dan sederhana sehingga belum memerlukan konfigurasi parameter yang kompleks. Proses pelatihan berjalan dengan cepat dan efisien karena jumlah fitur yang digunakan tidak terlalu banyak.

## Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan data testing. Metrik yang digunakan adalah akurasi, presisi, recall, dan confusion matrix. Hasil evaluasi menunjukkan bahwa model dapat memprediksi kelulusan mahasiswa dengan performa yang cukup baik. Confusion matrix menunjukkan prediksi tepat untuk sebagian besar data, meskipun terdapat sedikit kesalahan pada beberapa sampel.

## Alasan Pemilihan Model Final

Model Logistic Regression dipilih sebagai model final pada pertemuan ini karena performa yang stabil serta kemudahan interpretasi. Selain itu, model ini dapat memberikan estimasi kontribusi masing-masing fitur terhadap probabilitas kelulusan, sehingga dapat menjadi bahan analisis lanjutan.

## Kesimpulan

Secara keseluruhan, tahap Modeling Machine Learning pada pertemuan ini berjalan dengan baik. Pemilihan model Logistic Regression dianggap tepat berdasarkan karakteristik data dan tujuan analisis. Hasil evaluasi menunjukkan bahwa model memiliki performa yang cukup baik dan dapat digunakan sebagai dasar prediksi kelulusan mahasiswa. Pada pertemuan selanjutnya, model dapat dikembangkan lebih lanjut dengan metode lain seperti Random Forest untuk perbandingan performa.