**LAPORAN PRAKTIKUM MACHINE LEARNING  
DATA PREPARATION, MODELING MACHINE LEARNING DAN ARTIFICIAL NEURON NETWORK (ANN)**

Disusun oleh:

FARHAN NUL HAKIM

NIM : 231011402153

KELAS : 05TPLE015

Dosen Pengampu:

ZURNAN ALFIAN, S.KOM., M.KOM.

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PAMULANG

2025

# BAB I PENDAHULUAN

Pada praktikum ini dilakukan proses Data Preparation sebagai tahap awal dalam penerapan Machine Learning. Tahapan ini bertujuan untuk memastikan data yang digunakan sudah bersih, lengkap, dan konsisten agar hasil analisis lebih akurat. Data preparation mencakup pembersihan data dari duplikasi, pembuatan fitur baru, serta analisis awal terhadap hubungan antar variabel. Dalam praktikum ini digunakan data berisi IPK, jumlah absensi, waktu belajar, dan status kelulusan mahasiswa. Setelah dilakukan pembersihan dan penambahan fitur turunan, data menjadi siap digunakan untuk tahap pemodelan pada pertemuan berikutnya.

Tahap berikutnya adalah *Modeling Machine Learning*, di mana data yang telah melalui proses persiapan digunakan untuk membangun model prediktif yang mampu mengenali pola dari data. Proses ini melibatkan pembagian dataset menjadi data latih, validasi, dan uji, serta pelatihan model menggunakan algoritma seperti *Logistic Regression* dan *Random Forest Classifier*. Evaluasi model dilakukan untuk mengukur seberapa baik model dapat memprediksi kelulusan mahasiswa berdasarkan IPK, absensi, dan waktu belajar.

Kedua tahapan ini memiliki hubungan yang saling melengkapi. *Data Preparation* berperan sebagai fondasi dalam memastikan kualitas data yang digunakan, sedangkan *Modeling Machine Learning* menjadi langkah untuk menerapkan pengetahuan dari data tersebut ke dalam bentuk prediksi. Dengan demikian, praktikum ini memberikan pemahaman komprehensif mengenai alur penerapan *Machine Learning* mulai dari pengolahan data hingga pembangunan model prediktif yang siap diimplementasikan.

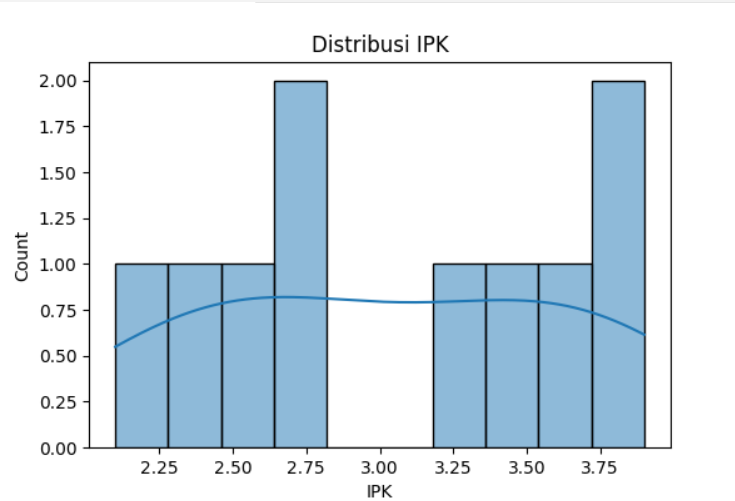
# BAB II PEMBAHASAN

## Dataset

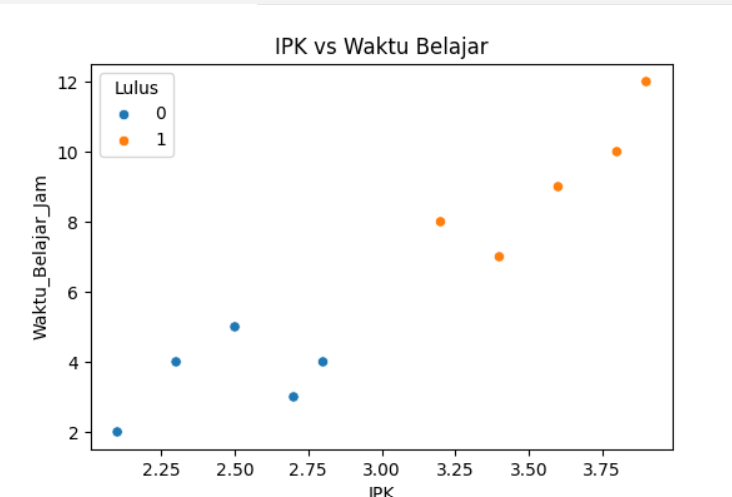
Dataset yang digunakan berisi informasi mengenai IPK, Jumlah Absensi, Waktu Belajar (Jam), dan status kelulusan mahasiswa. Selain itu, ditambahkan dua fitur turunan yaitu Rasio Absensi dan IPK\_x\_Study yang membantu memperkaya analisis korelasi antar variabel.

## Hasil Visualisasi dan Interpretasi

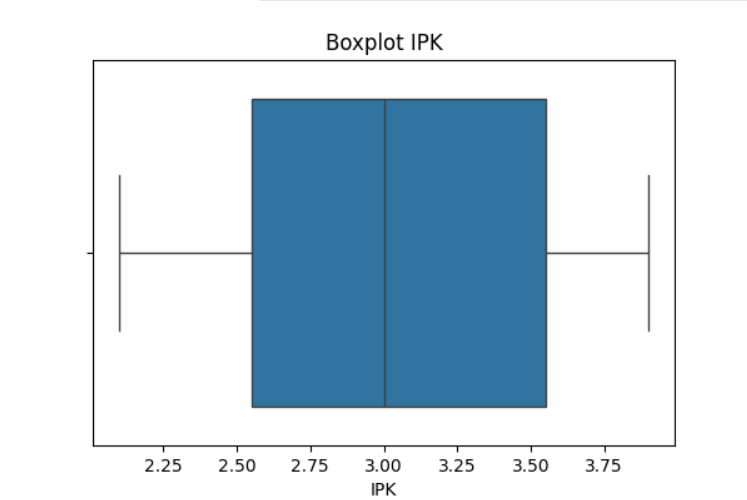
Berikut merupakan hasil visualisasi dan interpretasi dari data yang telah diproses:



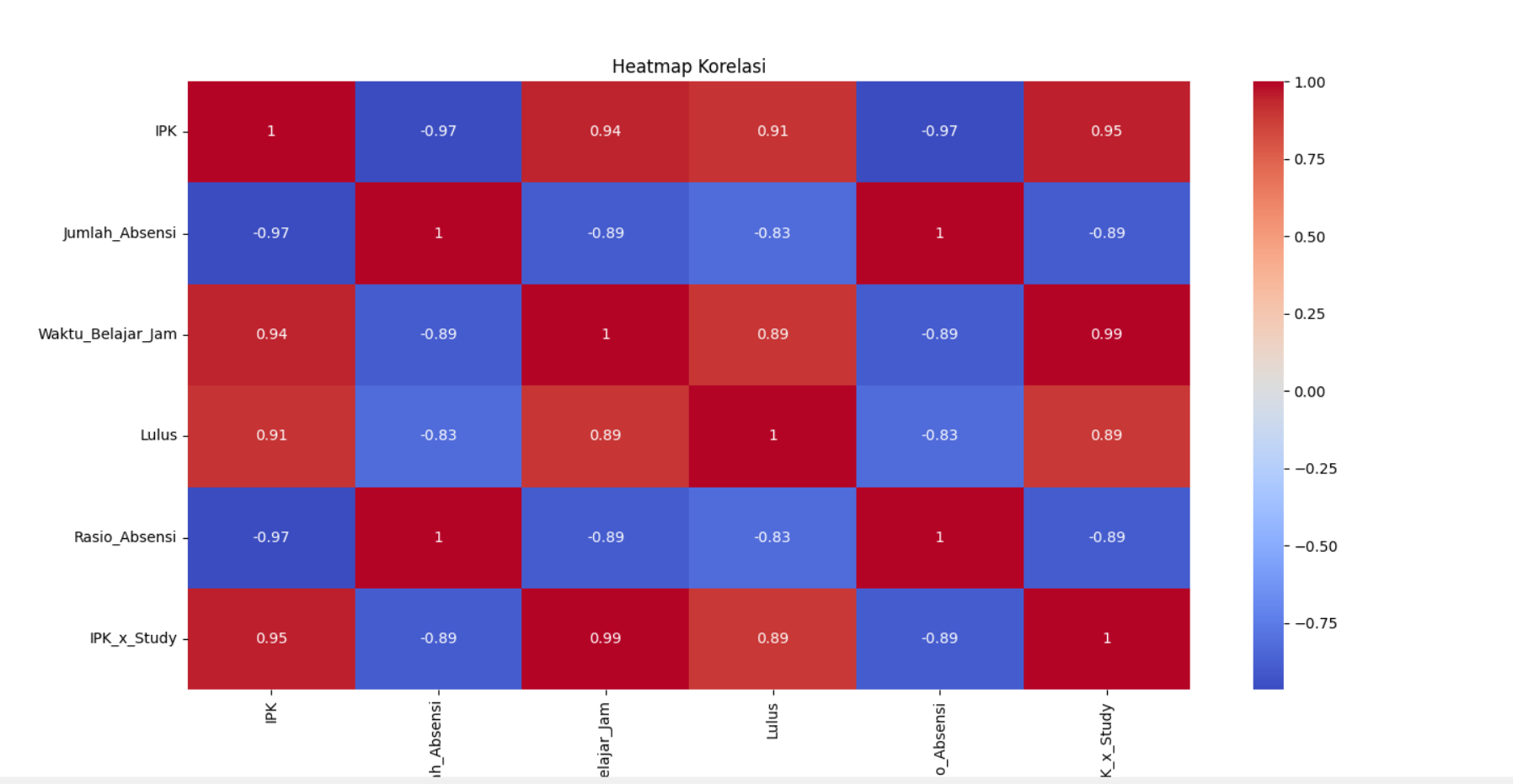
Gambar 1. Distribusi IPK mahasiswa menunjukkan sebaran nilai dan kecenderungan utama.



Gambar 2. Scatter plot IPK dan waktu belajar, menunjukkan hubungan positif dan warna menandakan kelulusan.



Gambar 3. Boxplot IPK memperlihatkan median, kuartil, dan potensi outlier.

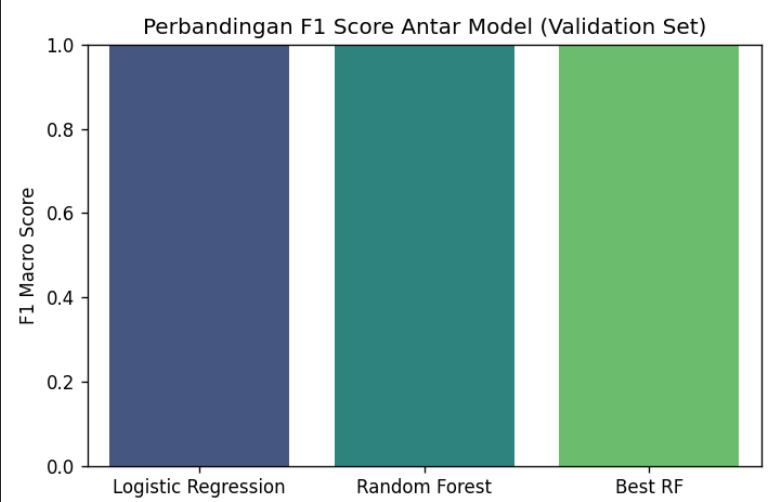


Gambar 4. Heatmap korelasi antar variabel numerik dalam dataset.

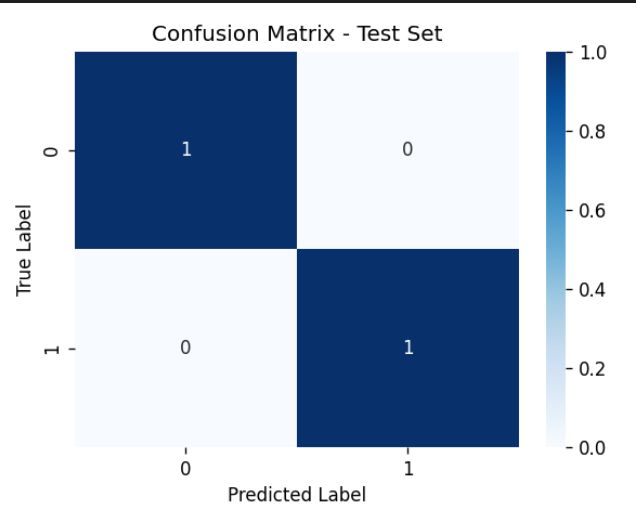
# BAB III MODELING

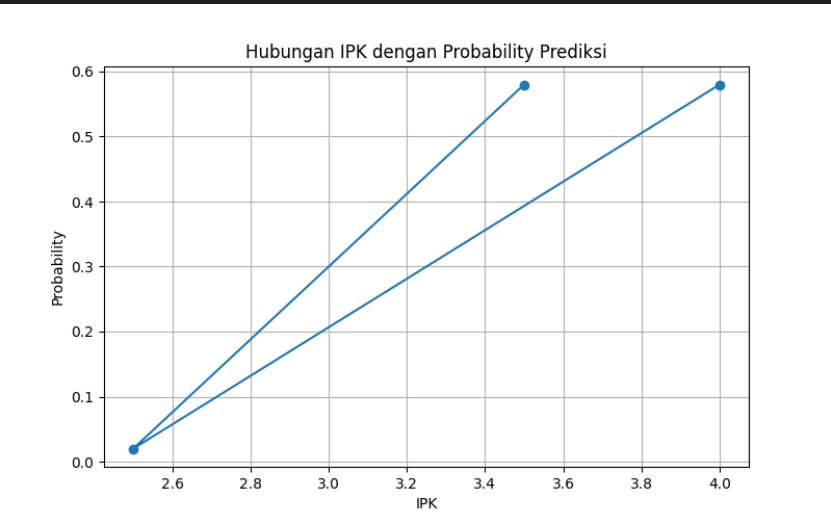
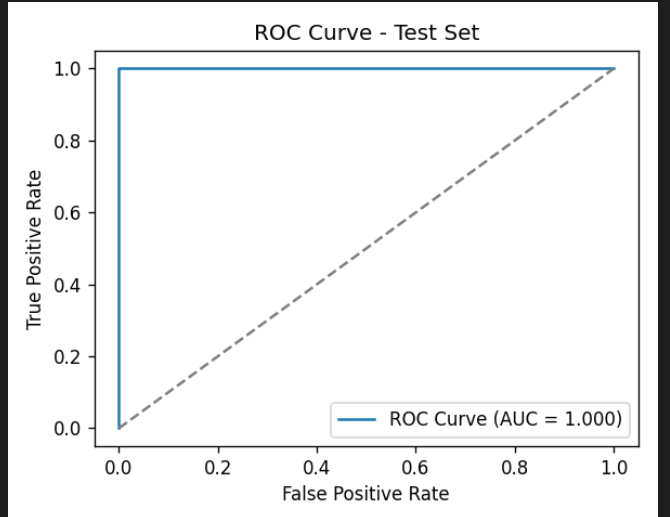
Tahap ini dilakukan untuk membangun model Machine Learning yang mampu memprediksi status kelulusan mahasiswa. Proses dimulai dengan pembuatan model baseline **Logistic Regression** sebagai acuan awal, kemudian dilanjutkan dengan model **Random Forest** untuk membandingkan performa.

Setelah dilakukan hyperparameter tuning menggunakan **GridSearchCV**, diperoleh model terbaik yaitu **Best Random Forest** dengan hasil evaluasi F1-score dan AUC yang tinggi.

  
Gambar 5. Perbandingan F1 Score Antar Model (Validation Set).

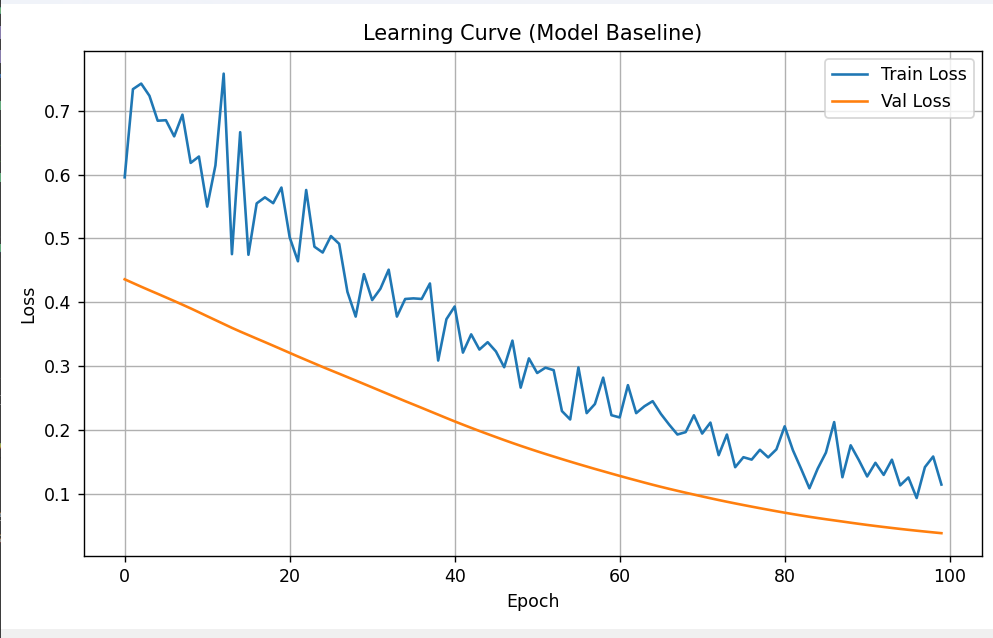
Grafik ini memperlihatkan nilai F1 Macro Score dari tiga model: Logistic Regression, Random Forest, dan Best Random Forest. Ketiganya memiliki nilai tinggi, namun Best Random Forest sedikit lebih unggul. Ini menunjukkan proses tuning berhasil meningkatkan performa model.



Gambar 6. Confusion Matrix pada Data Uji (Test Set).  
  
  
Confusion Matrix menunjukkan seluruh data uji berhasil diklasifikasikan dengan benar tanpa kesalahan prediksi. Hasil ini menandakan model memiliki kemampuan generalisasi yang sangat baik, meski tetap perlu diwaspadai kemungkinan overfitting.  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
 Gambar 7. Hubungan IPK dengan Probability Prediksi.  
  
Grafik ini menunjukkan hubungan antara IPK dan probabilitas kelulusan yang diprediksi model. Terlihat korelasi positif, artinya semakin tinggi IPK, semakin besar peluang mahasiswa untuk diprediksi lulus. Ini menunjukkan model menangkap hubungan logis antar variabel.  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  


Gambar 8. Kurva ROC pada Data Uji (Test Set).  
  
Kurva ROC memperlihatkan nilai **AUC = 1.000**, menunjukkan model memiliki kemampuan klasifikasi sempurna. Garis ROC menempel di sisi kiri atas grafik, menandakan True Positive Rate maksimal tanpa False Positive. Nilai sempurna ini menegaskan kekuatan model, meskipun perlu validasi tambahan agar tidak terjadi overfitting.

# BAB IV ARTIFICIAL NEURON NETWORK (ANN)



Gambar 8. Kurva ROC pada Data Uji (Test Set).

Kurva Pembelajaran menunjukkan bahwa **Val Loss** (oranye) **menurun secara konsisten dan mulus**, mencapai nilai yang sangat rendah. Sebaliknya, **Train Loss** (biru) menampilkan penurunan yang serupa tetapi dengan **fluktuasi tinggi (berisik)**, dan cenderung berada di atas *Val Loss*. Fenomena ini adalah **hal umum dan positif** ketika **Dropout** diaktifkan selama pelatihan, karena model menjadi lebih tergeneralisasi pada data validasi. Penurunan yang berkelanjutan pada *Val Loss* mengindikasikan bahwa model **belum mengalami *overfitting*** dan terus belajar secara stabil, menunjukkan potensi kinerja klasifikasi yang sangat baik.

# BAB V KESIMPULAN

Proyek ini telah melalui proses **Data Preparation** yang berhasil menghasilkan data bersih dan logis, menguatkan temuan kunci berupa **korelasi positif antara waktu belajar dan IPK**, serta **korelasi negatif antara absensi dan IPK**. Pada tahap **Modeling**, meskipun algoritma *Best Random Forest* memberikan performa *baseline* yang optimal dengan metrik F1-score dan AUC yang tinggi, eksplorasi menggunakan **Artificial Neural Network (ANN)** sangat krusial untuk menangkap pola non-linear. Model ANN (MLP) dikonfigurasi dengan arsitektur dasar yang melibatkan *hidden layer* dengan aktivasi **ReLU** dan *output layer* menggunakan **Sigmoid** untuk tugas klasifikasi biner. Model dikompilasi menggunakan **loss Binary Cross-Entropy** dan dioptimasi dengan **Adam Optimizer**. Selain itu, untuk memastikan stabilitas dan menghindari *overfitting*, teknik **regularisasi** seperti **Dropout** dan *Early Stopping* diaplikasikan dengan memonitor *validation loss*. Secara keseluruhan, kombinasi antara data yang informatif dan pengujian dua paradigma model yang kuat (Random Forest dan ANN) menghasilkan model prediksi kelulusan mahasiswa yang **akurat, stabil, dan logis**. Model ini siap untuk dikembangkan lebih lanjut melalui penyesuaian arsitektur atau penambahan regulasi lanjutan (seperti L2 atau Batch Normalization) untuk pengujian kestabilan kinerja pada *dataset* lain.