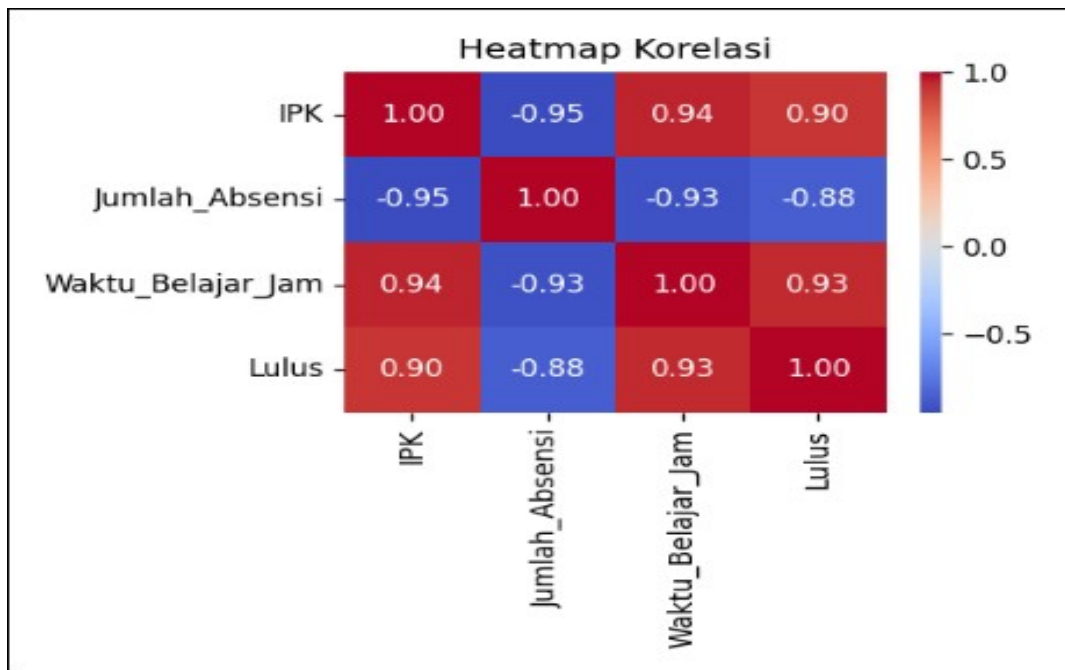


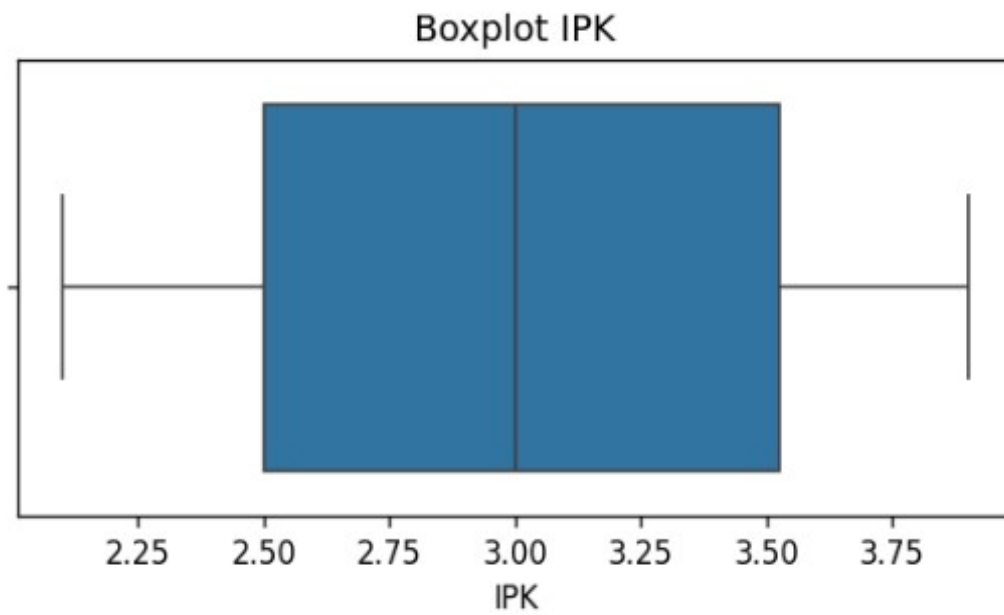
■ Laporan Analisis Data — Pertemuan 4

Laporan ini dibuat berdasarkan hasil analisis data dari file **kelulusan_mahasiswa.csv** yang berisi data akademik mahasiswa. Tahapan yang dilakukan meliputi: 1. **Data Collection**: Mengumpulkan data awal dengan 4 variabel utama. 2. **Data Cleaning**: Mengecek missing values dan duplikasi data. 3. **Exploratory Data Analysis (EDA)**: Melakukan visualisasi dan analisis korelasi antar variabel. 4. **Feature Engineering**: Membuat fitur tambahan seperti Rasio_Absensi dan IPK_x_Study . 5. **Splitting Dataset**: Membagi data menjadi train, validation, dan test set (70/15/15).

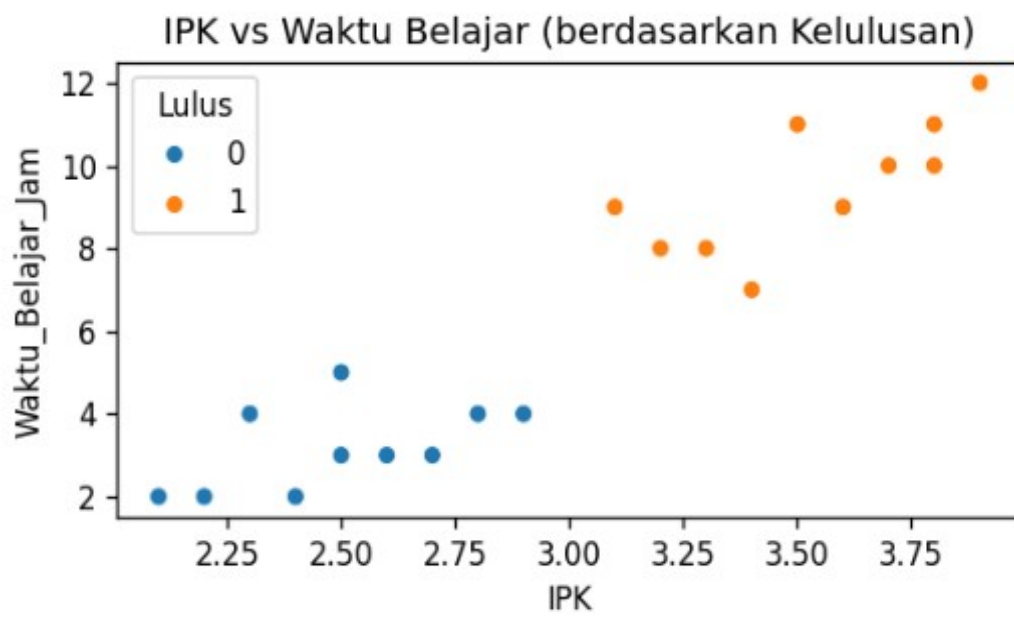
■ Visualisasi Data:



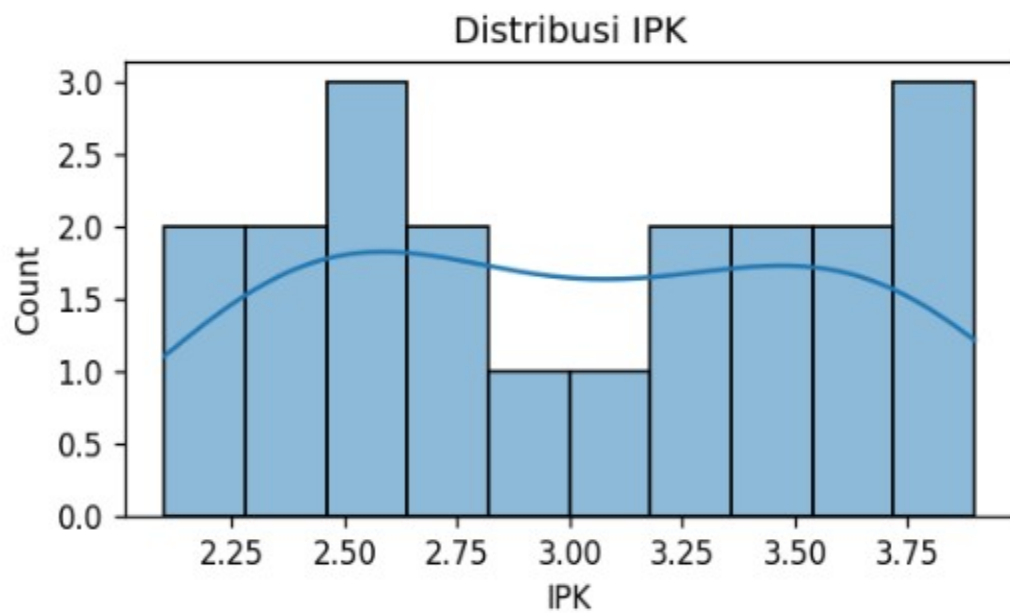
Gambar 1. Heatmap Korelasi antar variabel



Gambar 2. Boxplot distribusi IPK



Gambar 3. Hubungan IPK vs Waktu Belajar berdasarkan Kelulusan



Gambar 4. Distribusi nilai IPK mahasiswa

Kesimpulan:

Hasil analisis menunjukkan bahwa IPK dan waktu belajar memiliki korelasi positif dengan kelulusan mahasiswa, sementara jumlah absensi memiliki korelasi negatif. Mahasiswa dengan IPK tinggi dan waktu belajar lebih lama cenderung memiliki peluang kelulusan yang lebih besar. Dataset telah dibersihkan dan siap digunakan untuk tahap modeling selanjutnya.