

# LAPORAN ANALISIS DATA MAHASISWA

Mata Kuliah: Data Science

## 1. Exploratory Data Analysis (EDA)

Analisis data eksplorasi dilakukan untuk memahami struktur dan karakteristik data. Dataset yang digunakan adalah student-mat.csv, berisi 395 data siswa dengan 33 variabel. Data ini memuat informasi demografi, status sosial, aktivitas, kebiasaan belajar, absensi, dan nilai akademik.

Temuan EDA:

- Tidak ada missing values di dalam dataset.
- Distribusi nilai akhir (G3) terkonsentrasi di nilai tengah, antara 8–12.
- Korelasi kuat ditemukan antara nilai G1 dan G2 terhadap G3, menunjukkan bahwa nilai sebelumnya menjadi prediktor utama nilai akhir.
- Variabel seperti studytime (waktu belajar) dan absences (ketidakhadiran) memiliki korelasi rendah terhadap G3.

Fungsi Setiap Kolom:

Dataset memiliki variabel seperti:

- Demografi siswa: school, sex, age, address
- Keluarga dan orang tua: famsize, Pstatus, Medu, Fedu, Mjob, Fjob
- Akademik: studytime, failures, schoolsup, higher, G1, G2, G3
- Aktivitas sosial: activities, romantic, goout, freetime

## 2. Analisis Regresi Linier

Regresi linier digunakan untuk mengetahui apakah ada hubungan antara waktu belajar (studytime) dan nilai akhir (G3).

Hasil analisis:

- Intercept: sekitar 9.56
- Slope: sekitar 0.44
- R-Squared: sekitar 0.017

Interpretasi:

- R-Squared sangat kecil, berarti waktu belajar hanya menjelaskan sekitar 1.7% variasi nilai akhir.
- Kesimpulan: Tidak ada hubungan signifikan antara waktu belajar dan nilai akhir.
- Faktor lain seperti nilai sebelumnya lebih berpengaruh terhadap hasil akademik siswa.

## 3. Clustering Segmentasi Siswa Berdasarkan Absensi

Clustering dilakukan dengan metode K-Means menggunakan variabel absences.

Tujuan analisis adalah untuk mengelompokkan siswa berdasarkan tingkat kehadiran mereka.

Hasil clustering menunjukkan:

- Terbentuk 3 kelompok siswa dengan pola absensi berbeda.
- Siswa dengan absensi tinggi cenderung memiliki nilai akhir yang lebih rendah.
- Segmentasi ini dapat membantu dalam monitoring siswa dengan kehadiran buruk.

#### 4. Klasifikasi Berdasarkan 3 Variabel

Klasifikasi dilakukan menggunakan Decision Tree Classifier untuk memprediksi kelulusan siswa.

Variabel yang digunakan sebagai prediktor:

- studytime — Waktu belajar
- failures — Jumlah kegagalan akademik
- absences — Jumlah ketidakhadiran

Target klasifikasi adalah Pass, yaitu siswa dianggap lulus jika  $G3 \geq 10$ .

Hasil klasifikasi:

- Akurasi Model: sekitar 72%
- Faktor utama keputusan model adalah jumlah kegagalan sebelumnya, diikuti oleh absences dan studytime.

Kesimpulan dari Klasifikasi:

- Model cukup efektif dalam memisahkan siswa yang lulus dan tidak lulus.
- Faktor akademik dan perilaku (absensi) lebih berpengaruh dibandingkan waktu belajar.

#### 5. Kesimpulan Umum

- Dataset sudah siap digunakan tanpa perlu pembersihan data.
- EDA mengungkap hubungan kuat antara nilai sebelumnya dengan nilai akhir.
- Regresi linier menunjukkan bahwa waktu belajar bukan faktor dominan penentu nilai.
- Clustering berhasil memetakan siswa berdasar absensi untuk segmentasi akademik.
- Klasifikasi mampu memberikan prediksi kelulusan dengan akurasi cukup baik.

Rekomendasi:

- Monitoring khusus bagi siswa dengan tingkat absensi dan kegagalan tinggi.
- Pengembangan sistem prediksi akademik berbasis data untuk intervensi dini.