

LAPORAN SINGKAT UJIAN DATA ANALYST

Wahyu Agustino
Teknik Informatika

1. Lakukan EDA (Exploratory Data Analysis) pada data tersebut dan jelaskan juga fungsi masing-masing kolom

Kolom	Tipe	Deskripsi
school	object	Sekolah siswa (GP atau MS)
sex	object	Jenis kelamin (F = Female, M = Male)
age	int	Umur siswa dalam tahun
address	object	Tempat tinggal (U = Urban, R = Rural)
famsize	object	Ukuran keluarga (LE3 \leq 3, GT3 $>$ 3)
Pstatus	object	Status hidup orang tua (T = together, A = apart)
Medu	int	Pendidikan ibu (0=none, 4=higher education)
Fedu	int	Pendidikan ayah (0=none, 4=higher education)
Mjob	object	Pekerjaan ibu (teacher, health, dll.)
Fjob	object	Pekerjaan ayah
reason	object	Alasan memilih sekolah (home, reputation, dll.)
guardian	object	Wali siswa (mother, father, other)
traveltime	int	Waktu tempuh ke sekolah (1= \leq 15min, 4= $>$ 1jam)

studytime	int	Waktu belajar mingguan (1= \leq 2h, 4= \geq 10h)
failures	int	Jumlah kegagalan mata pelajaran
schoolsup	object	Dukungan tambahan dari sekolah (yes/no)
famsup	object	Dukungan belajar dari keluarga (yes/no)
paid	object	Kursus tambahan bayar (yes/no)
activities	object	Aktivitas ekstrakurikuler (yes/no)
nursery	object	Apakah mengikuti TK/PAUD (yes/no)
higher	object	Ingin melanjutkan ke pendidikan tinggi (yes/no)
internet	object	Akses internet di rumah (yes/no)
romantic	object	Dalam hubungan romantis (yes/no)
famrel	int	Hubungan dalam keluarga (1-5)
freetime	int	Waktu luang setelah sekolah (1-5)
goout	int	Frekuensi pergi keluar dengan teman (1-5)
Dalc	int	Konsumsi alkohol di hari kerja (1-5)
Walc	int	Konsumsi alkohol di akhir pekan (1-5)
health	int	Status kesehatan diri (1=sangat buruk, 5=sangat baik)
absences	int	Jumlah absen
G1, G2, G3	int	Nilai siswa (masing-masing pada 3 periode berbeda, G3 adalah akhir)

Keterangan:

Usia siswa: 15–22 tahun (rata-rata ~16.7)

Rata-rata nilai akhir (G3): 10.4 dari maksimal 20

Kebanyakan siswa memiliki akses internet, ingin lanjut kuliah, dan tinggal di kota (urban)

2. Carilah 2 variabel (bebas) yang dapat dilakukan analisa menggunakan Regresi Linear.

- G2 (Nilai ujian kedua): Korelasi = 0.90

- G1 (Nilai ujian pertama): Korelasi = 0.80

2 variabel diatas cocok dilakukan regresi linear dengan nilai akhir siswa (G3) karena memiliki korelasi sangat kuat sehingga sangat cocok untuk dalam analisis regresi linear untuk memprediksi nilai akhir siswa

3. Lakukan Clustering segmentasi siswa berdasarkan absensi atau waktu belajar

Tujuan:

Mengelompokkan siswa berdasarkan ketidakhadiran (absences) dan waktu belajar (studytime) menggunakan KMeans.

Hasil Cluster:

Cluster 0 (Rajin): Study time tinggi tetapi absensi rendah

Cluster 1 (Rata-rata): Study time dan absensi dalam skala sedang

Cluster 2 (Kurang Rajin): Studytime rendah tetapi absensi tinggi

4. Lakukan klasifikasi berdasarkan 3 variabel (bebas)

Tujuan:

Mengklasifikasikan apakah siswa akan mendapatkan nilai akhir baik ($G3 \geq 10$) berdasarkan:

- studytime
- failures
- absences

Evaluasi:

Akurasi tinggi (di atas 80%)

Matrix menunjukkan model efektif dalam membedakan siswa "baik" dan "tidak baik"

Kesimpulan:

Nilai G1 dan G2 sangat menentukan nilai akhir G3.

Absensi dan waktu belajar berperan penting dalam segmentasi siswa.

Model klasifikasi mampu mengidentifikasi siswa yang berpotensi berhasil secara akurat.