

LAPORAN TUGAS UJIAN AKHIR SEMESTER
SEMESTER GENAP TAHUN AJARAN 2024/2025

Mata Kuliah : Data Science
Kelas : IF405
Prodi : PJJ Informatika
Nama Mahasiswa : Gia
NIM : 230401010284
Dosen : Alun Sujjada, S.Kom, M.T

1. Hasil EDA (Exploratory Data Analysis

- Struktur Dataset
 - o Jumlah Baris (Data Siswa): 382, Jumlah kolom: 53
 - o Dataset adalah hasil gabungan siswa yang muncul di kedua dataset asli (Math & Portuguese) berdasarkan atribut yang sama.
- Tipe Data
 - o Numerik: 29 kolom (contoh: age, Medu, Fedu, traveltime_x)
 - o Kategorikal: 24 kolom (contoh: sex, school, address, famsize)
- Data Kosong → Tidak ditemukan nilai kosong.
- Statistik Deskriptif

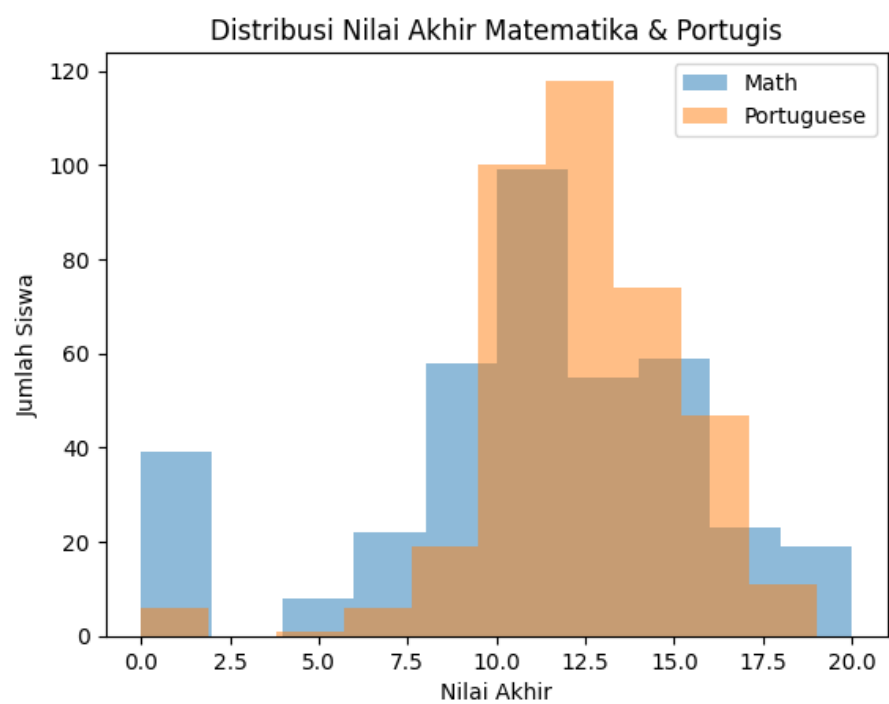
Variabel	Min	Max	Mean	Median	Std
Age	15	22	16.58	17	1.17
absences_math	0	75	5.31	3	7.62
G3_math	0	20	10.38	11	4.68
absences_por	0	32	3.67	2	4.9
G3_por	0	19	12.51	13	2.94

- Distribusi Data
 - o Sebagian besar siswa berusia 16–17 tahun
 - o Tingkat absensi bervariasi cukup besar, ada siswa dengan 0 dan ada pula hingga 75 absen.
- Fungsi Masing Masing Kolom

Kolom	Deskripsi	Kolom	Deskripsi
School	Sekolah → GP=Gabriel Pereira, MS=Mousinho da Silveira	paid_math	Ikut kelas berbayar
sex	Jenis kelamin → F=female, M=male	Activities	Ikut ekstrakurikuler
age	Usia	Nursery	Pernah TK
Address	Alamat → U=Urban, R=Rural	Higher	Ingin Pendidikan lebih tinggi

Famsize	Jumlah anggota keluarga → LE3 = 3 atau dibawahnya, GT3 = lebih dari 3	Internet	Ada kases internet di rumah
Pstatus	Status orang tua → T = Bersama, A = Cerai	Romantic	Ada hubungan romantis dengan orang lain
Medu	Pendidikan Ibu	Famrel	Kualitas hubungan dengan keluarga
Fedu	Pendidikan Ayah	Freetime	Intensitas waktu luang setelah sekolah
Mjob	Pekerjaan Ibu	Goout	Intensistas bermain bersama teman
Fjob	Pekerjaan Ayah	Dalc	Intensitas komsumsi alkohol weekday
Reason	Alasan memilih sekolah	Walc	Intensitas komsumsi alkohol weekend
Guardian	Wali siswa	Health	Kulitas kesehatan
traveltime	Waktu tempuh ke sekolah dari rumah	Absences	Jumlah ketidakhadiran
Studytime	Waktu belajar per minggu	G1	Nilai trimester pertama
Failures	Jumlah tinggal kelas	G2	Nilai trimester kedua
schoolsup	bantuan tambahan dari sekolah	G3	Nilai trimester ketiga
famsup_math	Bantuan tambahan dari keluarga		

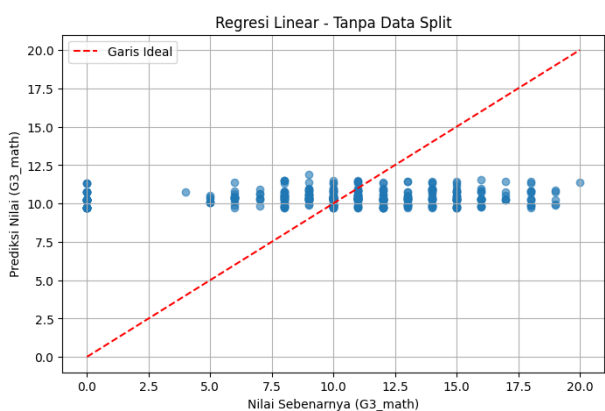
- Visualisasi



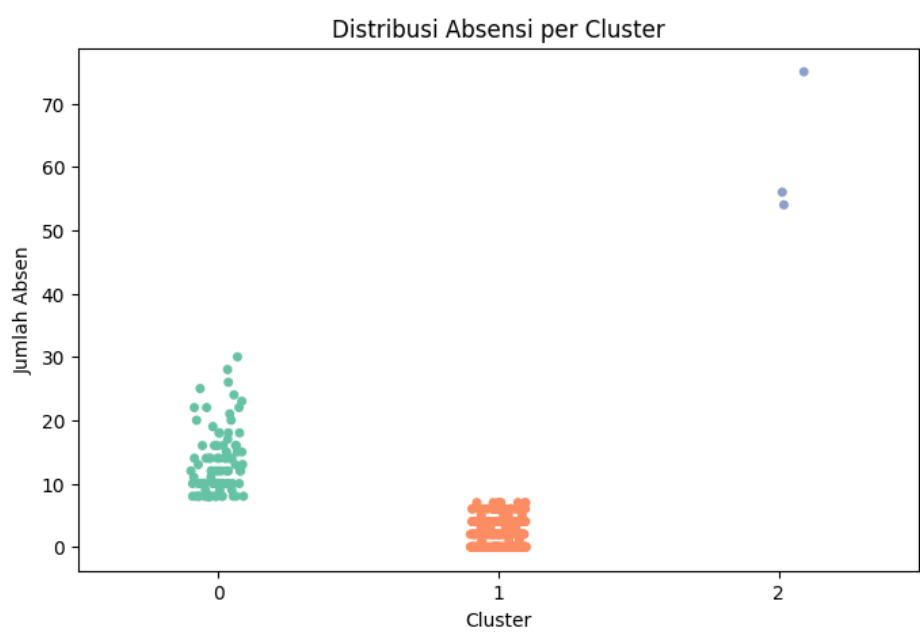
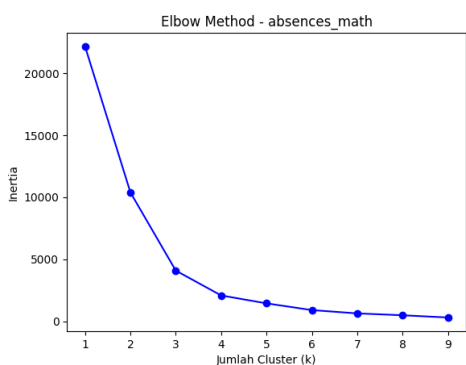
2. Regresi Linear → memprediksi nilai akhir matematika (G3_math) berdasarkan studytime_math dan absences_math. Hasil:

Koefisien	[0.52075709 0.02205333]	Mean Squared Error	21.70185631840226
Intercept	9.210888382028381	R ² Score	0.009622965842255082

Model sangat lemah dalam memprediksi nilai akhir. Waktu belajar dan absensi tidak cukup menjelaskan variasi nilai siswa.



3. Lakukan clustering segmentasi siswa berdasarkan absensi
- Digunakan Elbow Method dengan menghitung nilai inertia untuk k dari 1 hingga 9.
 - Dari grafik terlihat bahwa k = 3 adalah titik yang baik.
 - Siswa dikelompokkan ke dalam 3 kategori absensi: Cluster 0 = absensi rendah, Cluster 1 = absensi sedang, Cluster 2 = absensi tinggi.



4. Lakukan klasifikasi berdasarkan 3 variabel (bebas)!
- Model Logistic Regression digunakan untuk memprediksi apakah seorang siswa lulus matematika ($G3_math \geq 10$) berdasarkan:
- studytime_math: waktu belajar mingguan
 - failures_math: jumlah kegagalan sebelumnya

- absences_math: jumlah ketidakhadiran

Hasil Prediksi 5 siswa:

Studytime	Failures	Absences	Prediksi Lulus
2	0	6	1
2	0	4	1
2	3	10	0
3	0	2	1
2	0	4	1

Interpretasi:

Model memprediksi bahwa siswa dengan waktu belajar lebih tinggi dan tanpa kegagalan cenderung lulus. Sebaliknya, siswa dengan banyak kegagalan (seperti siswa ke-3) diprediksi tidak lulus meskipun absennya tidak terlalu tinggi.