



YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN

UNIVERSITAS SIBER ASIA

Kampus Menara, Jl. RM. Harsono, Ragunan - Jakarta Selatan. Daerah Khusus Ibukota Jakarta
12550. Telp. (+6221) 27806189. asiacyberuni@acu.ac.id. www.unsia.ac.id

LEMBAR JAWABAN

UJIAN AKHIR SEMESTER

SEMESTER GENAP TAHUN AJARAN 2024/2025

Mata Kuliah : Data Science

Kelas : IF405

Prodi : PJJ Informatika


Nama Mahasiswa : Fitri Rahayu Ningsih

NIM : 230401070301

Dosen : Alun Sujjada, S.Kom, M.T

Jawaban Ujian

Link drive file: <https://colab.research.google.com/drive/16tBwERtyQghnfR1r-dakgzpEY8rgoC2-?usp=sharing>

Nilai	Tanda Tangan Dosen Pengampu / Tutor	Tanda Tangan Mahasiswa
	(.....)	 (Fitri Rahayu Ningsih)
Diserahkan pada Tanggal:		Tanggal Mengumpulkan:

DATA SCIENCE
LAPORAN ANALYSIS STUDENT PERFORMANCE

Dosen Pengampu : Alun Sujjada, S.Kom, M.T



Disusun oleh:

Fitri Rahayu Ningsih (230401070301)

PROGRAM STUDI PJJ INFORMATIKA

UNIVERSITAS SIBER ASIA

JAKARTA

2025



YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN

UNIVERSITAS SIBER ASIA

Kampus Menara, Jl. RM. Harsono, Ragunan - Jakarta Selatan. Daerah Khusus Ibukota Jakarta
12550. Telp. (+6221) 27806189. asiacyberuni@acu.ac.id. www.unsia.ac.id

1. PENDAHULUAN

Proyek ini bertujuan untuk menganalisis data kinerja siswa dari dataset 'student-mat.csv' yang diperoleh dari UCI Machine Learning Repository. Data ini mencakup informasi demografis, dukungan pendidikan, kebiasaan belajar, dan nilai akhir siswa pada mata pelajaran matematika.

2. PEMBAHASAN

A. EDA & Deskripsi Dataset

Dataset berisi 395 siswa dan 33 kolom yang mencakup data demografi, pendidikan orang tua, dukungan belajar, aktivitas, dan nilai akademik.

Kolom	Deskripsi
school	Sekolah siswa (GP = Gabriel Pereira atau MS = Mousinho da Silveira)
sex	Jenis kelamin (F = Female, M = Male)
age	Usia siswa (15–22 tahun)
address	Alamat (U = Urban, R = Rural)
famsize	Ukuran keluarga (LE3 = ≤3 anggota, GT3 = >3 anggota)
Pstatus	Status tinggal bersama orang tua (T = bersama, A = terpisah)
Medu/Fedu	Pendidikan ibu/ayah (0=tidak sekolah, 4=pendidikan tinggi)
Mjob/Fjob	Pekerjaan ibu/ayah (teacher, health, services, at_home, other)
reason	Alasan memilih sekolah (home, reputation, course, other)
guardian	Wali utama (mother, father, other)
traveltime	Waktu tempuh ke sekolah (1–4: dari <15m sampai >1 jam)
studytime	Waktu belajar per minggu (1–4: dari <2 jam sampai >10 jam)
failures	Jumlah gagal dalam pelajaran sebelumnya
schoolsup	Dukungan tambahan dari sekolah (yes/no)
famsup	Dukungan dari keluarga (yes/no)
paid	Kursus tambahan untuk pelajaran tersebut (yes/no)
activities	Aktivitas ekstrakurikuler (yes/no)
nursery	Mengikuti TK (yes/no)
higher	Ingin melanjutkan ke pendidikan tinggi (yes/no)
internet	Akses internet di rumah (yes/no)
romantic	Memiliki hubungan romantis (yes/no)
famrel	Hubungan keluarga (1–5)
freetime	Waktu luang (1–5)
goout	Frekuensi keluar dengan teman (1–5)
Dalc	Konsumsi alkohol di hari kerja (1–5)
Walc	Konsumsi alkohol di akhir pekan (1–5)
health	Kondisi kesehatan (1=buruk, 5=sangat baik)
absences	Jumlah ketidakhadiran
G1, G2, G3	Nilai pada awal, pertengahan, dan akhir semester

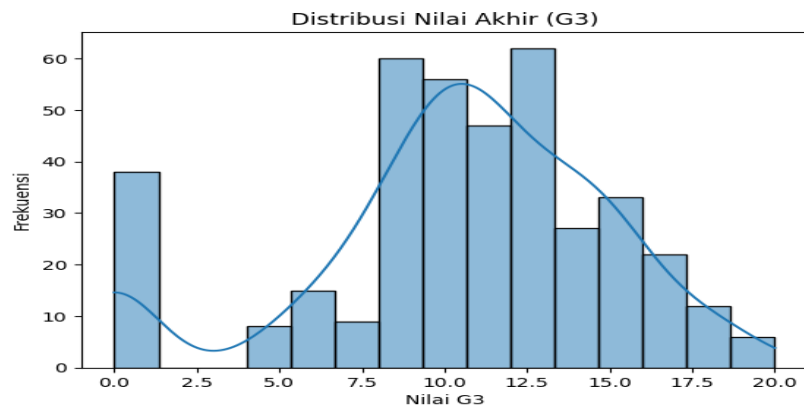
B. Visualisasi Distribusi Nilai Akhir (G3)

Sebagian besar siswa memiliki nilai akhir (G3) 10–14, menunjukkan distribusi terpusat di nilai menengah. Jumlah siswa menurun pada nilai ekstrem (>16 atau <6). Distribusi sedikit miring ke kiri, dengan lebih banyak siswa di nilai tinggi. Terdapat outlier di nilai 0, mungkin karena ketidakhadiran atau performa sangat buruk.



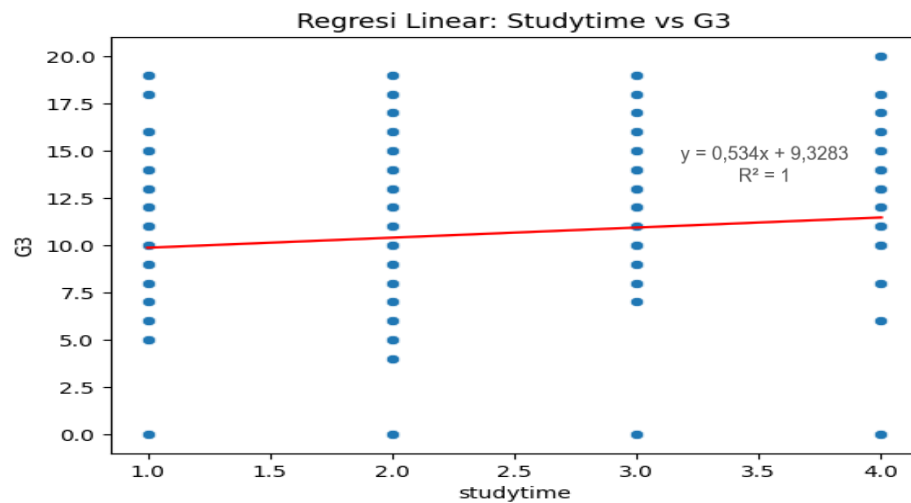
YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN UNIVERSITAS SIBER ASIA

Kampus Menara, Jl. RM. Harsono, Ragunan - Jakarta Selatan. Daerah Khusus Ibukota Jakarta
12550. Telp. (+6221) 27806189. asiacyberuni@acu.ac.id. www.unsia.ac.id



C. Regresi Linear

Analisis regresi dilakukan antara 'studytime' dan 'G3'. Hasil regresi menunjukkan adanya tren positif lemah, meskipun hubungan tidak terlalu kuat. Ini menunjukkan bahwa semakin banyak waktu belajar, nilai cenderung meningkat meskipun sedikit. Setiap peningkatan 1 level studytime, nilai G3 naik 0.534 poin.

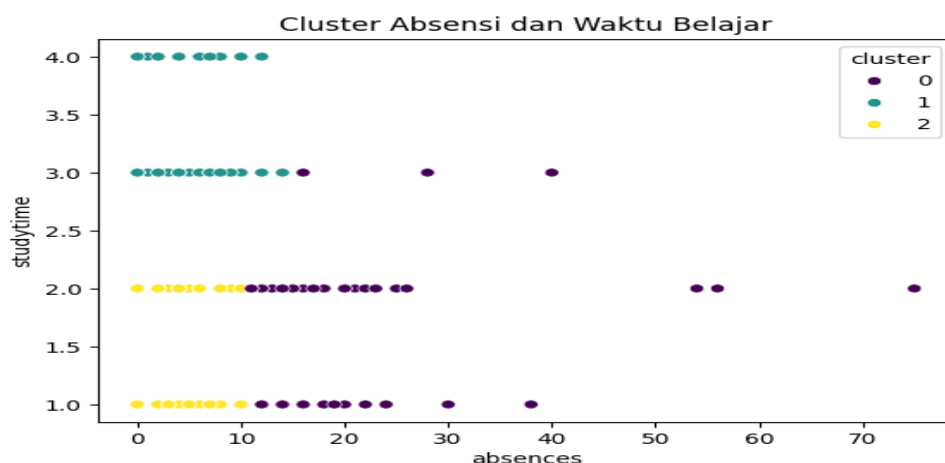


D. Clustering

Segmentasi siswa dilakukan berdasarkan 'absences' dan 'studytime' menggunakan metode KMeans. Hasil clustering menunjukkan tiga klaster berbeda

- Klaster 0: Absensi tinggi, waktu belajar rendah
- Klaster 1: Absensi sedang dan waktu belajar sedang ke tinggi
- Klaster 2: Absensi rendah, waktu belajar tinggi

Segmentasi ini membantu memahami perilaku siswa berdasarkan kebiasaan belajar dan kehadiran.

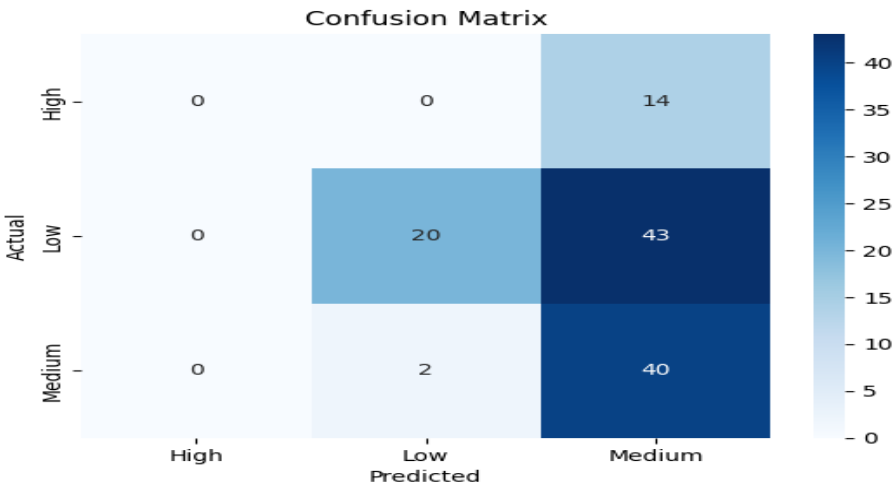




E. Klasifikasi

Pada analisis klasifikasi menggunakan model Decision Tree dengan tiga variabel bebas (studytime, failures, dan absences), model hanya mencapai akurasi sekitar 50%, yang menunjukkan performa rendah. Kelas "High" tidak berhasil diprediksi sama sekali (precision dan recall = 0), sementara kelas "Medium" mengalami overfitting dengan recall tinggi (95%) namun precision rendah (41%). Hal ini menunjukkan ketidakseimbangan kelas yang signifikan dan kurangnya fitur prediktif. Disarankan untuk menggunakan algoritma lain seperti Random Forest atau SVM, menambahkan fitur seperti G1 dan G2, serta menerapkan teknik penyeimbangan data seperti oversampling atau SMOTE.

Classification Report:				
	precision	recall	f1-score	support
High	0.00	0.00	0.00	14
Low	0.91	0.32	0.47	63
Medium	0.41	0.95	0.58	42
accuracy			0.50	119
macro avg	0.44	0.42	0.35	119
weighted avg	0.63	0.50	0.45	119



3. KESIMPULAN

Berdasarkan analisis data siswa, ditemukan bahwa sebagian besar siswa memiliki nilai akhir pada kisaran menengah (10–14), dengan waktu belajar dan tingkat absensi menjadi faktor yang berpengaruh. Model regresi linear menunjukkan bahwa studytime memiliki pengaruh terhadap nilai akhir, meskipun tidak cukup kuat jika berdiri sendiri. Melalui clustering, siswa berhasil dikelompokkan berdasarkan kebiasaan belajar dan kehadiran, yang dapat dimanfaatkan untuk intervensi yang lebih tepat sasaran. Namun, model klasifikasi Decision Tree menunjukkan performa yang rendah, terutama dalam memprediksi siswa dengan nilai tinggi. Oleh karena itu, disarankan untuk menambahkan fitur prediktif seperti G1 dan G2, menggunakan model ensemble seperti Random Forest, serta melakukan analisis lebih lanjut terhadap outlier seperti siswa dengan absensi ekstrem.