



YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN

UNIVERSITAS SIBER ASIA

Kampus Menara, Jl. RM. Harsono, Ragunan - Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550.
Telp. (+6221) 27806189. pmb@unsia.ac.id. <https://unsia.ac.id>

UJIAN AKHIR SEMESTER
SEMESTER GENAP TAHUN AJARAN 2024/2025

Nama : Alik Shodikin
NIM : 240401020026
Mata Kuliah : Data
Science Kelas : IF 405
Prodi : PJJ Informatika
Waktu : 14 Juli 2025 s/d 27 Juli
2025 Dosen : Alun Sujjada, S.Kom, M.T
Sifat Ujian : **Mandiri – Project Base**

SOAL UJIAN

KERJAKAN SOAL BERIKUT INI:

Unduhlah data set tentang “Student Performance” di:
<https://archive.ics.uci.edu/static/public/320/student+performance.zip>

Instruksi:

1. Lakukan EDA (Exploratory Data Analysis) pada data tersebut dan jelaskan juga fungsi masing-masing kolom!
2. Carilah 2 variabel (bebas) yang dapat dilakukan analisa menggunakan Regresi Linear.
3. Lakukan clustering segmentasi siswa berdasarkan absensi atau waktu belajar!
4. Lakukan klasifikasi berdasarkan 3 variabel (bebas)!

Output yang Diharapkan:

- File Jupyter Notebook atau Python script.
- Laporan singkat (maks. 3 halaman) dalam format PDF.
- Visualisasi yang menarik dan informatif.

-----Selamat Mengerjakan, Jangan Lupa Berdoa -----

JAWABAN :

1. EDA (Exploratory Data Analysis) dan Fungsi Masing2 Tabel :

a) **Ukuran Data Set :**

- Data Matematika: 395 siswa
- Data Bahasa Portugis: 649 siswa
- Data gabungan: 382 siswa (terlihat di skrip merge)

b) **Jenis Kelamin :**

- F (Perempuan) : 208 siswa /52,7 %
- M (Laki-laki) : 187 siswa / 47,3 %
- jumlah siswa perempuan sedikit lebih banyak dari laki-laki dalam pelajaran Matematik

c) **Umur :**

- Mayoritas siswa berusia 16–18 tahun, yang merupakan usia umum untuk jenjang SM

d) **Status Sosial :**

- Sebagian besar siswa tinggal dengan orang tua yang masih bersama.
- Banyak yang berasal dari keluarga besar.
- Orang tua sebagian memiliki pendidikan menengah–tinggi, terutama ibu.
- Mayoritas siswa mendapatkan dukungan pendidikan dari keluarga dan sekolah.
- Beberapa siswa berasal dari keluarga dengan pekerjaan formal (guru, kesehatan, layanan publik), yang menandakan kelas menengah

e) **Distribusi Nilai :**

- Sebagian siswa mengalami kesulitan akademik ($G3 < 10$)
- Pola kemajuan dari $G1 \rightarrow G2 \rightarrow G3$ dapat diamati
- Nilai bisa dipengaruhi oleh absensi, waktu belajar, dan dukungan keluarga

f) **Nilai dan Absensi :**

- Nilai Absensi berkisar dari 0 hingga 93 hari
- Sebagian besar siswa memiliki absen 0–10 hari
- Ada sebagian kecil siswa dengan absen ekstrem (30+ hari), yang kemungkinan mengalami gangguan belajar
- Kehadiran sangat mempengaruhi performa akademik.
- Siswa dengan tingkat absen rendah cenderung memiliki nilai akhir lebih tinggi.
- Analisis ini bisa digunakan untuk intervensi dini pada siswa yang sering absen

❖ **Fungsi Masing-Masing Tabel :**

No	Kolom	Tipe Data	Fungsi / Penjelasan
1	school	Kategorikal	Asal sekolah siswa: 'GP' (Gabriel Pereira) atau 'MS' (Mousinho da Silveira)
2	sex	Kategorikal	Jenis kelamin siswa: 'F' (perempuan) atau 'M' (laki-laki)
3	age	Numerik	Umur siswa (15–22 tahun)
4	address	Kategorikal	Jenis tempat tinggal: 'U' (urban) atau 'R' (rural)
5	famsize	Kategorikal	Ukuran keluarga: 'LE3' (≤ 3 anggota) atau 'GT3' (> 3 anggota)
6	Pstatus	Kategorikal	Status tinggal bersama orang tua: 'T' (together) atau 'A' (apart)
7	Medu	Numerik	Pendidikan ibu: 0 (tidak sekolah) sampai 4 (perguruan tinggi)
8	Fedu	Numerik	Pendidikan ayah: 0 sampai 4
9	Mjob	Kategorikal	Pekerjaan ibu: 'teacher', 'health', 'services', 'at_home', atau 'other'
10	Fjob	Kategorikal	Pekerjaan ayah: seperti 'Mjob'
11	reason	Kategorikal	Alasan memilih sekolah: 'home', 'reputation', 'course', atau 'other'
12	guardian	Kategorikal	Wali siswa: 'mother', 'father', atau 'other'
13	traveltime	Numerik	Waktu perjalanan ke sekolah: 1–4

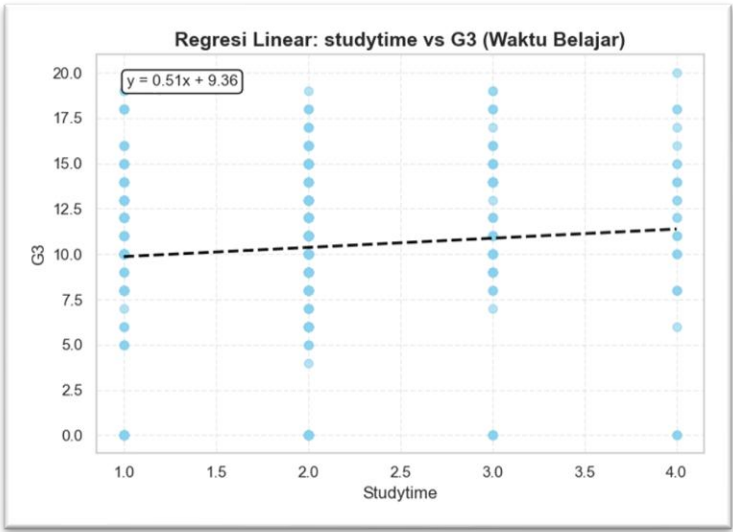
No	Kolom	Tipe Data	Fungsi / Penjelasan
14	studytime	Numerik	Waktu belajar per minggu: 1–4
15	failures	Numerik	Jumlah kegagalan kelas sebelumnya: 0–3 (lebih dari 3 digabung di 3)
16	schoolsup	Biner	Dukungan tambahan di sekolah: 'yes' atau 'no'
17	famsup	Biner	Dukungan dari keluarga: 'yes' atau 'no'
18	paid	Biner	Les tambahan berbayar: 'yes' atau 'no'
19	activities	Biner	Mengikuti kegiatan ekstrakurikuler: 'yes' atau 'no'
20	nursery	Biner	Pernah ikut pendidikan prasekolah: 'yes' atau 'no'
21	higher	Biner	Ingin melanjutkan pendidikan tinggi: 'yes' atau 'no'
22	internet	Biner	Akses internet di rumah: 'yes' atau 'no'
23	romantic	Biner	Menjalin hubungan romantis: 'yes' atau 'no'
24	famrel	Skala 1–5	Kualitas hubungan keluarga: 1 (buruk) hingga 5 (baik)
25	freetime	Skala 1–5	Waktu luang di luar sekolah
26	goout	Skala 1–5	Frekuensi keluar dengan teman
27	Dalc	Skala 1–5	Konsumsi alkohol di hari kerja
28	Walc	Skala 1–5	Konsumsi alkohol di akhir pekan
29	health	Skala 1–5	Status kesehatan saat ini
30	absences	Numerik	Jumlah ketidakhadiran
31	G1	Nilai (0–20)	Nilai ujian periode pertama
32	G2	Nilai (0–20)	Nilai ujian periode kedua
33	G3	Nilai (0–20)	Nilai akhir siswa (target)

2. Berikut dua pasang variabel dari dataset siswa yang cocok untuk dianalisis dengan Regresi Linea :

a) Studytime dan G3(Nilai Akhir)

Penjelasan :

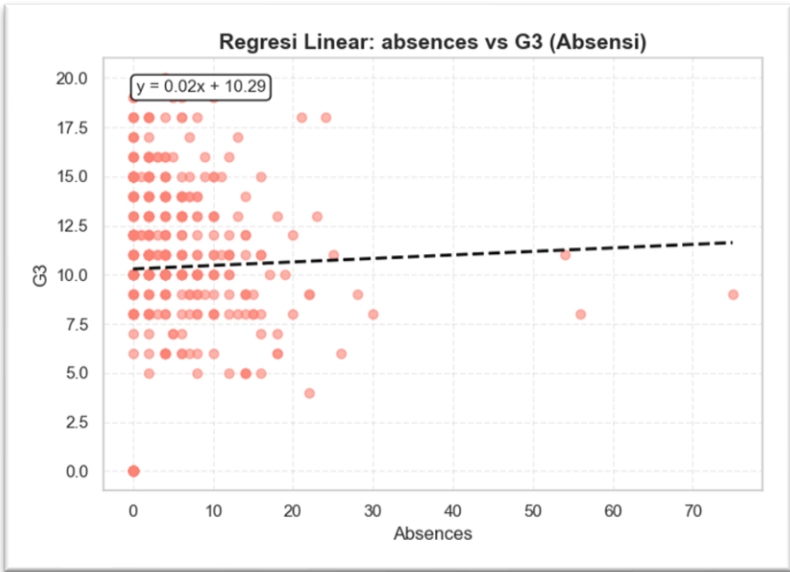
- Semakin besar nilainya, semakin lama waktu yang diinvestasikan siswa untuk belajar.
- Secara logis dan empiris, waktu belajar memang mempengaruhi hasil belajar.
- Jadi, Semakin lama waktu belajar siswa setiap minggu, semakin tinggi kemungkinan mereka mendapatkan nilai akhir yang baik



b) Absensi dan G3(Nilai Akhir)

Penjelasan :

- Kehadiran berhubungan langsung dengan paparan materi pelajaran, latihan, dan interaksi kelas.
- Ketidakhadiran yang tinggi sering menjadi indikator masalah akademik, pribadi, atau kesehatan.
- Jadi, Ketidakhadiran merupakan faktor risiko nyata terhadap kegagalan akademi



3. Berikut clustering segmentasi siswa berdasarkan absensi atau waktu belajar:

a) Cluster 0 / Siswa disiplin dan tekun

Penjelasan :

- Absensi Rendah (Jarang Bolos)
- Waktu Belajar tinggi
- Kemungkinan memiliki nilai akhir tinggi
- Disiplin dan terstruktur
- Menjadi contoh bagi siswa lain

b) Cluster 1 / Siswa Beresiko

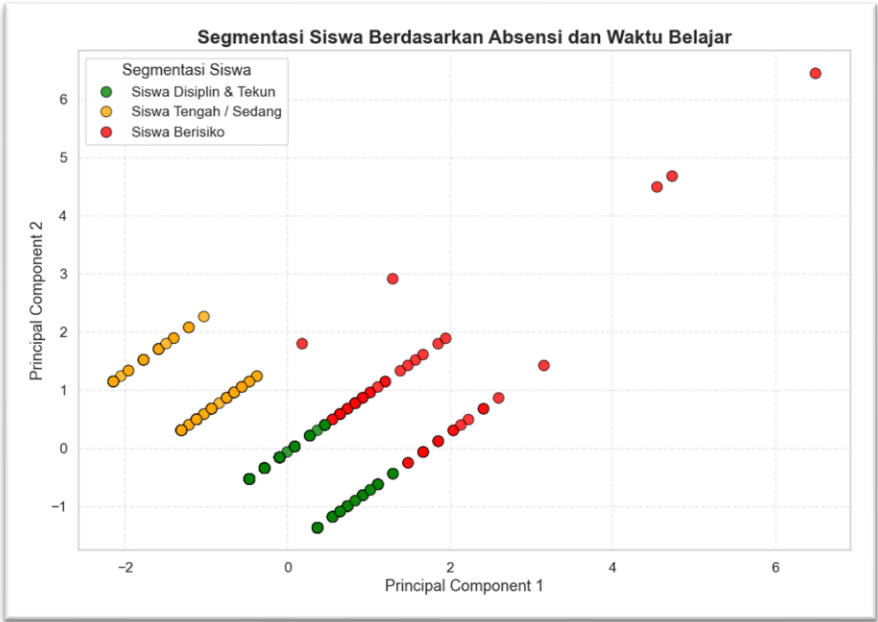
Penjelasan :

- Absensi tinggi (Sering Bolos)
- Waktu Belajar rendah
- Kemungkinan memiliki nilai akhir rendah
- Perlu intervensi khusus, seperti bimbingan belajar, konseling, atau dukungan keluarga

c) Cluster 2 / Siswa Tengah atau Sedang

Penjelasan :

- Absensi sedang
- Waktu Belajar sedang
- Performa naik atau turun tergantung faktor pendukung lain.



4. Berikut 3 Klasifikasi yang telah saya buat :

a) **Studytime / waktu lama belajar**

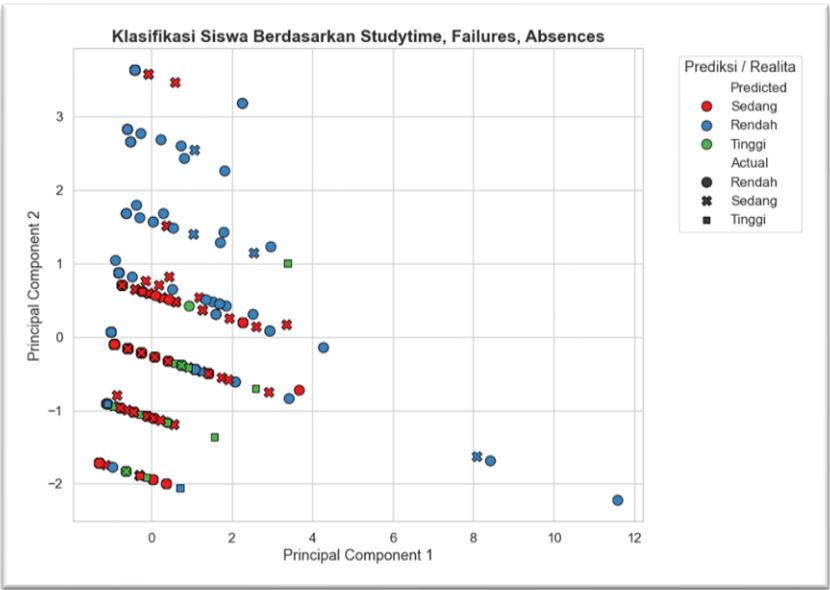
mendukung prediksi nilai tinggi jika digunakan bersama dengan failures.


b) **Failures / jumlah kegagalan sebelumnya**

kemungkinan adalah fitur paling kuat untuk memprediksi apakah siswa akan mendapat nilai rendah.

c) **Absences / jumlah ketidakhadiran**

memberi informasi tambahan, tapi efeknya bisa bervariasi tergantung konteks (misalnya: alasan absen)



Diverifikasi Oleh:			Disusun Oleh:
Penjaminan Mutu  Abdu Rahman, S.E., M.Ak.	Ketua Program Studi  (Syahid Abullah, S.Si., M.Kom)	Koordinator Mata Kuliah  (Syahid Abdullah, S.Si, M.Kom)	Dosen Pengampu  (Alun Sujjada, S.Kom, M.T)
Disahkan Tanggal: 13 Juli 2025			Dibuat Tanggal: 13 Juli 2025