



YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN  
**UNIVERSITAS SIBER ASIA**

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id). <https://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** : Catur Wibowo  
**NIM** : 240401020025  
**Dosen** : Alun Sujjada, S.Kom., M.T

## 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.



# YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN UNIVERSITAS SIBER ASIA

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id). <https://unsia.ac.id>

---

## JAWABAN

### 1. EDA dan Penjelasan Kolom

Dataset berisi 395 siswa sekolah menengah dengan 33 kolom, mencakup data demografi, sosial-ekonomi, perilaku belajar, dan hasil akademik (G3). Contoh distribusi nilai akhir (G3): Mayoritas nilai berada di tengah (8-14). Kolom penting antara lain: absences (absensi), studytime (waktu belajar), failures (kegagalan), G1-G3 (nilai)..

#### ▪ Ringkasan Data :

- **Jumlah data:** 395 siswa
- **Jumlah kolom:** 33 variabel
- **Tipe data:** 16 numerik, 17 kategorik

#### ▪ Fungsi Masing masing kolom

- **school:** Sekolah siswa ("GP"/"MS")
- **sex:** Jenis kelamin siswa ("F"/"M")
- **age:** Usia siswa (15-22 tahun)
- **address:** Tipe alamat rumah ("U"/"R")
- **famsize:** Ukuran keluarga ("LE3"/"GT3")
- **Pstatus:** Status tinggal orang tua ("T"/"A")
- **Medu:** Pendidikan ibu (0-4, semakin tinggi semakin terdidik)
- **Fedu:** Pendidikan ayah (0-4)
- **Mjob:** Pekerjaan ibu (teacher/health/services/at\_home/other)
- **Fjob:** Pekerjaan ayah (teacher/health/services/at\_home/other)
- **reason:** Alasan memilih sekolah (home/reputation/course/other)
- **guardian:** Wali utama siswa (mother/father/other)
- **traveltime:** Waktu tempuh ke sekolah (1-4, makin besar makin lama)
- **studytime:** Waktu belajar per minggu (1-4, makin besar makin lama)
- **failures:** Jumlah kegagalan kelas sebelumnya (0-4)
- **schoolsup:** Dukungan pendidikan tambahan di sekolah (yes/no)
- **famsup:** Dukungan pendidikan dari keluarga (yes/no)
- **paid:** Kelas tambahan membayar pada mata pelajaran ini (yes/no)
- **activities:** Kegiatan ekstrakurikuler (yes/no)
- **nursery:** Pernah ikut prasekolah (yes/no)
- **higher:** Keinginan melanjutkan pendidikan tinggi (yes/no)
- **internet:** Akses internet di rumah (yes/no)
- **romantic:** Memiliki hubungan romantis (yes/no)
- **famrel:** Kualitas hubungan keluarga (1-5)
- **freetime:** Waktu luang setelah sekolah (1-5)
- **goout:** Frekuensi keluar bersama teman (1-5)
- **Dalc:** Konsumsi alkohol hari kerja (1-5)
- **Walc:** Konsumsi alkohol akhir pekan (1-5)



# YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN UNIVERSITAS SIBER ASIA

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id). <https://unsia.ac.id>

- **health:** Status kesehatan saat ini (1-5)
- **absences:** Jumlah ketidakhadiran
- **G1:** Nilai ujian periode 1 (0-20)
- **G2:** Nilai ujian periode 2 (0-20)
- **G3:** Nilai ujian akhir (0-20)
  
- **Statistik Deskripsi Singkat**
  - **Rata-rata usia:** ~16.7 tahun
  - **Mayoritas tinggal di kota ("U")**
  - **Rata-rata absensi:** 5.7 kali, maksimum 75
  - **Rata-rata nilai akhir (G3):** 10.4 (skala 0-20)
  - **Mayoritas punya keinginan melanjutkan ke perguruan tinggi**
  - **Mayoritas hubungan keluarga baik (famrel  $\geq 4$ )**

## 2. Dua Variabel untuk Analisis Regresi Linear

Dua variabel bebas yang sangat relevan:

- **studytime (waktu belajar mingguan) dan G3 (nilai akhir)**
  - Hipotesis: Semakin lama waktu belajar, semakin tinggi nilai akhir.
- **absences (absensi) dan G3 (nilai akhir)**
  - Hipotesis: Semakin banyak absen, semakin rendah nilai akhir.
- **Script cek korelasi**

```
import matplotlib.pyplot as plt

plt.scatter(df['studytime'], df['G3'])

plt.xlabel('Study Time')

plt.ylabel('Final Grade (G3)')

plt.title('Study Time vs Final Grade')

plt.show()

plt.scatter(df['absences'], df['G3'])

plt.xlabel('Absences')

plt.ylabel('Final Grade (G3)')

plt.title('Absences vs Final Grade')

plt.show()
```

## 3. Clustering Segmentasi Siswa Berdasarkan Absensi atau Waktu Belajar

### a. Segmentasi Berdasarkan Absensi

Menggunakan KMeans (dengan 3 cluster misal):

```
from sklearn.cluster import KMeans

X = df[['absences']].values kmeans = KMeans(n_clusters=3, random_state=0).fit(X)

df['cluster_absen'] = kmeans.labels_
```



# YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN UNIVERSITAS SIBER ASIA

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id). <https://unsia.ac.id>

- **Cluster 0:** Siswa dengan absensi rendah (rajin hadir)
- **Cluster 1:** Siswa dengan absensi sedang
- **Cluster 2:** Siswa dengan absensi tinggi (sering absen )

**Rata-rata nilai G3 per cluster dapat dibandingkan untuk analisa perilaku siswa.**

## b. Segmentasi Berdasarkan Waktu Belajar

Alternatif lain pakai kolom studytime, proses sama, dan bisa mengelompokkan:

- **Cluster 0:** Siswa dengan waktu belajar rendah
- **Cluster 1:** Sedang
- **Cluster 2:** Tinggi

## 4. Klasifikasi Berdasarkan 3 Variabel Bebas

### Contoh Klasifikasi: Prediksi Lulus/Gagal Berdasarkan 3 Variabel

Target: **Lulus** jika  $G3 \geq 10$ , **Gagal** jika  $G3 < 10$

Fitur: **studytime, failures, absences**

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

df['pass'] = (df['G3'] >= 10).astype(int)
X = df[['studytime', 'failures', 'absences']]
y = df['pass']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
clf = RandomForestClassifier(random_state=42)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
print(classification_report(y_test, y_pred))
```

### Penjelasan:

- Klasifikasi dilakukan berdasarkan 3 variabel: waktu belajar, jumlah kegagalan, dan absensi.
- Hasil model dapat memberikan metrik akurasi, precision, recall, dan f1-score

## 5. Output

**Hasil Script Jupyter Notebook : ( student\_analysis.py) ketika dijalankan akan mengeksport file .png untuk tampilan visualisasi**



YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN

# UNIVERSITAS SIBER ASIA

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id), <https://unsia.ac.id>

```
Administration Command Prompt
C:\Users\user\Desktop\jupyter project\data science\python student_analysis.py
===== EDA SINGKAT =====
In [ ]:
class 'pandas.core.frame.DataFrame'
RangeIndex: 395 entries, 0 to 394
Data columns (total 33 columns):
#   Column      Non-Null Count  Dtype
---  -
0   school      395 non-null    object
1   sex         395 non-null    object
2   age         395 non-null    int64
3   address     395 non-null    object
4   families    395 non-null    object
5   status      395 non-null    object
6   medu        395 non-null    int64
7   fedu        395 non-null    int64
8   mjob        395 non-null    object
9   fjob        395 non-null    object
10  reason      395 non-null    object
11  guardian    395 non-null    object
12  traveltime  395 non-null    int64
13  studytime   395 non-null    int64
14  failures    395 non-null    int64
15  schoolsup   395 non-null    object
16  famsup      395 non-null    object
17  paid        395 non-null    object
18  activities  395 non-null    object
19  nursery     395 non-null    object
20  higher      395 non-null    object
21  internet    395 non-null    object
22  romantic    395 non-null    object
23  famrel      395 non-null    int64
24  freetime    395 non-null    int64
25  genst       395 non-null    int64
26  dalc        395 non-null    int64
27  walc        395 non-null    int64
28  health      395 non-null    int64
29  absences    395 non-null    int64
30  G1          395 non-null    int64
31  G2          395 non-null    int64
32  G3          395 non-null    int64
dtypes: int64(16), object(17)
memory usage: 102.0+ KB
None
None
None
count 395.000000 395.000000 395.000000 395.000000 ... 395.000000 395.000000 395.000000 395.000000
mean 16.696203 2.740367 2.525519 1.448101 ... 5.798861 10.908861 10.713924 10.415196
std 1.276641 1.094735 1.088201 0.697505 ... 0.800896 3.319101 3.781501 4.581441
min 15.000000 0.000000 0.000000 1.000000 ... 0.000000 3.000000 0.000000 0.000000
25% 16.000000 2.000000 2.000000 1.000000 ... 0.000000 0.000000 0.000000 0.000000
50% 17.000000 3.000000 2.000000 1.000000 ... 4.000000 11.000000 11.000000 11.000000
75% 18.000000 4.000000 2.000000 2.000000 ... 8.000000 13.000000 13.000000 14.000000
max 22.000000 4.000000 4.000000 4.000000 ... 75.000000 15.000000 19.000000 20.000000

[8 rows x 16 columns]
# school sex age address families status medu fedu ... genst dalc walc health absences G1 G2 G3
0 GP F 18 U GT3 A 4 4 ... 4 1 1 2 6 6 6 6
1 GP F 17 U GT3 T 1 1 ... 3 1 1 2 4 5 5 6
2 GP F 15 U LT3 T 1 1 ... 2 2 1 3 10 7 9 10
3 GP F 15 U GT3 T 4 2 ... 2 1 1 5 2 15 14 15
4 GP F 16 U GT3 T 3 3 ... 2 1 2 5 4 6 10 10

[5 rows x 33 columns]
Regresi Studytime-G3, koefisien: 0.53, Intercept: 9.33
Regresi Absences-G3, koefisien: 0.02, Intercept: 10.30
Confusion Matrix:
[[ 9 18]
 [ 4 46]]

Classification Report:
      precision    recall  f1-score   support

      0       0.60      0.33      0.43       27
      1       0.72      0.88      0.79       52

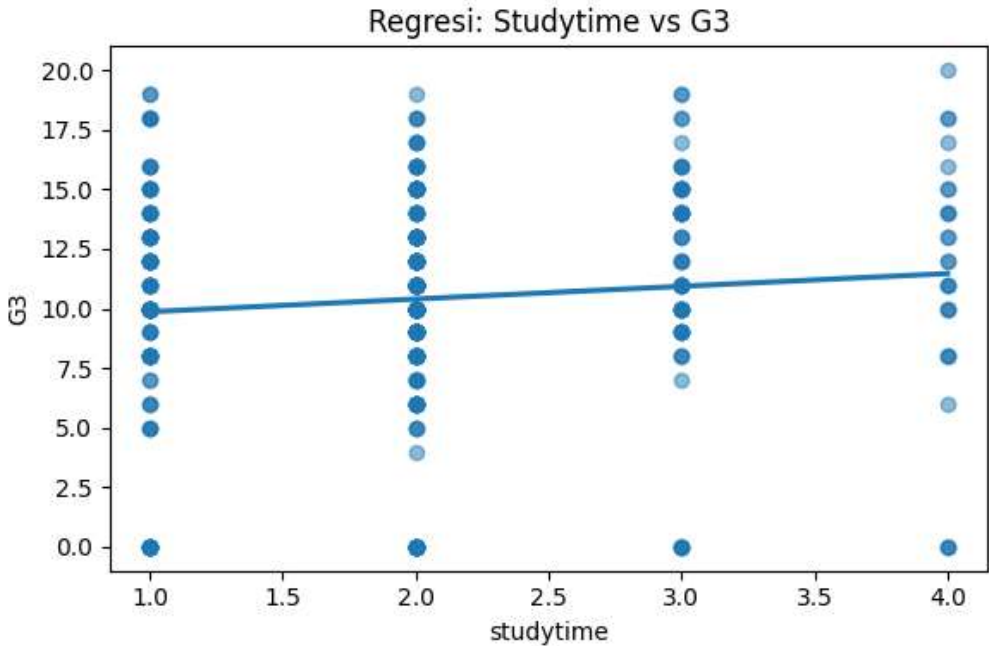
 accuracy      0.66      0.61      0.70       79
 macro avg     0.66      0.61      0.61       79
weighted avg     0.68      0.78      0.67       79

Salah satu plot sudah tersimpan ke file PNG.
C:\Users\user\Desktop\jupyter project\data science\

```

Hasil ekspor dari scrypt python menjadi gambar visualisasi:

a. Studytime vs G3 ( reg\_studytime\_g3.png )



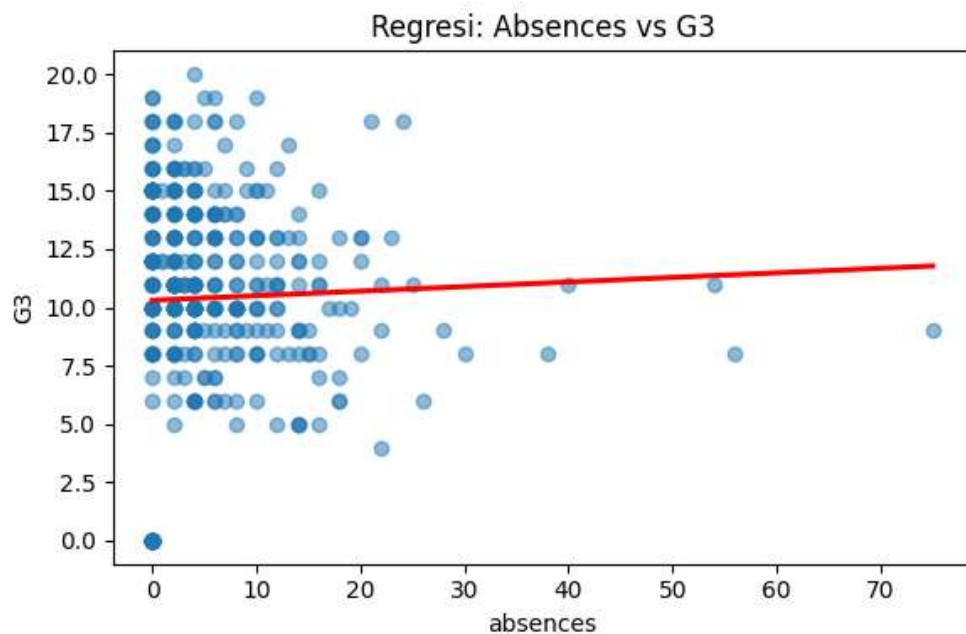


## YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN UNIVERSITAS SIBER ASIA

Menara UNAS, Jl. Harsono RM No.1 9, RT.5/RW.4, Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12550. [pmb@unsia.ac.id](mailto:pmb@unsia.ac.id). <https://unsia.ac.id>

- Visualisasi menggambarkan hubungan antara waktu belajar mingguan (studytime) dengan nilai akhir (G3).
- Garis regresi biru memperlihatkan kecenderungan naik, meski tidak terlalu tajam, menunjukkan hubungan positif antara studytime dan G3.
- Interpretasi:  
Semakin tinggi kategori waktu belajar, cenderung nilai akhir lebih tinggi. Namun, efeknya tidak sangat kuat, karena sebaran data masih cukup besar. Artinya, waktu belajar memang berpengaruh positif, tetapi bukan satu-satunya faktor penentu nilai akhir

### b. Absences vs G3 ( reg\_absences\_g3.png )



- Visualisasi menunjukkan hubungan antara jumlah absensi (absences) dengan nilai akhir siswa (G3).
- Garis regresi merah hampir datar atau sedikit menanjak, artinya tidak ada korelasi negatif yang signifikan antara absensi dan nilai akhir pada data ini. Bahkan, secara statistik sederhana, garis cenderung naik (korelasi sangat lemah positif), namun hal ini biasanya tidak bermakna secara praktis—kemungkinan besar disebabkan oleh outlier (misal, ada siswa yang absen sangat banyak namun nilainya tetap baik).
- Interpretasi:  
Dalam data ini, jumlah absensi bukan prediktor yang kuat terhadap penurunan nilai akhir. Namun, perlu analisis lebih dalam apakah ada faktor lain yang mempengaruhi.



**Kesimpulan :**

- Waktu belajar (studytime) lebih konsisten berkorelasi positif dengan nilai akhir, walaupun pengaruhnya sedang/lemah.
- Absensi (absences) pada data ini tidak berkorelasi negatif signifikan terhadap nilai akhir—bisa jadi faktor motivasi, outlier, atau kualitas belajar di luar kelas turut mempengaruhi.
- Saran: Fokus untuk mendorong peningkatan waktu belajar efektif pada siswa, serta memonitor faktor lain (misal kegagalan sebelumnya, dukungan keluarga, dll.) dalam upaya meningkatkan prestasi

**Link File script utuh ( student\_analysis.py ) dan file pendukung sebagai berikut :**

**Gdrive :**

<https://drive.google.com/drive/folders/12fSeKASBstBSgiEegNU1hw0a70mUInQG?usp=sharing>

**Github :**

<https://github.com/catoermela/student-performance-eda>

Nilai	Tanda Tangan Dosen Pengampu	Tanda Tangan Mahasiswa
	( Alun Sujjada, S.Kom., M.T)	(Catur Wibowo)
Diserahkan pada Tanggal:		Tanggal Mengumpulkan: