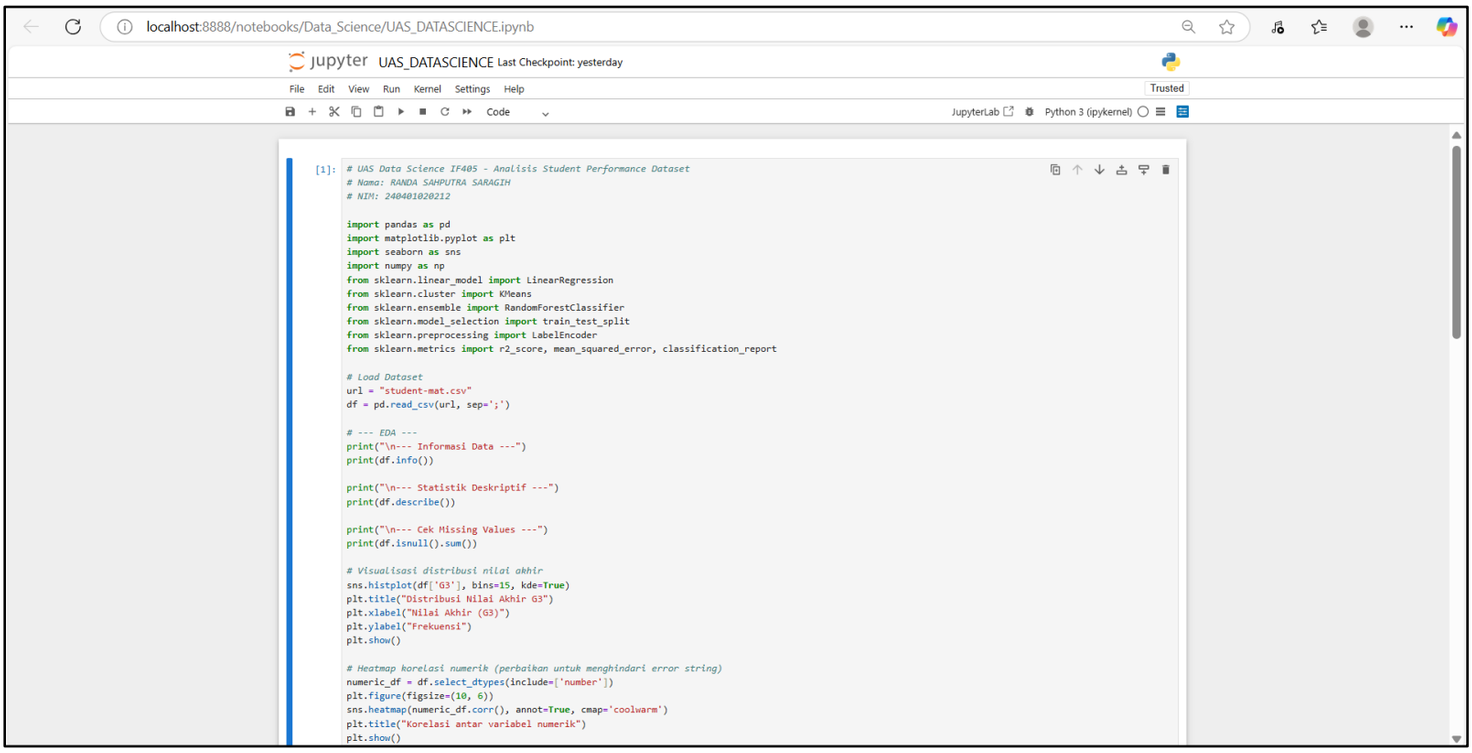
YAYASAN MEMAJUKAN ILMU DAN KEBUDAYAAN
UNIVERSITAS SIBER ASIA
Kampus Menara, Jl. RM. Harsono, Ragunan - Jakarta Selatan.Daerah Khusus Ibukota Jakarta 12550. Telp. (+6221) 27806189. asiacyberuni@acu.ac.id. www.unsia.ac.id

LEMBAR JAWABAN
UJIAN AKHIR SEMESTER
SEMESTER GENAP TAHUN AJARAN 2024/2025

Mata Kuliah	: Data Science
Kelas	: IF405
Prodi	: PJJ Inforamatika
Nama Mahasiswa	: Randa Sahputra Saragih
NIM	: 240401020212
Dosen	: Alun Sujjada, S.Kom., M.T

LAMPIRAN KODE PROGRAM




```
# --- REGRESI LINEAR ---
X = df[['G1', 'studytime']]
y = df['G3']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
pred = model.predict(X_test)

print("\n--- Regresi Linear ---")
print("R^2 Score:", r2_score(y_test, pred))
mse = mean_squared_error(y_test, pred)
rmse = np.sqrt(mse)
print("RMSE:", rmse)

plt.scatter(X_test['G1'], y_test, color='blue', label='Actual')
plt.scatter(X_test['G1'], pred, color='red', label='Predicted')
plt.xlabel("G1")
plt.ylabel("G3")
plt.title("Regresi Linear G1 vs G3")
plt.legend()
plt.show()
```

```python
# --- CLUSTERING ---
X_cluster = df[['absences', 'studytime']]
kmeans = KMeans(n_clusters=3, random_state=0, n_init=10)
df['cluster'] = kmeans.fit_predict(X_cluster)

plt.figure(figsize=(8, 6))
sns.scatterplot(x='absences', y='studytime', hue='cluster', data=df, palette='viridis')
plt.title("Clustering Siswa berdasarkan Absensi dan Waktu Belajar")
plt.xlabel("Absences")
plt.ylabel("Study Time")
plt.show()

# --- KLASIFIKASI ---
df['pass'] = (df['G3'] > 10).astype(int)
df['sex'] = LabelEncoder().fit_transform(df['sex'])

X_class = df[['sex', 'studytime', 'failures']]
y_class = df['pass']

X_train, X_test, y_train, y_test = train_test_split(X_class, y_class, test_size=0.3, random_state=1)
model = RandomForestClassifier(random_state=1)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print("\n--- Klasifikasi ---")
print(classification_report(y_test, y_pred))
```

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

LAMPIRAN OUTPUT

jupyter UAS_DATASCIENCE Last Checkpoint: yesterday

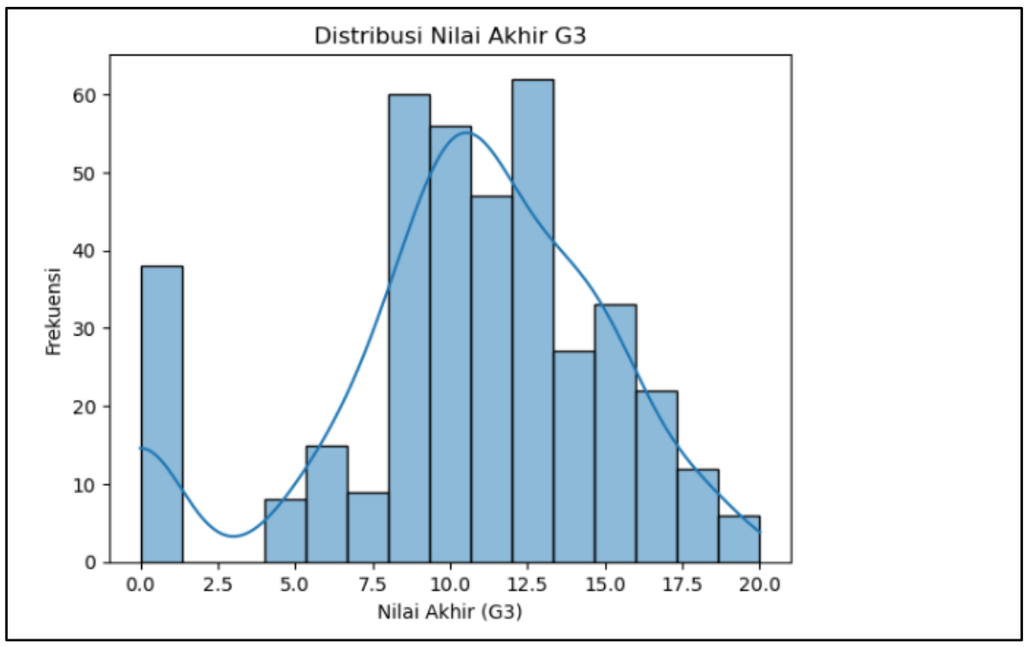
File Edit View Run Kernel Settings Help

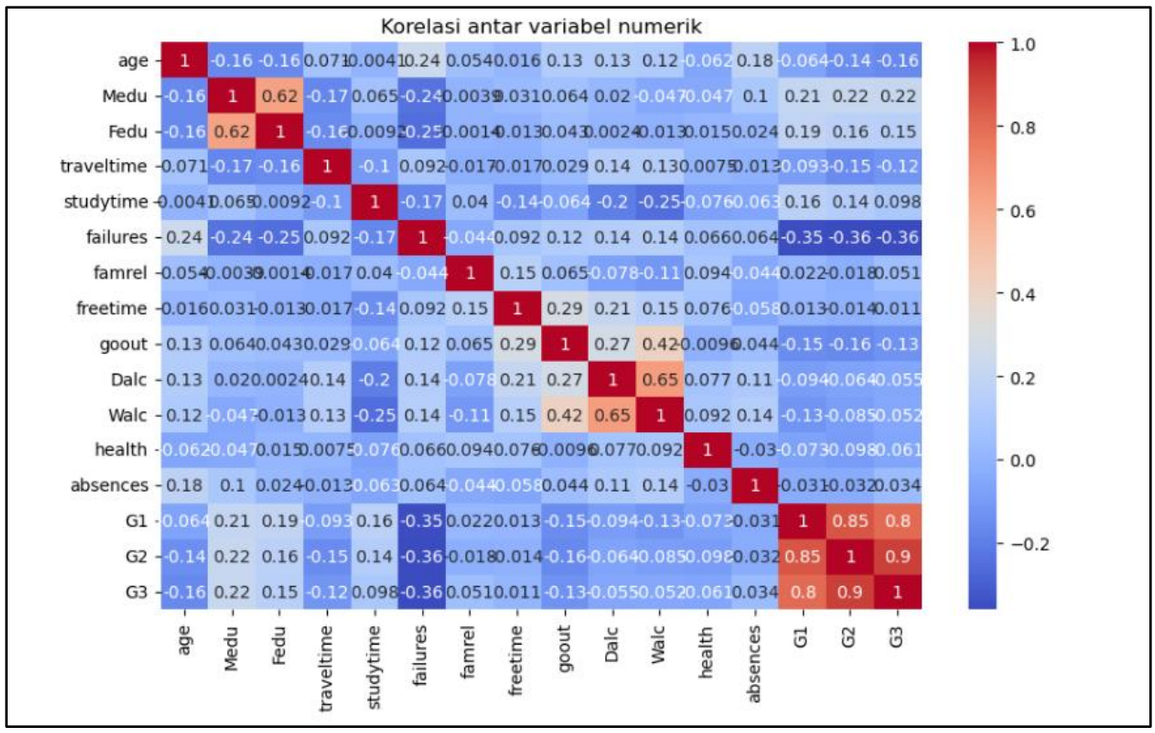
Code

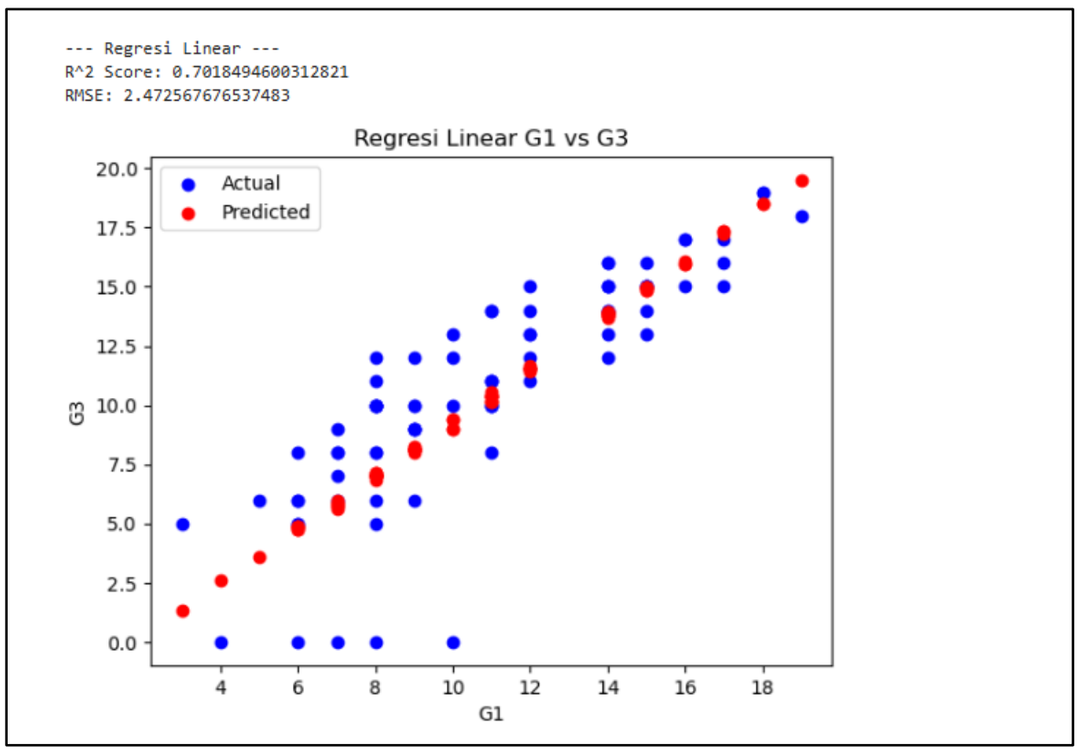
```
--- Informasi Data ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395 entries, 0 to 394
Data columns (total 33 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   school      395 non-null    object
 1   sex         395 non-null    object
 2   age         395 non-null    int64
 3   address     395 non-null    object
 4   famsize     395 non-null    object
 5   Pstatus     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  goout       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
```

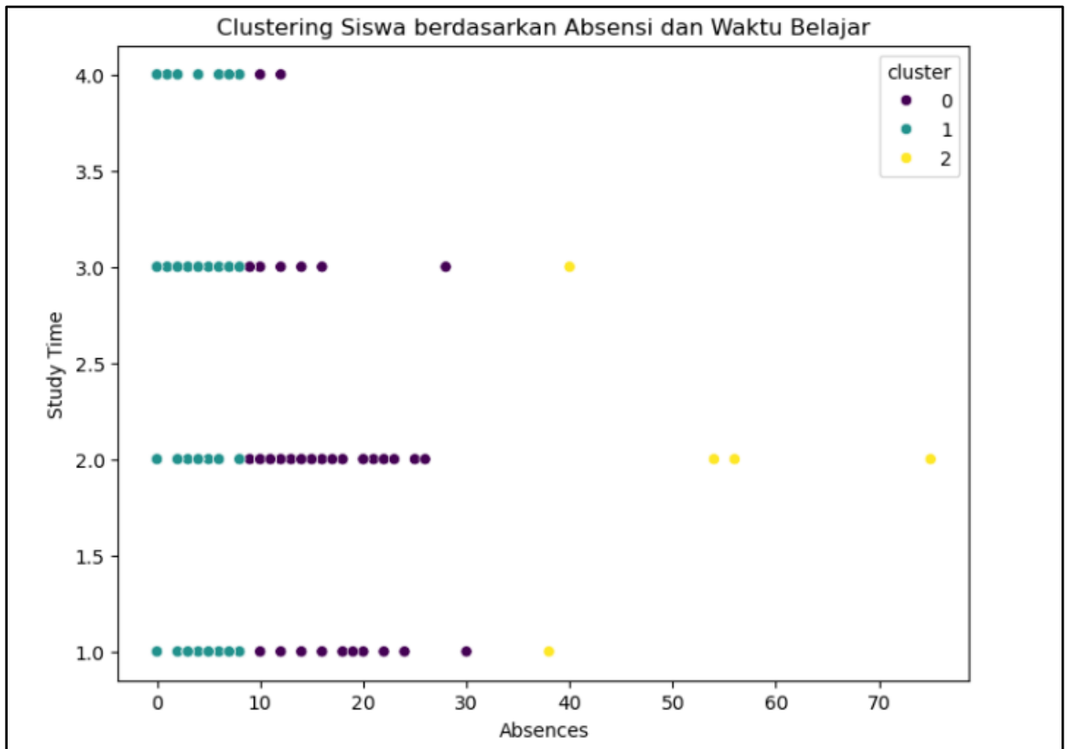
```
--- Statistik Deskriptif ---
              age        Medu        Fedu  traveltime   studytime    failures  \
count  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000
mean    16.696203    2.749367    2.521519    1.448101    2.035443    0.334177
std      1.276043    1.094735    1.088201    0.697505    0.839240    0.743651
min     15.000000    0.000000    0.000000    1.000000    1.000000    0.000000
25%     16.000000    2.000000    2.000000    1.000000    1.000000    0.000000
50%     17.000000    3.000000    2.000000    1.000000    2.000000    0.000000
75%     18.000000    4.000000    3.000000    2.000000    2.000000    0.000000
max     22.000000    4.000000    4.000000    4.000000    4.000000    3.000000

           famrel    freetime       goout        Dalc        Walc      health  \
count  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000
mean     3.944304    3.235443    3.108861    1.481013    2.291139    3.554430
std      0.896659    0.998862    1.113278    0.890741    1.287897    1.390303
min      1.000000    1.000000    1.000000    1.000000    1.000000    1.000000
25%      4.000000    3.000000    2.000000    1.000000    1.000000    3.000000
50%      4.000000    3.000000    3.000000    1.000000    2.000000    4.000000
75%      5.000000    4.000000    4.000000    2.000000    3.000000    5.000000
max      5.000000    5.000000    5.000000    5.000000    5.000000    5.000000

         absences          G1          G2          G3
count  395.000000  395.000000  395.000000  395.000000
mean     5.708861   10.908861   10.713924   10.415190
std      8.003096    3.319195    3.761505    4.581443
min      0.000000    3.000000    0.000000    0.000000
25%      0.000000    8.000000    9.000000    8.000000
50%      4.000000   11.000000   11.000000   11.000000
75%      8.000000   13.000000   13.000000   14.000000
max     75.000000   19.000000   19.000000   20.000000
```

```
--- Cek Missing Values ---
school        0
sex           0
age           0
address       0
famsize       0
Pstatus       0
Medu          0
Fedu          0
Mjob          0
Fjob          0
reason        0
guardian      0
traveltime    0
studytime     0
failures      0
schoolsup     0
famsup        0
paid          0
activities    0
nursery       0
higher        0
internet      0
romantic      0
famrel        0
freetime      0
goout         0
Dalc          0
Walc          0
health        0
absences      0
G1            0
G2            0
G3            0
dtype: int64
```




Korelasi antar variabel numerik
age
Medu
Fedu
traveltime
studytime
failures
famrel
freetime
goout
Dalc
Walc
health
absences
G1
G2
G3
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


--- Regresi Linear ---
R^2 Score: 0.7018494600312821
RMSE: 2.472567676537483
Regresi Linear G1 vs G3
Actual
Predicted
G3
G1
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
4
6
8
10
12
14
16
18




```
--- Klasifikasi ---
              precision    recall  f1-score   support

           0       0.63      0.54      0.58        59
           1       0.60      0.68      0.64        60

    accuracy                           0.61       119
   macro avg       0.62      0.61      0.61       119
weighted avg       0.62      0.61      0.61       119
```