Tugas UTS Machine Learning

Nama : Egidius Edi Putrawan Halawa

Nim : 231011403453

Kelas : 05TPLE005

1. Pilih dataset dengan variabel target kategorikal (misal: Iris, Titanic, atau dataset pilihan sendiri).
   Titanic dataset

titanic.csv > data

```
PassengerId,Survived,Pclass,Lname,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,Braund, Mr. Owen Harris,male,22,1,0,A/5 21171,7.25,,S
2,1,1,Cumings, Mrs. John Bradley (Florence Briggs Thayer),female,38,1,0,PC 17599,71.2833,C85
3,1,3,Heikkinen, Miss. Laina,female,26,0,0,STON/O2. 3101282,7.925,,S
4,1,1,Futrelle, Mrs. Jacques Heath (Lily May Peel),female,35,1,0,113803,53.1,C123,S
5,0,3,Allen, Mr. William Henry,male,35,0,0,373450,8.05,,S
6,0,3,Moran, Mr. James,male,,0,0,330877,8.4583,,Q
7,0,1,McCarthy, Mr. Timothy J,male,54,0,0,17463,51.8625,E46,S
8,0,3,Palsson, Master. Gosta Leonard,male,2,3,1,349909,21.075,,S
9,1,3,Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg),female,27,0,2,347742,11.1333,,S
10,1,2,Nasser, Mrs. Nicholas (Adele Achem),female,14,1,0,237736,30.0708,,C
11,1,3,Sandstrom, Miss. Marguerite Rut,female,4,1,1,PP 9549,16.7,G6,S
12,1,1,Bonnell, Miss. Elizabeth,female,58,0,0,113783,26.55,C103,S
13,0,3,Saundercock, Mr. William Henry,male,20,0,0,A/5. 2151,8.05,,S
14,0,3,Andersson, Mr. Anders Johan,male,39,1,5,347082,31.275,,S
15,0,3,Vestrom, Miss. Hulda Amanda Adolfina,female,14,0,0,350406,7.8542,,S
16,1,2,Hewlett, Mrs. (Mary D Kingcome) ,female,55,0,0,248706,16,,S
17,0,3,Rice, Master. Eugene,male,2,4,1,382652,29.125,,Q
18,1,2,Williams, Mr. Charles Eugene,male,,0,0,244373,13,,S
19,0,3,Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele),female,31,1,0,345763,18,,S
20,1,3,Masselmani, Mrs. Fatima,female,,0,0,2649,7.225,,C
21,0,2,Fynney, Mr. Joseph J,male,35,0,0,239865,26,,S
22,1,2,Beesley, Mr. Lawrence,male,34,0,0,248698,13,D56,S
```

2. Lakukan EDA dan preprocessing seperti pada bagian regresi.
   Eda
   ===== INFORMASI DATASET =====
   <class 'pandas.core.frame.DataFrame'>
   RangeIndex: 156 entries, 0 to 155
   Data columns (total 13 columns):
    #   Column   Non-Null Count Dtype
   --- ------   -------------- -----

```
 0   PassengerId  156 non-null    int64
 1   Survived     156 non-null    int64
 2   Pclass       156 non-null    int64
 3   Lname        156 non-null    object
 4   Name         156 non-null    object
 5   Sex          156 non-null    object
 6   Age          126 non-null    float64
 7   SibSp        156 non-null    int64
 8   Parch        156 non-null    int64
 9   Ticket       156 non-null    object
 10  Fare         156 non-null    float64
 11  Cabin        31 non-null     object
 12  Embarked     155 non-null    object
dtypes: float64(2), int64(5), object(6)
memory usage: 16.0+ KB
None

===== 5 DATA PERTAMA =====
   PassengerId  Survived  Pclass      Lname  ...            Ticket     Fare  Cabin  Embarked
0            1         0       3     Braund  ...         A/5 21171   7.2500    NaN         S
1            2         1       1    Cumings  ...          PC 17599  71.2833    C85         C
2            3         1       3  Heikkinen  ...  STON/O2. 3101282   7.9250    NaN         S
3            4         1       1   Futrelle  ...            113803  53.1000   C123         S
4            5         0       3      Allen  ...            373450   8.0500    NaN         S

[5 rows x 13 columns]

===== DESKRIPSI STATISTIK =====
        PassengerId    Survived      Pclass      Lname  ... Ticket        Fare  Cabin  Embarked
count    156.000000  156.000000  156.000000        156  ...    156  156.000000     31       155
unique          NaN         NaN         NaN        141  ...    145         NaN     28         3
top             NaN         NaN         NaN  Andersson  ... 113803         NaN   C123         S
freq            NaN         NaN         NaN          4  ...      2         NaN      2       110
mean      78.500000    0.346154    2.423077        NaN  ...    NaN   28.109587    NaN       NaN
std       45.177428    0.477275    0.795459        NaN  ...    NaN   39.401047    NaN       NaN
min        1.000000    0.000000    1.000000        NaN  ...    NaN    6.750000    NaN       NaN
25%       39.750000    0.000000    2.000000        NaN  ...    NaN    8.003150    NaN       NaN
50%       78.500000    0.000000    3.000000        NaN  ...    NaN   14.454200    NaN       NaN
75%      117.250000    1.000000    3.000000        NaN  ...    NaN   30.371850    NaN
NaN
```

```
max       156.000000    1.000000    3.000000       NaN  ...     NaN  263.000000    NaN
NaN

[11 rows x 13 columns]

===== JUMLAH DATA KOSONG =====
PassengerId      0
Survived         0
Pclass           0
Lname            0
Name             0
Sex              0
Age             30
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          125
Embarked         1
dtype: int64

 Processing:
===== DATA AWAL =====
   PassengerId  Survived  Pclass      Lname  ...           Ticket     Fare  Cabin  Embarked
0            1         0       3     Braund  ...        A/5 21171   7.2500    NaN         S
1            2         1       1    Cumings  ...         PC 17599  71.2833    C85         C
2            3         1       3  Heikkinen  ...  STON/O2. 3101282   7.9250    NaN         S
3            4         1       1   Futrelle  ...           113803  53.1000   C123         S
4            5         0       3      Allen  ...           373450   8.0500    NaN         S

===== DATASET HASIL PROCESSING =====
Jumlah data training : 124
Jumlah data testing  : 32

Contoh data setelah scaling:
   PassengerId    Pclass       Sex       Age     SibSp     Parch      Fare  Embarked
0    -1.720983  0.727607  0.748331 -0.437089  0.365311 -0.458217 -0.531122  0.607468
1    -1.698777 -1.794764 -1.336306  0.783472  0.365311 -0.458217  1.099279 -1.853960
2    -1.676571  0.727607 -1.336306 -0.131949 -0.584497 -0.458217 -0.513935  0.607468
3    -1.654365 -1.794764 -1.336306  0.554617  0.365311 -0.458217  0.636300  0.607468
```

4   -1.632158 0.727607 0.748331 0.554617 -0.584497 -0.458217 -0.510753 0.607468

titanic_processed.csv > data

```
PassengerId,Pclass,Sex,Age,SibSp,Parch,Fare,Embarked,Survived
-1.7209832769351496,0.7276068751089992,0.7483314773547882,-0.43708918814008696,0.36531076810
-1.6987770411037284,-1.794763625268864,-1.336306209562122,0.7834724551227902,0.3653107681076
-1.6765708052723072,0.7276068751089992,-1.336306209562122,-0.1319487773243677,-0.58449722897
-1.654364569440886,-1.794763625268864,-1.336306209562122,0.5546171470110007,0.36531076810761
-1.6321583336094645,0.7276068751089992,0.7483314773547882,0.5546171470110007,-0.584497228972
-1.6099520977780433,0.7276068751089992,0.7483314773547882,-0.1319487773243677,-0.58449722897
-1.587745861946622,-1.794763625268864,0.7483314773547882,2.004034098385667,-0.58449722897217
-1.5655396261152006,0.7276068751089992,0.7483314773547882,-1.9627912422186833,2.264926762267
-1.5433333902837794,0.7276068751089992,-1.336306209562122,-0.055663674620437864,-0.584497228
-1.5211271544523581,-0.5335783750799323,-1.336306209562122,-1.0473700097715255,0.36531076810
-1.498920918620937,0.7276068751089992,-1.336306209562122,-1.8102210368108236,0.3653107681076
-1.4767146827895155,-1.794763625268864,-1.336306209562122,2.3091745092013864,-0.584497228972
-1.4545084469580942,0.7276068751089992,0.7483314773547882,-0.5896593935479466,-0.58449722897
-1.432302211126673,0.7276068751089992,0.7483314773547882,0.8597575578267199,0.36531076810761
-1.4100959752952518,0.7276068751089992,-1.336306209562122,-1.0473700097715255,-0.58449722897
-1.3878897394638303,-0.5335783750799323,-1.336306209562122,2.0803192010895972,-0.58449722897
-1.365683503632409,0.7276068751089992,0.7483314773547882,-1.9627912422186833,3.2147347593469
-1.3434772678009879,-0.5335783750799323,0.7483314773547882,-0.1319487773243677,-0.5844972289
-1.3212710319695666,0.7276068751089992,-1.336306209562122,0.2494767361952814,0.3653107681076
-1.2990647961381452,0.7276068751089992,-1.336306209562122,-0.1319487773243677,-0.58449722897
-1.276858560306724,-0.5335783750799323,0.7483314773547882,0.5546171470110007,-0.584497228972
-1.2546523244753027,-0.5335783750799323,0.7483314773547882,0.4783320443070709,-0.58449722897
```

3. Gunakan minimal dua algoritma klasifikasi, misalnya: a. Logistic Regression b. Decision Tree c. K-Nearest Neighbors d. Support Vector Machine

===== DATASET AWAL =====

	PassengerId	Survived	Pclass	Lname	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund	Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings	Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen	Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle	Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12

Data berhasil diproses dan dibagi menjadi train/test.

===== Logistic Regression =====

Akurasi : 0.6875

Classification Report:

	precision	recall	f1-score	support
0	0.76	0.76	0.76	21
1	0.55	0.55	0.55	11
accuracy			0.69	32
macro avg	0.65	0.65	0.65	32
weighted avg	0.69	0.69	0.69	32

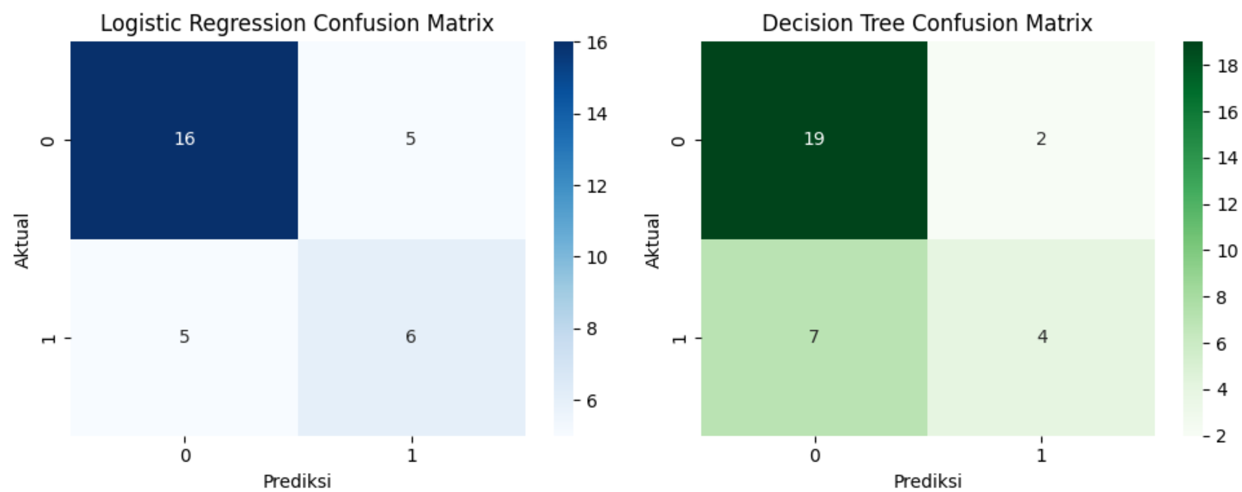
===== Decision Tree =====

Akurasi : 0.7188

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.90	0.81	21
1	0.67	0.36	0.47	11

	precision	recall	f1-score	support
accuracy			0.72	32
macro avg	0.70	0.63	0.64	32
weighted avg	0.71	0.72	0.69	32



	Model	Akurasi
0	Logistic Regression	0.68750
1	Decision Tree	0.71875

4. Lakukan evaluasi model menggunakan: a. Confusion Matrix b. Accuracy, Precision, Recall, F1-score c. ROC Curve (jika memungkinkan)
   ===== Logistic Regression =====
   Accuracy : 0.6875
   Precision : 0.5455
   Recall : 0.5455
   F1-score : 0.5455

Confusion Matrix:
[[16 5]
[5 6]]

Classification Report:

	precision	recall	f1-score	support
0	0.76	0.76	0.76	21
1	0.55	0.55	0.55	11
accuracy			0.69	32
macro avg	0.65	0.65	0.65	32
weighted avg	0.69	0.69	0.69	32

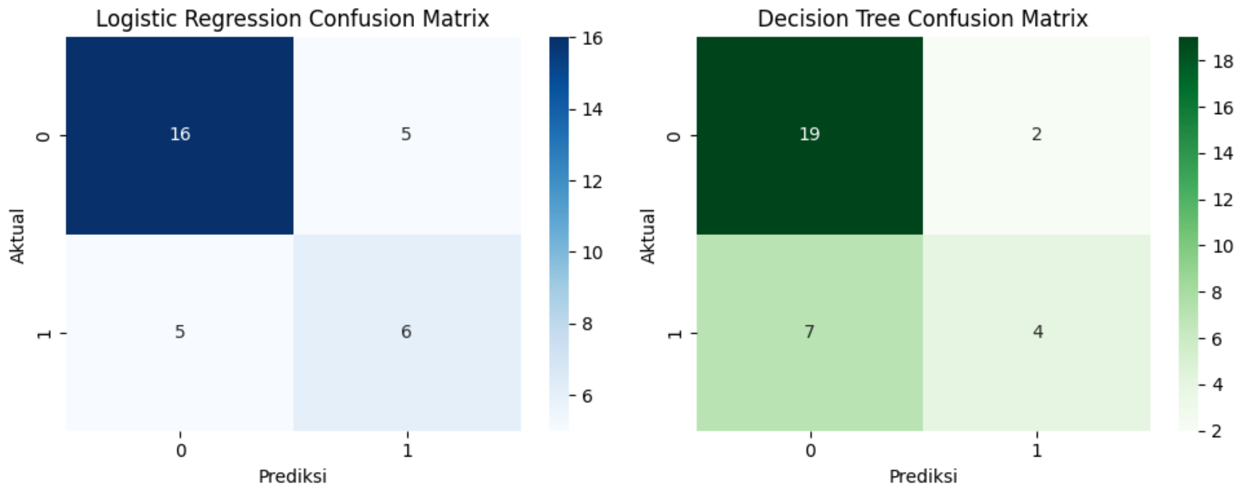
======================================

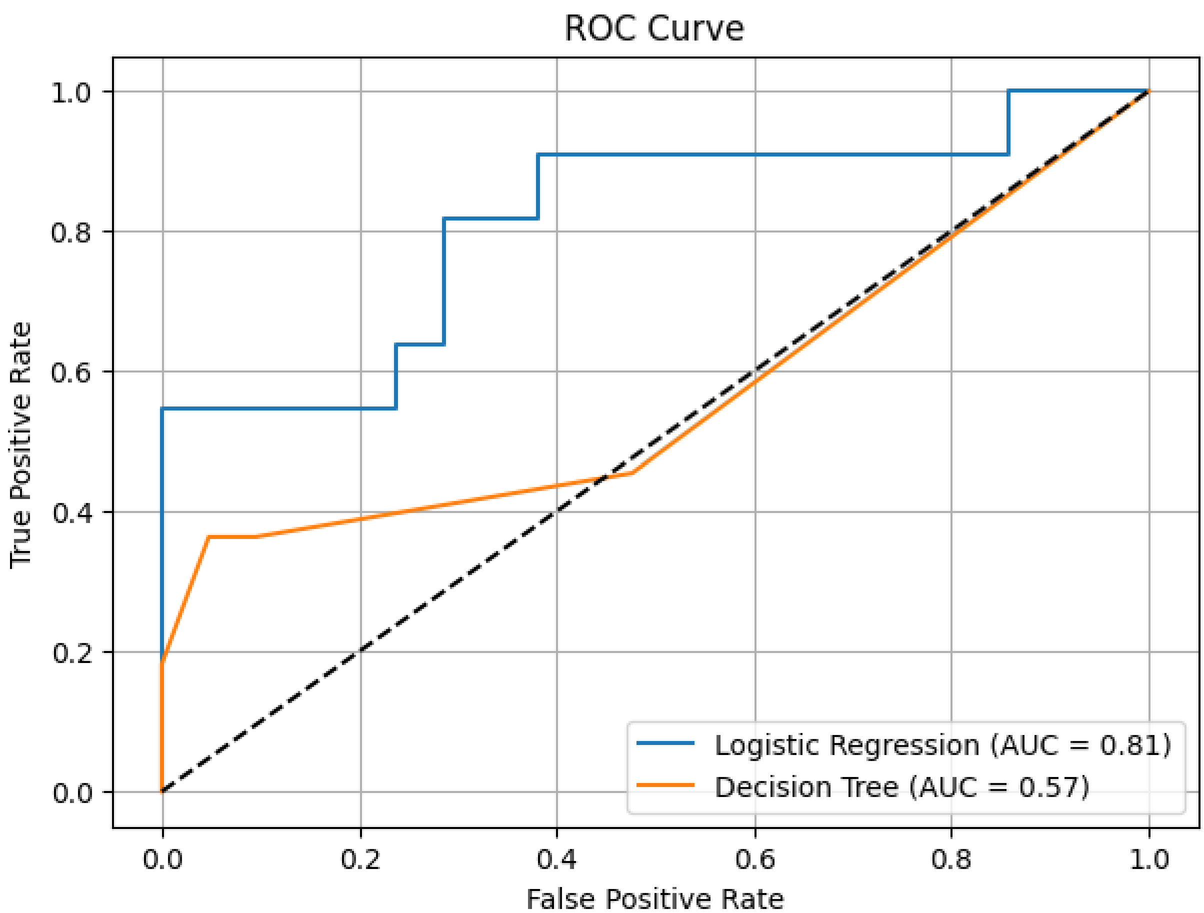
===== Decision Tree =====
Accuracy : 0.7188
Precision : 0.6667
Recall   : 0.3636
F1-score : 0.4706

Confusion Matrix:
[[19 2]
[7 4]]

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.90	0.81	21
1	0.67	0.36	0.47	11
accuracy			0.72	32
macro avg	0.70	0.63	0.64	32
weighted avg	0.71	0.72	0.69	32


Logistic Regression Confusion Matrix
Decision Tree Confusion Matrix
Aktual
Prediksi
16
5
5
6
19
2
7
4


ROC Curve
True Positive Rate
False Positive Rate
Logistic Regression (AUC = 0.81)
Decision Tree (AUC = 0.57)

5. Bandingkan hasil antar model dan tulis kesimpulan.
   Logistic Regression
   Hasil evaluasi:
   - Accuracy = 0.6875
   - Precision = 0.5455
   - Recall = 0.5455
   - F1-score = 0.5455
   - AUC = 0.81

   Interpretasi:
   - Logistic Regression memiliki performa cukup seimbang antara prediksi benar untuk kelas 0 (tidak selamat) dan kelas 1 (selamat).
   - Nilai AUC 0.81 menunjukkan model ini cukup baik membedakan antara penumpang yang selamat dan tidak.
   - Namun, nilai accuracy 68.75% masih menunjukkan beberapa kesalahan prediksi (terutama false negatives & false positives yang cukup banyak di confusion matrix).

   Decision Tree
   Hasil evaluasi:
   - Accuracy = 0.7188
   - Precision = 0.6667
   - Recall = 0.3636
   - F1-score = 0.4706
   - AUC = 0.57

   Interpretasi:
   - Model ini memiliki akurasi sedikit lebih tinggi (71.88%) dibanding Logistic Regression.
   - Namun, recall rendah (0.36) artinya banyak kasus “selamat” yang tidak berhasil dikenali (False Negative tinggi).
   - AUC 0.57 menunjukkan kemampuan pemisahan kelasnya lemah, mendekati prediksi acak.

   Perbandingan Kedua Model

Metrik	Logistic Regression	Decision Tree
Accuracy	0.6875	0.7188
Precision	0.5455	0.6667
Recall	0.5455	0.3636
F1-Score	0.5455	0.4706

Metrik	Logistic Regression	Decision Tree
AUC	0.81	0.57

Interpretasi umum:

- Decision Tree lebih akurat secara keseluruhan, tapi lebih buruk dalam mendeteksi penumpang yang selamat (Recall rendah).
- Logistic Regression lebih seimbang, terutama karena memiliki AUC lebih tinggi (0.81) yang menunjukkan kemampuan klasifikasi lebih baik secara keseluruhan.
- Decision Tree kemungkinan overfitting ringan terhadap data pelatihan (karena akurasi tinggi tapi AUC rendah).

Kesimpulan Akhir

🔹 Logistic Regression lebih direkomendasikan untuk kasus Titanic ini,
karena meskipun akurasinya sedikit lebih rendah, model ini memiliki kemampuan generalisasi yang lebih baik (AUC tinggi dan metrik seimbang).

🔹 Decision Tree cocok jika ingin interpretasi yang lebih mudah, tapi hasilnya kurang stabil dan kurang baik dalam mendeteksi kelas minoritas (selamat).

UNIVERSITAS PAMULANG
KARTU UJIAN TENGAH SEMESTER GANJIL 2025/2026
NOMOR UJIAN : 01258024135336

FAKULTAS / PRODI : ILMU KOMPUTER / TEKNIK INFORMATIKA S1

NAMA MAHASISWA : EGIDIUS EDI PUTRAWAN HALAWA

NIM : 231011403453

SHIFT : REGULER C

No	Hari/ Tanggal	Waktu	Ruang	Kelas	Mata Kuliah	Paraf
1	Sabtu, 1 Nov 2025	07.40 - 09.20	V.314	05TPLE005	KECERDASAN BUATAN	1
2	Sabtu, 1 Nov 2025	07.40 - 09.20	V.314	05TPLE005	METODE PENELITIAN	2
3	Sabtu, 1 Nov 2025	09.20 - 11.00	V.314	05TPLE005	SISTEM INFORMASI MANAJEMEN	3
4	Sabtu, 1 Nov 2025	09.20 - 11.00	V.314	05TPLE005	PEMROGRAMAN WEB I	4
5	Sabtu, 1 Nov 2025	11.00 - 13.50	V.314	05TPLE005	PENGOLAHAN CITRA DIGITAL	5
6	Sabtu, 1 Nov 2025	13.50 - 15.30	V.314	05TPLE005	DIGITAL ENTREPRENEURSHIP	6
7	Sabtu, 1 Nov 2025	13.50 - 15.30	V.314	05TPLE005	MACHINE LEARNING	7
8	Sabtu, 1 Nov 2025	16.00 - 17.40	V.314	05TPLE005	TEKNIK RISET OPERASIONAL	8

Peraturan dan Tata Tertib Peserta Ujian

1. Peserta ujian harus berpakaian rapi, sopan dan memakai jaket Almamater
2. Peserta ujian sudah berada di ruangan sepuluh menit sebelum ujian dimulai
3. Peserta ujian yang terlambat diperkenankan mengikuti ujian setelah mendapat ijin, tanpa perpanjangan waktu
4. Peserta ujian hanya diperkenankan membawa alat-alat yang ditentukan oleh panitia ujian
5. Peserta ujian dilarang membantu teman, mencontoh dari teman dan tindakan-tindakan lainnya yang mengganggu peserta ujian lain
6. Peserta ujian yang melanggar tata tertib ujian dikenakan sanksi akademik

Tangerang Selatan, 28 Oktober 2025
Ketua Panitia Ujian

Dr. Ubaid Al Faruq, S.Pd., M.Pd.
NIDN. 0418028702