

TUGAS UTS MACHINE LEARNING

NAMA : NURDIANSYAH

NIM : 231011400211

KELAS : 05TPLE005

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

Titanic csv

```
titanic.csv > data
1   PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
2   1,0,3, "Braund, Mr. Owen Harris", male, 22, 1, 0, A/5 21171, 7.25, , S
3   2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thayer)", female, 38, 1, 0, PC 17599, 71.2833
4   3,1,3, "Heikkinen, Miss. Laina", female, 26, 0, 0, STON/O2. 3101282, 7.925, , S
5   4,1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", female, 35, 1, 0, 113803, 53.1, C123, S
6   5,0,3, "Allen, Mr. William Henry", male, 35, 0, 0, 373450, 8.05, , S
7   6,0,3, "Moran, Mr. James", male, , 0, 0, 330877, 8.4583, , Q
8   7,0,1, "McCarthy, Mr. Timothy J", male, 54, 0, 0, 17463, 51.8625, E46, S
9   8,0,3, "Palsson, Master. Gosta Leonard", male, 2, 3, 1, 349909, 21.075, , S
10  9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)", female, 27, 0, 2, 347742, 11.1333, S
11  10,1,2, "Nasser, Mrs. Nicholas (Adele Achem)", female, 14, 1, 0, 237736, 30.0708, , C
12  11,1,3, "Sandstrom, Miss. Marguerite Rut", female, 4, 1, 1, PP 9549, 16.7, G6, S
13  12,1,1, "Bonnell, Miss. Elizabeth", female, 58, 0, 0, 113783, 26.55, C103, S
14  13,0,3, "Saundercok, Mr. William Henry", male, 20, 0, 0, A/5. 2151, 8.05, , S
15  14,0,3, "Andersson, Mr. Anders Johan", male, 39, 1, 5, 347082, 31.275, , S
16  15,0,3, "Vestrom, Miss. Hulda Amanda Adolfina", female, 14, 0, 0, 350406, 7.8542, , S
17  16,1,2, "Hewlett, Mrs. (Mary D Kingcome) ", female, 55, 0, 0, 248706, 16, , S
18  17,0,3, "Rice, Master. Eugene", male, 2, 4, 1, 382652, 29.125, , Q
19  18,1,2, "Williams, Mr. Charles Eugene", male, , 0, 0, 244373, 13, , S
20  19,0,3, "Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)", female, 31, 1, 0, 345763, 18, ,
21  20,1,3, "Masselmani, Mrs. Fatima", female, , 0, 0, 2649, 7.225, , C
22  21,0,2, "Fynney, Mr. Joseph J", male, 35, 0, 0, 239865, 26, , S
23  22,1,2, "Beesley, Mr. Lawrence", male, 34, 0, 0, 248698, 13, D56, S
```

1. Lakukan EDA dan preprocessing seperti pada bagian regresi.

Eda:

== 1. INFORMASI DATASET ==

<class 'pandas.core.frame.DataFrame'>

Rанг индекса: 891 записей, 0 до 890

Данные столбцы (всего 12 столбцов):

#	Column	Non-Null Count	Dtype
0	PassengerId	891	non-null
1	Survived	891	non-null
2	Pclass	891	non-null
3	Name	891	non-null
4	Sex	891	non-null
5	Age	714	non-null
6	SibSp	891	non-null
7	Parch	891	non-null
8	Ticket	891	non-null
9	Fare	891	non-null
10	Cabin	204	non-null
11	Embarked	889	non-null

dtypes: float64(2), int64(5), object(5)

память использования: 83.7+ KB

None

PassengerId ... Embarked

0	1	...	S
1	2	...	C
2	3	...	S
3	4	...	S
4	5	...	S

[5 rows x 12 columns]

==== 2. STATISTIK DESKRIPTIF ===

```
PassengerId ... Embarked  
count    891.000000 ...     889  
unique      NaN ...      3  
top        NaN ...      S  
freq       NaN ...     644  
mean    446.000000 ...     NaN  
std     257.353842 ...     NaN  
min     1.000000 ...     NaN  
25%   223.500000 ...     NaN  
50%   446.000000 ...     NaN  
75%   668.500000 ...     NaN  
max    891.000000 ...     NaN
```

[11 rows x 12 columns]

==== 3. CEK MISSING VALUES ===

```
PassengerId    0  
Survived      0  
Pclass        0  
Name          0  
Sex           0  
Age         177  
SibSp        0
```

```
Parch      0  
Ticket     0  
Fare       0  
Cabin     687  
Embarked   2  
dtype: int64
```

== 4. DISTRIBUSI VARIABEL TARGET (Survived) ==

```
Survived  
0  0.616162  
1  0.383838  
Name: proportion, dtype: float64
```

/home/titanic_eda.py:30: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown

```
plt.show()
```

== 5. SURVIVAL RATE BERDASARKAN JENIS KELAMIN ==

```
Sex  
female  0.742038  
male    0.188908  
Name: Survived, dtype: float64
```

== 5b. SURVIVAL RATE BERDASARKAN KELAS (Pclass) ==

```
Pclass  
1  0.629630  
2  0.472826
```

3 0.242363

Name: Survived, dtype: float64

==== 6. VISUALISASI TAMBAHAN ===

/home/titanic_eda.py:51: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown

plt.show()

==== EDA SELESAI ===

Preposseing:

X_test.csv

☰ X_test.csv > □ data

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked_Q	Embarked_S
1	3,0	28.0	1,1	15.2458	False	False		
2	2,0	31.0	0,0	10.5	False	True		
3	3,0	20.0	0,0	7.925	False	True		
4	2,1	6.0	0,1	33.0	False	True		
5	3,1	14.0	1,0	11.2417	False	False		
6	1,1	26.0	0,0	78.85	False	True		
7	3,1	28.0	0,0	7.75	True	False		
8	3,0	16.0	2,0	18.0	False	True		
9	3,1	16.0	0,0	7.75	True	False		
10	1,1	19.0	0,2	26.2833	False	True		
11	1,0	37.0	1,0	53.1	False	True		
12	3,0	44.0	0,0	8.05	False	True		
13	3,1	28.0	3,1	25.4667	False	True		
14	3,0	30.0	0,0	7.225	False	False		
15	2,0	36.0	0,0	13.0	False	True		
16	1,1	16.0	0,1	39.4	False	True		
17	1,0	42.0	1,0	52.5542	False	True		
18	3,1	28.0	0,0	7.8292	True	False		
19	2,0	27.0	0,0	13.0	False	True		
20	1,0	47.0	0,0	52.0	False	True		
21	3,0	24.0	0,0	7.8958	False	True		
22	1,0	34.0	0,0	26.55	False	True		
23	2,1	10.0	1,0	7.8542	False	True		
24	2,1	10.0	1,0	7.8542	False	True		

X_train.csv

y_test.csv

grid y_test.csv > data

	Survived
1	1
2	0
3	0
4	1
5	1
6	1
7	1
8	1
9	0
10	1
11	1
12	0
13	0
14	0
15	0
16	0
17	1
18	1
19	1
20	0
21	0
22	0
23	1

y_train.csv

y_train.csv > data

	Survived
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	1
12	1
13	0
14	0
15	0
16	1
17	1
18	1
19	1
20	0
21	1
22	0
23	0

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

==== Logistic Regression ===

Akurasi: 0.8100558659217877

Confusion Matrix:

[[90 15]

[19 55]]

Classification Report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.83	0.86	0.84	105
---	------	------	------	-----

1	0.79	0.74	0.76	74
---	------	------	------	----

accuracy		0.81		179
----------	--	------	--	-----

macro avg	0.81	0.80	0.80	179
-----------	------	------	------	-----

weighted avg	0.81	0.81	0.81	179
--------------	------	------	------	-----

==== Decision Tree ===

Akurasi: 0.7988826815642458

Confusion Matrix:

[[96 9]

[27 47]]

Classification Report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.78	0.91	0.84	105
---	------	------	------	-----

1	0.84	0.64	0.72	74
---	------	------	------	----

accuracy		0.80		179
----------	--	------	--	-----

macro avg	0.81	0.77	0.78	179
-----------	------	------	------	-----

weighted avg	0.80	0.80	0.79	179
--------------	------	------	------	-----

3. Lakukan evaluasi model menggunakan:

a. Confusion Matrix

b. Accuracy, Precision, Recall, F1-score

c. ROC Curve (jika memungkinkan)

==== Informasi Dataset ===

```
<class 'pandas.core.frame.DataFrame'>
```

RangelIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
---	--------	----------------	-------

```
---
```

0	PassengerId	891	non-null	int64
---	-------------	-----	----------	-------

1	Survived	891	non-null	int64
---	----------	-----	----------	-------

2	Pclass	891	non-null	int64
---	--------	-----	----------	-------

3	Name	891	non-null	object
---	------	-----	----------	--------

4	Sex	891	non-null	object
---	-----	-----	----------	--------

5	Age	714	non-null	float64
---	-----	-----	----------	---------

6	SibSp	891	non-null	int64
---	-------	-----	----------	-------

7	Parch	891	non-null	int64
---	-------	-----	----------	-------

8	Ticket	891	non-null	object
---	--------	-----	----------	--------

9	Fare	891	non-null	float64
---	------	-----	----------	---------

10	Cabin	204	non-null	object
----	-------	-----	----------	--------

11	Embarked	889	non-null	object
----	----------	-----	----------	--------

dtypes: float64(2), int64(5), object(5)

memory usage: 83.7+ KB

==== Jumlah Missing Values ===

PassengerId 0

...

==== Data Setelah Preprocessing ====

Jumlah data latih: (712, 8)

Jumlah data uji: (179, 8)

==== Logistic Regression ====

Akurasi: 0.8100558659217877

Confusion Matrix:

[[90 15]

[19 55]]

Classification Report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.83	0.86	0.84	105
---	------	------	------	-----

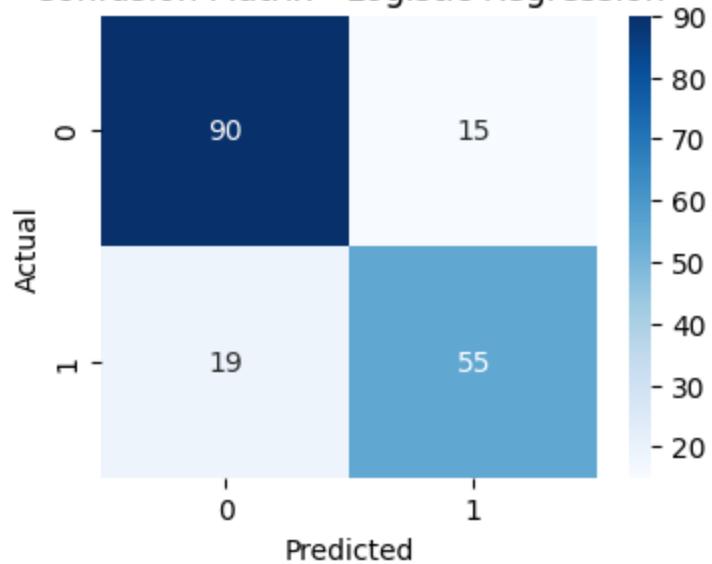
1	0.79	0.74	0.76	74
---	------	------	------	----

accuracy		0.81		179
----------	--	------	--	-----

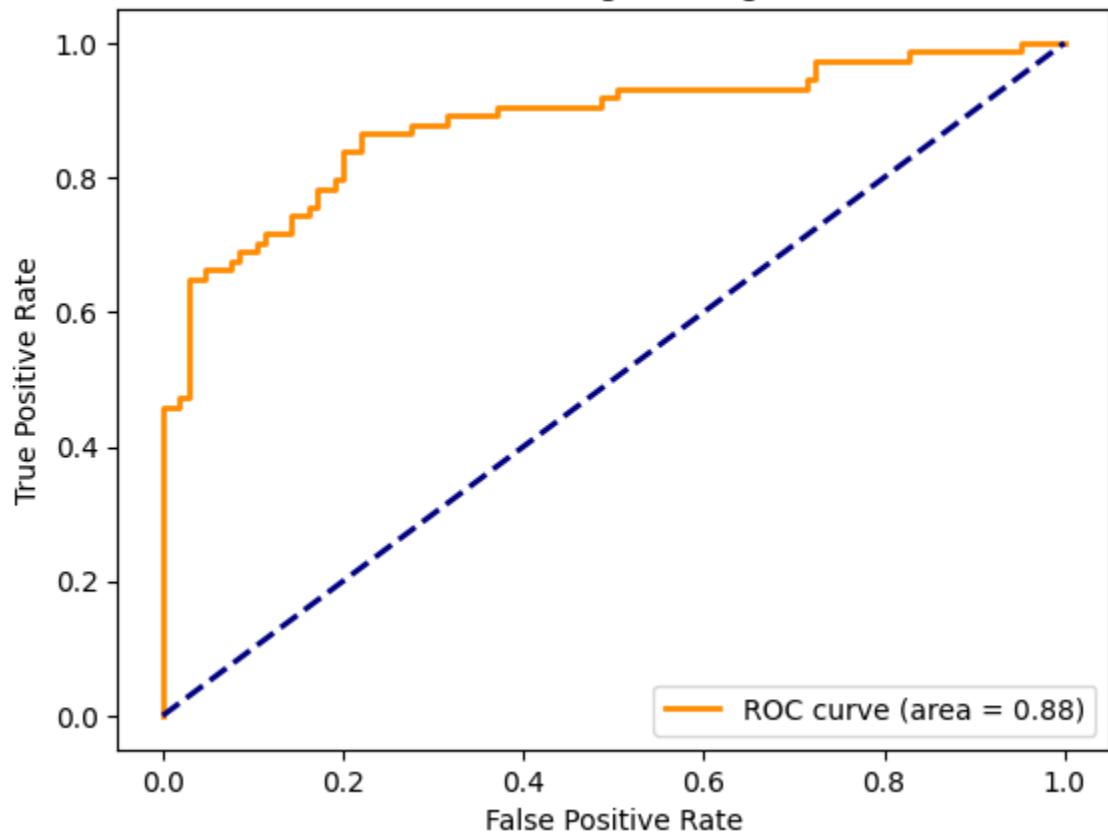
macro avg	0.81	0.80	0.80	179
-----------	------	------	------	-----

weighted avg	0.81	0.81	0.81	179
--------------	------	------	------	-----

Confusion Matrix - Logistic Regression



ROC Curve - Logistic Regression



==== Decision Tree ====

Akurasi: 0.7988826815642458

Confusion Matrix:

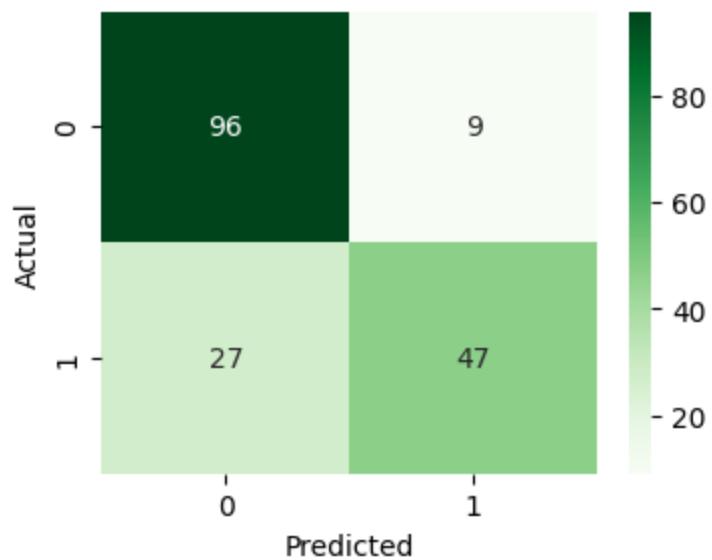
[[96 9]

[27 47]]

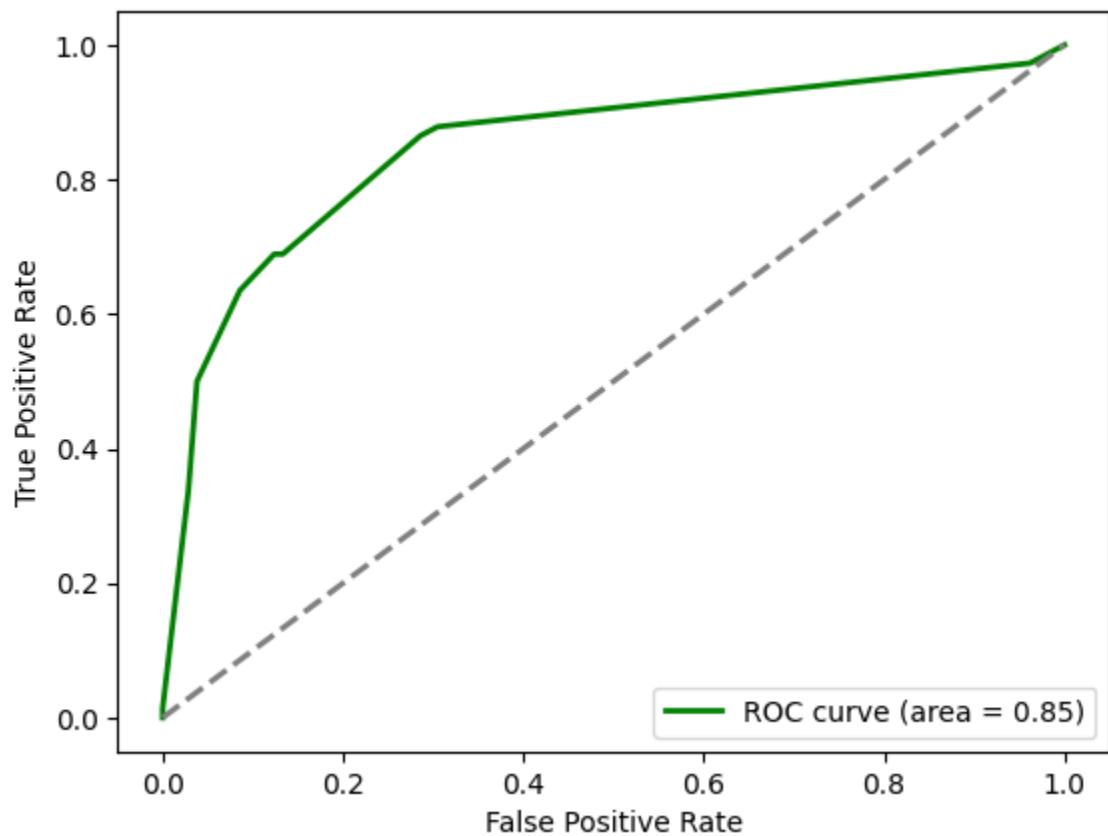
Classification Report:

	precision	recall	f1-score	support
0	0.78	0.91	0.84	105
1	0.84	0.64	0.72	74
accuracy		0.80		179
macro avg	0.81	0.77	0.78	179
weighted avg	0.80	0.80	0.79	179

Confusion Matrix - Decision Tree



ROC Curve - Decision Tree



==== Perbandingan Model ===

```

Model Accuracy ROC AUC
0 Logistic Regression 0.810056 0.881982
1 Decision Tree 0.798883 0.848713

```

4. Bandingkan hasil antar model dan tulis kesimpulan.

Perbandingan Hasil Model

Model	Akurasi	ROC AUC	Ciri Utama
Logistic Regression	0.81	0.88	Linear, lebih stabil terhadap noise dan generalisasi baik
Decision Tree	0.80	0.85	Non-linear, mudah menjelaskan pola, tapi sedikit lebih rawan overfitting

Analisis Kinerja Model

Logistic Regression

- **Accuracy:** 81%
- **Precision (kelas 1 / selamat):** 0.79
- **Recall (kelas 1 / selamat):** 0.74
- **F1-score:** 0.76
- Model ini **cukup seimbang** dalam mengenali penumpang yang selamat dan tidak selamat.
- Nilai **ROC AUC = 0.88** menunjukkan kemampuan diskriminatif yang sangat baik (semakin mendekati 1, semakin bagus).

Decision Tree

- **Accuracy:** 79.9%
- **Precision (kelas 1 / selamat):** 0.84
- **Recall (kelas 1 / selamat):** 0.64

- **F1-score:** 0.72
- Model ini lebih **tepat dalam memprediksi yang selamat (precision tinggi)**, tapi **lebih banyak meleset dalam mendekripsi semua yang selamat (recall rendah)**.
- Nilai **ROC AUC = 0.85** juga tergolong baik, meski sedikit di bawah Logistic Regression.

Kesimpulan

- Kedua model memiliki performa yang **mirip (selisih ±1%)**, tetapi **Logistic Regression unggul sedikit** pada hampir semua metrik.
- **Logistic Regression** lebih stabil, generalisasi lebih baik, dan cocok untuk dataset dengan variabel numerik & kategorikal sederhana seperti Titanic.
- **Decision Tree** tetap bermanfaat karena memberikan interpretasi yang lebih mudah (misalnya, kita bisa melihat aturan keputusan seperti “penumpang perempuan kelas 1 memiliki peluang selamat lebih tinggi”).
- Untuk implementasi nyata, Logistic Regression lebih direkomendasikan **jika tujuan utama adalah akurasi dan generalisasi**, sedangkan Decision Tree cocok **jika fokusnya pada interpretabilitas**.
- **Rangkuman Akhir**
- **Model terbaik untuk kasus Titanic ini adalah Logistic Regression** karena memberikan akurasi (81%) dan ROC AUC (0.88) tertinggi, serta performa yang seimbang antara precision dan recall.



UNIVERSITAS PAMULANG
KARTU UJIAN TENGAH SEMESTER GANJIL 2025/2026
NOMOR UJIAN : 01242862797869

FAKULTAS / PRODI : ILMU KOMPUTER / TEKNIK INFORMATIKA S1

NAMA MAHASISWA : NURDIANSYAH

NIM : 231011400211

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