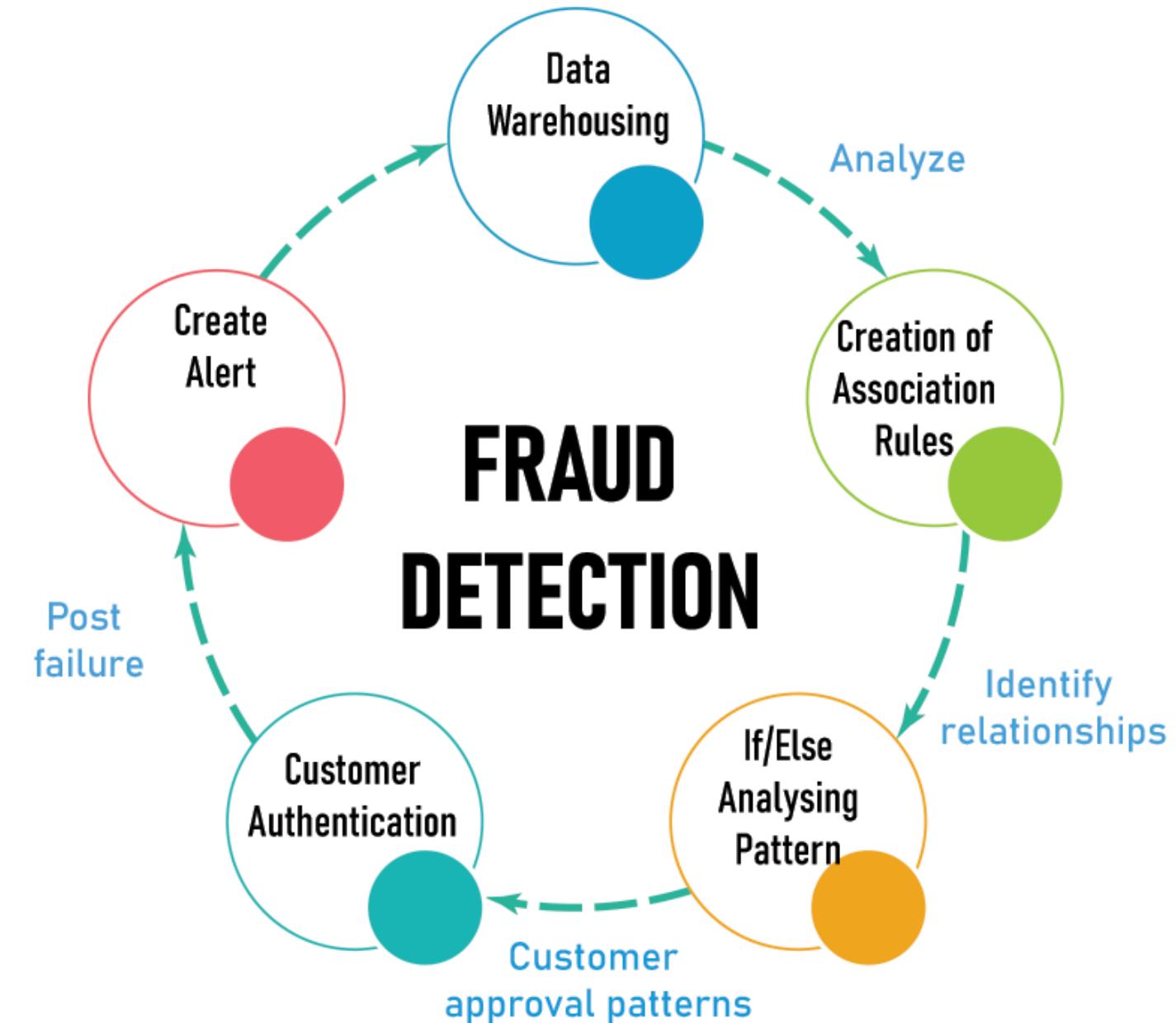


# ONLINE FRAUD PREDICTION

MACHINE LEARNING PROJECT - KELOMPOK 2

1. JIMMIE HENDERSON GUNAWAN – 2602164685
2. MELLISA ANGELINE – 2602077862
3. TIARA INTAN KUSUMA – 2602172220
4. ANGEL EODIA – 2602192140
5. BENNY STRATA WIJAYA – 2540128682



## APA ITU FRAUD PREDICTION?

Prediksi penipuan (fraud prediction) adalah proses menggunakan data dan teknik analisis untuk mengidentifikasi kemungkinan terjadinya penipuan atau kegiatan ilegal lainnya dalam suatu sistem atau transaksi.

## TABLE OF CONTENT

- 01 APA YANG KAMI BUAT?
- 02 DATASET
- 03 MODEL
- 04 PREPROCESSING
- 05 HASIL TRAINING
- 06 IMPORT MODEL



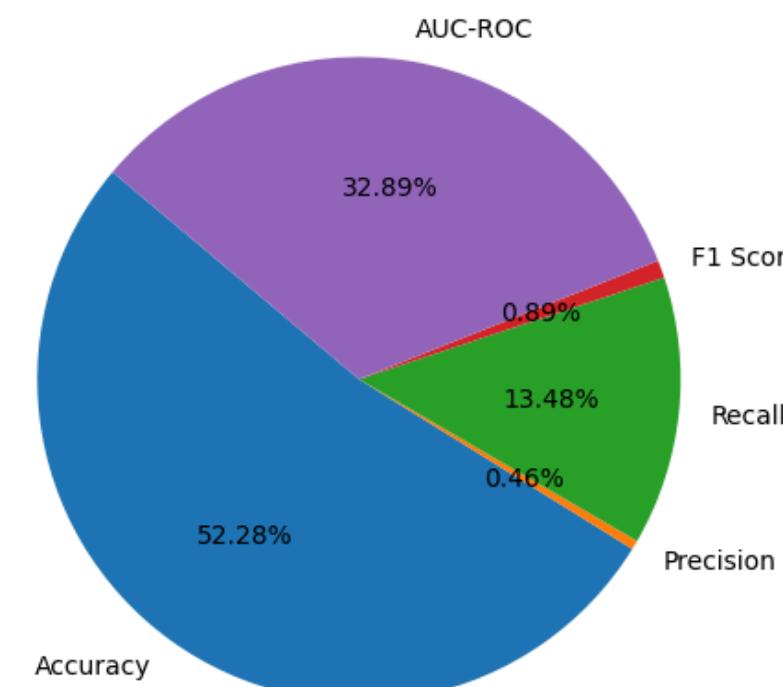
## **DATASET - ONLINE PAYMENT FROM KAGGLE**

step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.00	160296.36	M1979787155	0.0	0.00	0
1	1	PAYMENT	1864.28	C1666544295	21249.00	19384.72	M2044282225	0.0	0.00	0
2	1	TRANSFER	181.00	C1305486145	181.00	0.00	C553264065	0.0	0.00	1
3	1	CASH_OUT	181.00	C840083671	181.00	0.00	C38997010	21182.0	0.00	1
4	1	PAYMENT	11668.14	C2048537720	41554.00	29885.86	M1230701703	0.0	0.00	0
5	1	PAYMENT	7817.71	C90045638	53860.00	46042.29	M573487274	0.0	0.00	0
6	1	PAYMENT	7107.77	C154988899	183195.00	176087.23	M408069119	0.0	0.00	0
7	1	PAYMENT	7861.64	C1912850431	176087.23	168225.59	M633326333	0.0	0.00	0
8	1	PAYMENT	4024.36	C1265012928	2671.00	0.00	M1176932104	0.0	0.00	0
9	1	DEBIT	5337.77	C712410124	41720.00	36382.23	C195600860	41898.0	40348.79	0

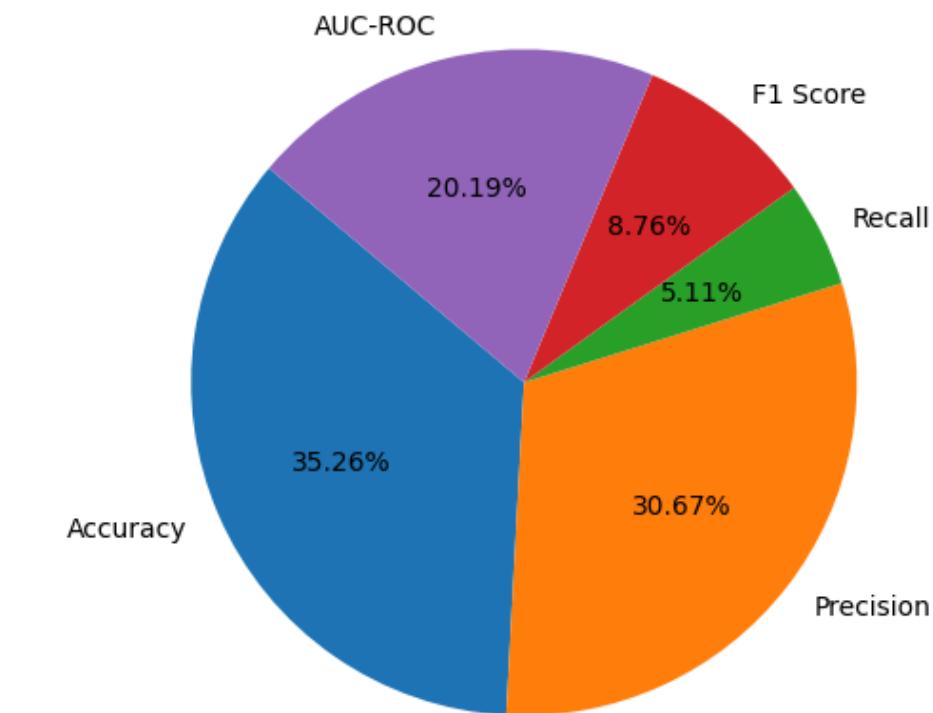
SOURCED FROM:

[HTTPS://WWW.KAGGLE.COM/DATASETS/RUPAKROY/ONLINE-PAYMENTS-FRAUD-DETECTION-DATASET](https://www.kaggle.com/datasets/rupakroy/online-payments-fraud-detection-dataset)

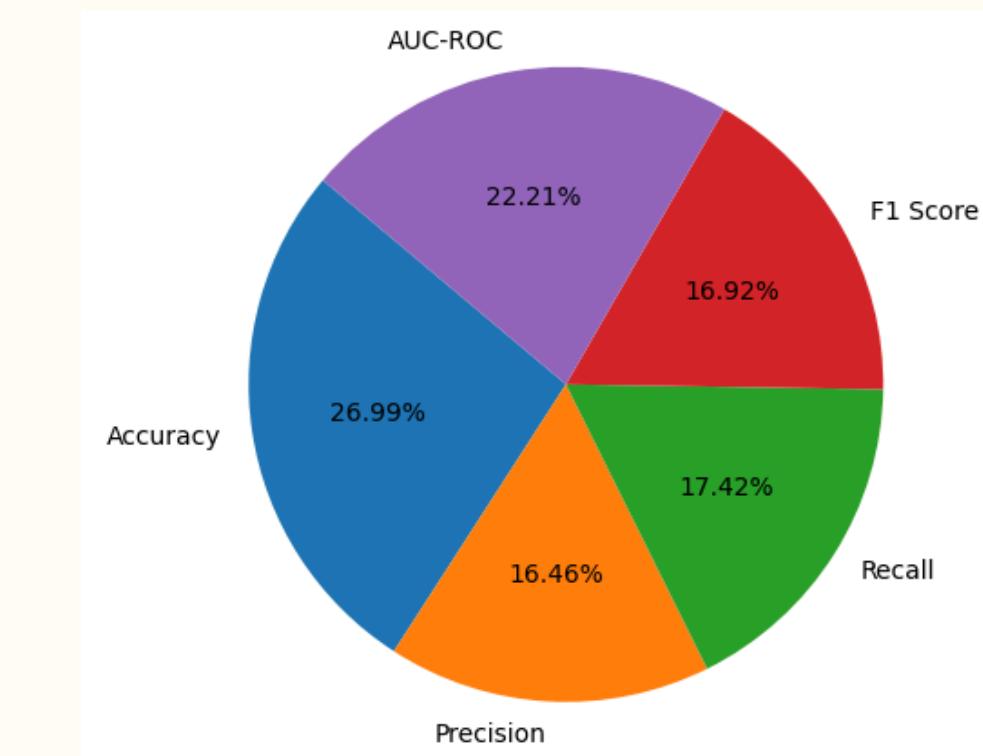
## 3 MODEL YANG DIGUNAKAN



Naive Bayes Classifier



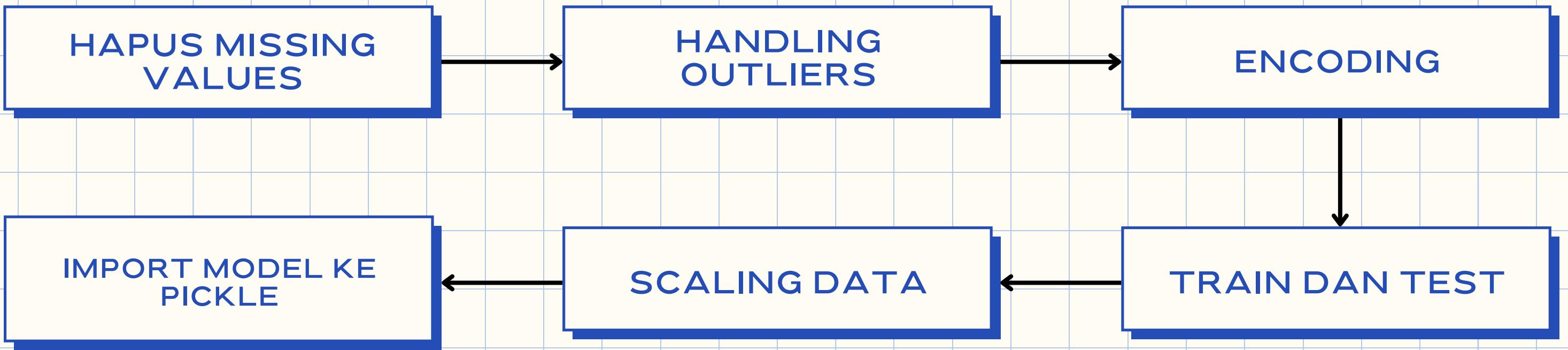
Logistic Regression



Decision Tree



# PREPROCESSING



## 1.

# DATA CLEANING

Proses mempersiapkan dan membersihkan data mentah sebelum dimasukkan ke dalam model

## HANDLING MISSING VALUE

### Handling Missing Values

```
if data.isnull().values.any():
    print('Unfortunately, there are missing values in the dataset\n')
    data.dropna(inplace=True)
    print('Shape : ', data.shape)
else:
    print('Fortunately, there aren\'t missing values in the dataset.')
[7]   ✓ 1.5s
... Fortunately, there aren't missing values in the dataset.
```

## ENCODING

### Label Encoding

```
[13] data['type'].replace({'CASH_OUT':0, 'PAYMENT':1, 'CASH_IN':2, 'TRANSFER':3, 'DEBIT':4}, inplace=True)
    ✓ 2.3s
```



## 2. SPLITTING DATA



### SPLIT THE DATA AND TARGET

#### Splitting the data and target

```
# X Data
X = data.drop(['isFraud'], axis=1)
print('X shape is : ', X.shape)
print()

# y Data
y = data['isFraud']
print('y shape is : ', y.shape)

✓ 0.1s

X shape is : (6362620, 6)

y shape is : (6362620,)
```

### TRAIN TEST SPLIT

#### Data Partitioning / Train Test Split

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42, shuffle=True)

# Splitted Data
print('X_train shape is ', X_train.shape)
print('X_test shape is ', X_test.shape)
print('y_train shape is ', y_train.shape)
print('y_test shape is ', y_test.shape)

✓ 2.1s

X_train shape is (4453834, 6)
X_test shape is (1908786, 6)
y_train shape is (4453834,)
y_test shape is (1908786,)
```

Data dibagi menjadi training dan testing set. Set pelatihan digunakan untuk melatih model machine learning, sedangkan set pengujian digunakan untuk menguji seberapa baik model tersebut menggeneralisasi data

### 3. DATA SCALING

## Data Scaling / Feature Scaling

```
# Standardization (Z-Score Normalization)
# StandardScaler for Data

scaler = StandardScaler()

# Fit the scaler on the training data
scaler.fit(X_train)

# Transform the training and testing data
X_train_scaled = scaler.transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

✓ 1.2s

# ACCURACY

## LOGISTIC REGRESSION

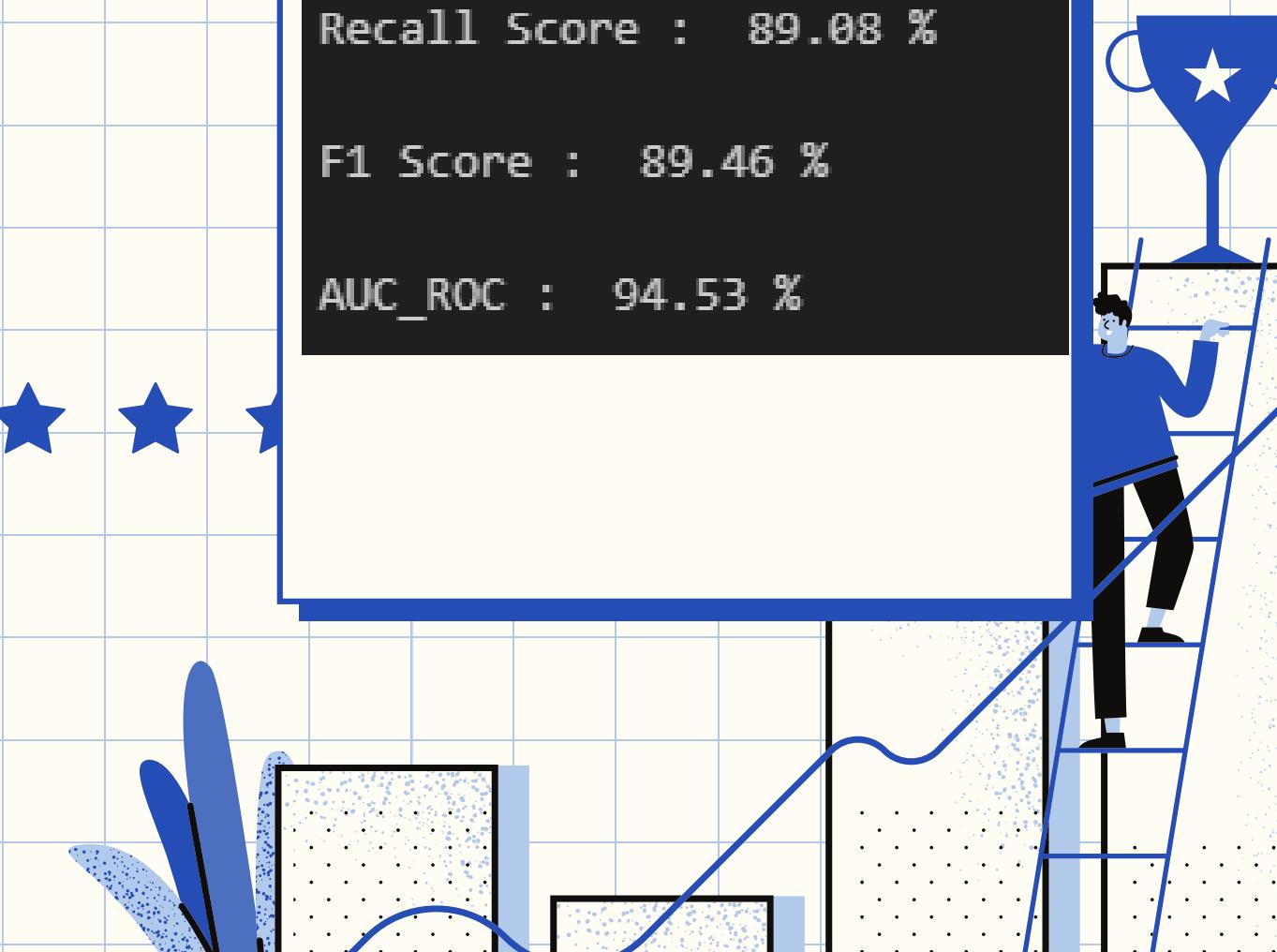
Accuracy Score : 99.92 %  
Precision Score : 86.10 %  
Recall Score : 40.70 %  
F1 Score : 55.27 %  
AUC\_ROC : 70.34 %

## NAIVE-BAYES

Accuracy Score : 99.17 %  
Precision Score : 2.77 %  
Recall Score : 16.18 %  
F1 Score : 4.73 %  
AUC\_ROC : 57.73 %

## DECISION TREE

Accuracy Score : 99.97 %  
Precision Score : 89.85 %  
Recall Score : 89.08 %  
F1 Score : 89.46 %  
AUC\_ROC : 94.53 %



# IMPORTING THE MODELS INTO PICKLE

```
import pickle
Model_DTT = DecisionTreeClassifier()
Model_DTT.fit(X_train_scaled, y_train)
# y_pred_DT = Model_DT.predict(X_test_scaled)
with open('decision_tree_model.pickle', 'wb') as model_file:
    model = pickle.dump(Model_DTT, model_file)
Model_LRR = LogisticRegression()
Model_LRR.fit(X_train_scaled, y_train)
with open('logistic_regression_model.pickle', 'wb') as model_file:
    model = pickle.dump(Model_LRR, model_file)
Model_NBB = GaussianNB()
Model_NBB.fit(X_train_scaled, y_train)
with open('naive_bayes_model.pickle', 'wb') as model_file:
    model = pickle.dump(Model_NBB, model_file)
```

✓ 46.7s

Deploy



# Online Fraud Detection

## Navigation

Go to

- Home
- Prediction
- EDA
- About

## Navigation

Go to

- Home
- Prediction
- EDA
- About

## Online Fraud Payment Detection

Predict whether a transaction is fraudulent

Choose the model for prediction

Decision Tree

### Select Transaction Type

Transaction Type

Cash Out

### Input Transaction Details

Transaction Amount

0

Your Balance Before Transaction

Welcome to the Online Fraud Payment Detection app. This application helps you predict whether a financial transaction is fraudulent or not.

Use the sidebar to navigate through the different sections of the app.

Get Started

Developed with ❤ by Kelompok 2

Deploy :

# 🔍 Online Fraud Payment Detection

Predict whether a transaction is fraudulent

Choose the model for prediction

Decision Tree

## Select Transaction Type

Transaction Type

Cash Out

## Input Transaction Details

Transaction Amount

0

Deploy



# Exploratory Data Analysis

## Exploring the transaction dataset

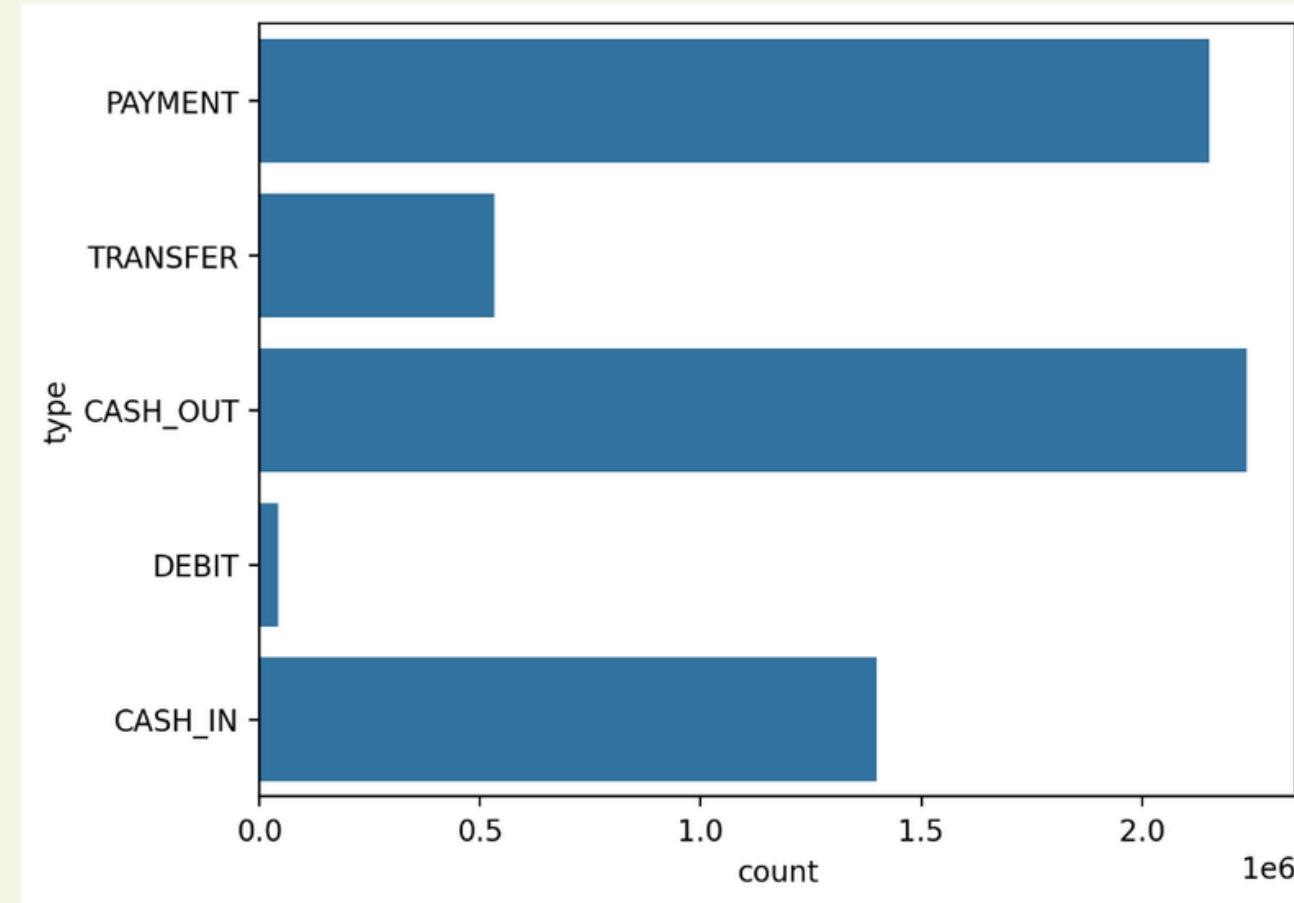
### Dataset Overview

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbal
0	1	PAYOUT	9,839.64	C1231006815	170,136	160,296.36	M1979787155	
1	1	PAYOUT	1,864.28	C1666544295	21,249	19,384.72	M2044282225	
2	1	TRANSFER	181	C1305486145	181			
3	1	CASH_OUT	181	C840083671	181			
4	1	PAYOUT	11,668.14	C2048537720	41,554			

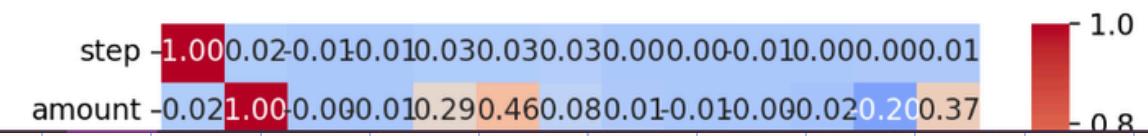
### Basic Statistics

	step	amount	oldbalanceOrg	newbalanceOrig	oldbal
count	6,362,620	6,362,620	6,362,620	6,362,620	6,362,620
mean	243.3972	179,861.9035	833,883.1041	855,113.6686	1,100,700
std	142.332	603,858.2315	2,888,242.673	2,924,048.503	3,399,100
min	1	0	0	0	0
25%	156	13,389.57	0	0	0

### Distribution of Transaction Types



### Correlation Matrix



Deploy :

## Navigation

Go to

- Home
- Prediction
- EDA
- About

# About - Online Fraud Detection from kelompok 2

This application helps to predict whether a financial transaction is fraudulent or not. It uses three different machine learning models:

- **Decision Tree:** Known for its interpretability and ability to handle non-linear relationships.
- **Logistic Regression:** A simple and effective baseline model providing probability estimates.
- **Naive Bayes:** Computationally efficient and handles high-dimensional data well.
- **Our Team:**

Jimmie Henderson Gunawan (2602164685)

Mellisa angeline (2602077862)

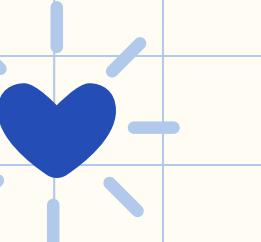
tiara intan kusuma (2602172220)

angel eodia (2602192140)

benny strata wijaya (2540128682)

**DEMO OF THE APP USING STREAMLIT:**

**HTTPS://FRAUD-DETECTION-MACHINE-LEARNING.STREAMLIT.APP/**



**THANK YOU**