FASTAG FRAUD DETECTION

Presented by Maria Mohamed Bassam

INTRODUCTION

Objective: Detect fraudulent transactions in Fastag data using machine learning

Dataset: Contains various features related to toll transactions, including a fraud indicator.

DATASET OVERVIEW

1. Transaction_ID: Unique identifier for each transaction.
2. Timestamp: Date and time of the transaction.
3. Vehicle_Type: Type of vehicle involved in the transaction.
4. FastagID: Unique identifier for Fastag.
5. TollBoothID: Identifier for the toll booth.
6. Lane_Type: Type of lane used for the transaction.
7. Vehicle_Dimensions: Dimensions of the vehicle.
8. Transaction_Amount: Amount associated with the transaction.
9. Amount_paid: Amount paid for the transaction.
10. Geographical_Location: Location details of the transaction.
11. Vehicle_Speed: Speed of the vehicle during the transaction.
12. Vehicle_Plate_Number: License plate number of the vehicle.
13. Fraud_indicator: Binary indicator of fraudulent activity (target variable

Rows: 5000

Columns: 13

HANDLING MISSING VALUES

1. **FastagID:** 10.98% missing

2. **Strategy:** Impute missing FastagID with 'Unknown'

3. **Verification**: No missing values left.

FEATURE ENGINEERING

- Convert Timestamp to datetime

```
data['Timestamp'] = pd.to_datetime(data['Timestamp'])
```

- Encode Categorical Variables

```
from sklearn.preprocessing import LabelEncoder

lbe = LabelEncoder()
for col in data.columns:
    if data[col].dtype == 'object':
        data[col] = lbe.fit_transform(data[col])
```

MODEL DEVELOPMENT

Logistic Regression Model

Train-Test Split: 70% training, 30% testing

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

X = data.drop(['Transaction_ID', 'Timestamp', 'FastagID', 'TollBoothID',
               'Geographical_Location', 'Vehicle_Plate_Number', 'Fraud_indicator'], axis=1
y = data['Fraud_indicator']

xtrain, xtest, ytrain, ytest = train_test_split(X, y, random_state=41, test_size=0.30)

LR = LogisticRegression()
LR.fit(xtrain, ytrain)
```

MODEL PERFORMANCE

```
from sklearn.metrics import classification_report


print(classification_report(ytest, LR.predict(xtest)))
```

- **Accuracy:** 99%
- **Precision & Recall**: High for both classes

HANDLING CLASS IMBALANCE

- SMOTE (Synthetic Minority Over-sampling Technique)
- Improved Performance

CONCLUSION

- **Effective Detection**: Logistic regression model identifies fraudulent transactions accurately.
- **Next Steps**: Explore advanced models and fine-tune parameters.
- **Impact**: Enhances toll collection integrity and reduces fraud-related losses.

THANK YOU