Online Payments Fraud Detection using Machine Learning

1. Introduction

This project focuses on detecting fraudulent online payment transactions using Machine Learning techniques. The system analyzes transaction patterns and predicts whether a transaction is legitimate or fraudulent based on historical financial data.

2. Technologies Used

- Python
- Pandas & NumPy
- Scikit-learn
- Random Forest Classifier
- Flask (Web Framework)
- HTML
- Git & GitHub

3. Data Collection

Dataset was downloaded from Kaggle: Online Payments Fraud Detection Dataset. The dataset contains transaction details such as amount, balance before and after transaction, transaction type, and fraud label.

4. Data Preprocessing

The following preprocessing steps were performed:

- Removed unnecessary columns (nameOrig, nameDest, isFlaggedFraud)
- Label encoded categorical variable (transaction type)
- Handled imbalanced dataset using class_weight='balanced'
- Split dataset into training and testing sets (80%-20%)

5. Model Building

Multiple classification algorithms were tested. Random Forest Classifier was selected as the final model due to high performance and robustness.

Model Accuracy Achieved: ~99%

6. Model Saving

The trained model was saved using pickle:

```
pickle.dump(model, open('payments.pkl', 'wb'))
```

7. Flask Web Application

A Flask web application was developed to allow users to input transaction details and receive fraud prediction results in real-time.

- Home Page
- Prediction Page
- Result Page
- Fraud Probability Threshold Tuning

8. Test Cases

Legitimate Transaction Example:

```
Step: 1
Type: 3
Amount: 9.19
OldbalanceOrg: 170136
NewbalanceOrig: 160296
OldbalanceDest: 0
NewbalanceDest: 0
```

Fraud Transaction Example:

```
Step: 94
Type: 1
Amount: 500000
OldbalanceOrg: 500000
NewbalanceOrig: 0
OldbalanceDest: 0
NewbalanceDest: 0
```

9. Conclusion

The project successfully demonstrates an end-to-end Machine Learning pipeline from data preprocessing to model deployment using Flask. The system can identify potentially fraudulent transactions and provide probability-based fraud detection.