ChatGPT

Major Project Report

Title:

Online Payments Fraud Detection using Machine Learning

Objective:

To build a machine learning model that classifies online transactions as either fraudulent or non-fraudulent using a real-world dataset. This solution can help financial institutions automatically detect suspicious activity and reduce fraud losses.

Dataset Description:

The dataset used for this project was sourced from Kaggle and contains historical data of transactions. It includes the following columns:

1. **step** – Represents a unit of time (in hours)
2. **type** – Type of transaction (TRANSFER, CASH_OUT, etc.)
3. **amount** – Transaction amount
4. **nameOrig** – Customer initiating the transaction
5. **oldbalanceOrg** – Sender's balance before transaction
6. **newbalanceOrig** – Sender's balance after transaction
7. **nameDest** – Receiver of the transaction
8. **oldbalanceDest** – Receiver's balance before transaction
9. **newbalanceDest** – Receiver's balance after transaction
10. **isFraud** – Label (1 if fraudulent, 0 otherwise)

Methodology:

Step 1: Data Loading & EDA

- Loaded data using Pandas
- Checked shape, null values, and class distribution

Step 2: Feature Engineering

- Derived new feature: `hour` from `step`
- Encoded `type` using LabelEncoder
- Dropped identifiers like `nameOrig` and `nameDest`

Step 3: Handling Class Imbalance

- Fraudulent transactions were highly underrepresented
- Applied **SMOTE** (Synthetic Minority Oversampling Technique) to balance the dataset

Step 4: Model Training

- Used **Random Forest Classifier** for prediction
- Split data into 70% training and 30% testing

Step 5: Model Evaluation

- Evaluated model using:
- Confusion Matrix
- Classification Report (Precision, Recall, F1-score)
- Visualized **feature importances** using a horizontal bar chart

---

Results:

- Model accurately classifies fraudulent transactions with high recall and precision
- Top contributing features include: `amount`, `oldbalanceOrg`, and `type`
- Feature importance graph provides explainability for model decisions

---

Tools & Technologies Used:

- Python 3
- Pandas, NumPy
- Scikit-learn
- imbalanced-learn (SMOTE)
- Matplotlib, Seaborn

Future Work:

- Deploy the model using Streamlit or Flask as a real-time web app
- Integrate time-series and geolocation-based features
- Use deep learning techniques for anomaly detection

Conclusion:

This project demonstrates how machine learning can be used effectively to combat fraud in online payments. With further enhancements, it can be deployed as a full-scale fraud monitoring tool.

Submitted by: Preethi Kethireddy
GitHub: https://github.com/Masscoders00/fraud-detection

Institution: Raghu Engineering College
Course: B.Tech – Computer Science