FRAUD DETECTION IN CREDIT CARD TRANSACTIONS

Introduction

With the exponential rise in online transactions, credit card fraud has become a major concern for financial institutions and users. Traditional rule-based systems are often inadequate in detecting complex fraudulent patterns. This project aims to develop a machine learning-based system to detect anomalous and fraudulent transactions using both unsupervised and supervised learning techniques.

Abstract

This project implements anomaly detection and classification techniques to identify fraudulent credit card transactions. Using the publicly available Kaggle credit card dataset, the project involves data preprocessing, balancing imbalanced classes, applying anomaly detection algorithms (Isolation Forest and Local Outlier Factor), and building an XGBoost classifier. A web-based interface was developed using Streamlit to allow real-time predictions. The model's performance was evaluated using a confusion matrix and ROC curve. The end product is a user-friendly dashboard capable of detecting fraud based on transaction features.

Tools Used

- **Programming Language:** Python
- **Libraries:**
  - **Data Handling:** Pandas, NumPy
  - **Visualization:** Matplotlib, Seaborn
  - **Machine Learning:** Scikit-learn, XGBoost
- **Deployment:** Streamlit
- **Dataset:** Kaggle - Credit Card Fraud Detection Dataset (284,807 transactions)

Steps Involved in Building the Project

1. **Data Loading & Preprocessing**
   - Loaded CSV dataset from Kaggle.
   - Performed scaling on anonymized features (V1 to V28) using StandardScaler.
   - Detected class imbalance (fraudulent transactions ≈ 0.17%).
2. **Handling Imbalance**
   - Used RandomUnderSampler to balance the dataset before training the classifier.

3. Anomaly Detection

- Applied Isolation Forest and Local Outlier Factor to flag suspicious transactions.
- Evaluated anomaly detection precision/recall using confusion matrix.

4. Model Building with XGBoost

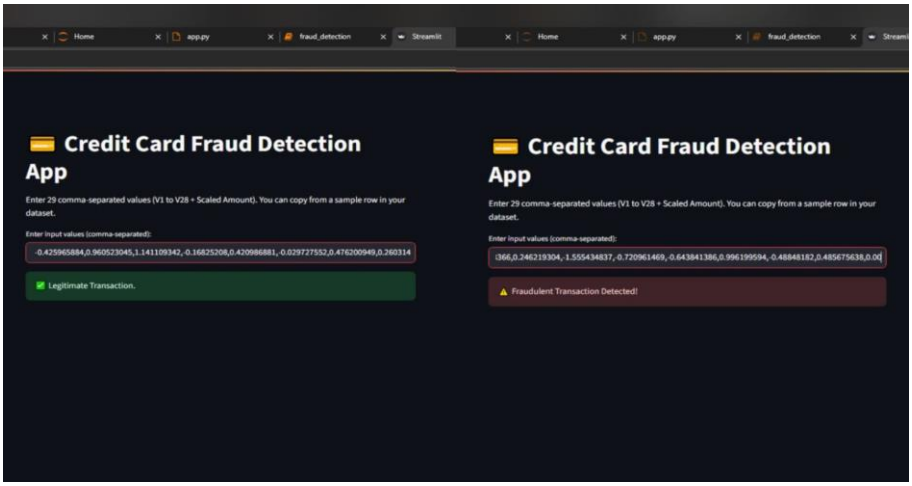
- Trained an XGBoost classifier on the balanced dataset.
- Achieved high accuracy and F1 score on test data.
- ROC Curve plotted to evaluate model's ability to distinguish fraud.

5. Model Evaluation

- Used Confusion Matrix, Precision, Recall, F1-Score, and ROC-AUC to evaluate model.
- Found significant improvement compared to unsupervised methods alone.

6. Web Deployment

- Built a Streamlit dashboard to:
- Upload transaction data
- Predict fraud status
- Display results interactively



Conclusion

The project successfully implemented both anomaly detection and supervised learning techniques for identifying fraudulent credit card transactions. By leveraging machine learning models like Isolation Forest, Local Outlier Factor, and XGBoost, the system was able to achieve a robust level of accuracy. The web-based interface provides an easy-to-use platform for users to test new data. This solution can be integrated into real-world systems to enhance fraud detection capabilities and ensure secure transactions.

DONE BY:N.DESHMA