Fraud Detection in Credit Card Transactions

Introduction

Fraudulent transactions pose a significant threat to financial institutions and consumers alike. This project focuses on identifying fraudulent credit card transactions using anomaly detection techniques to enhance the security of online transactions.

Abstract

This report details the development of a credit card fraud detection system utilizing Python and various machine learning tools. The primary objective was to accurately identify fraudulent transactions from a dataset obtained from Kaggle. Techniques such as data preprocessing, anomaly detection with Isolation Forest and Local Outlier Factor, and classification using XGBoost were employed. The final output is a user-friendly dashboard that allows users to input transaction details and receive real-time fraud predictions.

Tools Used

- **Programming Language:** Python
- **Libraries:**
  - Scikit-Learn for machine learning algorithms
  - XGBoost for classification
  - Pandas for data manipulation
  - Streamlit for web application development
  - Matplotlib and Seaborn for data visualization

Steps Involved in Building the Project

1. **Data Acquisition:**
   - The Kaggle credit card dataset containing 1,000,000 transactions was utilized.
2. **Preprocessing:**
   - The dataset was loaded, and checks for null values and data types were performed.
   - Class distribution was analyzed to identify imbalances in fraudulent transactions.
3. **Balancing the Dataset:**
   - SMOTE (Synthetic Minority Over-sampling Technique) was applied to balance the training dataset.
4. **Anomaly Detection:**
   - Isolation Forest and Local Outlier Factor were implemented to detect anomalies in the dataset.
5. **Model Training:**
   - The XGBoost classifier was used for training the model on the balanced dataset.

    - Cross-validation was performed to evaluate the model's performance, achieving a mean F1 score of approximately 0.9988.
6. **Model Evaluation:**
    - A confusion matrix and classification report were generated to assess model accuracy, achieving an accuracy of 100%.
    - The ROC curve was plotted to visualize the model's performance.
7. **Dashboard Development:**
    - An interactive dashboard was created using Streamlit to allow users to input transaction details for fraud prediction.
8. **Deployment:**
    - The Streamlit app was deployed on Streamlit Community Cloud, obtaining a permanent URL for public access.

Conclusion

The developed fraud detection system demonstrates an effective approach to identifying fraudulent transactions in real-time. By utilizing advanced machine learning techniques and creating an intuitive user interface, the system can potentially save financial institutions from significant losses due to fraud. Future enhancements may involve incorporating additional features and exploring other machine learning algorithms for improved accuracy.