Fraud Detection in Credit Card Transactions

Introduction

Credit card fraud is a major concern in the financial sector, costing billions each year. As fraudulent transactions become more sophisticated, detecting them in real-time is essential. This project aims to build a machine learning model capable of identifying fraudulent credit card transactions using various detection techniques and evaluation metrics.

Abstract

The dataset used for this project contains anonymized credit card transactions, with a significant class imbalance: fraudulent transactions make up less than 0.2% of the total. To address this, the project explores both unsupervised anomaly detection (using Local Outlier Factor) and supervised learning (using XGBoost classifier). Key techniques such as SMOTE (for handling class imbalance), StandardScaler (for normalization), and evaluation metrics like confusion matrix, F1-score, and ROC-AUC were employed. The final model achieved high performance and provides a robust basis for real-time fraud detection systems.

Tools Used

- Language: Python
- Libraries:
  - pandas, numpy for data handling
  - matplotlib, seaborn for visualization
  - scikit-learn for preprocessing and evaluation
  - imbalanced-learn for SMOTE
  - xgboost for the supervised model
  - LocalOutlierFactor for anomaly detection

Steps Involved in Building the Project

1. Dataset Loading: The dataset was imported from Kaggle and inspected for structure and class imbalance.
2. Exploratory Data Analysis: Class distribution, feature correlations, and value distributions were visualized to understand the dataset.
3. Data Preprocessing:
   - Normalized the Amount feature using StandardScaler

- Removed duplicate rows to improve anomaly detection accuracy
- Dropped irrelevant columns like Time

4. Anomaly Detection:
   - Applied Local Outlier Factor (LOF) to detect rare patterns without using class labels.
5. Supervised Learning:
   - Applied SMOTE to oversample the minority class.
   - Trained an XGBoost Classifier on the resampled data.
   - Evaluated using confusion matrix, classification report, and ROC-AUC score.
6. Model Evaluation:
   - Achieved a high AUC score indicating strong separation between fraud and non-fraud.
   - Visualized feature importance and ROC curve for interpretation.

Conclusion

This project successfully demonstrates a practical application of machine learning in detecting fraudulent transactions. The use of both anomaly detection and supervised learning ensures flexibility and robustness. With further steps such as deployment via Flask or Streamlit, the model can be integrated into real-time systems to assist banks and fintech platforms in preventing fraud.