Fraud Detection in Credit Card Transactions

Abstract

Credit card fraud poses a significant threat in todays digital economy. This project presents a machine learning-based approach to detect fraudulent transactions using a real-world dataset. The aim is to develop a model that can accurately classify transactions as fraudulent or legitimate, thereby minimizing financial losses. Various anomaly detection methods like Isolation Forest and Local Outlier Factor, along with supervised learning using XGBoost, were employed. The models were evaluated based on precision, recall, and ROC-AUC score, with a focus on minimizing false negatives.

Introduction

With the rapid increase in online transactions, detecting fraud has become crucial. The dataset used in this project contains anonymized credit card transactions, including 492 frauds out of 284,807 transactions. Due to this class imbalance, detecting fraud is challenging. This project applies both unsupervised and supervised machine learning techniques to build a reliable fraud detection system.

Tools Used

- Python (Jupyter Notebook)

- Libraries: pandas, numpy, scikit-learn, xgboost, matplotlib, seaborn

- Machine Learning Algorithms: Isolation Forest, Local Outlier Factor, XGBoost

- Visualization Tools: Seaborn, Matplotlib

Steps Involved in Building the Project

1. Data Loading and Preprocessing:

   - Loaded the dataset using Pandas.

   - Dropped non-informative features like Time.

   - Normalized features using StandardScaler.

2. Handling Imbalanced Data:

   - Employed undersampling of non-fraud cases to match the fraud cases for balanced training.

3. Model Building:

   - Used Isolation Forest and Local Outlier Factor for anomaly detection.

   - Trained an XGBoost classifier for supervised learning.

4. Model Evaluation:

  - Evaluated using confusion matrix, classification report, and ROC-AUC curve.

  - Focused on improving recall to minimize undetected fraud.

5. Saving the Model:

  - Final model serialized using joblib for reuse.

Conclusion

This project demonstrates that machine learning can significantly enhance fraud detection capabilities. XGBoost showed promising results with a high recall rate, making it suitable for critical financial applications. Future improvements could include SMOTE-based oversampling, real-time fraud detection systems, and integration with business dashboards.