Fraud Detection in Credit Card Transactions

This project focuses on identifying fraudulent credit card transactions using anomaly detection and supervised learning. The system combines unsupervised models like Isolation Forest and Local Outlier Factor (LOF) for anomaly detection, and XGBoost for classification. The solution includes a web-based dashboard for real-time prediction of transaction legitimacy.

Objectives

1. Detect fraudulent transactions efficiently with minimal false positives.
2. Use both unsupervised and supervised learning models for robust classification.
3. Build a user-friendly interface to test transaction inputs interactively.
4. Provide key visualizations (ROC curve, confusion matrix) for evaluation.

Tools and Libraries

- Python 3
- Scikit-learn
- XGBoost
- Pandas, NumPy
- Matplotlib, Seaborn
- Joblib
- Streamlit for deployment

Dataset

Dataset: Kaggle Credit Card Fraud Detection Dataset

- 284,807 transactions
- 492 frauds (highly imbalanced)
- Features: anonymized V1 to V28, Time, Amount, and Class (label)

Data preprocessing includes balancing via under/oversampling, scaling, and feature engineering.

Methodology

1. Data Preprocessing: Cleaning, scaling, and balancing.
2. Anomaly Detection: Isolation Forest and Local Outlier Factor to identify potential frauds.
3. Classification: XGBoost model trained on processed data for accurate prediction.
4. Evaluation: ROC curve, confusion matrix, and F1-score.
5. Deployment: Streamlit web app for real-time transaction testing.

Results and Discussion

The combined approach of anomaly detection followed by classification significantly improves fraud detection accuracy. The XGBoost model achieves an F1-score above 0.95 on the balanced dataset, with ROC-AUC greater than 0.98. The web app allows non-technical users to easily test transactions and visualize model confidence.

Conclusion

This project provides an end-to-end fraud detection system that is interpretable, deployable, and effective at handling imbalanced datasets. The dashboard makes fraud prediction accessible to users in real-time, serving as a practical proof of concept for credit card fraud monitoring.