

Credit Card Fraud Detection Report

1. Introduction

Credit card fraud is a significant financial issue faced by banking institutions and customers worldwide. With the rapid growth of online transactions, the frequency of fraudulent activities has increased. This project aims to build a machine learning-based solution that can detect fraudulent credit card transactions using a real-world dataset. The dataset is highly imbalanced, which mirrors the actual scenario where fraud cases are rare but critical to identify.

2. Abstract

This project explores the use of supervised and unsupervised learning algorithms for the detection of fraudulent credit card transactions. We applied data preprocessing techniques, performed extensive exploratory data analysis, and trained various models, including Isolation Forest, Local Outlier Factor, and XGBoost. Due to the significant class imbalance, special consideration was given to metrics like precision, recall, and AUC. The final model was deployed using a Flask application with a front-end interface for real-time fraud detection.

3. Tools Used

- Python
- Pandas and NumPy for data manipulation
- Matplotlib and Seaborn for visualization
- Scikit-learn for preprocessing and modeling
- XGBoost for robust classification
- Flask for backend deployment
- HTML/CSS with Bootstrap for the frontend dashboard
- Kaggle Notebook for development and experimentation

4. Steps Involved in Building the Project

1. Data Collection

- Dataset sourced from Kaggle: Credit card Fraud Dataset.

2. Data Preprocessing

- Checked for missing values and standardized the 'Time' and 'Amount' features using StandardScaler.

3. Exploratory Data Analysis

- Identified class imbalance and visualized patterns in amount and time.
- Explored feature correlations to understand their relationship with fraud.

4. Model Training

- Applied and evaluated Isolation Forest and Local Outlier Factor for anomaly detection.
- Trained and tuned an XGBoost model using adjusted weights to handle class imbalance.

5. Evaluation

- Used classification reports, confusion matrices, and ROC-AUC scores.

6. Deployment

- Built a Flask application with an interactive HTML dashboard.
- Integrated the trained model to provide real-time fraud detection results.

5. Conclusion

Detecting credit card fraud is a challenging yet crucial task due to the imbalance in datasets and the evolving nature of fraud tactics. This project successfully demonstrated the implementation of a robust fraud detection system using machine learning. The integration of data science techniques with web technologies allowed us to create a user-friendly interface for real-time fraud prediction. With further enhancements, such as deep learning or real-time streaming data analysis, the system could be scaled for production use in financial institutions.

Key Takeaways

- Fraud detection is a class imbalance problem.
- Feature scaling and careful modeling are crucial.
- XGBoost provided strong performance and interpretability.
- Full pipeline developed from data preprocessing to deployment.