Credit Card Fraud Detection using Machine Learning

Introduction

In the digital era, credit card fraud has become a significant concern for financial institutions and customers alike. With a growing number of transactions happening online, detecting fraudulent activities is critical. This project aims to develop a robust machine learning-based solution for identifying fraudulent credit card transactions effectively and efficiently.

Abstract

The project leverages real-world transactional data to build a fraud detection system using machine learning. It addresses class imbalance through SMOTE (Synthetic Minority Over-sampling Technique) and applies the XGBoost algorithm for classification. A series of preprocessing, exploratory data analysis (EDA), and performance evaluation steps were followed. The model achieved high accuracy and interpretability, and the solution is optionally deployable using a Streamlit app for real-time predictions.

Tools Used

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- XGBoost
- imbalanced-learn (SMOTE)
- Joblib (for model saving)
- Streamlit (optional UI)

Steps Involved in Building the Project

1. Data Preprocessing
- Loaded the dataset (creditcard.csv)
- Checked for null values and cleaned the data
- Scaled the Time and Amount features

2. Exploratory Data Analysis (EDA)
- Visualized fraud vs. non-fraud counts using pie and bar charts
- Generated a correlation heatmap for deeper feature insights
- Analyzed the transaction amount distribution

3. Handling Class Imbalance
- Used SMOTE to balance the dataset, increasing fraud class samples
- Split the data into training and testing sets

4. Model Building & Evaluation
- Trained the model using XGBoost Classifier
- Evaluated using confusion matrix, classification report, and ROC-AUC
- Saved the model using joblib

5. (Optional) Deployment
- Created a simple Streamlit app to predict fraud based on user inputs

Credit Card Fraud Detection using Machine Learning

Conclusion

This project demonstrates an effective approach to detecting fraudulent transactions using advanced machine learning methods. By addressing data imbalance and utilizing a powerful classifier like XGBoost, the solution is both accurate and practical. The added ability to deploy through Streamlit makes the model interactive and real-time ready. This project serves as a valuable asset for internship or academic showcase purposes.

Author: Rupa Sri Rudroju
Project Type: Internship Showcase