

Introduction

In the digital age, credit card transactions have become a cornerstone of modern commerce. However, the rise in online payments has also led to a surge in fraudulent activities, costing billions of dollars annually. Detecting fraudulent transactions is a complex challenge due to the sheer volume of data, evolving fraud patterns, and the highly imbalanced nature of fraud datasets—where genuine transactions vastly outnumber fraudulent ones.

This project explores the use of machine learning to identify credit card fraud effectively. We leverage both anomaly detection techniques and supervised classification models to analyse transaction data and flag suspicious behaviour. Specifically, we use Isolation Forest for unsupervised anomaly detection, followed by an *XGBoost* classifier for final prediction. The goal is to build a system that not only achieves high accuracy but is also deployable in real-time through a user-friendly *Streamlit* dashboard, allowing for practical application in fraud prevention.

Abstract

With this project we aim to detect fraudulent credit card transactions using machine learning techniques. Working with the highly imbalanced Kaggle credit card dataset, we apply anomaly detection methods like *Isolation Forest* to highlight suspicious patterns. An *XGBoost* classifier is then trained to classify transactions as fraudulent or legitimate, with model performance evaluated using ROC curves, confusion matrix, and classification metrics. To demonstrate real-time fraud detection, we built an interactive dashboard using *Streamlit* that allows users to input or sample transaction data and receive instant predictions. This solution showcases how machine learning can effectively support fraud detection systems by enabling faster, data-driven decisions and minimizing financial risk.

Tools and technologies used

- Python
- Pandas & NumPy
- Scikit-Learn:
- XGBoost.
- Matplotlib & Seaborn
- Streamlit
- Joblib:

Steps Involved in Building the Project

- Dataset Collection
- Data Preprocessing

- Anomaly Detection
- Model Building with XGBoost
- Model Saving
- Dashboard Development (Streamlit)
- Deployment

Conclusion

This project demonstrates the practical application of machine learning techniques in detecting credit card fraud. By combining anomaly detection methods (Isolation Forest) with a powerful supervised classifier (XGBoost), we successfully built a robust system capable of identifying suspicious transactions with high accuracy.

To address the major challenge of class imbalance, appropriate data handling techniques were used to improve model performance and reduce false negatives—crucial in fraud detection scenarios. Furthermore, the development of a user-friendly Streamlit dashboard allowed for real-time testing and prediction, making the model accessible and interactive for practical use.

Overall, the project highlights the effectiveness of intelligent systems in reducing financial fraud risk and showcases how data science can be integrated into real-world applications for impactful results.