

Credit Card Fraud Detection using Machine Learning – Project Report

1. Introduction

Credit card fraud is a serious issue affecting individuals and institutions. This project builds a machine learning pipeline that uses both unsupervised and supervised learning techniques to detect fraudulent credit card transactions.

2. Abstract

This project uses Isolation Forest, Local Outlier Factor, and XGBoost to detect fraud in an imbalanced dataset. Using the Kaggle credit card fraud dataset, the process includes:

- data cleaning,
- feature imputation,
- model training,
- evaluation, and
- deployment using a Streamlit dashboard.

The trained model is stored and used for live predictions.

3. Tools Used

- Python
- Pandas, NumPy
- scikit-learn, imbalanced-learn
- XGBoost
- Seaborn, Matplotlib
- Google Colab (for model development)

- Streamlit (for app deployment)

4. Steps Involved in Building the Project

1. Loaded dataset (creditcard.csv) in Google Colab.
2. Cleaned the data by removing rows with missing labels and handling missing values using `SimpleImputer`.
3. Applied Isolation Forest and Local Outlier Factor (unsupervised anomaly detection).
4. Split data using `train_test_split` and trained XGBoost classifier with imbalance handling.
5. Evaluated all models using confusion matrix and classification report.
6. Saved the trained model (`xgboost_model.pkl`) inside the `/model` folder.
7. Built an interactive frontend with Streamlit (`app.py`) to predict fraud from user inputs.

Ran the app locally using:

```
streamlit run app.py
```

5. Folder Structure

```
credit_card_fraud_detection/  
├── data/  
│   └── creditcard.csv  
├── model/  
│   └── xgboost_model.pkl  
├── streamlit_app/  
│   └── app.py  
├── fraud_detection.ipynb  
└── project_report.pdf
```

6. Conclusion

The XGBoost classifier showed strong fraud detection capability when combined with class balancing. Unsupervised models provided fast anomaly flags. The working Streamlit app allows real-time fraud predictions by accepting 30 feature inputs. This end-to-end solution demonstrates an effective and interactive ML-based fraud detection system