Credit Card Fraud Detection using Machine Learning

1. Introduction

Credit card fraud is a major issue in the financial sector. The aim of this project is to develop a system that can detect fraudulent transactions by analyzing credit card transaction data using machine learning algorithms. Detecting anomalies in large datasets is crucial for minimizing financial losses.

2. Abstract

In this project, both supervised and unsupervised learning techniques were applied to a real-world credit card transaction dataset. Unsupervised models like Isolation Forest and Local Outlier Factor were used to detect anomalies, while XGBoost, a robust supervised model, was trained to classify transactions. The model achieved high precision and recall, especially for the minority class (fraud cases). The system is capable of accurately identifying suspicious transactions and can be deployed in real-time environments.

3. Tools Used

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- scikit-learn (Isolation Forest, LOF)
- XGBoost
- Google Colab

4. Steps Involved in Building the Project

1. Loaded and explored the dataset using Pandas.
2. Cleaned the dataset and handled missing values using SimpleImputer.
3. Applied Isolation Forest and Local Outlier Factor to identify potential frauds.

4. Trained XGBoost on the cleaned dataset using stratified splitting.
5. Evaluated performance using confusion matrix and classification report.
6. Saved the trained model as a `.pkl` file for future deployment.

5. Conclusion

The project successfully built a machine learning pipeline that detects credit card fraud with high accuracy. XGBoost performed exceptionally well in differentiating fraud from non-fraud transactions. This model can be integrated into a financial system to enhance fraud detection efforts.