Credit Card Fraud Detection Using Machine Learning

Pavan Kotian

1️⃣ Objective

The objective of this project is to build and evaluate machine learning models to detect fraudulent credit card transactions in a highly imbalanced dataset, while maximizing fraud detection performance using appropriate evaluation metrics.

2️⃣ Dataset Overview

- Dataset size: **284,807 transactions**
- Fraudulent transactions: **492**
- Legitimate transactions: **284,315**
- No missing values present
- Strong class imbalance (~0.17% fraud)

This imbalance makes accuracy an unreliable metric and necessitates the use of precision, recall, ROC-AUC, and class balancing techniques.

3️⃣ Methods Used

- Data preprocessing and feature scaling
- Train–test split
- Models:
    - Logistic Regression
    - Random Forest Classifier
- Techniques for imbalance handling:
    - Class weights
    - SMOTE (Synthetic Minority Oversampling Technique)
- Evaluation metrics:
    - Precision
    - Recall

- F1-score
- Confusion Matrix
- ROC-AUC
- Precision-Recall Curve

4️⃣ Model Experiments & Results

Logistic Regression (with SMOTE)

- Fraud Precision: **0.06**
- Fraud Recall: **0.92**
- Good recall but very low precision, resulting in many false positives.

Random Forest (with SMOTE)

- Fraud Precision: **0.87**
- Fraud Recall: **0.83**
- ROC-AUC: **0.968**

This model provided the best balance between detecting fraud and minimizing false alarms.

5️⃣ Key Insights

- Accuracy is misleading for imbalanced datasets.
- Recall is critical in fraud detection but must be balanced with precision.
- Random Forest significantly outperformed Logistic Regression.
- SMOTE helped improve minority-class learning.
- Threshold tuning impacts precision-recall tradeoff.

6️⃣ Final Conclusion

Random Forest combined with SMOTE proved to be the most effective approach for credit card fraud detection in this dataset. The model achieved strong recall while maintaining high precision, making it suitable for real-world fraud detection systems where minimizing missed fraud cases is critical.