Credit Card Fraud Detection Using Random Forest

By Aravind Suresh

Introduction

- Objective: Detect fraudulent credit card transactions using a machine learning model.
- Dataset: Credit card transaction data with features V1 to V28, Time, Amount, and Class.

Dataset Overview

- The dataset consists of 31 columns, including 28 anonymized features (V1 to V28), transaction Time, Amount, and the target variable Class (0 for non-fraud, 1 for fraud).
- Total Records: 284,807 transactions

Data Preprocessing

Steps:

1. Load and inspect the dataset.
2. Ensure the target column Class is of integer type.
3. Handle missing values using forward fill.
4. Scale numerical features using StandardScaler.

Handling Class Imbalance

- Technique: SMOTE (Synthetic Minority Over-sampling Technique)
- Purpose: Generate synthetic samples for the minority class (fraudulent transactions) to balance the dataset.

Model Selection

1. Algorithm: Random Forest Classifier
2. Parameters:

- 1. n_estimators: 100 (number of trees in the forest)
- 2. random_state: 42 (for reproducibility)

Model Training

Process:

1. Split the data into training and testing sets (80/20 split, stratified).
2. Apply SMOTE to the training set.
3. Train the Random Forest model on the resampled dataset.

Model Evaluation Metrics

Metrics:

1. Accuracy: 1.00
2. Precision: 0.87
3. Recall: 0.83
4. F1 Score: 0.85
5. ROC AUC Score: 0.98

Confusion Matrix

Description :

- True Positives (TP): 81 (Fraudulent transactions correctly identified as fraud)
- True Negatives (TN): 56852 (Non-fraudulent transactions correctly identified as non-fraud)
- False Positives (FP): 17 (Non-fraudulent transactions incorrectly identified as fraud)
- False Negatives (FN): 12 (Fraudulent transactions incorrectly identified as non-fraud)

56852	12
12	81

Results Interpretation

- Accuracy: High overall correctness.
- Precision: High precision indicates few false positives.
- Recall: Improved recall indicates better detection of fraudulent transactions.
- F1 Score: Good balance between precision and recall.
- ROC AUC Score: Excellent ability to distinguish between classes.

Conclusion

Summary:

- The Random Forest model, combined with SMOTE, provides a robust credit card fraud detection solution with high accuracy and balanced precision and recall.
- Future Work: Exploring further hyperparameter tuning, additional features, and other algorithms to enhance performance.