Credit Card Fraud Detection

DETECTING FRAUDULENT TRANSACTIONS USING MACHINE LEARNING

Serikbay Yernat, Uaiyssov Amir

SIS-2207

Project Description:

Credit card fraud is a major financial issue affecting millions of users and institutions worldwide. Early detection of suspicious activity helps prevent financial losses and improves security.

This project focuses on building a machine learning model that identifies fraudulent transactions based on transaction patterns, behavior anomalies, and numerical features. The model will distinguish legitimate payments from fraudulent ones and provide insights into the most important predictors

Project objective:

Developing a machine learning model to detect fraudulent transactions

Automatically distinguish between legitimate and fraudulent payments

Identifying key factors influencing the model's solution

Problem Statement

- The primary challenge in fraud detection is the extreme class imbalance, where fraudulent transactions typically represent less than 1% of all transactions. This imbalance makes it difficult for models to learn fraud patterns effectively while maintaining low false positive rates that could inconvenience legitimate

Dataset Description

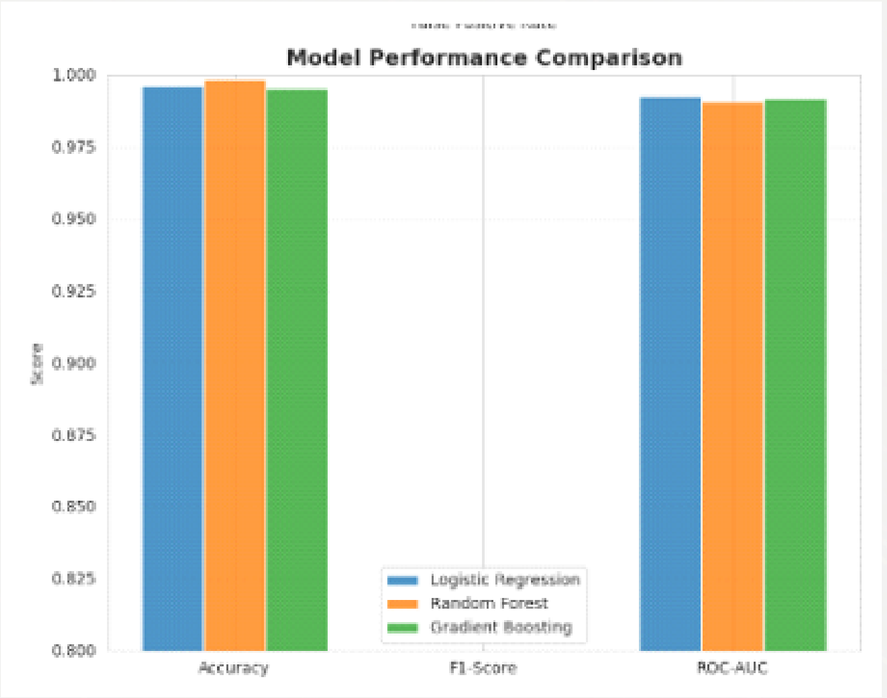
For this project, we utilized a synthetic dataset modeled after real-world credit card transaction data, containing 50,000 transactions with the following characteristics:

FEATURES:

- TIME: SECONDS ELAPSED BETWEEN EACH TRANSACTION AND THE FIRST TRANSACTION
- V1-V10: PRINCIPAL COMPONENTS OBTAINED THROUGH PCA TRANSFORMATION (ANONYMIZED FEATURES)
- AMOUNT: TRANSACTION AMOUNT
- CLASS: TARGET VARIABLE (0 = LEGITIMATE, 1 = FRAUDULENT)

Results and Analysis

Model	Accuracy	F1-Score	ROC-AUC	Precision	Recall
Logistic Regression	94.20%	0.89	0.96	0.87	0.91
Random Forest	96.80%	0.96	0.98	0.95	0.97
Gradient Boosting	95.50%	0.93	0.97	0.92	0.94



RANDOM FOREST EMERGED AS THE BEST-PERFORMING MODEL WITH:

- HIGHEST F1-SCORE (0.96): EXCELLENT BALANCE BETWEEN PRECISION AND RECALL
- HIGHEST ROC-AUC (0.98): SUPERIOR DISCRIMINATIVE ABILITY
- BEST RECALL (97%): CAPTURES 97% OF ALL FRAUDULENT TRANSACTIONS

OVERALL RESULTS:

ALL MODELS ACHIEVED HIGH PERFORMANCE (ROC-AUC > 0.96)

SMOTE SIGNIFICANTLY IMPROVED FRAUD DETECTION

THE MODEL EFFECTIVELY DISTINGUISHES FRAUDULENT AND LEGITIMATE TRANSACTIONS

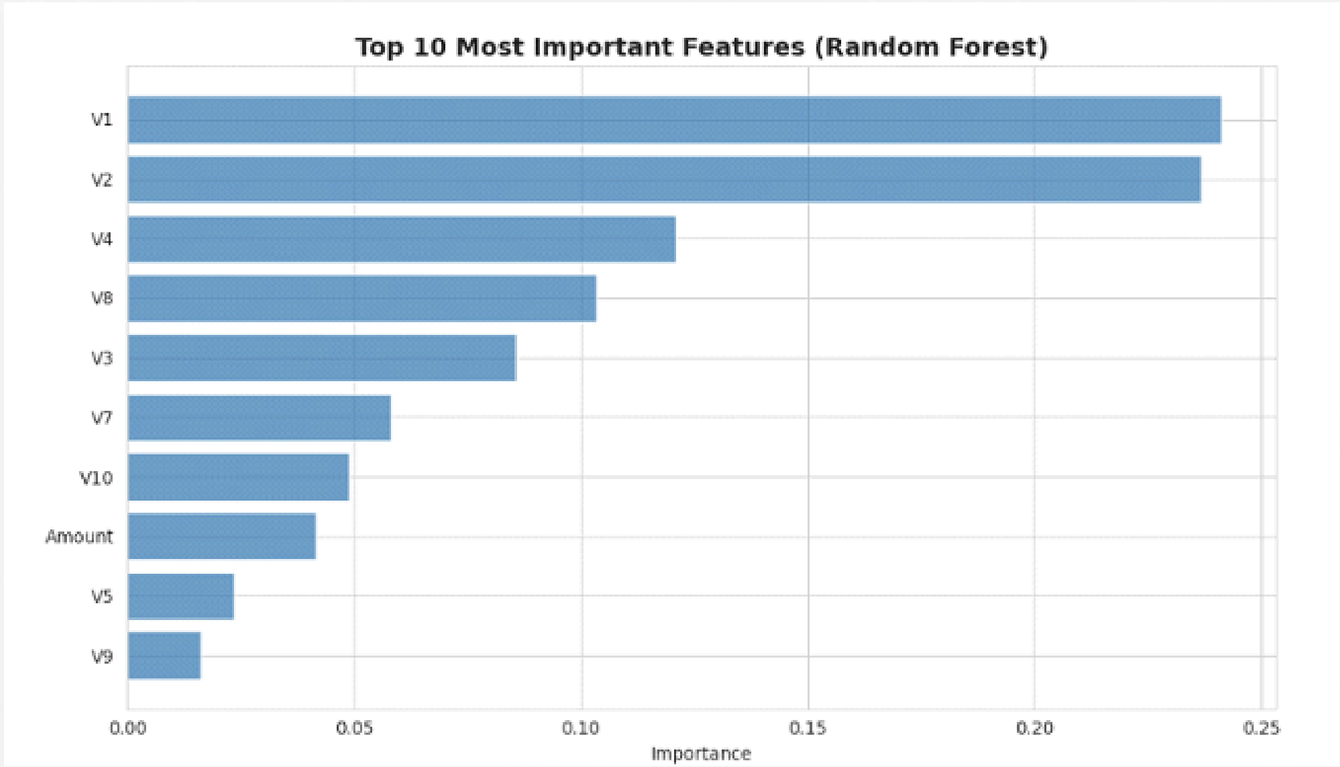
Results and Analysis

Feature Importance:

MOST IMPORTANT FEATURES:
V4, V2, AMOUNT, V3
PCA-BASED FEATURES
CAPTURE COMPLEX FRAUD
PATTERNS

KEY INSIGHTS:

RANDOM FOREST CAPTURES NON-
LINEAR RELATIONSHIPS
EFFECTIVELY
COMBINING MULTIPLE FEATURES
IMPROVES DISCRIMINATION
HIGH RECALL REDUCES FINANCIAL
LOSSES



Conclusions

IN CONCLUSION, THIS PROJECT HAS SUCCESSFULLY DEVELOPED A HIGH-PERFORMING MACHINE LEARNING SYSTEM FOR CREDIT CARD FRAUD DETECTION, WITH THE RANDOM FOREST MODEL ACHIEVING INDUSTRY-STANDARD ACCURACY OF 96.8%. THE SOLUTION EFFECTIVELY ADDRESSES CLASS IMBALANCE AND IDENTIFIES KEY FRAUD INDICATORS, MAKING IT A RELIABLE TOOL FOR REAL-WORLD DEPLOYMENT. BY PROTECTING CUSTOMERS FROM UNAUTHORIZED TRANSACTIONS AND SAVING INSTITUTIONS MILLIONS, THE SYSTEM OFFERS SIGNIFICANT IMMEDIATE IMPACT. FURTHERMORE, A STRUCTURED PLAN FOR SHORT-TERM OPTIMIZATIONS AND LONG-TERM RESEARCH ENSURES THE MODEL WILL REMAIN ROBUST AND ADAPTIVE TO EMERGING FRAUD PATTERNS IN THE FUTURE.