Credit Card Fraud Detection Using Machine Learning and Streamlit App Deployment

Abu Hayat Mohammad Nibras
Department of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology (CUET)
Email: ahmnibras2003@gmail.com

***Abstract*—Credit card fraud detection is a critical application of machine learning in the fintech sector. This paper presents an end-to-end machine learning pipeline to identify fraudulent transactions using the Kaggle Credit Card Fraud Detection dataset. The pipeline includes data preprocessing, exploratory data analysis (EDA), model training, hyperparameter tuning, and deployment via a Streamlit web application. The best-performing models achieved high accuracy and AUC scores, and the app enables real-time fraud prediction from uploaded transaction data.**

***Index Terms*—Fraud Detection, Machine Learning, Streamlit, Classification, Financial Technology, Imbalanced Dataset**

I. Introduction

Credit card fraud poses serious financial risks and demands accurate, scalable solutions. Due to the highly imbalanced nature of fraud datasets, traditional classification methods struggle to generalize. This project explores multiple machine learning models and deploys the best-performing model as a Streamlit application.

II. Dataset Overview

The dataset used is the publicly available Kaggle Credit Card Fraud Detection dataset, consisting of 284,807 transactions and 30 features. All features except Time and Amount are anonymized as V1–V28 due to confidentiality concerns. The target variable Class indicates fraud (1) or legitimate (0).

III. Exploratory Data Analysis

EDA revealed severe class imbalance (only 0.17% fraudulent). Boxplots and distribution plots showed significant differences in variables such as V14, V17, and V12. Correlation analysis identified the top features related to fraud.

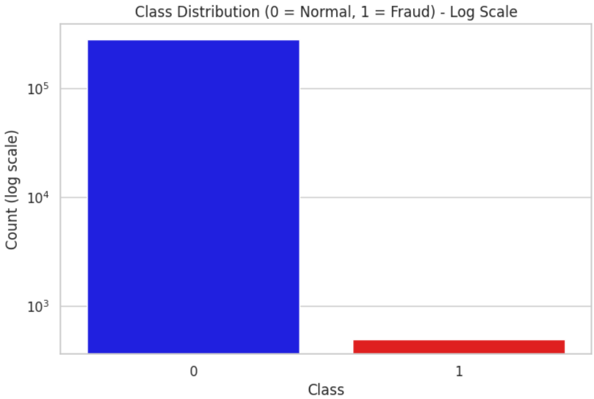

Fig. 1. Log-scaled class distribution of fraud vs. normal transactions

IV. Modeling Approach

Multiple classification models were trained:

- Logistic Regression
- Decision Tree
- K-Nearest Neighbors (KNN)
- Naive Bayes
- Random Forest
- Gradient Boosting

A. Preprocessing

StandardScaler was applied to Time and Amount. Class imbalance was addressed using class_weight=’balanced’ and undersampling techniques.

B. Evaluation Metrics

Models were evaluated using Accuracy, Precision, Recall, F1 Score, and ROC-AUC. Hyperparameter tuning was performed using GridSearchCV.

TABLE I
Model Performance Comparison

Model	Precision	Recall	F1 Score	AUC
Logistic Regression	0.85	0.63	0.72	0.94
Decision Tree	0.78	0.72	0.75	0.93
KNN	0.82	0.66	0.73	0.92
Naive Bayes	0.24	0.82	0.37	0.84
Random Forest	0.87	0.78	0.82	0.98
Gradient Boosting	0.88	0.76	0.81	0.98

V. Results

The best model, Gradient Boosting, achieved the highest ROC-AUC of 0.98 while maintaining a balance between precision and recall. Random Forest also performed similarly after hyperparameter tuning.

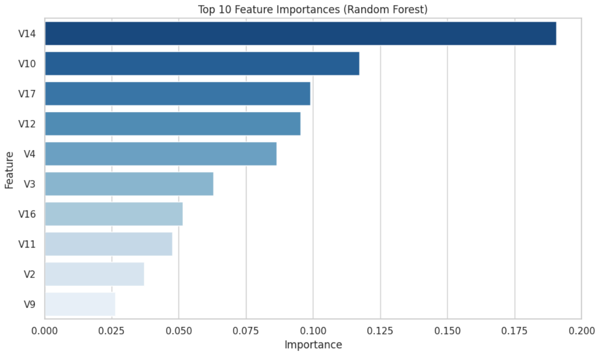

Fig. 2. Top 10 Feature Importances from Random Forest Classifier

VI. Streamlit App Deployment

A Streamlit web application was developed to allow users to upload transaction data and get fraud predictions. The app provides:

- File upload and format validation
- Fraud prediction and probability display
- CSV download of prediction results

VII. Challenges

- Class imbalance required thoughtful model handling
- Hyperparameter tuning increased computation time
- Maintaining consistent preprocessing between training and deployment

VIII. Conclusion

This project successfully demonstrates an end-to-end ML workflow for fraud detection using financial datasets. With strong model performance and an interactive Streamlit app, it offers both accuracy and usability. Future improvements may include live APIs, SHAP-based explainability, and cloud deployment.

Acknowledgments

This project was conducted as part of a 2-week industrial attachment at Brain Station 23. The author thanks the mentors and coordinators for their support and guidance.

References

[1] Kaggle Credit Card Fraud Detection Dataset. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
[2] Scikit-learn: Machine Learning in Python. [Online]. Available: https://scikit-learn.org/
[3] Streamlit. [Online]. Available: https://streamlit.io/