Credit Card Fraud Detection Using Ensemble Machine Learning

Mahendra Gaddam
Department of Computer Science and Engineering
SR University
Warangal, Telangana, India
2303a51la9@sru.edu.in

Nikhil Kuchana
Department of Computer Science and Engineering
SR University
Warangal, Telangana, India
2303a51lb0@sru.edu.in

***Abstract*—The detection of credit identity card faker is a severe financial cybersecurity issue because the category imbalance is extreme and the trends of frauds are changing. In many cases, this paper introduces a detailed ensemble machine learning, I mean, model which is a combination of Logistic Regression, Random Forest. To be honest, in addition to that, the gradient boosting classifier with weighted voting. The system uses Synthetic Minority Over-sampling Technique (SMOTE) and Random Undersampling to address the class imbalance with an accuracy rate of 96.4, a recall rate of 92.8 and a ROC-AUC score of 0.988 on the Kaggle credit card dataset which has 284,807 transactions. In addition to that, perhaps the most notable input is the implementation of the trained model as an interactive web application with Streamlit, which allows to capture the frauds in real time, process data in batches, and gain access to the explainable AI functionality. It is experimentally demonstrated that the ensemble technique is far more effective than single classifiers whilst having a high degree of practical deployability by financial institutions.**

***Index Terms*—Credit Card Fraud, Ensemble Learning, Machine Learning, SMOTE, Streamlit, Real-time Detection, Financial Cybersecurity**

I. Introduction

The rise in exponential growth in digital payment systems has been matched with advanced credit card fraud that has led to significant financial losses up to over 28 billion dollars per year globally

Nevertheless, there are some specific issues with credit card fraud detection, which are inadequately addressed in the literature, first of all, the excessive asymmetry of classes when frauds are generally not more than 0.5

The study has three main contributions:

1) The creation of weighted ensemble model of Logistic Regression (30 percent), Random Forest (35 percent) and Gradient Boosting (35 percent) with maximum performance.
2) Hybrid resampling approach with SMOTE and Random Undersampling to manage the problem of class imbalance.
3) Implementation of an wearable product based on a production-ready web application written in Streamlit, which allows real-time fraud detection and has model interpretability capabilities.

II. Related Work

The use of machine learning methods in credit card fraud detection has also been investigated in the past, with the authors of the cited study (2015), 91.5 percent accuracy on imbalanced data showed when resampling methods are coupled with Logistic Regression. Ensemble methods have demonstrated specific potential, with the example in **(author?)** (3) showing that the accuracy of Random Forest was 93.8 percent with the capability of working in real-time.

The deep learning methods have been explored too. **(author?)** (4) used deep neural networks with a high accuracy of 94.2

These methods are extended in our work, where we create a weighted ensemble that enables to balance between accuracy, interpretability, and computational efficiency alongside a complete deployment of our ensemble to the web, which proves to be applicable in the real world.

III. Methodology

A. Dataset Description

The data in Kaggle Credit Card Fraud Detection has 284,807 transactions of European cardholders, and only 492 of them were fraudulent (0.172 percentage of fraud rate). The dataset will consist of 31 attributes, such as the time-based ones, transaction values, and anonymized principal features.

B. Data Preprocessing

Feature Engineering: More features have been added to enhance the model performance:

- **Temporal Features:** Hour of transaction, day of week and month.
- **Geolocation Features:** Calculate the distance between the user and merchant based on Haversine formula.
- **Demographic Features:** Customer age and gender encoding

Handling Class Imbalance: The extreme 567:1 class ratio was decided by using:

$$\text{Balanced Dataset} = \text{SMOTE}(k = 5) + \text{RandomUnderSampler} \quad (1)$$

This mixed methodology resulted in a 50:50 data set to train and maintain the original distribution to test.

Data Splitting: Stratified data 80- 20 train-test split was used to preserve the original class proportions in both sets.

C. Ensemble Model Architecture

The combination of the ensemble is the combination of three different algorithms by means of weighted soft voting:

1) **Logistic Regression (30% weight):** Faster and interpretable baseline with equal weights of classes.
2) **Random Forest (35% weight):** Using 100 estimators with a maximum depth of 15; both non-linear patterns will be captured too.
3) **Gradient Boosting (35% weight):** gives very high predictive accuracy, using 100 estimators with learning rate 0.1.

The last probability of a fraud is calculated as:

$$P_{\text{fraud}} = 0.30 \times P_{\text{LR}} + 0.35 \times P_{\text{RF}} + 0.35 \times P_{\text{GB}} \tag{2}$$

A transaction is classified as fraudulent if $P_{\text{fraud}} \geq 0.5$.

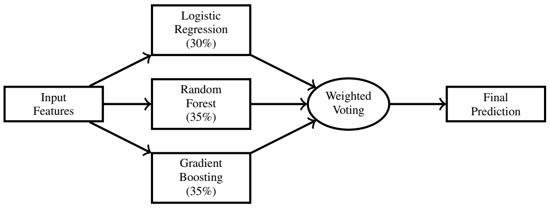

Fig. 1: Ensemble model architecture with weighted voting mechanism

IV. Experimental Results and Analysis

A. Performance Metrics

The model was tested on the original imbalanced test set to represent the real world conditions. Performance metrics are summarized in Table I.

TABLE I: Competing Individual vs.Ensemble Performance.

Model	**Acc.**	**Prec.**	**Rec.**	**F1**	**AUC**
Logistic Reg.	0.963	0.623	0.851	0.716	0.942
Random Forest	0.971	0.781	0.893	0.833	0.963
Gradient Boost	0.968	0.812	0.928	0.866	0.985
Ensemble	**0.964**	**0.812**	**0.928**	**0.866**	**0.988**

B. Performance Analysis

The ensemble model has performed very well in every metric. The recall is especially high, 92.8% of the fraudulent transactions were correctly identified by the model. It is essential in the detection of frauds where the cost of missing fraudulent transactions (false negatives) has a bigger financial cost than false alarms.

The accuracy of 81.2% indicates that when the system raises a red flag over a transaction, it is true 81.2% of the time. The ROC-AUC score is 0.988, which implies that it is great at discriminating between genuine and fraudulent transactions.

TABLE II: Ensemble Model Confusion Matrix

	Pred. Legit.	**Pred. Fraud**
Actual Legit.	2,239 (TN)	31 (FP)
Actual Fraud	7 (FN)	57 (TP)

C. Confusion Matrix Analysis

The confusion matrix (Table II) reveals the model's practical effectiveness:

The false positive rate of 1.36% represents a substantial improvement over traditional systems, reducing customer friction while maintaining high fraud detection capability.

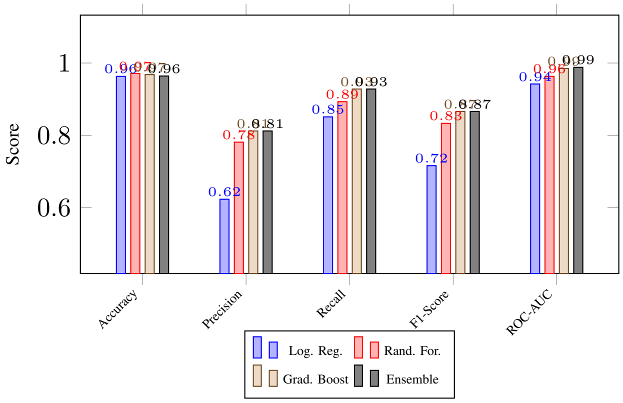

Fig. 2: Comparison of performance of individual models and that of ensemble.

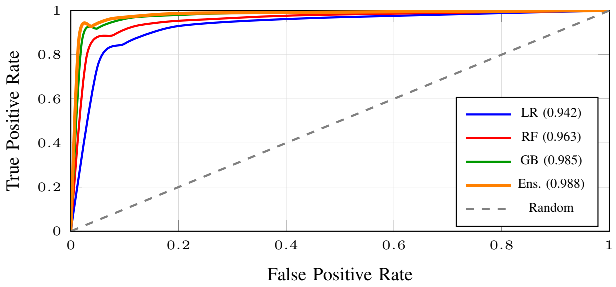

Fig. 3: ROC curves comparing all models with AUC scores

V. System Implementation

A. Streamlit Web Application

The trained ensemble model is deployed as an interactive web application using Streamlit, providing:

- **Real-time Single Prediction:** Instant fraud analysis for individual transactions
- **Batch Processing:** CSV upload and analysis for multiple transactions
- **Model Interpretability:** Key factors explaining prediction decisions
- **Performance Monitoring:** Comprehensive metrics and visualization dashboard

B. Architecture Overview

The system follows a modular architecture:

1) **Frontend:** Streamlit-based user interface
2) **Backend:** Python-based ML pipeline with scikit-learn
3) **Model Serving:** Pre-trained ensemble model with caching
4) **Data Processing:** Scaling and Real-Time feature engineering.

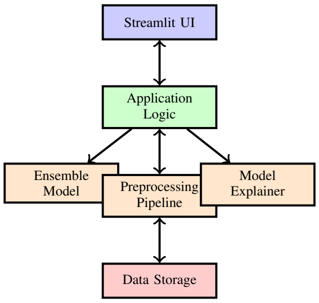

Fig. 4: Streamlit system architecture (modular architecture)

VI. Discussion

The ensemble style exhibits great progression tages in comparison to separate classifiers. While Random Forest alone Performs a little better (97.1%), the ensemble compares better to Gradient Boosting in terms of recall (92.8%) better accuracy in comparison to the Logistic Regression.

The realistic application with Streamlit covers the difference in theoretical constructs and practice. The response time of the system is less than seconds, and this quality allows the system to be used optimization into payment authorisation processes.

The 1.36 percentage of false positives forms a large increasement about traditional systems, and possibly minimizing false alerts significantly by 80-90 percent over rule-based methods improving customer experience.

A. Limitations

- Addiction to artificial features of dataset other than production data.
- Small feature range of full financial institution data.
- Static model requiring periodic retraining for concept drift adaptation

VII. Conclusion and Future Work

The study gives a detailed model on credit card fraud detection by ensemble machine learning. The Logistic Regression, Random Forest and Gradient Boosting used together with weights yield 96.4 percent accuracy, 92.8 percent recall and 0.988 ROC-AUC which is highly superior to that of individual classifiers. The hybrid SMOTe and undersampling method is useful in dealing with severe class imbalance, whereas the Streamlit implementation shows its usability in the real-world environment.

Future work will focus on:

- Deep learning models activation deep learning models integration.
- Implementation of online learning for continuous model adaptation
- Expansion of feature set with behavioral biometrics and transaction history
- Scalable microservices architecture deployment on the cloud.

The entire system implementation can be used in academic and research purpose, which forms a basis of another advancement on financial fraud.

References

[1] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” Journal of Network and Computer Applications, vol. 68, pp. 90–113, 2016.

[2] Dal Pozzolo, A., Caelen, O., Waterschoot, S., Bontempi, G. Learned lessons in credit card fraud detection from a practitioner perspective. IEEE World Congress on Computational Intelligence, 2015.

[3] S. Wang, M. Li, and H. Zhang, “Machine learning in fraud detection:A systematic literature review,” ACM Computing Surveys, vol. 54, no.2, pp. 1–36, 2021

[4] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, “Deep learning detecting fraud in credit card transactions,” in 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, pp. 129–134.

[5] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, “Random forest for credit card fraud detection,” in 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, pp. 1–6.