# Credit Card Fraud Detection Using Ensemble Learning Methods

[1]K Tirumala Achari, [2]Aayush Chandra Pattnaik, [3]Swayamshree Acharya, [4]Aditya Kumar Sahu, [*5]Ashalata Panigrahi

1,2,3,& 4 UG Scholars, Department of Computer Science and Engineering, NIST University, Berhampur, Odisha, India

[*5] Associate Professor, Department of Computer Science and Engineering, NIST University, Berhampur, Odisha, India

## Abstract

Credit cards offers financial flexibility and allow to make payments easily for both online and offline purchase. The creation of digital payment platforms has made it easier for criminals to commit fraud with credit cards at the same time. Credit card fraud is a serious problem that affects different companies, financial organizations, consumers, and banks all over the world. Due to the increasing complexity and variety of fraudulent operations, intelligent fraud detection systems are necessary. This study proposes an efficient credit card fraud detection system based on ensemble machine learning techniques such as random forest, XGBoost, LightGBM, CatBoost, and hybrid ensemble methods. Credit card fraud detection dataset used for experimental study. The developed system includes an interactive web-based dashboard which enables users to upload transaction data and visualize fraud patterns while predicting fraudulent transactions in real time. The performance of the models was evaluated using metrics such as accuracy, precision, detection rate, and F1-score. XGBoost achieves highest accuracy of 0.999544, precision of 0.876923, and F1-score of 0.814286. Random forest achieves lowest accuracy of 0.998982, precision of 0.582524, and F1-score of 0.674157.

**Keywords**: Ensemble learning, machine learning, fraud detection, Random Forest, XGBoost, LightGBM, CatBoost

## 1. Introduction

Credit card fraud detection (CCFD) is an essential area of research in the field of financial security and machine learning. In credit card fraud, an unauthorized person uses another person's credit card information for making transactions without the person's permission or authorization. The various types of fraudulent activities include the theft of credit cards, counterfeit credit cards, and unauthorized transactions through the Internet. The e-commerce sector and the Internet have grown rapidly in recent years, resulting in more credit card transactions and simultaneously increasing the rate of worldwide fraud activities. The research area of credit card fraud detection (CCFD) focuses on the methods for detecting credit card fraud using machine learning techniques for the protection of financial security. The conventional fraud detection systems depend on two primary methods which include rule-

based systems and manual monitoring. The systems can detect specific fraud types but they struggle to catch sophisticated and changing fraud patterns which exist in contemporary financial systems. Organizations now use machine learning methods to detect fraudulent transactions because these systems can automatically identify patterns from past transaction records. However, credit card fraud detection remains challenging due to several factors such as highly imbalanced datasets, evolving fraud patterns, and the need for real-time detection. In most financial transaction datasets, fraudulent transactions represent only a very small percentage of the total transactions, which makes it difficult for machine learning models to accurately detect fraud cases without generating false alarms. To address these challenges, this study proposes an ensemble learning approach that combines multiple machine learning algorithms to improve fraud detection accuracy and reliability. The proposed system integrates Random Forest, XGBoost, LightGBM and CatBoost models into a hybrid ensemble framework. The system uses different models to examine transaction patterns which produce the final prediction through a unified decision-making process. In addition to the machine learning model, this work also presents an interactive web-based fraud detection system that enables users to analyze transaction data, visualize fraud patterns, and perform fraud predictions using trained models. The integration of data visualization and machine learning prediction provides a comprehensive platform for understanding and detecting fraudulent activities in financial transaction.

In [1] the authors have developed the model to detect fraud using logistic regression technique. The results reported that sensitivity 97%, , error rate 2.8%, and accuracy 97.2%. In [2] authors have proposed credit card fraud detection model using machine learning techniques namely, SVM, ANN, and KNN. They reported that ANN achieves highest accuracy: 99.71%, precision:99.68% false alarm rate: 0.12% . SVM model achieves accuracy:94.65%, precision:85.45%, and false alarm rate: 5.2%. KNN model achieves accuracy: 97.15%, precision: 96.84%, false alarm rate: 2.88%. The authors have proposed the model in [3] using three classification techniques namely, logistic regression, k-nearest neighbour, and Naïve Bayes. The achieved accuracy values are 83% for Naïve Bayes, 97.69% for logistic regression, 54.86% for K-NN. Dighe er al. [4] proposed models using different classification techniques namely, KNN, Naïve Bayes, Logistic Regression and Neural Network, Multi-Layers Perceptron and Decision Tree and reported accuracy of each model. KNN achieves highest value of 99.13% and logistic regression achieves lowest value with 96.27%.

The main contributions of this study are

- Development of a machine learning-based fraud detection system using multiple ensemble models including Random Forest, XGBoost, LightGBM, and CatBoost.
- Apply different data preprocessing techniques, including feature scaling and handling class imbalance, to improve classification performance.
- Design of a hybrid ensemble model that combines predictions from multiple classifiers to enhance fraud detection accuracy.
- Development of a web-based dashboard system that provides visualization of transaction data, statistical analysis, and real-time fraud prediction.
- Evaluate different models using metrics such as accuracy, precision, recall, and F1-score.

The rest of this study is organized as follows: Section 2 describes proposed methodology, different ensemble classification techniques, and experimental setup. Section 3 discuss the results. Finally the paper is concluded in Section 4.

# 2. Materials and Methods

## 2.1 Proposed Methodology

The proposed methodology follows the following activities as depicted in Figure 1.

Step 1 : Load the dataset

Step 2: Data pre-processing includes data cleaning, feature scaling and feature engineering.

Step 3: Handle class imbalance using SMOTE.

Step 4: Built models using ensemble learning techniques. The performance of the model are evaluated using accuracy, precision, recall, Fi-score.

Step 5: Feature importance analysis.

Step 6. Deployment of the Credit Card Fraud Detection System

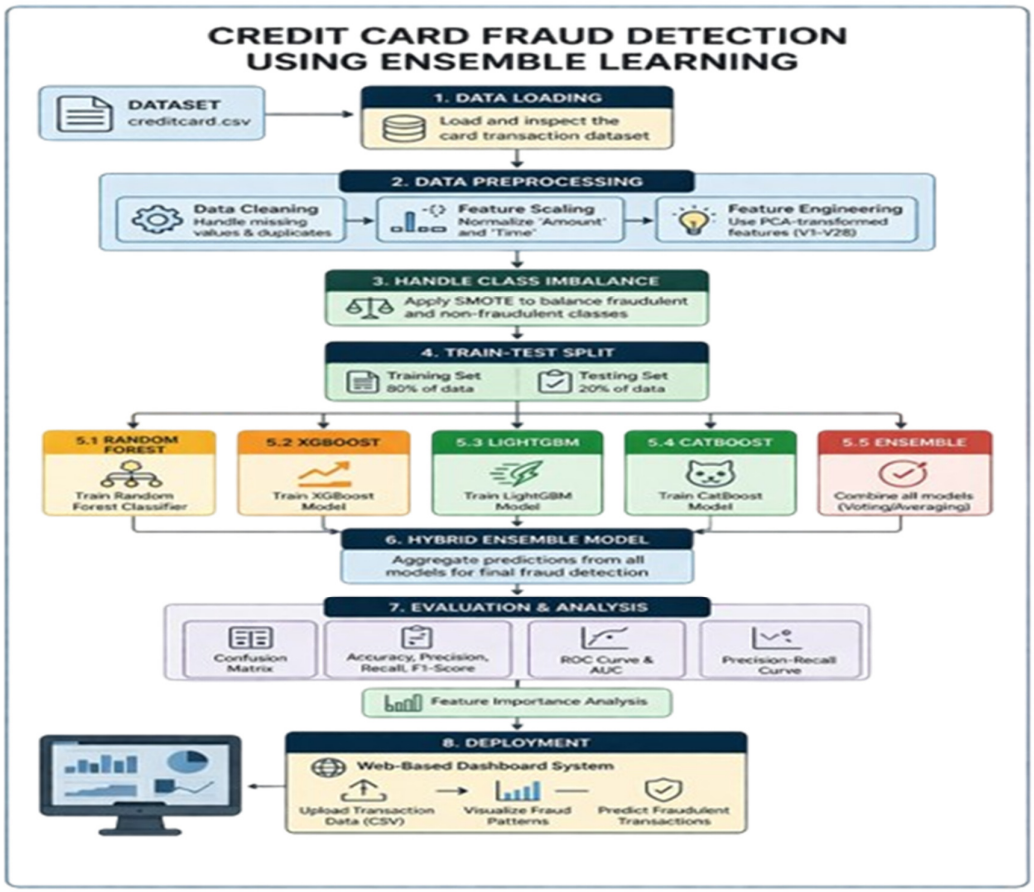


Fig. 1 Proposed Model

## 2.2 Classification Techniques

### Random Forest (RF)

Random Forest [5] handles imbalanced classification tasks, robustness to noisy features, and ability to provide meaningful feature importance scores. RF's bagging approach — training multiple decision trees on random data subsets — creates diverse and uncorrelated trees whose combined prediction is more reliable than any individual tree. The classifier is trained with 100 decision trees (nestimators=100), maximum tree depth of 10 (max_depth=10), random state of 42 for reproducibility, and parallel processing using all available CPU cores (n_jobs=-1). The max_depth parameter of 10 prevents individual trees from becoming too deep and overfitting to the training data

### XGBoost

XGBoost's gradient boosting approach [6] sequentially correcting the errors of previous trees , allows it to focus learning resources on the most difficult-to-classify transactions, which are often the fraudulent ones in an imbalanced dataset. The XGBoost classifier is trained with 100 boosting rounds (n_estimators=100), maximum tree depth of 6 (max_depth=6), learning rate of 0.1, logloss evaluation metric, random state of 42, and parallel processing (n_jobs=-1). The learning rate of 0.1 provides a good balance between training speed and model quality.

### LightGBM

LightGBM uses decision trees [7] that grow efficiently by minimizing memory usage and optimizing training time. The classifier is trained with 100 estimators, maximum depth of 6, and learning rate of 0.1.

### CatBoost

CatBoost's symmetric decision trees [8] produce fast and consistent predictions. The CatBoost classifier is trained with 100 iterations, depth of 6, learning rate of 0.1, and verbose output disabled (verbose=0) to suppress training progress output. CatBoost's ordered boosting technique processes the training examples in a random permutation during each iteration, calculating gradients for each example using only the model trained on the preceding examples.

## 2.3 Experimental Setup

### Data Preprocessing

Data preprocessing prepares the dataset for machine learning algorithms by executing its different tasks. The process includes three main tasks which involve data cleaning, applied PCA ( Principal component analysis) for feature extraction, and scaling on numerical features ( min-max normalization) of the dataset.

### Confusion Matrix

To evaluate the performance of different machine learning models, confusion matrices were generated for each classifier. The confusion matrix provides detailed information about the prediction results by comparing actual and predicted values. The confusion matrix consists of

four components.

True Positive (TP) indicates the count of fraud transactions which were correctly detected.
True Negative (TN) indicates the count of legitimate transactions which were correctly detected.
The False Positive (FP) definition describes a situation where actual transactions were incorrectly detected as suspicious activities.
The definition of False Negative (FN) describes a situation in which criminal transactions were incorrectly recognized as authentic transactions.
The performance metrices are:

Accuracy = (TP + TN) / [ (TP+TN+FP+FN)]
Sensitivity or TPR or Recall = (TP ) / (TP+FN )
FAR = (FP) / ( TN+FP)
Precision = PPV = (TP ) / (TP+FP )
F-value = 2 × ( PPV × TPR / ( PPV + TPR)

**Description of Dataset**

The dataset contains transactions made using credit cards in September 2013 by European cardholders [9]. The dataset holds transactions made by a cardholder over a two-days period, i.e., September 2013. There were 284,807 transactions in total, of which 492, or 0.172 percent, were fraudulent. It contains only numerical input variables which are the result of a PCA transformation. Because disclosing a consumer's transaction details is considered a problem of confidentiality, the main component analysis is applied to the majority of the dataset's features using principal component analysis (PCA). PCA is a standard and widely used technique in the relevant literature for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance. The dataset contains 31 columns including time, v1 to v28, amount and class label ( value 1 in case of fraud and 0 otherwise).

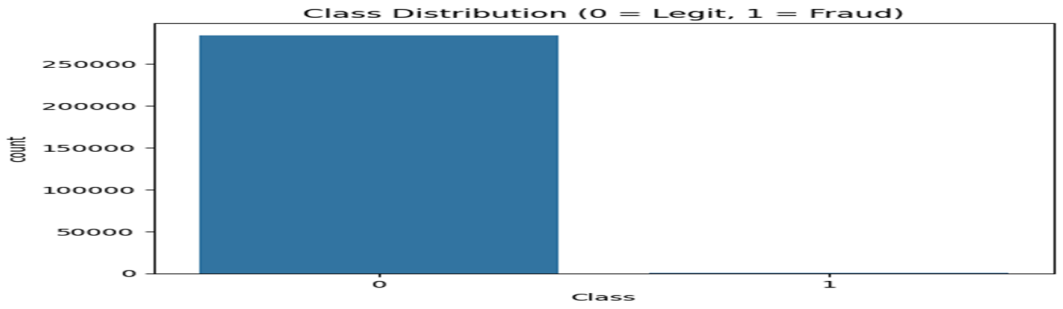


Fig. 2 Distribution of Records

## 3. Results and Discussions

Experimental results were obtained using five ensemble learning methods namely, Random Forest, XGBoost, LightGBM, CatBoost, and hybrid ensemble methods. The performance of the five models are compared using confusion matrix and the evaluation metrics are accuracy,

precision, recall, F1-score and AUC-ROC depicted in table 1. Figures 2, 3, 4, and 5 illustrates confusion matrix of different models.

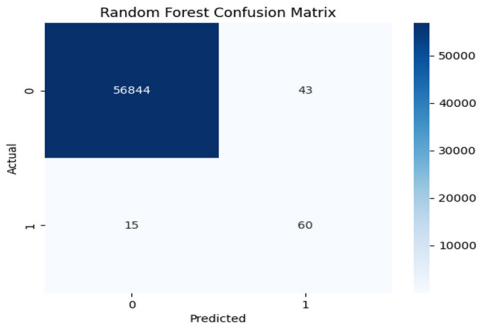


Fig. 2 Confusion matrix of RF

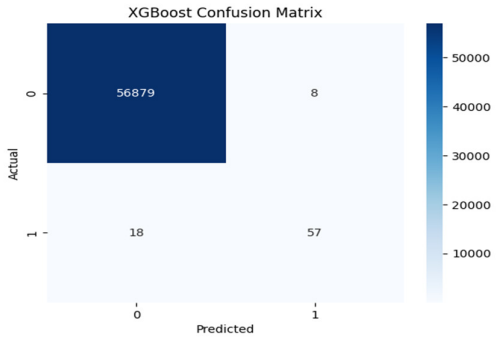


Fig. 3 Confusion matrix of XGBoost

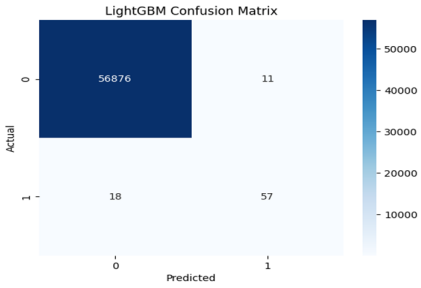


Fig. 4 Confusion matrix of LightGBM

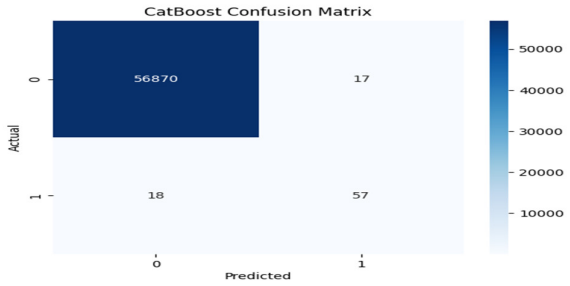


Fig. 5 Confusion matrix of CatBoost

XGBoost model achieves highest accuracy of 0.999544, precision of 0.876923, and F1-score of 0.814286. Both hybrid and random forest model achieves highest recall of 0.80. Hybrid model achieves highest AUC-ROC of 0.993. Random forest model achieves lowest accuracy of 0.998982, precision of 0.582524, F1-score of 0.674157, and AUC-ROC of 0.989. The comparison shows that XGBoost. learning methods achieves very high accuracy in fraud detection. Hybrid model achieves high detection rate.

The ROC-AUC scores for all models are very high, ranging from 0.989 for Random Forest to 0.993 for the Hybrid Model, confirming that all models have excellent discriminative ability. The Hybrid Model's AUC of 0.993 represents the best threshold-independent discrimination performance, meaning it achieves the best trade-off between true positive rate and false positive rate across all possible classification thresholds. Figure 6 represent comparison of accuracy values of different models.

**Table 1 Model Performance Comparison**

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| XGBoost | 0.999544 | 0.876923 | 0.76 | 0.814286 | 0.991 |
| LightGBM | 0.999491 | 0.838235 | 0.76 | 0.797203 | 0.990 |
| Hybrid Model | 0.999421 | 0.769231 | 0.80 | 0.784314 | 0.993 |
| CatBoost | 0.999386 | 0.770270 | 0.76 | 0.765101 | 0.990 |
| Random Forest | 0.998982 | 0.582524 | 0.80 | 0.674157 | 0.989 |

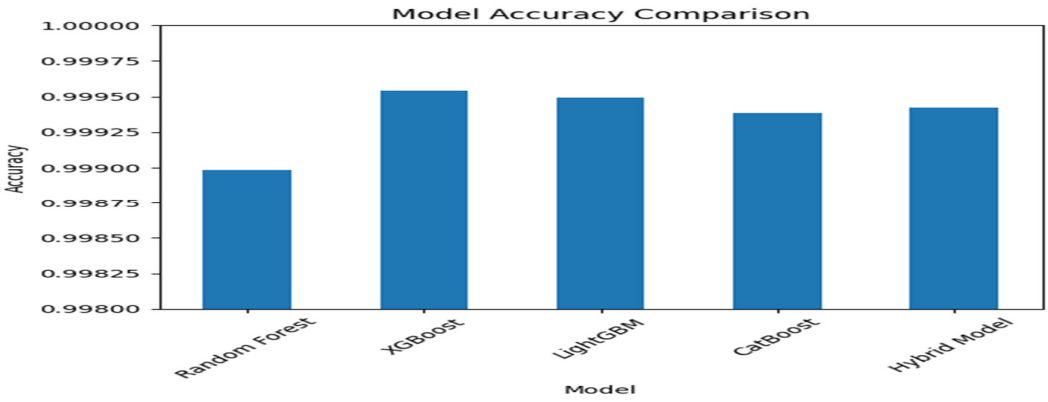


Fig. 6 Comparison of Accuracy

Since fraud detection datasets are highly imbalanced, precision–recall curves provide a better evaluation of model performance compared to accuracy alone. Precision measures how many predicted fraud transactions are actually fraud, while recall measures how many actual fraud transactions are correctly detected. Figure 7 and 8 represents precision-recall curve ROC curve respectively.

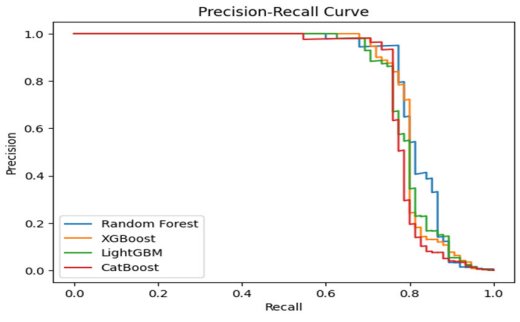


Fig. 7 Precision-Recall Curve

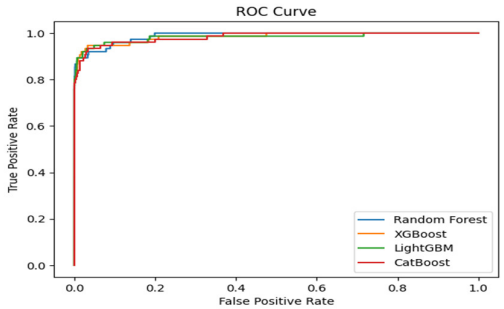


Fig. 8 ROC Curve

### Feature Importance Analysis

Feature importance analysis was performed to determine which attributes contribute most to fraud detection. Each model identifies the most influential features used during classification.

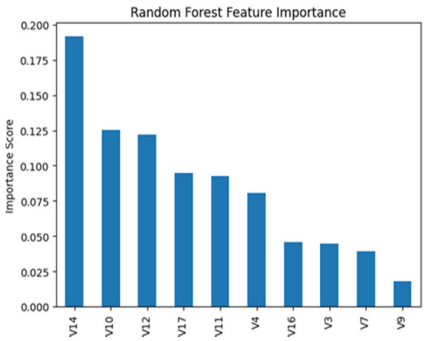


Fig.9 Random Forest Feature Importance

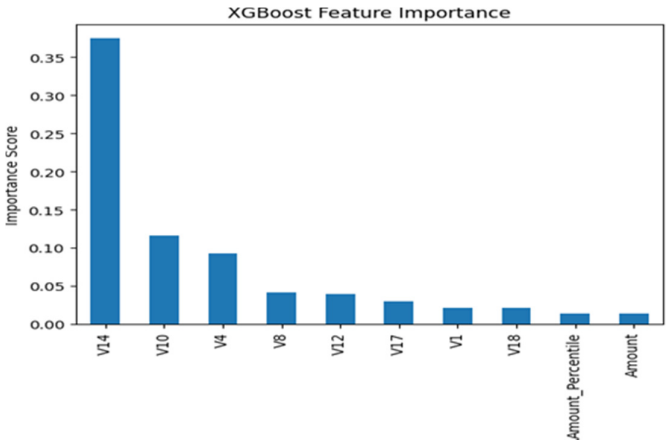


Fig. 10 XGBoost Feature Importance

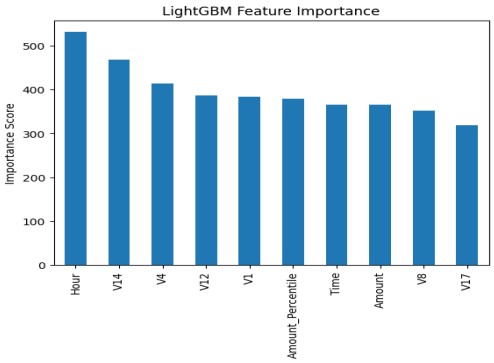


Fig. 11 LightGBM Feature Importance

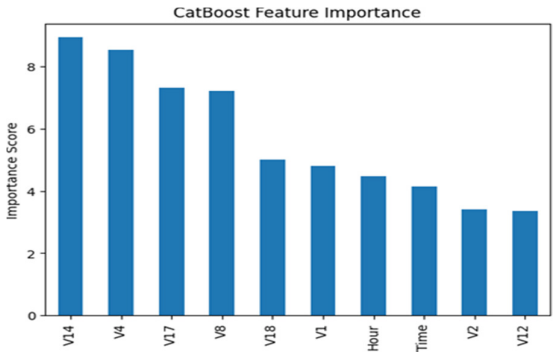


Fig. 12 CatBoost Feature Importance

The deployment of the Credit Card Fraud Detection System as a real-time web application represents a significant advancement in making sophisticated fraud detection technology accessible and practical for real-world use. Each transaction be evaluated and classified within milliseconds to support immediate authorization or decline decisions. The Flask-based deployment architecture of this study supports real-time single transaction prediction through the /predict API endpoint.

## 4. Conclusion

The credit card fraud detection systems was built using different ensemble learning methods such as Random Forest and XGBoost and LightGBM, CatBoost, and hybrid ensemble methods to enhance its ability to detect fraudulent activities. The model was analysed using splitting approach ( 80% training and 20% teasting). Compared the performance of the classifiers using confusion matrix. XGBoost achieves highest accuracy of 0.999544, precision of 0.876923, and F1-score of 0.814286. Random forest achieves lowest accuracy of 0.998982,

precision of 0.582524, and F1-score of 0.674157.
Our future work will focus on deep learning models and autoencoders for improved sequential patten detection. Integration of federated learning to enable collaborative model training across multiple financial institutions without sharing sensitive transaction data.

# References

[1] Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12).

[2] Jain, Y., Tiwari, N., Dubey, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering, 7(5), 402-407.

[3] Safa, M. U., & Ganga, R. M. (2019). Credit Card Fraud Detection Using Machine Learning. International Journal of Research in Engineering, Science and Management, 2(11).

[4] Dighe, D., Patil, S., & Kokate, S. (2018, August). Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.

[5] Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

[6] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[7] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

[8] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

[9] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud