Metadata of the chapter that will be visualized in SpringerLink

Book Title	Intelligent Computing	
Series Title		
Chapter Title	An Ensemble-Based Machine Learning for Predicting Fraud of Credit Card Transactions	
Copyright Year	2022	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Baabdullah**
	Particle	
	Given Name	**Tahani**
	Prefix	
	Suffix	
	Role	
	Division	Data Science and Cybersecurity Center (DSC2), Department of Electrical Engineering and Computer Science
	Organization	Howard University
	Address	Washington, D.C., 20059, USA
	Email	Tahani.baabdullah@bison.howard.edu
Author	Family Name	**Rawat**
	Particle	
	Given Name	**Danda B.**
	Prefix	
	Suffix	
	Role	
	Division	Data Science and Cybersecurity Center (DSC2), Department of Electrical Engineering and Computer Science
	Organization	Howard University
	Address	Washington, D.C., 20059, USA
	Email	Danda.Rawat@howard.edu
Author	Family Name	**Liu**
	Particle	
	Given Name	**Chunmei**
	Prefix	
	Suffix	
	Role	
	Division	Data Science and Cybersecurity Center (DSC2), Department of Electrical Engineering and Computer Science
	Organization	Howard University
	Address	Washington, D.C., 20059, USA
	Email	Chuliu@howard.edu
Author	Family Name	**Alzahrani**
	Particle	
	Given Name	**Amani**

Prefix	
Suffix	
Role	
Division	Data Science and Cybersecurity Center (DSC2), Department of Electrical Engineering and Computer Science
Organization	Howard University
Address	Washington, D.C., 20059, USA
Email	Amani.Alzahrani@bison.howard.edu

Abstract

Recently, using credit cards has been considered one of the essential things of our life due to its pros of being easy to use and flexible to pay. The critical impact of the increment of using credit cards is the occurrence of fraudulent transactions, which allow the illegal user to get money and free goods via unauthorized usage. Artificial Intelligence (AI) and Machine Learning (ML) have become effective techniques used in different applications to ensure cybersecurity. This paper proposes our fraud detection system called Man-Ensemble CCFD using an ensemble-learning model with two stages of classification and detection. Stage one, called ML-CCFD, utilizes ten machine learning (ML) algorithms to classify credit card transactions to class 1 as a fraudulent transaction or class 0 as a legitimate transaction. As a result, we compared their classification reports together, precisely precision, recall (sensitivity), and f1-score. Then, we selected the most accurate ML algorithms based on their classification performance and prediction accuracy. The second stage, known Ensemble-learning CCFD, is an ensemble model that applies the Man-Ensemble method on the most effective ML algorithms from stage one. The output of the second stage is to get the final prediction instead of using common types of ensemble learning, such as voting, stacking, boosting, and others. Our framework's results showed the effectiveness and efficiency of our fraud detection system compared to using ML algorithms individually due to their weakness issues, such as errors, overfitting, bias, prediction accuracy, and even their robustness level.

Keywords (separated by '-')

Fraud detection - Imbalanced datasets - Credit card fraud - Fraudulent transaction - Ensemble learning

An Ensemble-Based Machine Learning for Predicting Fraud of Credit Card Transactions

Tahani Baabdullah(✉), Danda B. Rawat, Chunmei Liu, and Amani Alzahrani

Data Science and Cybersecurity Center (DSC2), Department of Electrical Engineering and Computer Science, Howard University, Washington, D.C. 20059, USA
{Tahani.baabdullah,Amani.Alzahrani}@bison.howard.edu,
{Danda.Rawat,Chuliu}@howard.edu

AQ1

Abstract. Recently, using credit cards has been considered one of the essential things of our life due to its pros of being easy to use and flexible to pay. The critical impact of the increment of using credit cards is the occurrence of fraudulent transactions, which allow the illegal user to get money and free goods via unauthorized usage. Artificial Intelligence (AI) and Machine Learning (ML) have become effective techniques used in different applications to ensure cybersecurity. This paper proposes our fraud detection system called Man-Ensemble CCFD using an ensemble-learning model with two stages of classification and detection. Stage one, called ML-CCFD, utilizes ten machine learning (ML) algorithms to classify credit card transactions to class 1 as a fraudulent transaction or class 0 as a legitimate transaction. As a result, we compared their classification reports together, precisely precision, recall (sensitivity), and f1-score. Then, we selected the most accurate ML algorithms based on their classification performance and prediction accuracy. The second stage, known Ensemble-learning CCFD, is an ensemble model that applies the Man-Ensemble method on the most effective ML algorithms from stage one. The output of the second stage is to get the final prediction instead of using common types of ensemble learning, such as voting, stacking, boosting, and others. Our framework's results showed the effectiveness and efficiency of our fraud detection system compared to using ML algorithms individually due to their weakness issues, such as errors, overfitting, bias, prediction accuracy, and even their robustness level.

Keywords: Fraud detection · Imbalanced datasets · Credit card fraud · Fraudulent transaction · Ensemble learning

1 Introduction

1.1 Credit Card Fraud

The majority of people of different ages have their credit cards and use them daily in most of their purchases. Using credit cards is considered one of the essential things of our life due to its pros of being easy to use and flexible to

K. Arai (Ed.): SAI 2022, LNNS 507, pp. 1–16, 2022.
https://doi.org/10.1007/978-3-031-10464-0_14

pay. Figure 1 shows the approval procedure for a credit card transaction from swipe/use a credit card by cardholder until complete or cancel that transaction, as the following steps: cardholder, merchant's payment system (point-of-sale (POS) terminal/software or e-commerce website), merchant's bank, payment brand, cardholder's bank to authorize that transaction then route it back to the same points with its authorization number/code until arriving the merchant to finalize the transaction with the customer. Therefore, using our credit cards frequently everywhere, including online shopping, will increase the probability of fraud risk and for being used by an unauthorized party. The critical impact of the increment of using credit cards is the occurrence of fraudulent transactions, which allow the illegal user to get money and free goods via unauthorized usage. Credit card fraud (CCF) has become the main issue for financial institutions, the credit card industry, the community, and cardholders. Thus, governments, businesses and companies, and financial institutions pay more attention to this security issue and apply different security detection systems to detect and suspend fraudulent transactions [1,13,17,21].

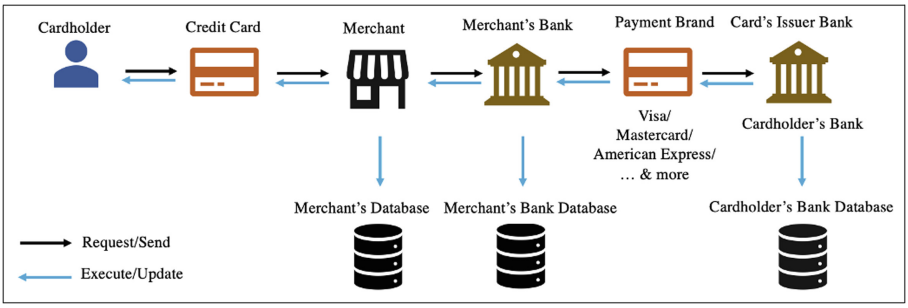

Fig. 1. Credit card transaction approval procedure

Many academics and researchers proposed fraud detection systems and technical methods to solve credit card fraud. Still, there are a lot of issues and challenges with these detection systems and solutions, such as skewed datasets and imbalance classification, lack of public datasets, anomaly/outliers detection, and concept drift. Artificial Intelligence (AI) and Machine Learning (ML) have become effective techniques used in different applications to ensure cybersecurity and build security prevention and detection systems. Thus, credit card issuers should improve their security systems by implementing advanced Credit Card Fraud Prevention and Fraud Detection methods. Machine Learning-based techniques can constantly improve their performance and prediction accuracy of fraud prevention and detection based on cardholder's behavior analysis. However, the changing of the cardholder's behavior and profile and fraud techniques used negatively affect the classification performance and prediction accuracy of CCFD systems [3]. There are many challenges faced by CCFD systems, such

as lack of public real-world datasets, skewed data and imbalanced classification, anomaly/outlier detection, supports real-time detection, concept drift, and reduction of a large number of data [1,7,9,10].

1.2 Ensemble Learning

As known, each ML technique has its pros and cons, such as errors, overfitting, bias, prediction accuracy, and even their robustness level. Also, the classification performance and prediction accuracy for each ML technique vary depending on the datasets. Thus, we can not generalize the best ML-CCFD system or method used with any datasets. During our research to find the single best-performing model to detect fraudulent transactions, we realized the benefit of combining several ML techniques to ensure the performance and accuracy of detection, as mentioned in many research papers as in [14]. Hence, ensemble learning combines many ML classifier models trained on the same datasets to get the final prediction by all of them instead of depending on the single best-performing model, as shown in Fig. 2.

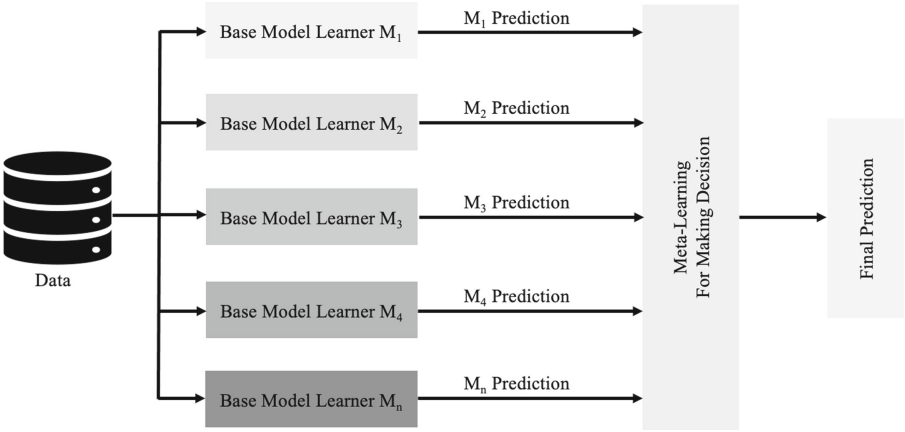

Fig. 2. Ensemble learning

The main advantages of using ensemble learning algorithms are better robustness and predictions. Many ML techniques have variance in the predictions or the model stability. Therefore, ensemble learning techniques provide stable predictions than an individual model; also, ensemble learning presents the best predictive skill than an individual model.[1] Using ensemble learning to combine simple ML algorithms is better than applying complex algorithms with multi-layers.

[1] https://machinelearningmastery.com/ensemble-learning-algorithms-with-python/.

This paper proposes our credit card fraud detection Man-Ensemble CCFD system using an ensemble-learning model with two stages of classification and detection. Stage one (ML-CCFD) utilizes ten machine learning (ML) algorithms then trains them individually to classify credit card transactions to either class 1 as a fraudulent transaction or class 0 as a legitimate transaction. As a result, we compared their classification reports together, precisely precision, recall (sensitivity), and f1-score. Therefore, we selected the most accurate ML algorithms based on their classification performance and prediction accuracy. The second stage is an ensemble-learning CCFD that applies our method Man-Ensemble on the most effective ML algorithms chosen from stage one. The output of this stage is to get the final prediction instead of using common types of ensemble learning, such as voting, stacking, boosting, and others. The results of our framework showed the effectiveness and efficiency of our fraud detection system compared to using ML algorithms individually due to their weakness issues, such as errors, overfitting, bias, prediction accuracy, and even their robustness level. To summarize, the contributions of this paper are as follows:

- Applying our proposed CCFD system on two different datasets, real-world and synthetic datasets, to ensure the accuracy of the final prediction of our framework.
- Training the most common ML algorithms used in credit card fraud detection individually.
- Using our detection method to find the final prediction of our ensemble-learning model.

The rest of this paper is organized as follows; Sect. 2 presents the literature review of some related works that used ensemble-learning in credit card fraud detection. In Sect. 3, we describe our methodology and proposed framework. Section 4 displays the results assessment and performance evaluation. Section 5 summarizes the conclusion of our work and our final results.

2 Literature Review

Many academics and researchers presented detection systems and methods to detect fraudulent credit card transactions using ML techniques. Using ensemble methods (EMs) for classification purposes is considered one of the interesting areas of research in ML. Hence, a lot of recent research mentions the importance of using EMs to improve the classification performance of classifier models by predicting suitable classes.

Using ensemble strategies in unsupervised outlier detection is a common method to improve the estimation of the outlier scores [22]. Also, combining supervised and unsupervised outlier detection algorithms were performed by using sequential [15], and parallel [18] ensemble strategies. An ensemble machine learning approach on real-world datasets was presented by Sohony et al. in [16], which was a combination of random forest and neural network. It worked appropriately to predict the label of a new sample with high accuracy and confidence.

Carcillo et al. proposed a Hybrid technique that combines supervised and unsupervised methods to improve fraud detection accuracy. Thus, unsupervised learning techniques support the fraud detection systems to find anomalies. They computed unsupervised outlier scores in various levels of granularity. They used a real-life credit card dataset for fraud detection. The final results showed the effectiveness of the combination and its improvement of the detection accuracy [6]. Carcillo et al. in [5], a fraud detection open-source platform (SCARFF) was designed, implemented, and tested to detect fraudulent transactions of credit card transactions. The framework used big data tools (Kafka, Spark, and Cassandra) with two ML classifiers (Feedback Random Forest classifier and Delayed classifier). The results displayed the framework's scalability, efficiency, and accuracy over a significant stream of transactions. Motwani et al. applied several ML techniques and evaluated them on actual credit card datasets; then, they proposed a predictive ensemble model for credit risk detection. The proposed model was evaluated based on different performance metrics, and the results of the proposed model showed the improvement of the prediction accuracy and correlation coefficient [11].

Polikar et al. introduced the definition of ensemble learning as an ML paradigm that combines multiple base learners (individual ML algorithms) to resolve the same problem [12]. Arya et al. proposed a predictive framework (DEAL) based on extra-tree ensemble and deep neural network that represents each transaction as a tensor to reveal the latent and inherent relations between spending patterns and fraudulent transactions [2]. Young et al. presented a deep super-learning approach with high log loss and accuracy results than deep neural networks. The results of their deep super-learner showed that the performance of the deep super-learner was better than the performance of the individual base learners and, in some cases, deep neural networks [19]. Hamori et al. made a study to compare the effectiveness of using ensemble learning or deep learning of default payment data and analysis their prediction accuracy and classification ability. The study included three ensemble-learning methods (bagging, random forest, and boosting) and different neural network methods with varying activation functions. The results illustrated that the classification ability to boost ensemble learners is better than other ML methods, including neural networks [8]. Zareapoora and Shamsolmoalia proposed their experiment of training various data mining techniques performed on real-life credit card transactions dataset, and they evaluated each methodology based on specific design criteria. The observed results showed that the bagging ensemble classifier is the best classifier to construct the fraudulent transaction detection model [20].

3 Proposed Method

3.1 CCFD System

ML algorithms are considered the most valuable cybersecurity and fraud detection techniques, but many have weaknesses, such as errors, overfitting, bias, prediction accuracy, and even their robustness level. Ensemble learning is an

assemble method of multi ML algorithms and allows them learned, then use common types of ensemble learning to get the final prediction. Hence, this paper uses many ML algorithms to ensure our model's classification performance, prediction accuracy, robustness, and results. We propose our credit card fraud detection, called the Man-Ensemble CCFD system, based on using an ensemble-learning model that has two stages of prediction, as illustrated in Fig. 3.

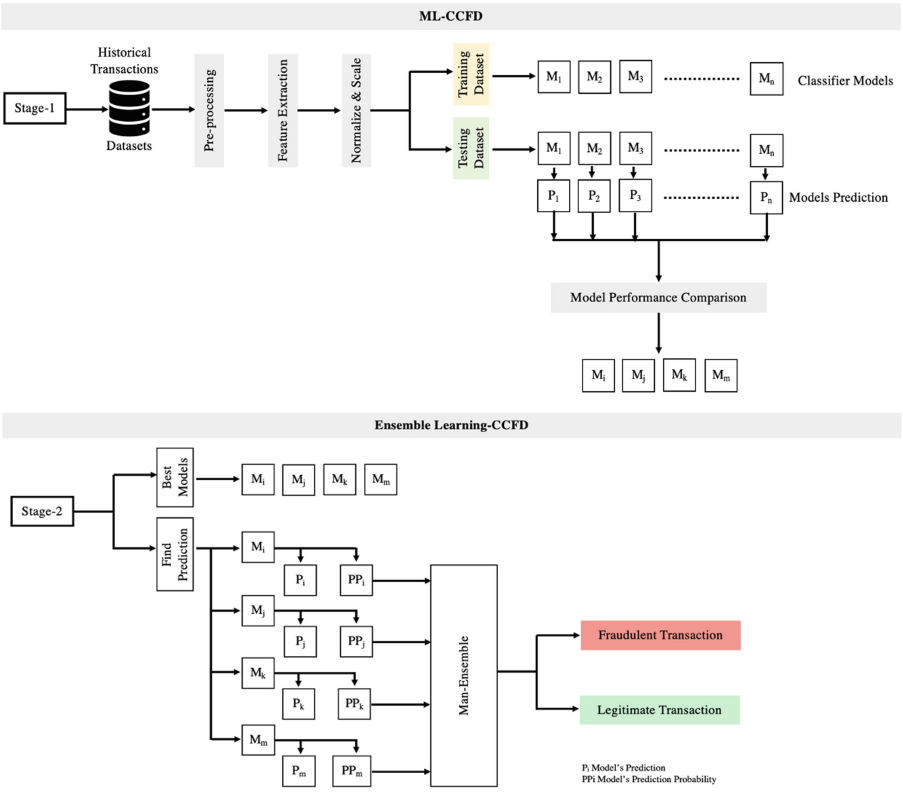

Fig. 3. Our proposed man-ensemble CCFD system

We utilized ten ML algorithms to classify transactions to fraud (class 1) and non-fraud (class 0) in Stage one (ML-CCFD). The ten ML algorithms used in stage one are as follows: Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Support vector clustering (SVC), eXtreme Gradient Boosting (XGBoost), Stochastic Gradient Descent (SGD), Gradient boosting classifiers (GBC), and Light Gradient Boosting (LGB). We used five of these ML techniques in our work [4] to compare between them and find the most accurate one with our data to detect fraudulent transactions accurately. Our old work showed that the RF classifier model was the best ML method due to its highest accuracy, sensitivity, and

AUPRC. Therefore, we have doubled the number of ML techniques used in this current experiment; also, we used an ensemble-learning technique to improve the performance and prediction accuracy.

3.2 Datasets and Features Extractions

These classifier models are trained individually on two datasets (CreditCard.csv, FraudTrain.csv and FraudTest.csv) available on Kaggle datasets, as described in Table 1. We applied our experiment on two different datasets to ensure the accuracy of our learner model's performance and prediction. The first dataset was a real-world dataset, CreditCard.csv, which has 31 features, 28 features (V1, V2, V3 ... V28) were transformed with Principal Component Analysis (PCA) for confidentiality issues. Thus, the rest of the features not transformed with PCA are 'Time' and 'Amount', representing the seconds elapsed between each transaction and the first transaction and the exact amount. The last feature is 'Class', which has the value 1 in case of a fraudulent transaction and 0 in the original transaction. This dataset has credit cards transactions that occurred in two days in September 2013 by European cardholders to contain 492 frauds out of 284,807 transactions. Therefore, this dataset is imbalanced because the fraudulent transactions (positive class) are 0.172% of all transactions.[2]

Table 1. Our datasets

CreditCard datasets	Fraud datasets
Real-world credit card transaction dataset	Simulated/synthetic credit card transaction dataset
Transformed data with PCA for confidentiality	Available to include 1000 customers and 800 merchants
Made by credit cards in September 2013 by European cardholders for researches in University Libre de Bruxelles (ULB)	Created by Brandon Harris from Jan 1st, 2019 to Dec 31th, 2020
492 frauds out of 284,807 transactions	9,651 frauds out of 1,852,394 transactions

The second dataset was the FraudTrain.csv and FraudTest.csv datasets; we merged them into 1,852,394 credit card transactions and 24 features. It is a synthetic (simulated) credit card transaction dataset from January 1, 2019, to December 31, 2020, generated using Sparkov Data Generation tool by Brandon Harris.[3] It contains 1,842,743 legitimate and 9,651 fraud transactions for 1000 customers dealing with a pool of 800 merchants. The features are as follows: transaction index, transaction date and time, credit card number, merchant

[2] https://www.kaggle.com/mlg-ulb/creditcardfraud.
[3] https://github.com/namebrandon/Sparkov_Data_Generation.

name, category of merchant, amount of transaction, cardholder's name, cardholder's gender, cardholder's address, cardholder's latitude and longitude location, cardholder's city population, cardholder's job, cardholder's date of birth, transaction number, UNIX time of the transaction, merchant's latitude and longitude location, and target class.[4] Thus, it is evident that our two datasets are imbalanced data since the positive class is the minority compared to the other class in binary classification example. Figure 4 and 5 show the data ratio before and after adjusting imbalanced data by undersampling it to solve the skewed distribution issue between the fraudulent transactions to the original transactions.

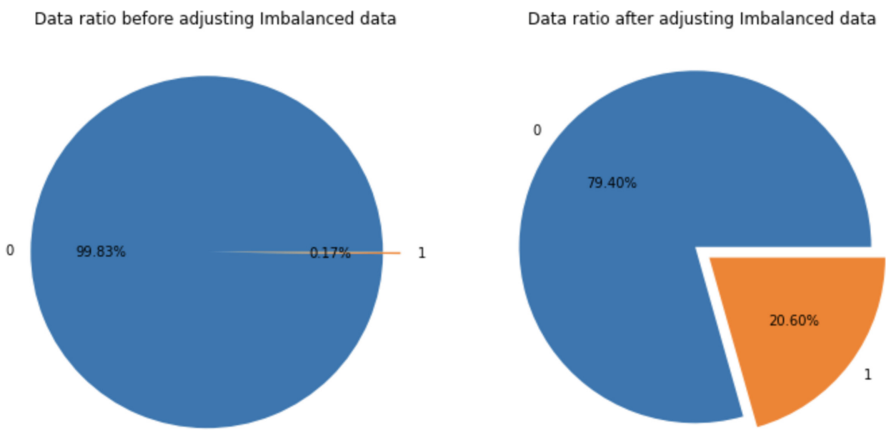

Fig. 4. CreditCard dataset

3.3 Machine Learning and Ensemble Techniques

As shown in Fig. 6, stage one (ML-CCFD) trains ten ML algorithms to classify transactions to fraud (class 1) and non-fraud (class 0) and compare their classification performance to choose the most accurate learner models. The ten ML algorithms used are as follows: Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Support vector clustering (SVC), eXtreme Gradient Boosting (XGBoost), Stochastic Gradient Descent (SGD), Gradient boosting classifiers (GBC), and Light Gradient Boosting (LGB). The stage processing will be as follows, importing datasets, preparing features, under-sampling our data, training models, evaluating classifier models, measuring performance, finally, selecting the most effective learners.

[4] https://www.kaggle.com/kartik2112/fraud-detection?select=fraudTrain.csv.

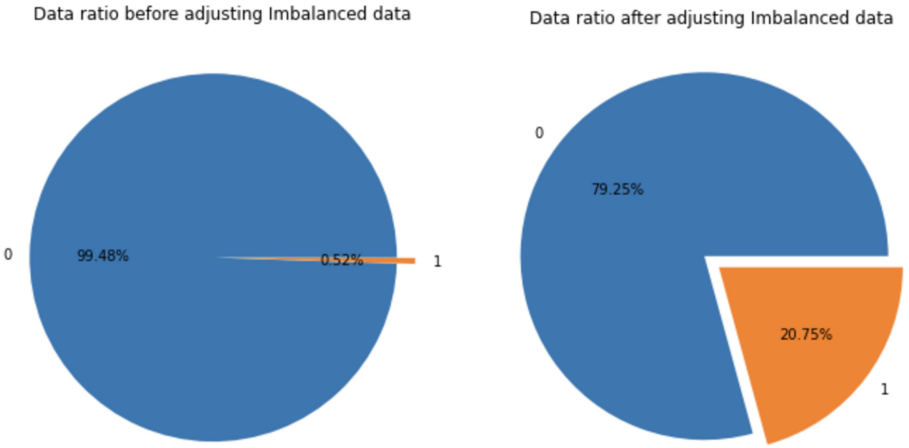

Fig. 5. Fraud dataset

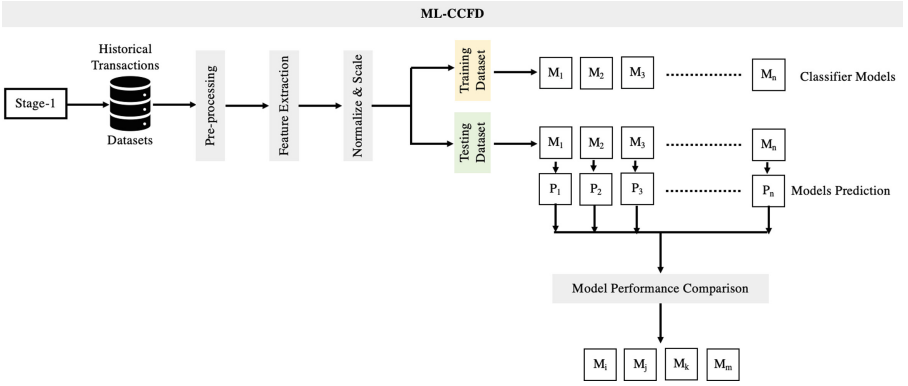

Fig. 6. Stage1 ML-CCFD

We compare their classification reports, precision, recall (sensitivity), f1-score, and false negative. Then, we choose the most accurate classifier models based on their performance and prediction accuracy to proceed to the next stage.

The second stage, as illustrated in Fig. 7, is the Ensemble-learning CCFD model that receives the effective learners from the previous stage.

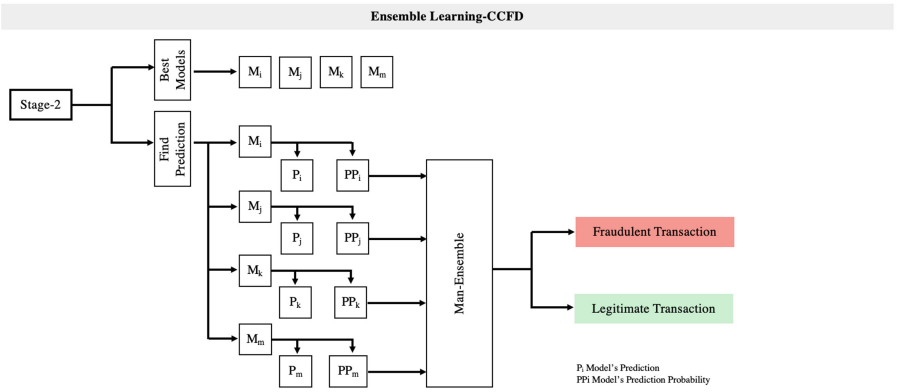

Fig. 7. Stage2 Ensemble-learning CCFD

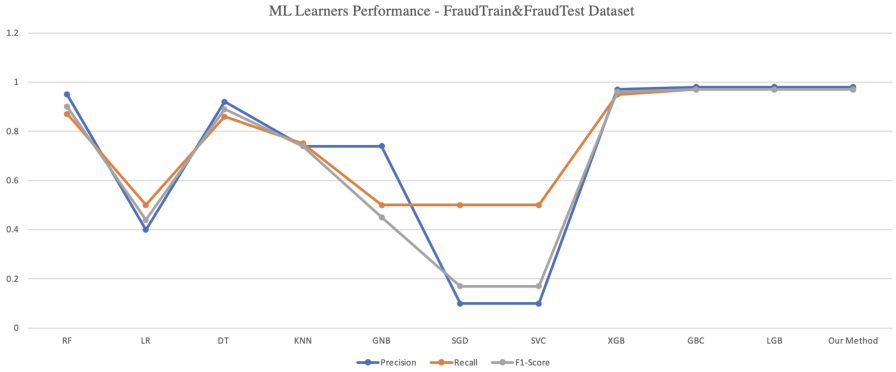

Fig. 8. Prediction accuracy for all ML algorithms for fraud datasets

A transaction will be classified as fraud or non-fraud based on the final prediction via our method instead of using common types of ensemble learning, such as voting, stacking, boosting, and others. Our method aims to find the final prediction depending on the prediction probabilities of the effective classifier models, as explained in algorithm 1. In our experiment, we got the efficient ML algorithms and accurate classifier models in stage one, RF, XGBoost, GBC, and LGB, which are the input of stage two to get the output as the final prediction of the transaction. By receiving the efficient learners in the second stage, the learner's prediction probability is computed for all efficient learners to find the average prediction probability, round it, find the predicted value, and then compare it to the actual value to evaluate our method's accuracy.

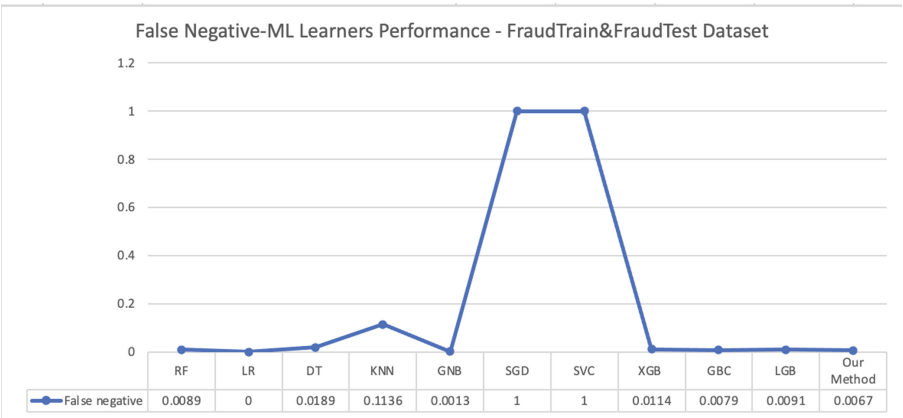

Fig. 9. False negative for all ML algorithms for fraud datasets

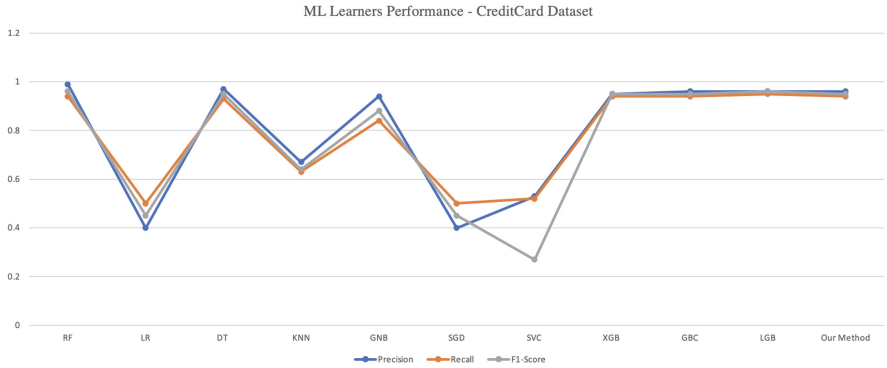

Fig. 10. Prediction accuracy for all ML algorithms for CreditCard datasets

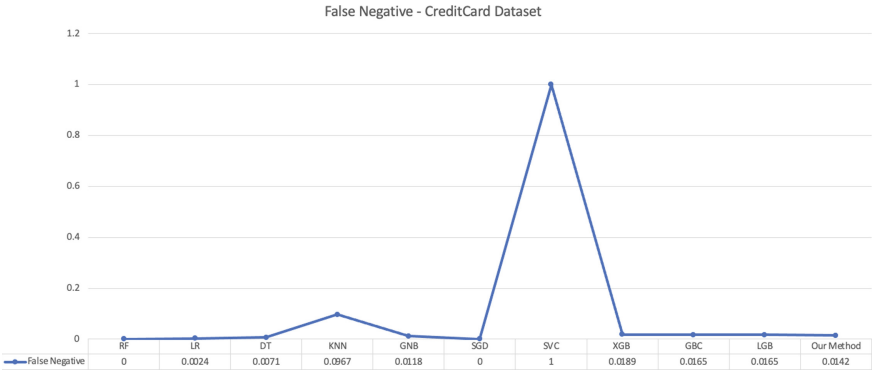

Fig. 11. False negative for All ML algorithms for CreditCard datasets

Algorithm 1. Man-Ensemble Method

Input: Most Accurate ML learners
Output: Final predication

$BestModels = [M_1, M_2, M_3]$
$totalPrediction = 0$
$n = numberofdatarecords$
$k = numberofbestmodels$
$threshold = 0.5$

for $i = 1 \rightarrow n$ **do**
    for $i = 1 \rightarrow k$ **do**
        $findpredict\_proba_{M_k}(X\_Test) for\ class\ 0$

        $totalPrediction+ = predict\_proba_{M_k}(X\_Test)$
    end for
    $avgPrediction = totalPrediction/k$
    if $avgPrediction \geq threshold$ **then**
        $avgPrediction = 1$
    else
        $avgPrediction = 0$
    end if
    $PredictedValue = 1 - avgPrediction$
    $allPredictedValue[i] = PredictedValue$
end for
$compare\ predictedValue == actualValue$
$find\ accuracy\ and\ other\ metrics$

4 Results Assessment and Performance Evaluation

The results of the first stage applied on our datasets are displayed in Tables 2 and 3 to compare the classification reports and the prediction accuracy among the ten ML algorithms, including Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Support vector clustering (SVC), eXtreme Gradient Boosting (XGBoost), Stochastic Gradient Descent (SGD), Gradient boosting classifiers (GBC), and Light Gradient Boosting (LGB). The stage one results show that the most effective ML algorithms and accurate classification models are RF, XGBoost, GBC, and LGB.

Then, the second stage of finding final prediction results is displayed in Tables 4 and 5 for the most effective learning models from stage one.

Table 2. Stage1 results for fraud datasets

Model	Precision	Recall	F1-score
RF	**0.95**	**0.87**	**0.90**
LR	0.40	0.50	0.44
DT	0.92	0.86	0.89
KNN	0.74	0.75	0.74
GNB	0.74	0.50	0.45
SVC	0.10	0.50	0.17
XGB	**0.97**	**0.95**	**0.96**
SGD	0.10	0.50	0.17
GBC	**0.98**	**0.97**	**0.97**
LGB	**0.98**	**0.97**	**0.97**

Table 3. Stage1 results for CreditCard datasets

Model	Precision	Recall	F1-score
RF	**0.99**	**0.94**	**0.96**
LR	0.40	0.50	0.45
DT	0.97	0.93	0.95
KNN	0.67	0.63	0.64
GNB	0.94	0.84	0.88
SVC	0.53	0.52	0.27
XGB	**0.95**	**0.94**	**0.95**
SGD	0.40	0.50	0.45
GBC	**0.96**	**0.94**	**0.95**
LGB	**0.96**	**0.95**	**0.96**

Table 4. Stage2 results for fraud datasets

Precision	Recall	F1-Score
0.98	0.97	0.97

Table 5. Stage2 results for CreditCard datasets

Precision	Recall	F1-score
0.96	0.94	0.95

The results of our ensemble model show the improvement of the prediction accuracy and models performance, as shown in Fig. 8, 9, 10 and 11.

Our method improves the number of false negatives (fraud transactions), which is very important to reduce cost and detect more fraud instances. Therefore, our framework's results emphasize the effectiveness and efficiency of our fraud detection system compared to other ML algorithms used individually due to their errors, overfitting, bias, prediction accuracy, and even their robustness level.

5 Conclusion

The critical impact of the increment of using credit cards is the occurrence of fraudulent transactions, which allow the illegal user to get money and free goods via unauthorized usage. Credit card fraud (CCF) has become the main issue for financial institutions, the credit card industry, the community, and cardholders. Thus, governments, businesses and companies, and financial institutions pay more attention to this security issue and apply different security detection systems to detect and suspend fraudulent transactions, such as Artificial Intelligence (AI) and Machine Learning (ML). Our paper aims to propose our credit card fraud detection (Man-Ensemble CCFD) system based on using an ensemble-learning model with two prediction stages. Stage one (ML-CCFD) utilizes ten machine learning (ML) algorithms to classify credit card transactions to class 1 as fraudulent or class 0 as a legitimate transaction. As a result, their classification reports were compared together, precisely precision, recall (sensitivity), and f1-score, the most accurate models will proceed to the second stage. These ML algorithms were selected based on their performance and prediction accuracy. The second stage is an Ensemble-learning CCFD that assembles the most effective ML algorithms chosen from stage one to get the final prediction instead of using common types of ensemble learning, such as voting, stacking, boosting, and others. The results of our framework showed the effectiveness and efficiency of our fraud detection system compared to using ML algorithms individually due to their weakness issues, such as errors, overfitting, bias, prediction accuracy, and even their robustness level. Indeed, the results of our ensemble method applied on two different datasets show the improvement of the prediction accuracy and classification performance. Also, it provides the minimum number of false negatives (fraud transactions) compared to single ML learners. It is essential to reduce errors and costs and to detect more fraud instances.

6 Future Work

Our method proved its accuracy and effectiveness in detecting fraudulent transactions on transformed real-world and synthetic credit card transaction datasets. Thus, in the future, it is recommended to apply it with neural network and deep learning techniques to check their accuracy and efficiency, also to use it with real-world credit card transactions datasets and in a real-time detection system. As known, most of the research studies and projects were offline fraud detection systems, on transformed data, or private datasets for confidentiality issues.

Acknowledgment. This work was supported by NSF under grant agreement DMS-2022448, and the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant Agreement CCF-0939370.

References

1. Abdallah, A., Maarof, M.A., Zainal, A.: Fraud detection system: a survey. J. Netw. Comput. Appl. **68**, 90–113 (2016)
2. Arya, M., Sastry, G.H.: DEAL-'deep ensemble algorithm' framework for credit card fraud detection in real-time data stream with google TensorFlow. Smart Sci. **8**(2), 71–83 (2020)
3. Awoyemi, J.O., Adetunmbi, A.O., Oluwadare, S.A.: Credit card fraud detection using machine learning techniques: a comparative analysis. In: 2017 International Conference on Computing Networking and Informatics (ICCNI), pp. 1–9 (2017)
4. Baabdullah, T., Alzahrani, A., Rawat, D.B.: On the comparative study of prediction accuracy for credit card fraud detection with imbalanced classifications. In: 2020 Spring Simulation Conference (SpringSim), pp. 1–12 (2020)
5. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., Bontempi, G.: SCARFF: a scalable framework for streaming credit card fraud detection with spark. Inf. Fusion **41**, 182–194 (2018)
6. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., Bontempi, G.: Combining unsupervised and supervised learning in credit card fraud detection. Inf. Sci. **557**, 317–331 (2021)
7. Dighe, D., Patil, S., Kokate, S.: 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 (2018)
8. Hamori, S., Kawai, M., Kume, T., Murakami, Y., Watanabe, C.: Ensemble learning or deep learning? Application to default risk analysis. J. Risk Financ. Manag. **11**(1), 12 (2018)
9. Jurgovsky, J., et al.: Sequence classification for credit-card fraud detection. Expert Syst. Appl. **100**, 234–245 (2018)
10. Modi, K., Dayma, R.: Review on fraud detection methods in credit card transactions. In: 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1–5 (2017)
11. Motwani, A., Bajaj, G., Mohane, S.: Predictive modelling for credit risk detection using ensemble method. Int. J. Comput. Sci. Eng. **6**(6), 863–867 (2018)
12. Polikar, R.: Ensemble based systems in decision making. IEEE Circ. Syst. Mag. **6**(3), 21–45 (2006)
13. Popat, R.R., Chaudhary, J.: A survey on credit card fraud detection using machine learning. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1120–1125 (2018)
14. Randhawa, K., Loo, C.K., Seera, M., Lim, C.P., Nandi, A.K.: Credit card fraud detection using AdaBoost and majority voting. IEEE Access **6**, 14277–14284 (2018)
15. Rayana, S., Zhong, W., Akoglu, L.: Sequential ensemble learning for outlier detection: a bias-variance perspective. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1167–1172 (2016)
16. Sohony, I., Pratap, R., Nambiar, U.: Ensemble learning for credit card fraud detection. In: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pp. 289–294 (2018)

17. Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., Singh, A.K.: Credit card fraud detection using machine learning: a study. arXiv preprint arXiv:2108.10005 (2021)
18. Veeramachaneni, K., Arnaldo, I., Korrapati, V., Bassias, C., Li, K.: AI^2: training a big data machine to defend. In: 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), pp. 49–54 (2016)
19. Young, S., Abdou, T., Bener, A.: Deep super learner: a deep ensemble for classification problems. In: Bagheri, E., Cheung, J.C.K. (eds.) Canadian AI 2018. LNCS (LNAI), vol. 10832, pp. 84–95. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89656-4_7
20. Zareapoor, M., Shamsolmoali, P., et al.: Application of credit card fraud detection: based on bagging ensemble classifier. Procedia Comput. Sci. **48**(2015), 679–685 (2015)
21. Zhang, X., Han, Y., Wei, X., Wang, Q.: HOBA: a novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Inf. Sci. **557**, 302–316 (2021)
22. Zimek, A., Campello, R.J.G.B., Sander, J.: Ensembles for unsupervised outlier detection: challenges and research questions a position paper. ACM SIGKDD Explor. Newsl. **15**(1), 11–22 (2014)

Author Queries

Chapter 14

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	