Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ENSEMBLE MACHINE LEARNING MODEL FOR CLASSIFICATION OF CREDIT CARD FRADULENT TRANSACTIONS

DR TINA ELIZABETH MATHEW
Assistant Professor in Computer Science
Government College Kariavattom
Thiruvananthapuram, Kerala, India
Email: tinamathew04@gmail.com

ABSTRACT

Digital payment systems such as bank credit cards, debit cards, wallets and more, allow users to make payments anywhere anytime without much hassle and at their convenience. On the other hand, digital payments like Credit card transactions, are vulnerable to many security issues including banking frauds. A credit card user will always prefer a highly reliable system that can detect and prevent banking frauds. Hence techniques that provide better security to these elements or during transactions, identification of genuine and fraudulent transactions are crucial. Machine learning is a promising field of study that can help deal with such critical problems of detection and classification of fraudulent transactions. In this study, the suitability of various machine learning classifiers is investigated in the detection of credit card frauds and, an ensemble machine learning framework which constitutes of a majority voting system implemented on selected classifiers is developed. The performance of the model with feature selection – Pearson Correlation Coefficient and without feature selection is also analyzed. To address the problem of heavy imbalance in the dataset, two class balancing techniques such as Random Under Sampling and Synthetic Minority Oversampling techniques are also applied. The results demonstrate the appropriateness of applying Machine Learning techniques in credit card fraud detection and classification.

Keywords: *Classification, Credit Card Fraud, Ensemble Learning, Machine Learning (ML), Synthetic Minority Over Sampling Technique (SMOTE), Random Under Sampling (RUS).*

1. INTRODUCTION

Financial transactions attribute to an integral and crucial part of our day to day lives. As a matter of fact, in today’s digital world financial transactions have gone electronic or online, and the countries of the world are already moving towards a cashless society. The convergence of three industrial giants’ telecommunication, banking and retail sectors have instigated the rapid growth as well as widespread use of digital platforms one being, ecommerce, which has led to a surge in online transactions. As a consequence of increased digital transactions, this has led to escalation in the number of fraudulent and illegitimate transactions(1). The ease in use and global accessibility of Internet and comprehensive increase in ecommerce as well as other digital platforms in recent years, brings about the need for increased safety and secureness in financial transactions. Thus, security poses a major concern for all digital transactions. Fraudulent activities can severely damage the financial sector(2) Among the various digital payment systems prevalent, card payment systems are the most generally accepted, convenient and widely employed means of payment. The swift expansion of communication technology has facilitated significant expansion of digital payment systems, especially in the bank card systems. Over the years, governments have pitched in steps that curtail typical banking and ATM transactions, leading to significant increase in transactions involving bank cards. The pandemic and post pandemic scenario too have added and aided in a significant amount of card usage. The surveys reported (3) are proof of the burgeoning digital transactions. The year 2022 alone shows that, over 4.5 billion individuals around the world rely on online platform purchases, with sales expenses totaling around 5.54 trillion US dollars. The worldwide E-commerce sales report shows an almost 5-fold increase in online payments from 2014 to 2022. The most predominant digital payment modes identified are Card payments, Mobile payments, Unified Payment Interface (UPI), Internet banking, Prepaid cards, Digital Wallets, and Micro ATM. This study focuses on the classification of fraudulent transactions made with credit cards from regular valid transactions.

Digital transactions have been advocated on various platforms such as e-commerce platforms because it makes handling and dealing with cash easier to both the users and the retailers. However, the tremendous volume of internet money transfers has led to increased attempts of fraud. The global financial industry is experiencing a significant threat from credit card fraud, making it a challenging subject for all businesses. As the technology involved in digital transactions are upgraded the fraudsters too evolve with newer techniques making it a challenging task.(4) Frauds can occur in many ways from Application Frauds, loss of cards, usage of imprints, ROC pumping, Magstrips to Phishing. Money Fraud is an illegal activity in which a person or a swindler directly or indirectly uses the money of the victim for and by fake transactions with information from the victim's credit card. As the usage of credit/debit card or net banking is experiencing significant growth, the possibility of numerous fraudulent activities is also increasing. These may include sharing of card details, personal details, and One Time Password (OTP) to anonymous and most probably fake calls (5). Credit card frauds are of mainly two categories - internal fraud and external fraud, which are based on the instigator of frauds, either a firsthand party or third party. Developing countries like India are a major marketplace for e-business based on the population and buying trends of customers.. It is predicted that by 2025, the country's digital transactions will grow by 71%. Thus, usage of digital payments such as credit cards is on a rise in the country This has led to a substantial increase in exploitation, misuse, and frauds of digital payments including credit cards. Lack of efficient Financial fraud detection systems has added to the woes in this field(6). Many countries like UK are expected to become cashless societies by 2026. Hence fraud detection and monitoring systems are the need of the hour.

A credit card is a small rectangular plastic card that is issued to a client or an account holder which permits them to purchase products and services within their available credit limit or enables them to withdraw cash in advance within the preapproved limit. Credit cards give users the benefit of time, allowing them to settle their debts in a defined amount of time by deferring payment to the next billing cycle. Due to worldwide acceptance and the ability to earn reward points for purchases, improvement of the credit score of the customers, providing hassle free shopping facility, and eliminating the need to carry cash around, credit cards are the most commonly used mode of payment in both online and offline e-commerce platforms. Credit card payments are projected to have tremendous growth in the near future. Payment with a credit card can be made by swiping it on a Point-of-Sale device (PoS) or Micro ATM or entering card information on the merchant's website. Online purchases necessitate providing extremely private information to merchant sites such as credit card numbers, cardholder names, Card Verification Value (CVV), and expiry date of card, and without appropriate secure systems the data can be misused at numerous levels.

Credit card payment mode is the most adaptive and convenient modes of payment to the customer. Even though credit card transactions are secure and quick way of payment, it has vulnerabilities that makes it susceptible to fraud activities. A major concern is online credit card frauds. This is a hot topic in the research domain and various techniques are utilized for fraud detection. Credit card fraud involves illegal use of information for credit card transactions(7). Detecting of fraud transactions in an efficient and accurate manner is challenging. Hence, it has high potential in research. Banks have categorized unauthorized activities involving credit cards into different types. Physical theft of the card or pickpocketing where the physical card is compromised. A second method is through skimming card information. Fraudsters steal card details either physically or through cracking details using fraud generation softwares, sniffers, by cloning sites or spyware that prey on users and collect genuine data. Fraud and scam activities may be perpetrated through the misuse of extremely private data shared by account holders during the time of credit card transactions. Owing to the huge volume of transactions happening every day, it is challenging to keep track of them all and spot the fraudulent ones. A third method is by Phishing or other scams where card information is solicited from the customers by posing as authorized agents where they convince customers to provide card information on the pretext of making transactions. A most serious method is by cyberattacking or carding where the secure systems are hacked and compromised, and data is leaked. Another method is through dumpster diving where customers discard card bills and information which can be retrieved by anyone and be misused. SIM swap is another technique used by cyber criminals where they pretend to be a credit card holder and request a duplicate sim card from the mobile operator. Hence, cybercriminals will stoop to any level to do fraudulent activities. Due to the increasing fraud activities in the credit card payment industry, Fraud Detection System (FDS) integrated with banking systems, electronic payment systems, fraud investigators and other telecommunication applications which help thwart fraudulent transactions from taking place are the need of the hour. Machine Learning techniques are seen to be effective for various kinds of classification problems. They have been found to be effective in detecting fraudulent transactions (8). The goal of this study is to develop a robust fraud detection system which can block the vulnerabilities of credit card transactions and spot a fraud effectively. It employs ensemble learning with majority voting to analyze the strengths and weaknesses of a set of 10 base classifiers. Owing to the extremely imbalanced nature of the credit card dataset, under sampling using Random Under Sampling (RUS) and over sampling using Synthetic Minority Oversampling Technique (SMOTE) are used to balance the dataset. The performances of classifiers were measured and analyzed. Six preeminent performing classifiers are selected to produce output for an ensemble learner. Majority voting system is used to generate final outputs. The comparative study between the results of the proposed method and other existing methods shows the efficacy of the proposed model. To improvise the results, the Ensembling of the base classifiers is done. Ensemble learning techniques generate a set of classifiers, which are then used to predict outcomes by applying majority voting or bagging of the predictions.

The remaining section of the paper is organised as follows. Section 2 discusses the background of credit card fraud detection. The concepts related to the proposed methodology are discussed in section 3. Section 4 illustrates the proposed method. Section 5 describes the experimental set up and datasets used. Section 6 is an analysis of the results of the proposed method. The proposed approach is compared with recent methods developed for the same purpose. With section 7, the paper is concluded.

2. BACKGROUND AND RELATED WORK

There are seven major stakeholders involved in credit card transactions. The: Card holder, Merchant, Payment Gate, Payment Processor, Card Payment System, Issuing Bank, and Acquiring Bank. A Cardholder is the customer or owner of the credit card used to make purchases. The customer presents the card either using a device connected with merchant system - Point-of-sale (POS) or via e-commerce site. A Merchant is the business owner who can sell products to the cardholder. The Payment gateway denotes the software that communicates the transaction details from merchant to payment processor. A Payment processor communicates with merchants, merchant banks, card networks and other entities to make card payments. The Credit card network signifies card payment networks deployed with payment processors to facilitate communication between merchant and issuing bank. Few popular credit card network providers are Visa, Master card, American Express, RuPay, and so on. The bank that provides credit card to the customers is designated as the Issuing bank. An Acquiring bank is the Merchant's bank which accepts credit card transactions of a card holder. When the customer presents the card to the merchant for payment through online or offline mode, the merchant sends the transaction details to the payment gateway, which passes the transaction to the payment processor using a secure channel. The payment processor sends the transaction authorization request to card payment network, where the details are verified by communicating with Customer or Issuing bank. Finally, the Issuing bank sends the acknowledgement to the customer. Several recent studies have utilized different machine learning and deep learning methods such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Multilayer Perceptron (MLP), AdaBoost, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Linear Regression, Ensemble learning, besides using deep learning techniques such as Autoencoders(9), Convolutional Neural Networks (CNN), LSTM, Recurrent Neural Networks (RNN),Generative Opponent networks (10) and more for credit card fraud detection. ML techniques on a whole, are seen useful for spam filtering, weather prediction, classification, prediction and diagnosis of diseases as proposed by (11), (12), forecasting problems (13)and many more. In the case of financial transactions ML based classifiers help in classifying fraudulent transactions from normal transactions, identifying abnormal transactions from patterns, differentiating fraudsters and regular customers through credit card profiling and so on. Fraudsters, on the other hand, employ different means to circumvent the detection mechanisms. (14) compared Logistic Regression and Random Forest for credit card fraud detection and their study illustrated the superior performance of Random Forest over Logistic regression.

On credit card fraud data, Siddhant (15) assessed the performance of LR, KNN, RF, NB, MLP, AdaBoost, quadrant discriminative analysis, pipelining, and ensemble learning. The pipeline method was employed, and the accuracy obtained was 99.99 %. The Ensemble Learning and Pipelining framework performed appreciably better than other classifiers. The imbalance of the dataset was not considered. (7) used ensemble learning with majority voting for fraud detection. Initially, standard models were analysed and later hybrid models with majority voting was applied. NB, RF, MLP, LR, SVM, Adaboost classifier were evaluated as standard models. The majority voting method achieved good accuracy rates in detecting credit card frauds than other standard models. The experiments were conducted on benchmark and real datasets. Bagging, Random Forest, LR, and Voting were utilised by (16), and the results were compared to various effective single classifiers such as KNN, Nave Bayes, SVM, RBF Classifier, MLP, and Decision Tree. The work was done in WEKA with 10-fold cross validation, and the Ensemble technique with Random Forest had an accuracy of 94.95 percent, Bagging with Random Forest had an accuracy of 94.78 percent, Voting had an accuracy of 93.05 percent, and SVM had an accuracy of 93.9 percent. (17) proposed an Ensemble Classification Method for Credit Card Fraud Detection in which Adaboost & Voting, KNN and Naive Bayes classifiers were used. They implemented a hybrid approach based on KNN and Naive Bayes. They identified the shortcomings of voting-based methods and stated them as more complex and time consuming. The imbalance of the dataset was not considered. (18) suggested a Credit card fraud detection using AdaBoost and majority voting. Initially, standard models RF, SVM, LOR, were used, RF was found to perform better. Then they used hybrid models - MLP, SVM, LOR, DT and Harmony Search (HS). HS was found useful for finding the best parameters for the classification models. Self-Organizing Map (SOM) was used as the visualization technique. AdaBoost and Majority voting were applied for standard and hybrid models. Feature selection techniques were seen effective in the classification. (19)] developed an Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier. They proposed a Bagging ensemble classifier based on decision tree algorithm which was a novel technique in credit card fraud detection systems. It worked well with imbalanced dataset by splitting the dataset into four groups. The process was seen to take much less time. The concept of Ensembling classifiers was seen effective in classification. In their study (20) used a GA based feature selection method with the classifiers LR, DT, RF, ANN, and NB on the European credit card dataset. They did not use any balancing techniques for the dataset. The test accuracy of ANN and RF classifiers was 99.94%

however performance in terms of recall and F-score was poor. (21) in their study used SMOTE along with machine learners RF, NB, LR and MLP on the European dataset. SMOTE was seen to improve the results. In their work (22) proposed that hybrid methods were suitable for the European dataset than the individual classifiers. They identified Light Gradient Boost Method and Adaboost as the best combination based on performance. (16) used an optimized Light gradient boost method to analyze credit card frauds and obtained an accuracy 0f 98.4%. (23) illustrated that Random Forest classifier was advantageous in predicting fraudulent credit card transactions. The MCC and accuracy obtained was best when compared with other classifiers such as LR and Adaboost.

A major concern raised in all the studies is the highly skewed nature of the datasets available besides the privacy concerns involved and sensitive nature of financial data. Other issues detected were concept drift where the habits, strategies, techniques and behavior of customers as well as fraudsters evolve over time and verification latency issues where only a small percentage of the transactions made are verified by investigators (6). The imbalance issue regarding the dataset is to be considered without which the performance cannot be accounted. To tackle the imbalance issue techniques are to be utilized. Studies suggest that hybrid and ensemble models can help address these issues (24). Each individual classifier has its own pros and cons and an ensemble will enable to overcome issues of individual classifiers. (25) suggested Majority Voting as a good option to deal with credit card fraud detection. Besides accuracy other performance metrics like F1- score. MCC are to be considered for evaluation.

3. ABOUT THE DATASET

The dataset for this investigation was obtained from Kaggle. The files include 284,807 credit card transactions done by European cardholders in September 2013. There are 492 fraudulent transactions and 284315 legal transactions in this database. The dataset was downloaded from the Kaggle website. (Itoo & Singh, 2021). The dataset is significantly imbalanced, with frauds accounting for 0.172 percent of all transactions and real transactions accounting for 99.827 percent. Owing to confidentiality concerns the original features and additional background information about the data are not shared. The data solely contains numerical values transformed using Principal Component Analysis (PCA). The features V1, V2, …, V28 have been transformed using PCA; the features 'Amount' and 'Time' have not been transformed using PCA. The seconds elapsed between each transaction and the first transaction in the dataset is stored in the feature 'Time.' The transaction Amount is represented by the feature 'Amount,' which can be employed as dependent cost-sensitive learning(26).

4. METHODOLOGIES USED

The objective of this paper is to examine the evaluation performance of several advanced data mining and machine learning techniques on credit card data and propose a suitable model for credit card fraud classification. According the literature survey done various models are seen to be imperative in credit card classification(27) .The most suitable machine learning models are selected for base learners of credit card fraud detections from literature study These classifiers are applied individually and the performance of each is evaluated using two sampling techniques RUS, SMOTE. Ensemble techniques are also seen to help in improving performance. Hence Bagging and Boosting ensembles are used. The sections below discuss some of the relevant classifiers used in this study. Each classifier has its own pros and cons (28)

4.1.1. K- Nearest Neighbor

The K nearest neighbor method is popularly used non-parametric supervised learning technique in many classification tasks such as disease classification(29) It is an instance-based method which classifies objects based on the closest feature set in the given data (30). Studies imply that this supervised learning algorithm is one of the best classifier algorithms in credit card fraud detection. KNN achieves consistently high performance. The K-NN algorithm compares the new data with existing predictions and places the new data in the most relevant output group.

4.1.2. Logistic Regression

Logistic regression is a statistical technique applied for classification and regression. Logistic regression models have illustrated good performance in classification problems(31), (32). To forecast the likelihood of counterfeit credit cards, a logistic regression approach is applied. Maximum likelihood estimate is used in logistic regression analysis to determine group membership (33). However, in order to evaluate the results of the group membership prediction with precision and accuracy, a preliminary analysis of the cleaned dataset is performed to see if the logistic regression assumptions were met.

4.1.3. Naive Bayes

Two assumptions underpin the Naive Bayes classifier. To begin, all features in an entry must be categorized so that they all contribute equally to the decision-making process, secondly, all characteristics should be statistically unrelated (8). The Bayes rule is used to classify an instance by applying it to each of the classes it belongs to. The model classifies the 2 categories of transactions fraudulent and valid transactions based on this rule. Naive Bayes models have seen to give good performance in classification tasks(34). (35) illustrated that Naïve Bayes was good for removing noise in credit card dataset.

4.1.4. Support Vector Machine

The SVM algorithm is a supervised machine learning technique used in credit card fraud detection. The essential concept behind the SVM classification algorithm is to create a hyper plane as the decision plane, with the largest distance between the positive and negative modes [16]. SVM uses a kernel function to transfer the data to a specified very high-dimensional space and identifies the hyper plane that optimizes the margin between the two classes. The usage of appropriate kernel functions in SVM enhances classification process(36). SVM usually illustrates good performance for large datasets and hence it is considered suitable for dealing with the credit card fraud detection problem (17)

4.1.5. Multi-Layer Perceptron

A feed forward artificial neural network called a multilayer perceptron (MLP) generates a set of outputs from a collection of inputs (15). Several layers of input nodes are interconnected as a directed graph between the input and output layers of an MLP. Back propagation is used by MLP to train the network (18) illustrated that MLP was suitable for detecting fraud transactions better than other classifiers.

4.1.6. Decision Trees

It is a supervised learning technique applied in classification, regression and prediction jobs. It is usually represented as a treelike structure in which the features of the dataset are represented as internal nodes and decision rules as branches and the outcome as leaf nodes. Instances are classified using decision trees by sorting them down the tree from the root to a leaf node, which provides the classification (33).

The instance is classified by starting at the root node of the tree, checking the attribute specified by this node, and then progressing along the tree branch according to the attribute value. This procedure is then performed for the new node's sub tree. Various categories of decision trees have been seen to be appropriate for classification problems(37), (12). Ensemble of trees are also seen effective for classification than individual classifiers (34)

4.1.7. Random Forest:

Random Forest classifier is considered as an ensemble of decision trees with bagging. This classifier has been seen to aid in classification tasks such as disease classification(11). Compared to individual decision trees it is a much sought after method owing to the accuracy of its predictions. It is based on a random selection of features; Random Forest models decide where to partition the data. Random Forest models incorporate differentiation because each tree splits based on different features, rather than splitting at comparable features at each node throughout the model. Because of the higher level of differentiation, there is a larger ensemble to aggregate over, resulting in a more accurate predictor. Random Forest Classifier takes much more time to execute when processing real-time credit card transactions data (30). It is a classifier that can handle large datasets, albeit unbalanced ones.

4.1.8. AdaBoost

AdaBoost is a machine learning algorithm that can be used to improve the performance of any other machine learning technique(27). It works well with weak learners and boosting. On a classification task, these are models that reach accuracy just above random chance. Decision trees with one level are the most suitable and hence most commonly used algorithm with AdaBoost. These trees are known as decision stumps because they are so short and only have one classification decision.

4.1.9. Gradient Boosting

The Gradient Boosting algorithm combines a number of weak learners into a single strong learner. Individual decision trees are the poor learners in this case. All of the trees are connected in a succession, with each tree attempting to reduce the inaccuracy of the one before it. Boosting algorithms are typically slow to learn but extremely precise due to this sequential relationship.

4.1.10. Extreme Gradient Boosting

XGBoost is a scalable and extremely accurate version of gradient boosting that extends the boundaries of computing power for boosted tree algorithms. It was designed primarily to increase machine learning model performance and computational speed. Unlike GBDT, XGBoost builds trees in parallel rather than sequential. It employs a level-wise technique, scanning over gradient values and evaluating the quality of splits at each feasible split in the training set using partial sums.

.4.2 Ensemble Learning

The Ensemble learner combines several different machine learning classifiers for classification. Hard voting classifies data based on most frequent (mode) predictions made by various classifiers. We can compute a weighted majority vote by connecting a weight w_j with the classifier C_j:

$$\hat{y} = \arg\max_i \sum_{j=1}^{m} w_j \chi_A\big(C_j(\mathbf{x}) = i\big),$$

Where χA is the characteristic function [C_j(x)=i∈A][C_j(x)=i∈A], and A is the set of unique class labels.

Soft voting classifies data based on the probabilities of all the predictions made by different classifiers. It predicts class labels based on expected probability p for classifiers.

$$\hat{y} = \arg\max_i \sum_{j=1}^{m} w_j p_{ij},$$

where w_j is the weight that can be assigned to the jth classifier.(13) from his studies on single classifiers concluded that single classifiers are weak in classifying data sets which are unbalanced and have overlapping classes. (27), (38) proposed that ensemble methods could be suitable methods for credit card fraud classification. In this study boosting and bagging ensemble models are used.

4.3 Feature selection

Feature selection is the process of choosing a subset of features from a larger dataset for accurate predictions (39). It removes irrelevant and repetitive features thereby reducing the computational cost and improving storage space (25)In the experiment, the highly relevant features were selected using WEKA. Among 30 distinctive features, the Positively co-related features to the target variable - V11, V4, V2, V21, V19, V20, V8, V27, V28, Amount, V26, V25, V22 were selected using the Pearson Correlation coefficient feature selection method. Several feature selection techniques have been proposed in literature and have illustrated performance enhancement of classification(20). A correlation based feature selection method helps in identifying relation between features(22).

4.4 Balancing of Data

Data balancing can be applied using various techniques. A study by (40) indicated that resampling techniques helped in better performance of models in the case of credit card fraud detection and classification. Two popular techniques in resampling are Random Under Sampling and Synthetic Minority Over-Sampling.

4.4.1 Random Under sampling (RUS)-

It utilizes a probability sampling method which randomly selects the samples of the majority class equally as that of the minority class and removes samples from the majority class from the training dataset .(41) in their work illustrated that RUS was a simple way to deal with imbalance of data besides being effective.

4.4.2 Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE is a widely used oversampling technique which creates synthetic data by selecting a minority class and constructing a new data point using K-Nearest Neighbor (KNN). SMOTE is pretty useful for dealing with unbalanced datasets (42), According to (41) SMOTE is a comparatively complex method yet it yields goods results.

4.5 Proposed Work

The proposed model works in two phases, Phase I and Phase II. In Phase I, the performance of ten simple base classifiers are assessed and six best models needed for Ensembling in Phase II are selected based on their performance in phase I. (43) in their work established that feature selection and dataset balancing are critical elements in achieving significant results. The models used in Phase I are LR, DT, KNN, NB, SVM, RF, AdaBoost, GB, XGB, MLP. Since, the dataset contains noise, intensive pre-processing is required to clean the data, the dataset is subjected to feature selection in order to eliminate features that are irrelevant or redundant. The performance of the models is evaluated on the actual data, data sampled using Random under sampling and data sampled using SMOTE. The ten classifiers are trained and tested using the data. Accuracy of the ten classifiers is given in table 1. Based on the performances in phase I, six classifiers are selected for phase II.

On the basis of performance in Phase I, six models-DT, KNN, LR, RF, MLP, AdaBoost were selected. The proposed methodology uses ensemble learning with majority voting- both hard voting and soft voting on the selected models The steps involved in the proposed ensemble model are depicted in Figure 1

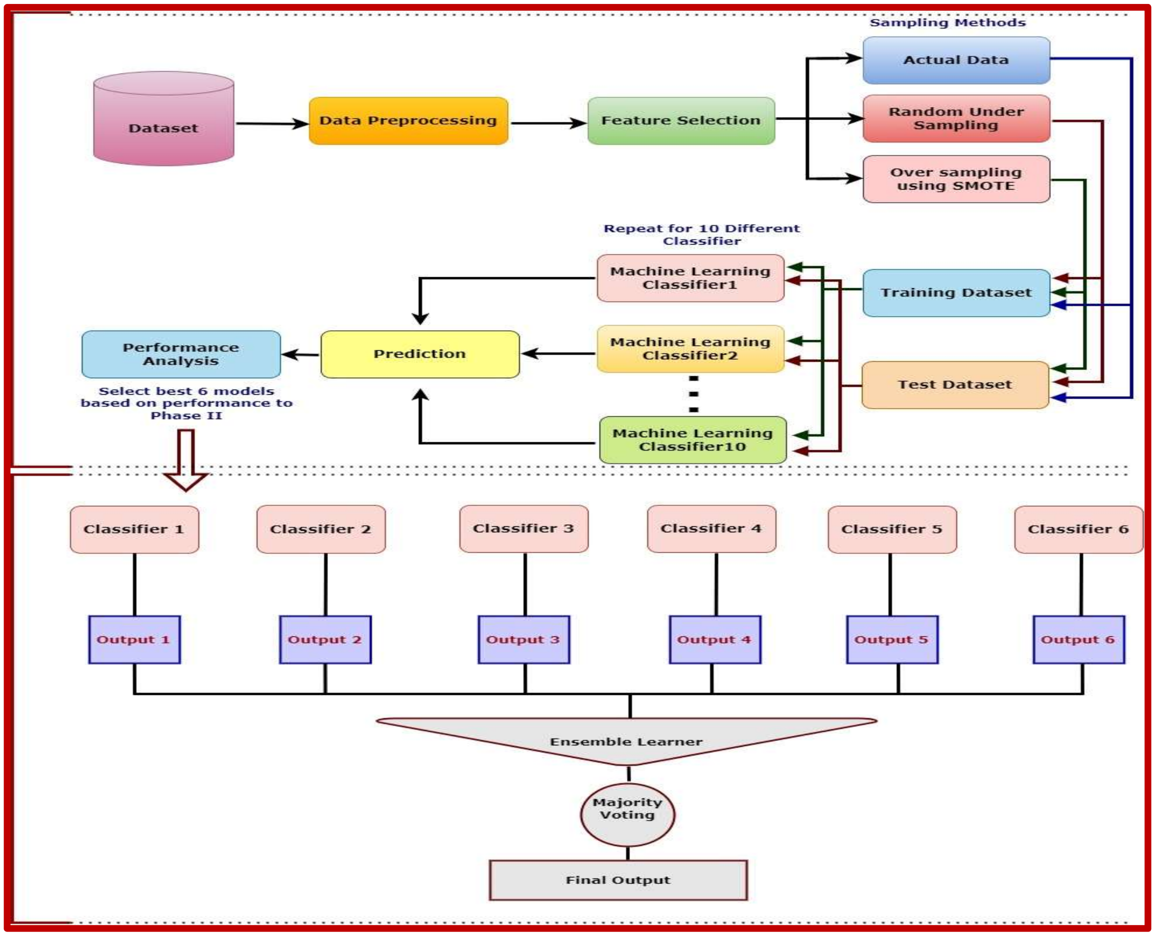

Figure 1 Working of Proposed Model

5. RESULTS & DISCUSSIONS

The different models are evaluated by applying feature selection and without feature selection on the normal data, and resampled data. The proposed model is a binary classification problem and various metrics are used to assess the performance of the binary classification models. The classification accuracy, precision, Recall, F1 score and MCC score are measured for actual data, data sampled using Random under sampling and SMOTE are measured with and without feature selection. The measures are defined as follows.

- Accuracy –Measures the correct predictions made by the model and is the ratio of correctly predicted cases against the Total cases.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$$

- Precision: Precision: Ratio of correctly predicted fraud cases to total fraud cases

$$\text{Precision} = \frac{TP}{TP+F}$$

- Sensitivity/Recall: Ratio between correctly identified fraud cases to total cases

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

- F1- Score: It is the weighted average of precision and recall.

$$\text{F1- Score} = \frac{2*\text{TP}}{2*\text{TP+F} \quad \text{+FN}}$$

- MCC Score represents the classification rate. If the rate is above 60, it is good classification.

MCC is considered to be a good metric for imbalanced classes(1)

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

The performance of the classifiers is illustrated in Table 1. Measures on the original data, under sampled and oversampled data are provided.

Table 1. Performance Metrics of ML classifiers

Machine Learning Models	Feature Selection	Accuracy	Precision	Sensitivity/ Recall	F1- Score	MCC Score	Sampling Method
KNN	Without FS	99.94	92.31	71.11	80.33	80.99	Actual Data
		89.86	95.59	84.42	89.66	80.39	RUS
		99.9	99.8	100	99.9	99.8	SMOTE
	With FS	99.94	91.75	65.93	76.72	77.75	Actual Data
		88.51	94.12	83.12	88.28	77.68	RUS
		99.72	99.45	100	99.72	99.45	SMOTE
Logistic Regression	Without FS	99.91	82.61	56.3	66.96	68.16	Actual Data
		91.22	92.67	90.26	91.45	82.45	RUS
		97.59	98.49	96.66	97.56	95.19	SMOTE
	With FS	99.9	79.55	51.85	62.78	64.18	Actual Data
		86.82	91.97	81.82	86.6	74.22	RUS
		91.27	95.72	86.38	90.81	82.93	SMOTE
Naive Bayes	Without FS	97.87	5.64	79.26	10.54	20.81	Actual Data
		90.2	94.33	86.36	90.17	80.75	RUS
		92.45	97.41	87.21	92.03	85.38	SMOTE
	With FS	97.78	4.49	64.44	8.39	16.61	Actual Data
		83.45	92	74.68	82.44	68.41	RUS
		86.61	95.69	76.63	85.1	74.7	SMOTE
SVM	Without FS	99.93	82.11	74.81	78.29	78.35	Actual Data
		93.58	95.92	91.56	93.69	87.26	RUS
		98.14	99.23	97.03	98.11	96.3	SMOTE
	With FS	99.9	82.72	49.63	62.04	64.03	Actual Data
		86.49	92.54	80.52	86.11	73.75	RUS

		91.58	97.83	85.02	90.98	83.88	SMOTE
MLP	Without FS	99.94	84.43	76.3	80.16	80.23	Actual Data
		90.88	93.2	88.96	91.03	81.85	RUS
		99.01	99.38	98.64	99.01	98.03	SMOTE
	With FS	99.92	78.23	71.85	74.9	74.93	Actual Data
		88.18	96.12	80.52	87.63	77.58	RUS
		93.57	96.1	90.8	93.38	87.27	SMOTE
Decision Tree	Without FS	99.91	71.01	72.59	71.79	71.75	Actual Data
		90.2	90.32	90.91	90.61	80.37	RUS
		99.82	99.75	99.9	99.82	99.64	SMOTE
	With FS	99.89	65.41	64.44	64.93	64.87	Actual Data
		85.47	86.27	85.71	85.99	70.91	RUS
		99.65	99.53	99.77	99.65	99.3	SMOTE
Random Forest	Without FS	99.96	91.45	79.26	84.92	85.12	Actual Data
		91.89	94.52	89.61	92	83.92	RUS
		99.99	99.98	100	99.99	99.98	SMOTE
	With FS	99.94	92.55	64.44	75.98	77.2	Actual Data
		88.85	92.91	85.06	88.81	78.05	RUS
		99.98	99.96	99.99	99.98	99.95	SMOTE
AdaBoost	Without FS	99.91	77	57.04	65.53	66.23	Actual Data
		85.81	91.18	80.52	85.52	72.25	RUS
		89.18	87.4	91.51	89.41	78.44	SMOTE
	With FS	99.91	77	57.04	65.53	66.23	Actual Data
		85.81	91.18	80.52	85.52	72.25	RUS
		89.18	87.4	91.51	89.41	78.44	SMOTE
Gradient Boost	Without FS	99.86	67.44	21.48	32.58	38.01.	Actual Data
		93.24	94.67	92.21	93.42	86.51	RUS
		98.72	99.35	98.07	98.71	97.44	SMOTE
	With FS	99.93	90.22	61.48	73.13	74.44	Actual Data
		88.51	92.86	84.42	88.44	77.42	RUS
		95.89	97.93	93.76	95.8	91.87	SMOTE
XG Boost	Without FS	99.95	92.79	76.3	83.74	84.12	Actual Data

		91.89	92.76	91.56	92.16	83.77	RUS
		99.98	99.97	100	99.98	99.97	SMOTE
	With FS	99.94	95.83	68.15	79.65	80.79	Actual Data
		90.2	93.1	87.66	90.3	80.57	RUS
		99.91	99.85	99.98	99.91	99.82	SMOTE

From Table 1 it can be seen that the individual models have high accuracy scores owing to the high number of genuine transactions. Thus, identifying the best models on accuracy alone is difficult. The model should be able to detect fraud transactions and hence recall score should be high, but at the same time genuine transactions should not be misclassified as fake ones. Therefore, to maintain the balance between the two, F1- score is also compared. F1-score as well as MCC is considered to be an appropriate measure for imbalanced datasets. Random Forest and XGBoost models are seen to perform better. Naïve bayes has an accuracy of 97% but the recall and precision values are very low.

Figure 2, Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 depicts a comparative result analysis of different classifiers when worked with normal data, Random Under Sampling (RUS), and SMOTE respectively with and without Feature Selection. The Measures used are Accuracy, Precision, Recall, F1 score and MCC score.

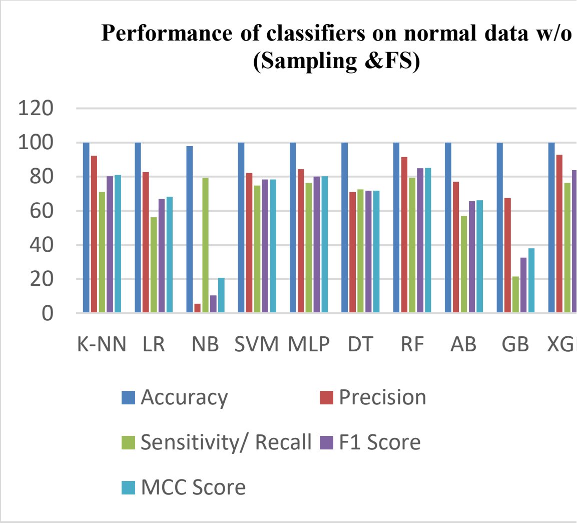

Figure 2 Comparison of ML classifiers using Normal Data only

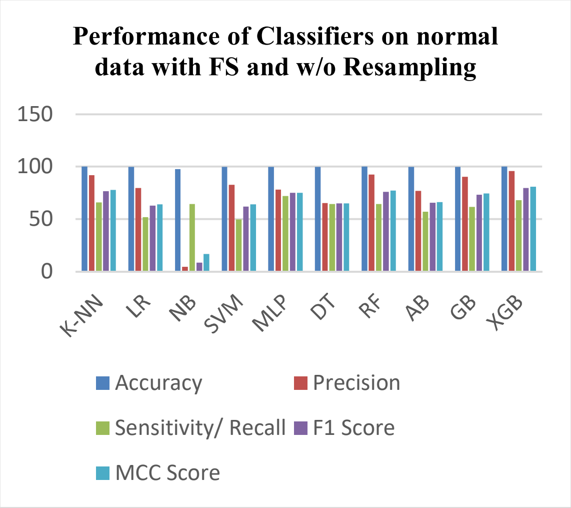

Figure 3 Comparison of ML classifiers using Normal Data and Feature Selection

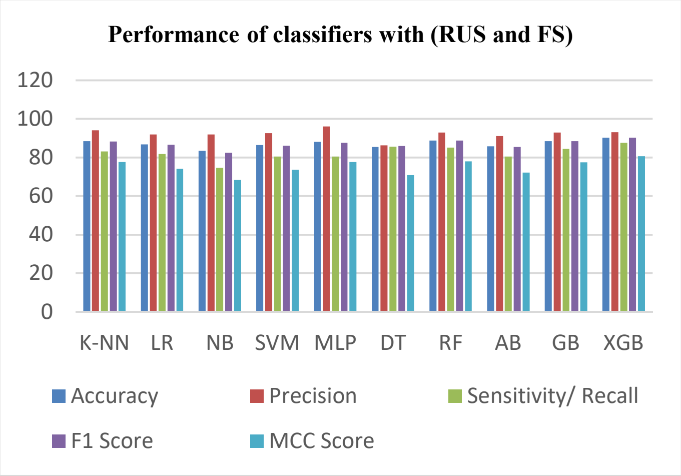

Figure 4 Comparison of ML classifiers using – RUS and Feature Selection

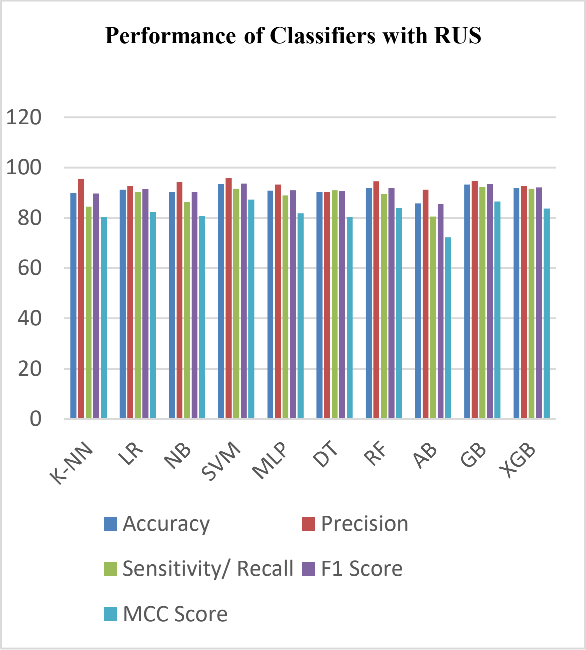

Figure 5 Comparison of Classifiers with RUS

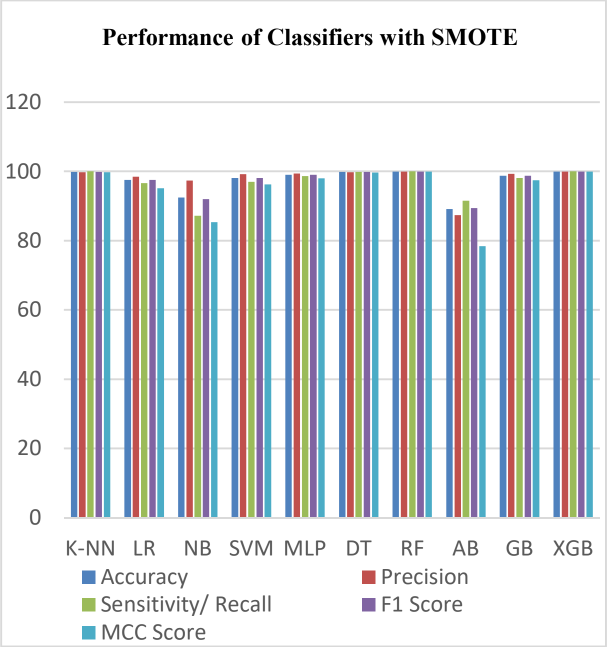

Figure 6 Comparison of ML classifiers using - SMOTE & FS

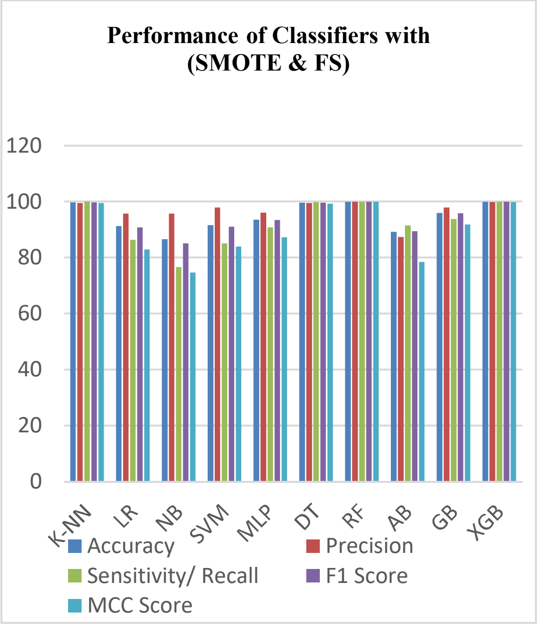

Figure 7 Comparison of ML classifiers with SMOTE

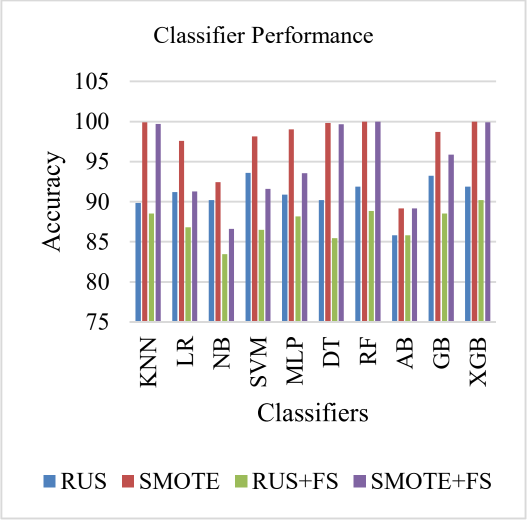

Figure 8 Accuracy Comparison of Classifiers

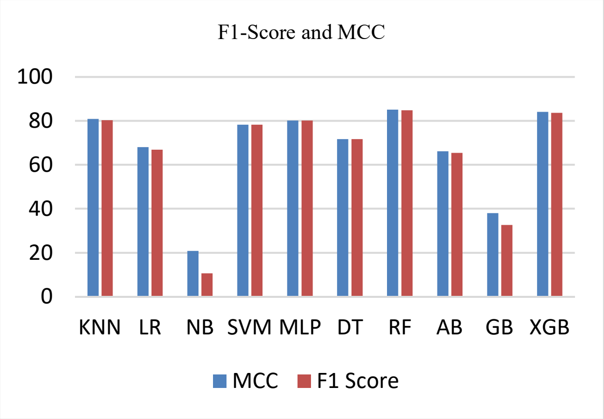

Figure 9 Comparison based on F1-Score and MCC.

Figure 8 and Figure 9 illustrates a comparison of the classifiers based on the performance metrics Accuracy, Matthews Correlation coefficient and F1-Score. From the experiments, the Random Forest classifier performed with best accuracy and F1-Score in both with and without feature selection models. SVM also performs well with Random under Sampling (RUS) but it takes very large execution time, particularly in SMOTE models. Boosting techniques (XG, GB, Adaboost) illustrated good performance. Among the boosting algorithms Adaboost is selected owing to the fact that it is less prone to overfitting and the timeliness of the results are not so acutely critical. MLP is seen to work well with large datasets and hence it was selected for the ensemble. Logistic regression yields robust and reliable results when a large dataset is used. K-NN is advantageous to be used as a base classifier in an ensemble. Usually, Decision trees are considered as classifiers that need less effort and less preprocessing and this was a basis for selection to the ensemble. Random Forests are good in handling overfitting and are usually considered as good predictors. The 6 classifiers are chosen for the ensemble are DT, RF, K-NN, LR, Adaboosr and MLP..

. *Table 2. Performance of ensemble learning with majority voting*

Ensemble Learning	DT	K N N	LR	RF	M L P	AdaBoos	Hard Voting	Soft Voting	Sampling Method
Accuracy	99.917	99.838	99.913	99.957	99.827	99.907	99.912	99.923	Normal Data
	99.891	99.912	99.900	99.935	99.827	99.906	99.914	99.920	Normal Data with FS
	99.916	99.838	99.913	99.957	99.827	99.907	99.912	99.924	RUS
	99.889	99.912	99.900	99.935	99.827	99.906	99.915	99.920	RUS with FS
	99.917	99.838	99.913	99.955	99.827	99.907	99.912	99.925	SMOTE
	99.920	99.838	99.913	99.957	99.827	99.907	99.912	99.924	SMOTE with FS

Table 2 depicts the results of the ensemble of 6 classifiers based on accuracy. It can be perceived from the results that both hard and soft voting provides stable and similar results without much variation. Soft voting classifiers performed better with normal data without applying feature selection. The highest accuracy generated for normal data is 99.923; and accuracy with Random under Sampling was obtained as 99.924% and with SMOTE it produced 99.925%. The ensemble was chosen based on classifiers showing comparatively moderate performance scores. A bagging classifier- Random Forest, A boosting classifier- Adaboost, and other classifiers of different capabilities- Decision Trees, k-NN, Logistic Regression and Multilayer Perceptron were selected. Feature selection using Pearson Coefficient did not bring any significant improvement in the model. This can be owed to the selection of features using PCA method in the dataset.

A general issue perceived is that the percentage of fraud transactions being very insignificant the accuracy scores seem to project the performance of the majority class. The performance of the minority class is a significant factor and is to be considered. Appropriate feature selection techniques are to be identified for better feature selection that improve performance of the classifiers. Other categories of techniques that help in dealing with imbalance issues of the dataset needs to be tested.

6. CONCLUSION

The study depicted the effectiveness of Machine learning algorithms in classifying credit card fraudulent transactions and regular valid transactions amongst the huge volume of credit card transactions. A comprehensive study of ten powerful machine learning methods was done and a model for classification of fraudulent transactions and valid transactions was developed. The proposed model was an ensemble of six machine learning classifiers. The six classifiers were selected based on performance and constituted as an ensemble which were combined using majority voting – both hard and soft voting. Ensemble learning methods were perceived to outperform individual learning methods. Both hard and soft voting illustrated stable results and soft voting classifiers were seen to perform better with normal data having no feature selection. A future work will to be see the performance provided by various techniques such as federated learning and deep learning in classification of genuine and fraudulent cases.

REFERENCES

[1]. Bin Sulaiman R, Schetinin V, Sant P. Review of Machine Learning Approach on Credit Card Fraud Detection. Hum-Centric Intell Syst. 2022;2(1–2):55–68.

[2]. Makki S, Assaghir Z, Taher Y, Haque R, Hacid MS, Zeineddine H. An experimental study with imbalanced classification approaches for credit card fraud detection. IEEE Access. 2019;7:93010–22.

[3]. Visa credit cards in circulation Statista 2020 [Internet]. Available from: https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

[4]. Itoo F, Singh S. Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. Int J Inf Technol. 2021;13:1503–11.

[5]. Asha RB, KR SK. Credit card fraud detection using artificial neural network. Glob Transit Proc. 2021;2(1):35–41.

[6]. Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G. Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Trans Neural Netw Learn Syst. 2017;29(8):3784–97.

[7]. Randhawa K, Loo CK, Seera M, Lim CP, Nandi AK. Credit card fraud detection using AdaBoost and majority voting. IEEE Access. 2018;6:14277–84.

[8]. Awoyemi JO, Adetunmbi AO, Oluwadare SA. Credit card fraud detection using machine learning techniques: A comparative analysis. In: 2017 international conference on computing networking and informatics (ICCNI). IEEE; 2017. p. 1–9.

[9]. Sanober S, Alam I, Pande S, Arslan F, Rane KP, Singh BK, et al. An enhanced secure deep learning algorithm for fraud detection in wireless communication. Wirel Commun Mob Comput. 2021;2021:1–14.

[10].Singh KD, Singh P, Kang SS. Ensembled-based credit card fraud detection in online transactions. In: AIP Conference Proceedings. AIP Publishing LLC; 2022. p. 050009.

[11].Mathew TE. An Improvised Random Forest Model for Breast Cancer Classification,. NeuroQuantology. 2022 May;20(5):713–22.

[12].MATHEW TE. AN OPTIMIZED EXTREMELY RANDOMIZED TREE MODEL FOR BREAST CANCER CLASSIFICATION. J Theor Appl Inf Technol. 2022;100(16).

[13].Kalid SN, Ng KH, Tong GK, Khor KC. A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. IEEE Access. 2020;8:28210–21.

[14].M. V. Krishna and J. Praveenchandar,. Comparative Analysis of Credit Card Fraud Detection using Logistic regression with Random Forest towards an Increase in Accuracy of Prediction,. In 2022. p. 1097–101.

[15].Bagga S, Goyal A, Gupta N, Goyal A. Credit card fraud detection using pipeling and ensemble learning. Procedia Comput Sci. 2020;173:104–12.

[16].A. A. Taha, S. J. Malebary. “Analysis of Credit Card Fraud Detection Using Fusion Classifiers”.,.

[17].Seeja KR, Zareapoor M. FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining. Zhang W, editor. Sci World J [Internet]. 2014 Sep 11;2014:252797. Available from: https://doi.org/10.1155/2014/252797

[18].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. 2018. p. 289–94.

[19].Saheed YK, Hambali MA, Arowolo MO, Olasupo YA. Application of GA feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. In: 2020 international conference on decision aid sciences and application (DASA). IEEE; 2020. p. 1091–7.

[20].Ileberi E, Sun Y, Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data. 2022;9(1):1–17.

[21].Varmedja D, Karanovic M, Sladojevic S, Arsenovic M, Anderla A. Credit card fraud detection-machine learning methods. In: 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE; 2019. p. 1–5.

[22].22. Malik EF, Khaw KW, Belaton B, Wong WP, Chew X. Credit card fraud detection using a new hybrid machine learning architecture. Mathematics. 2022;10(9):1480.

[23].Jain V, Kavitha H, Kumar SM. Credit Card Fraud Detection Web Application using Streamlit and Machine Learning. In: 2022 IEEE International Conference on Data Science and Information System (ICDSIS). IEEE; 2022. p. 1–5.

[24].P. Tomar, S. Shrivastava and U. Thakar. , “Ensemble Learning based Credit Card Fraud Detection System,.” Conf Inf Commun Technol CICT Kurnool India. 2021;1–5.

[25].Fadaei Noghani F, Moattar M. Ensemble classification and extended feature selection for credit card fraud detection. J AI Data Min. 2017;5(2):235–43.

[26].Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Frederic Oblé, Gianluca Bontemp. i Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection. Inf Sci 2019.

[27].Suryanarayana SV, Balaji GN, Rao GV. Machine learning approaches for credit card fraud detection. Int J Eng Technol. 2018;7(2):917–20.

[28].Tiwari P, Mehta S, Sakhuja N, Kumar J, Singh AK. Credit card fraud detection using machine learning: a study. ArXiv Prepr ArXiv210810005. 2021;

[29].Mathew, T. E., & Kumar, K.S A. A Modified-Weighted-K-Nearest Neighbour and Cuckoo Search Hybrid Model for Breast Cancer Classification. Indian J Comput Sci Eng IJCSE. 2021 Jan;12(1):166–77.

[30].Lim KS, Lee LH, Sim YW. A review of machine learning algorithms for fraud detection in credit card transaction. Int J Comput Sci Netw Secur. 2021;21(9):31–40.

[31].Mathew T. A logistic regression with recursive feature elimination model for breast cancer diagnosis. Int J Emerg Technol. 2019;10(3):55–63.

[32].Mathew TE, Kumar KA. A Logistic Regression Based Hybrid Model For Breast Cancer Classification. Indian J Comput Sci Eng. 2020;11(6):899–906.

[33].Dhankhad S, Mohammed E, Far B. Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In: 2018 IEEE international conference on information reuse and integration (IRI). IEEE; 2018. p. 122–5.

[34].Mathew TE. Simple and ensemble decision tree classifier based detection of breast cancer. Int J Sci Technol Res. 2019;8(11):1628–37.

[35].Akila S, Reddy US. Credit card fraud detection using non-overlapped risk based bagging ensemble (NRBE). In: 2017 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE; 2017. p. 1–4.

[36].Mathew TE. A comparative study of the performance of different Support Vector machine Kernels in Breast Cancer Diagnosis. Int J Inf Comput Sci. 2019;6(6):432–41.

[37].Mathew TE. Appositeness of hoeffding tree models in breast cancer... - Google Scholar. J Curr Sci. 2022 Sep;12(3):291–407.

[38].AL-FAQIR S, OUDA O. CREDIT CARD FRAUDS SCORING MODEL BASED ON DEEP LEARNING ENSEMBLE. J Theor Appl Inf Technol. 2022;100(14).

[39].Shukla S, Rakesh D. Dynamic ensemble based feature selection model for credit card fraud detection. In: 2020 IEEE 17th India Council International Conference (INDICON). IEEE; 2020. p. 1–6.

[40].Alfaiz NS, Fati SM. Enhanced credit card fraud detection model using machine learning. Electronics. 2022;11(4):662.

[41].Hordri NF, Yuhaniz SS, Azmi NFM, Shamsuddin SM. Handling class imbalance in credit card fraud using resampling methods. Int J Adv Comput Sci Appl. 2018;9(11):390–6.

[42].Kumari P, Mishra SP. Analysis of credit card fraud detection using fusion classifiers. In: Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017. Springer; 2019. p. 111–22.

[43].Marabad S. Credit Card Fraud Detection using Machine Learning. Asian J Converg Technol AJCT ISSN-2350-1146. 2021;7(2):121–7.