Credit Card Fraud Detection Using Adaboost And Majority Voting

Anusree.B[1], B.Ramesh Kumar[2]

M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu India[2]

Corresponding Author: Anusree.B

Abstract: The credit card fraud is mostly come in financial services. The credit card fraud is generated huge number of problems in every year. Lack of research on this credit card problem and submits the real-world credit card fraud analyzes, that is issues. In this paper is introduced best data mining algorithm called "machine learning algorithm", which is used to detect the credit card fraud, so initially use this algorithm and it is one of the standard model. Then, secondly apply the hybrid methods namely, "AdaBoost and majority vote method". Use this model efficacy, which is evaluated, and then use the credit card data set it is publicly available one. The financial institution included the real-world credit card data set, so it is taking and then analyzed. In this robustness algorithm additionally evaluate the noise added data samples. This concept is used in experiment and then produce the result positively indicate the hybrid method, that is majority voting, it provides good accuracy rates in credit card fraud detection.

I. INTRODUCTION

Fraud is a cheating or a wrongful or criminal activity, its main aim is focus financial or personal sign. In this proposed system is uses two mechanism namely, (i) fraud prevention and (ii) fraud detection, for avoiding loss from fraud, that detecting details from fraud. In the first fraud prevention mechanism is a most protective and proactive method, it is stops the fraud from beginning. Then, the second mechanism fraud detection is guessing the fraudster. This mechanism is needed for a fraudulent transaction, but it is guess the fraudster, in the time transaction attempted by fraudster.

Credit card fraud is related with illegal use of credit card information for purchases that is credit card amount are used in product purchases. In the purchasing time the user use the credit card, the fraudster trace out the password or user oriented important details, then it will be applied in our transaction easily use the credit card cash amount but cannot find out that person, that is fraudster. The credit card transaction completed through physically or digitally. The physical transactions based credit card is used in during transaction, but the digital transaction based credit card is used only the telephone or internet. The cardholders are basically provides the important details such as, card number, expiry date, and card verification number via telephone or website. But technological world currently use the credit card so increase the credit card transactions in every day and the rise of e-commerce field like that every second use this credit card. The number of credit card transactions is increased in every year. So the technology is mostly developed and gets more benefit in the people, but another side increases this credit card fraud cases. It is most effective problem in the world. Then, the logical and numerical authentication methods are applied in this credit card fraud cases, but this method is not most detected one, because the fraudsters are hidden their details like identity and location in the internet, so that problem is big impact of financial industry also. This credit card fraud problem affects both sides that mean admin and user side. It affects the (a) issuer fees, (b) charges, (c) administrative charges that is the fees are loss. So the merchants make the decision that is high rate fix in goods or discounts are reduced. In this proposed system is to reduce the loss from credit card fraud, to eliminate the fraud cases. In two machine learning techniques are used in (i) artificial networks, (ii) rule-detection techniques, (iii) decision trees, (iv) logistic regression, and (v) support vector machine (SVM). This above models are combining several methods that is, hybrid methods. The AdaBoost and majority voting methods are applied and to detect the credit card fraud.

Key Terms:
Fraud

The term, "Fraud" means wrongful or criminal activity or cheating or steals, so this aim is to focus on financial or personal sign. The fraud is classified in to several types based on their fields. But this proposed system based this fraud is only one type. This fraud is occurs in the credit card transactions, so this fraud is known as, "Credit card fraud". To point out the fraud is very difficult one. The characteristics of fraud are (i) it does not provide the direction; (ii) It is not discuss any requirements because it is hidden attacked one. The fraud is called in many terms are, "fraudulent activities", "manipulative behavior", "wrongful", "criminal". It is carefully organized the crime that is fraudsters, it don't operate independently, because dependence on fraudsters.

Credit card fraud

It is an unauthorized takes the amount from another's credit. So, this credit card fraud is divided in to sub types are, (i) Application fraud, (ii) behavioral fraud, which is lost or stolen the cards or take credit through mail.   (i) application fraud: it involves individuals that is includes the new credit cards from issuing companies, but it using the fake or false personal information and fill it shortly. (ii) Behavioral fraud: acts as cardholder but not present the real card holder, that is present the fraudster. Stealing the physical card and use it. Most of the credit card fraud is based on behavioral credit card fraud. It is a form of identity theft that steals the card.

Fraud Detection and Fraud Prevention

The term, fraud detection is to recognize or discover the fraudulent activities, that is focus only, fraud are occur or not to be identified. Another term, fraud prevention is to avoid or reduce the fraud that means. These two terms are shared the concept of "Fraud Reduction", that is credit card fraud detection or protection.

Fraud Cycle
(i)   Fraud Detection: Choose and applying the fraud detection techniques and then, assigning the fraud risk, then solve it.
(ii) Fraud Investigation: The human expert or investigator investigates the doubted or suspicious any error is occur, it gives the logical and complexity.
(iii) Fraud Confirmation: It means confirm the occurrence of fraud. That is, it determines the true fraud label.
(iv) Fraud Prevention: To identify before the fraudster exactly, it helps the future fraud detection techniques.

II.   LITERATURE REVIEW

In this paper [1] author has presented the concept namely, "Support Vector Machine (SVM)" for credit card fraud detection and false alarms reduction. In every day in life, that is daily the credit card is used for any amount oriented transactions like, purchasing goods and services in this time the credit card act as virtual card for credit card transactions. In this virtual card is used for online and offline transaction, that is online based transactions, it needs the internet, second one is the credit card act as physical card, this physical card is used for offline transaction.

The combination of virtual card and physical card is called "credit card". In physical-card based transactions, like purchase the product, in this time the physical card used by the customer or cardholder, then the card give to merchant for making a payment, then the merchant return the credit card, then over the credit card transaction. This physical transaction based fraud is come with the reason is attacker steals the credit card in purchasing time. Another way of this fraud is come with fraudster, the fraudster simply know the card details in purchasing time. This fraud is only come with the basic reason is not aware the credit card details, that is casually use the credit card by the genuine cardholder, so easily take card oriented information. So this kind of fraud is to detect to use the concept namely "usual spending patterns". This concept is to analyze the spending patterns in each and every card and to figure out any opposite activity with respect the usual pattern, that is the usual patterns changes are carried out.

The physical card transactions based on (i) time-consuming and (ii) resources-demanding task, so the issuers searching the efficient algorithm, so for this algorithm automatically set the incoming transactions. The data mining is a well-known technique for suitable solutions. The big data problems are involving the risks are (i) credit card risk modeling, (ii) churn prediction, (iii) survival analysis. The fraud detection is performed the prediction task, it which require the tailored approach, it is to address and predict future fraud. It is one of benefit. Sometimes, most of the fraud detection systems produce good results in detecting fraudulent transactions, but this system generates the false alarms, this is drawback of this system. The credit card company needs to minimize the loss, so the restricted features follow the company is available, but the user feel it is

restricted one. In this system is known as a novel credit card fraud detection system. It is based on the integration SVM.

The paper [2] "Fraud prediction for credit card using classification method" has presented by author. In the digital world every day meets new innovations such as, (i) credit cards, (ii) debit cards, (iii) mobile banking, (iv) internet managing, and this all above features included innovations are based on bank account. These features are used to exchange the cash for many purposes like, online purchases, pay the current bill, transfers money and so on. The credit card money is based on day by day which means online exchanges with expansion in online shopping, online charge payment, insurance premium and different charges, so this credit card transaction is give more benefits like, save time, save traveling amount, and many. In this paper take this credit card transaction problem and apply the data mining procedures are valuable. So it is to estimate and then, categorized the client's credit risk score that is, normal or fraud. The existing system to include the clients from online based money transactions that money exchanges by utilizing particular data mining techniques or classification methods. In another method is to break down the fake, is called "Naïve Bayes". This model provides great accuracy, recall more time and find out the precision.

The author kindly presented the paper [3] namely, "conditional weighted transaction aggregation for credit card fraud detection", which reduce the problem of substantial losses for credit card companies and consumers. In this system is to develop the robust and high protection that is to develop the fraud detection techniques that recognize the differences between fraudulent and legitimate transactions. The current protection methods are mainly operates the transaction level or account level is based on bank account. These transaction approaches involve the analysis and aggregation of previous transaction data based information are analyzed and then, to detect the credit card fraud. This approach handles all transactions attributes are same that is equally treated as importance. The conditional weighted transaction aggregation technique describes to identify this issue used the supervised machine learning techniques, so it is to identify the fraudulent transactions. This technique is effective and better than existing system.

This paper [4] author has presented the main concept are "AdaCost: Misclassification cost-sensitive boosting", which is used to the cost of misclassifications updated by training distribution based on successive boosting rounds. This AdaCost is differs from AdaBoost. This concept main goal is to reduce the cumulative misclassification cost, it is more than AdaBoost. AdaCost shows the reduced upper bound of cumulative misclassification cost of the given training set. It moreover significantly reduce the total (cumulative) misclassification cost is over the method of AdaBoost (without consuming the additional computing power).

The author generally presented the data mining concept and it is apply to credit card fraud detection, so this paper [5] based on the concept namely, "Distributed Data Mining", which reduce the credit card problems. The credit cards transactions can be used in many people, that ever-larger share this payment system and increase their credit card rate in every day, but meet the losses by bank. In this improved technique is very important to maintain the payment system. But already banks have before stage; use the fraud warning systems, this system provides simple protection so to move the protection developing level of this current stage like that.

The large-scale data-mining techniques are used and then improve the state-of-art. To analyze the huge amount of transaction data that importantly computes the fraud detectors, that problem are come in e-commerce field, this analyzation is used the "scalable techniques". The fraud detection task is performed this technical and logical problem, it includes the skewed distribution of training data and then, include the nonuniform cost per error, but this is not based on KDD (Knowledge Discovery Data), and this is not includes the (i) scalability (ii) efficiency. In the proposed system is combining multiple learned fraud detectors based on "cost model". It reduces the loss due to fraud via distributed data mining of fraud models. In these systems approach namely, scalable black-box is used for building efficient fraud detectors, so significantly reduce loss due to the illegitimate behavior.

This paper [6] author has presented the SVM for credit card fraud detection. The proposed system is to detect the credit card fraud that means behavior based analyze the fraud using the support vector machine (SVM). In the developing world every day use the credit card so it is an unavoidable one, but in this time increase the frauds are already known. In this approach is adopted for efficient feature extraction method. It analyze and predict the behavior transaction pattern, if suppose this pattern differs from other to find out the fraud, that is doubted pattern is occur in behavior pattern it is predicted. This detection system is affected by the large amount of data, so it is solved by the proposed system. The proposed system is to provide the accuracy, high fraud detection and catch the fraud rate, and then low false alarms are the main tasks of this system like.

In this paper [7] author has presented the data mining classification technique, then it applied in to the skewed data, that means it is an problem and it is fraud detection in finance data, and continuously submit the minority report in fraud detection. In this proposed system is carry the new innovative and extraordinary fraud detection methods and side by refer or catch their existing system problem, based on detection research and minority report. This minority report is related into the data mining problem of skewed data or skewed distributions. This paper refers and uses the (i) backpropagation (BP), (ii) Naïve Bayesian (NB), (iii) C4.5

algorithms. The above techniques are used as data partitions, and derived from the minority oversampling that is to adjust the data set distributions, with replacement. But this methods is not based on (a) Single meta-classifier that is to select the best base classifier (stacking), then it combine the base classifiers' prediction that is to improve or develop the cost savings, these method are called in to bagging and stacking-bagging. This method is demonstrates the automobile insurance fraud detection data set. This method is a common method so it is applied in to the industry, but BP use without both sampling and partitioning. This paper is compare the fraud detection techniques and to choose the best one, that is take the new fraud detection method "meta-learning approach" and then secondly take c4.5 algorithm which includes the trained data using undersampling and then, oversampling and SMOTEing without partitioning that is oversampling approaches. This concept based experimental results shows that the starting fixed decision and the cost matrix and partitioning and then multiple algorithms approaches is to achieve the higher cost savings, this cost savings are take and it is compared in to the entire training data set with different class distributions and carried out the variations. The proposed system result is combination of classifiers to provides the best cost savings these result is get from the above three algorithms. Finally, the proposed system produces correct and efficient result for fraud detection.

In this paper [8] author has presented the paper and improve the detection manner. In this paper introduce the concept of "Meta-Learning Strategy", which is used for improving credit card fraud detection. In this system is to reduce the big issues of the credit card frauds such as, (i) the transactions labeled or pointed out the fraudulent transaction it is fact manner, (ii) To predict the false alarms, (iii) The credit card transaction data from daily transaction and to determine the savings improvements, this concept based identifying the fraudulent transactions it is tested one, in this research use the meta-classifier model, obviously. The meta-classifier model is includes the three base classifiers are (i) k-nearest neighbor, (ii) decision tree, and (iii) naïve Bayesian algorithms. The naïve Bayesian algorithm is used the meta-level algorithm and then, to combine the base classifier predictions, this prediction is to produce the final classifier that is the result. It follows the two main processes are (a) modeling techniques (Neural Network-NN), (b) updated data set.

This paper [9] author has presented data mining concept, "classification technique", which is used for credit card fraud detection. The mining is a term, to predict some that means, but the data mining is a technique, is to find out the hidden predictive information and to find the useful information from raw data or rough data or collected data. The data mining is applied in the collected data and get the result is useful information, how to get this data, which means it eliminates the unwanted and non clear data, so get the wanted and needed information, that's all. The data mining introduce and describe the technique is called "Prediction analysis". The prediction analysis is predicts the new things from the current data or mined data. The proposed system concept, "classification" applied in to this prediction analysis. In this research paper is based on the prediction of credit card fraud detection, so the prediction analysis introduced by the proposed system. The main concept is based on the NN (Neural Network), which the system captured the details about the previous systems, so this proposed system is easily get the new into normal and fraud transactions, and then find out. The SVM is a classifier and it is proposed into the detecting the credit card fraud, it classified into the input data into normal and fraud transactions. In the proposed system is take the input data and it is divided into test and training sets, and it is predicted the precision and recall.

This paper [10] author has presented the decision tree algorithms for class imbalanced learning in credit card fraud detection, which is reduces the financial problems. This proposed system is to apply comparative method in decision tree techniques. This concept is developed for the property of class imbalance level problem by financial institutions, so to detect the credit card fraud. The imbalanced data is takes and it is consider the input of fraud detection, this fraud detection is used by many classifiers. The imbalanced data includes the performance of classifiers by analyze the overall accuracy of performance measures. This is makes the decision and then produce the results in misclassified minority class. Actually, the credit card data as consider the imbalanced data. But most of the classifiers perform poorly so the credit card imbalanced data are affected. The proposed system introduce the term, "Resampling", it is related to the imbalanced data. In the proposed system, aim is to find out the best classifier based on distribution. The credit card data includes the random under sampling (RUS), it includes the feature selection and finally, useful model is measured by the risks of credit card fraud, it is efficiently. In this proposed system is applied in to the RUS and feature selection for the family of classifier that is "Decision tree classifier". In this proposed system finally, provides the result is denotes the improved performance for the decision tree classifiers are already known, so for this system is very efficient to detect the fraud.

This paper [11] author has presented the decision tree induction for financial fraud detection by using the ensemble learning techniques. The credit card fraud problem is mostly affected the banking industries. The rise of web services give the advantages (banking) at the same time raises the banking frauds. The banking systems every second have robust, safe and secured one. It is order to detects and prevents the fraudulent activities of physical and virtual (any kind of) transactions. However, totally and entirely eliminating or avoiding banking fraud that is credit card fraud is impossible one. But the machine learning technique is to

minimize the frauds. In this paper aim is to reduce the banking frauds. Then, this system conducted the experiment for banking fraud detection by using the (i) ensemble tree learning techniques, and (ii) genetic algorithm. These two techniques are to indicate the ensemble of decision trees, so the bank transaction datasets are identified and to prevent the bank fraud. The ensemble of decision tree on the credit card dataset is to provide an evaluation and effectiveness.

The technological based presented this paper [12] "fraud detection in online payments using the Spark ML" by author. This proposed concept is to reduce the frauds in online payments. The standard fraud detection technique system is included the machine learning algorithms, but it requires the big size data that was handled by the distributed computing frameworks. So the proposed system is the fraud detection and addressed the problem by using the fraud detection classifier namely, "Spark ML", is using all payment data. Then, additionally solve the problems are, class imbalance of credit card and non-stationarity. The result of the method is reached by applying undersampling and oversampling. These sampling methods use the training data and to reduce the class imbalance. This model is updates the new information and decreased data that information of non-stationarity is provides. The proposed system is finally provides the problem, give the importance in to online payment provider (Nordics). This system every second sends the purchase orders for review, that information carry out the fraud investigators. Then, before catch the frauds because enable this model feature.

Table 1.0. Comparison table

Paper No	Technique	Advantages	Disadvantages
1	SVM for reduction.	To reduce credit card fraud and to predict future fraud.	It generates the false alarms.
2	Classification method, Naïve Bayes.	It provides great accuracy, recall more time, find out the precision.	It based on client based online transaction.
3	Conditional Weighted Transaction Aggregation	To develop the fraud detection and differences between fraudulent and legitimate transaction.	It only identify the fraudulent transactions.
4	AdaCost	To reduce the cumulative misclassification cost.	But, significantly reduce the misclassification cost.
5	Large-Scale data mining technique	To improve State-to-art	But, not based on KDD.
6	SVM for detection	To predict the behavior pattern, high fraud detection.	Low false alarms.
7	BP, NB, C4.5 algorithm	To detect the fraud in finance data.	Only take the skewed data.
8	Meta-learning	To improve the detection.	To reduce only big issues.
9	NN	Find out fraud transactions.	Only find out the fraud transaction, not reduce the fraud.
10	Class imbalance learning	To reduce the financial problems.	The poor classifier affects the imbalanced data.
11	Decision tree induction	To detect the financial fraud.	To minimize the fraud but not fully reduce the fraud.
12	Spark ML	To reduce the fraud in online payment.	It requires the big size data.

The above table 1.0. Depict the working methodologies of various data mining techniques, which can be used to achieve the Fraud detection and prevention of credit card.

III. CONCLUSION

The Data mining, best concept of machine learning algorithm is used for credit card fraud in this proposed system is proposed. Then, the number of standard models such as NB, SVM, and DL is used for evaluation terms. The credit card data is available in publically, it is used for evaluation that is, use the standard models and hybrid models. The hybrid models such as AdaBoost and majority voting, this models are combination methods, also. The MCC metrics are only calculates the performance measures and it takes the account, and it predicts the true or false outcomes of credit card transaction. The best MCC score majority voting is used the majority voting. The financial institution gives the credit card data set for evaluation. But the

perfect MCC score is get only the use of combination of AdaBoost and Majority voting, because that combination method is shows and give the robustness and strong performance. In this proposed concept is enhanced to online learning models. Use the online learning to enable the rapid detection of credit card fraud. The proposed system is help to detect and before prevent the fraudulent transaction and activities, so to reduce the number of losses in financial industry.

REFERENCES

[1]. Mehak Kamboj, Shankey Gupta. "Credit card Fraud Detection and False Alarms Reduction using Support Vector Machines". International Journal of Advance Research, ideas and innovations in technology, ISSN:2454-132X.

[2]. Er. Monika, Er. Amarpreet Kaur."Fraud Prediction for credit card using classification methid". International Journal of Engineering and Technology, (2018); 7(3) 1083-1086.

[3]. Wee-Yong Lim, Amit Sachan, Vrizlynn Thing. "Conditional Weighted Transaction Aggregation for Credit Card Fraud Detection". HAL ID: hal-01393754.

[4]. Wei Fan, Salvatore J.Stolfo, Junxin Zhang, Philip K.Chan. "AdaCost: Misclassification Cost-sensitive Boosting".

[5]. Philip K.Chan, Wei Fan, Andreas L.Prodromids, Salvatore J.Stolfo. "Distributed Data Mining in credit card fraud detection".

[6]. V.Dheepa, R.Dhanapal. "Behavior Based Credit Card Fraud Detection using Support Vector Machines". ISSN: 2229-6956 (Online).

[7]. Clifton Phua, Damminda Alahakoon, Vincent Lee. "Minority Report in Fraud Detection: Classification of Skewed Data".

[8]. Joseph King-Fung Pun. "Improving Credit Card Fraud Detection using a Meta-Learning Strategy".

[9]. Tamanna Chouhan, Ravi Kant Sahu. "Classification Technique for the credit card fraud detection". International Journal of Latest Trends in Engineering and Technology. e-ISSN: 2278-621X.

[10]. Maira Anis, Mohsin Ali, Amit Yadav. "A Comparative study of decision tree algorithms for class imbalanced learning in credit card fraud detection". International Journal of Economics, Commerce and Management, ISSN 2348 0386.

[11]. Vijayalakshmi Mahanra Rao, Yashwant Prasad Singh. "Decision Tree Induction for Financial fraud detection using ensemble learning techniques".

[12]. Ignacio Amaya De La Pena. "Fraud Detection in online payments using Spark ML".[2017]