

Finding Credit Card Fraud By Using Adaboost Concept And Majority Voting

Mande Dhanalakshmi¹, K.SKiran Kumar²

1PG Scholar, Dept of CSE, VelagaNageswaraRao College Of Engineering,
Ponnur(Post),Ponnur(Md)Guntur(D.T), Andhra Pradesh

2Asst Professor, Dept of CSE, VelagaNageswaraRao College Of Engineering,
Ponnur(Post),Ponnur(Md)Guntur(D.T)A. Andhra Pradesh

ABSTRACT_ *The credit card fraud is for the most part come in money related administrations. The charge card extortion is produced immense number of issues in consistently. Absence of research on this Visa issue and presents this present reality charge card extortion breaks down, that is issues. In this paper is presented best information mining calculation called "AI calculation", which is used to perceive the charge card extortion, so at first utilize this calculation and it is one of the standard model. At that point, also apply the half breed strategies in particular, "AdaBoost and lion's share vote strategy". Utilize this model adequacy, which is assessed, and afterward utilize the charge card informational index it is freely accessible one. The money related organization included genuine world informational collection, so it is taking and dissected. In this vigor calculation also assess the commotion included information tests. This idea is utilized in analysis and afterward produce the outcome emphatically show the mixture strategy, that is dominant part casting a ballot, it gives great exactness rates in Visa misrepresentation location.*

Keywords—*AdaBoost, Majority Voting.*

1.Introduction

Fraud is a wrongful or criminal deception aimed to bring financial or personal gain. In avoiding loss from fraud, two mechanisms can be used: fraud prevention and fraud detection. Fraud prevention is a proactive method, where it stops fraud from happening in the first place. On the other hand, fraud detection is needed when a fraudulent transaction is attempted by a fraudster.

Credit card fraud is concerned with the illegal use of credit card information for purchases. Credit card transactions can be accomplished either physically or digitally.

In physical transactions, the credit card is involved during the transactions. In digital transactions, this can happen over the telephone or the internet. Cardholders typically provide the card number, expiry date, and card verification number through telephone or website. With the rise of e-commerce in the past decade, the use of credit cards has increased dramatically. The number of credit card transactions in 2011 in Malaysia were at about 320 million, and increased in 2015 to about 360 million. Along with the rise of credit card usage, the number of fraud cases have been constantly increased. While numerous authorization techniques have been in place, credit card fraud cases have not hindered effectively. Fraudsters favour the internet as their identity and location are hidden. The rise in credit card fraud has a big impact on the financial industry. The global credit card fraud in 2015 reached to a staggering USD \$21.84 billion.

Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges. Since the merchants need to bear the loss, some goods are priced higher, or discounts and incentives are reduced. Therefore, it is imperative to reduce the loss, and an effective fraud detection system to reduce or eliminate fraud cases is important. There have been various studies on credit card fraud detection. Machine learning and related methods are most commonly used, which include artificial neural networks, rule-induction techniques, decision trees, logistic regression, and support vector machines [1]. These methods are used either standalone or by combining several methods together to form hybrid models.

In this paper, a total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models. They are evaluated using both benchmark and realworld credit card data sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper are extracted from actual credit card transaction information over three months.

2.Literature Survey

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 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.

3.System Analysis

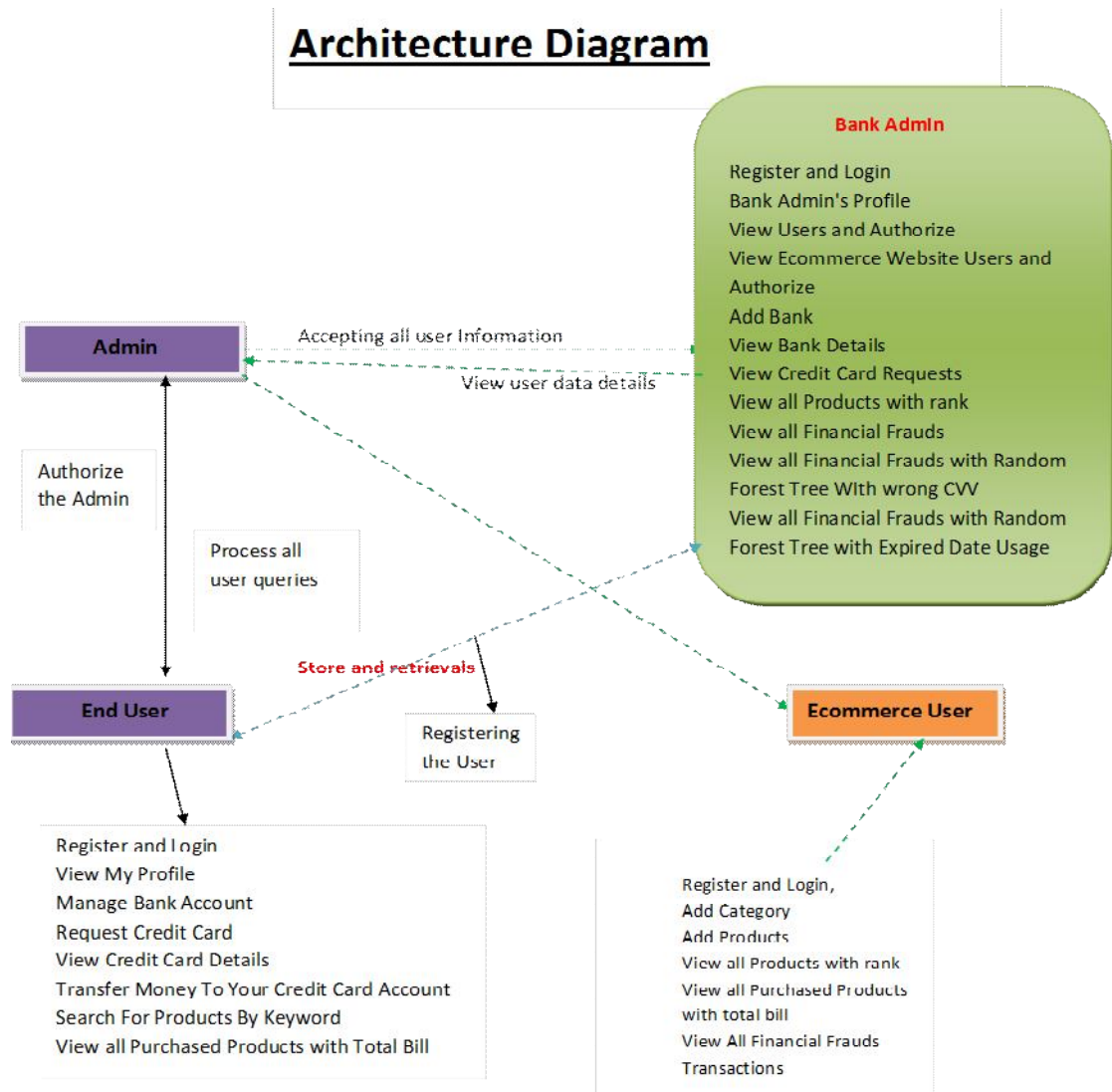
3.1 Existing model

Three methods to detect fraud are presented.

Firstly, clustering model is used to classify the legal and fraudulent transaction using data parameter value.

Secondly, Gaussian mixture model past behavior and current behavior can be calculated to detect any abnormalities from the past behavior.

Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios.



3.1.1 Drawbacks

The high amount of losses due to fraud and the awareness of the relation between loss and the available limit has to be reduced.

Testing credit card FDSs using real data set is a difficult task.

The fraud has to be deducted in real time and the number of false alert.

3.2 Proposed system

Total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection.

3.2.1 Advantages of proposed system

The system is very fast due to AdaBoost Technique.

Effective Majority Voting techniques.

4.Methodology

4.1 Machine learning algorithms

A sum of twelve calculations are utilized in this test examine. They are utilized related to the AdaBoost

Table 1.0. Comparison table

Paper No	Technique	Advantages	Disadvantages
1	SVM reduction	To reduce credit card fraud and to predict future fraud.	It generates the false alarms.
2	Classification method, Naïve Bays.	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 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 behaviour 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.

4.2 Majority voting

Dominant part casting a ballot is much of the time utilized in information grouping, which includes a joined model with something like two calculations. Every calculation makes its very own forecast for each test. The last yield is for the one that gets most of the ballot, as pursues.

Examine K selected classes (or marks), with C_i, K . K speaks to the i th target class anticipated by a classifier. specific info x , every classifier furnishes a forecast concerning the objective class, submit a sum of K expectation, i.e., P_K . Greater part casting a ballot expects to deliver a consolidated expectation for info x , $P(x)_j, j \in K$ from all the K forecasts, i.e., $p_k(x)_j$. A double capacity can be utilized to speak to the votes.

If $p_k(x) = i, i \in K, V_k(x \in C_i) = 0$

At that point, entirety the votes from all K classifiers for every C_i , and the name that gets the most elevated vote is the last (joined) anticipated category.

4.3. Adaboost

Versatile Boosting or AdaBoost is utilized related to various kinds of calculations to upgrade their execution. The yields are joined by utilizing a insignificance entirety, which speaks to the consolidated yield of the supported classifier, i.e.,

$FT(x) = f_t(x), t=1$

Where each f_t is a classifier (feeble student) that profits the anticipated class regarding input x . Each frail student gives a yield forecast, $h(x_i)$, for each preparation test. In each cycle t , the feeble student is picked, and is distributed a coefficient, α_t , with the goal that the preparation blunder aggregate, E_t , of the subsequent t -arrange helped classifier is limited,

$$E_t = E [F_{t-1}(x_i) + \alpha_t h(x_i)]$$

where $F_{t-1}(x)$ is the supported classifier worked in the past stage, $E(F)$ is the blunder capacity, and $f_t(x) \alpha_t h(x)$ is powerless student thought about for the last classifier.

Adaboost changes power less students for misclassified information tests. It is, nonetheless, touchy to commotion and outliers. For whatever length of time that the classifier execution isn't arbitrary, Adaboost can improve individuals outcome.

5.Conclusion

The Data mining, best concept of machine learning algorithm is used for credit card fraud in this proposed system is proposed. Then, the character of quality replica 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 replica such as AdaBoost and majority voting, this models are blend technique, 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 figure 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 internet instruction to allow the quick awareness of credit card fraud. The proposed system is help to detect and before prevent the fraudulent transaction and activities, so to decrease the unit of dropping in economic industry.

REFERENCES

- [1]. MehakKamboj, 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. AmarpreetKaur."Fraud Prediction for credit card using classification methid". International Journal of Engineering and Technology, (2018); 7(3) 1083-1086.
- [3]. Wee-Yong Lim, AmitSachan, 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, DamminaAlahakoon, 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]. TamannaChouhan, 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]. MairaAnis, Mohsin Ali, AmitYadav. "A Comparative study of decision tree algorithms for class imbalanced learning in credit card fraud detection". International Journal of Economics, Commerce andManagement, ISSN 2348 0386.

Author's Profile



MANDE DHANALAKSHMI pursuing M. Tech in Computer Science and Engineering from VelagaNageswaraRao College Of Engineering 2017-19 respectively.



K.S KIRAN KUMAR, Qualification : M.Tech, having 6 years experience in teaching field at present he is working as Asst professor in VelagaNageswaraRao College Of Engineering, Ponnur(Post),Ponnur(Md)Guntur(D.T), Andhra Pradesh kirancse@gmail.com, phone:9290091551