International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Optimized Random Forest for Credit Card Fraud Detection with User Interface

Aiswarya C J[1]

[1]*Department of Computer Sci. & Engg, Thejus Engineering College, Vellarakkad, Thrissur, Kerala, India*

Abstract - *Due to the rapid advancement in electronic commerce technology, the use of credit card has dramatically increased. With the development in information technology and improvements in communication channels, credit card fraud events are also spreading. Credit card frauds are those events in which the criminals make use of a stolen card to steal the confidential information of other peoples credit card. Due to its growing trend of transaction frauds, it has resulted in great loss of money every year. Currently in an online transaction environment the physical card is no longer required as if information can used to make a payment. This has made it much easier for criminals to conduct a fraud which has brought a negative influence on economy. Therefore ,fraud detection is very essential in order to identify fraud on time before the criminal uses stolen information.so here we are proposing an effective fraud detection model using the random forest ML algorithm in an optimized way with an user interface can capable of predicting the credit card fraud events more accurately. We are also going through the limitations of the machine learning algorithm that we selected for this prediction model and we are trying maximum to over come this limitation through the parameter tuning. We are training the system with the normal and abnormal behavior features. The prediction models classify the test transaction that we given is fraud or not. Finally we are also providing and user interface as an security enhancement, while any fraud events may occurred.*

Key Words: RF, UI, Parameter Tuning

1. INTRODUCTION

The fast and wide reach of the Internet has made it one of the major selling channels for the retail sector. In the last few years, there has been a rapid increase in the number of card issuers, card users and online merchants, giving very little time for technology to catch-up and prevent online fraud completely. Statistics show that on-line banking has been the fastest growing Internet activity with nearly 44% of the population in the US actively participating in it. As overall e-commerce volumes continued to grow over the past few years, the figure of losses to Internet merchants was projected to be between $5-$15 billion in the year 2005. Online banking and e- commerce have been experiencing rapid growth over the past few years and show tremendous promise of growth even in the future. This has made it easier for fraudsters to indulge in new and abstruse ways of committing credit card fraud over the Internet. Credit card fraud involves illegal use of card or card information without the knowledge of the owner and hence is an act of criminal deception. It is necessary to employ an effective and economical solution to combat fraud. So, our major objective is to implement an effective credit card fraud detection model. Here we are using an efficient machine learning algorithm which is very effective to detect the credit card fraud. We are also going through the limitations of the machine learning algorithm that we selected and trying maximum to overcome this limitation through the parameter tuning. Initially we need an historical transaction data related to the credit card transactions. We are training the system with the normal and abnormal behavior features. The prediction model will classify the test transaction that we give is fraud or not. Finally we are also providing an user interface (UI) as an security enhancement, while any fraud event occurred.

2. LITERATURE REVIEW

Behdad, M [1], reviews the most popular types of electronic fraud and the existing nature inspired detection methods that are used for them. This paper focus to implement a fraud detection method to adapt the changing environment. The paper finding solution by Using Nature-Inspired Techniques like Ant Colony Optimization, Evolutionary Algorithm, Etc. when new challenges like, click fraud, spam blog arises, It is inefficient to solve using the method proposed on this paper. Also the training is too expensive in this model.

Quah, J. T. S [2] proposed a model to detect the real time credit card fraud using computational intelligence. This paper focuses on real-time fraud detection and presents a new and innovative approach in understanding spending patterns to decipher potential fraud cases. It makes use of Self Organization Map to decipher, filter and analyze customer behavior for detection of fraud. The drawback of this model is that, more accurate fraud detect system needed for dynamic and voluminous E-Commerce system.

Srivastava [3] presents the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. An HMM is a double embedded stochastic process with two hierarchy levels. But the major drawback of this model is that, they

only consider the transaction amount as the feature in transaction process.

Kundu[4], in his work he presenting a BLAST-SSAHA Hybridization for Credit-Card fraud detection system for all credit card issuing banks to minimize their losses. This method use two-stage sequence alignment with profile analyzer (PA) and deviation analyzer (DA). The major limitation of this model is that it cannot detect frauds in real time.

Jakobsson[5], implements an implicit authentication through learning user behavior. This paper proposing a user behavior model which treats the transaction feature independently. This method using implicit authentication and it authenticating users based on behavior pattern. The limitation of this model is that it only deal with behavior patterns

3. METHODOLOGY

While literature survey, we analyzed various machine learning algorithms and finally concluded that, using the random forest machine learning algorithm in an optimized way with an user interface can capable of predicting the credit card fraud events more accurately. For the training purpose, in our proposed system we are using the random forest machine learning algorithm. Random forest, one of ensemble methods, is a combination of multiple tree predictors such that each tree depends on a random independent dataset and all trees in the forest are of the same distribution . The capacity of random forest not only depends on the strength of individual tree but also the correlation between different trees. The stronger the strength of single tree and the less the correlation of different tress, the better the performance of random forest. The variation of trees comes from their randomness which involves bootstrapped samples and randomly selects a subset of data attributes. Although there possibly exist some mislabeled instances in our dataset, random forest is still robust to noise and outliers. The major drawback of the random forest machine learning algorithm is its complexity and its increased execution time. In this project we are overcoming these drawbacks of the random forest algorithm by tuning the parameters that we are using in it. Tuning the parameters means that from the entire features available we are using only the most important features after the parameter tuning. For that we are analyzing dataset and its entire features and represents the importance of the each feature in a bar graph. From this graph we can see the importance of each features for this prediction. From that we choose only the important features for the random forest machine learning algorithm. we will more accurate and fast result by using the parameters in the RF algorithm after tuning rather than using the entire parameters in the algorithm. When we are giving the customer id, we can able to predict that the transaction of the customer is normal or fraud. If it is a fraud transaction we are also providing an security enhancement to give the user an alert via admin that a fraud event occurred through a email.

4. IMPLEMENTATION

Implementation procedure carried out on the proposed method describes in this section. The implementation is performed in eight steps. These steps are described detail in following section

4.1 DATA COLLECTION

The initial stage of our proposing system is the data collection. The data that we are collecting must be real and precise. For our model, we are collecting online dataset from UCI repository. The accuracy of the final result depends on the data that we collected. The dataset that we collected should be precise and must be a real data. The patterns of the pervious real data can only efficiently predict the future events.

4.2 DATA CREATION

The very next module after the data collection is data creation. Here we are normalizing the data's in the dataset that we are collected using an excel sheet. The each rows in the sheet represents the individual users and columns represents features related to every individual transaction. Here there is also having an additional column called label which carries either the value 0 represents the transaction is normal or 1 represents the transaction is fraud.

4.3 DATA PROCESSING

The next module in our system is data processing. The data that we created need to be processed. The data processing involves 3- sub phases, feature extraction, feature vector creation and label vector creation. Feature extraction starts from an initial set of measured data and builds derived values intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction. Feature vectors are used to represent numeric or symbolic characteristics, called features, of an object in a mathematical, easily analyzable way. Data labeling, in the context of machine learning, is the process of detecting and tagging data samples. The process can be manual but is usually performed or assisted by software. The major advantage of the labeling the data is, it Predictable good results and control over the process.

4.4 PARAMETER TUNING

For the training purpose, in our proposed system we are using the random forest machine learning algorithm. This algorithm contains a group of/or a forest of decision trees. Each features in the dataset that we are giving in the decision

tree. The major drawback of random forest is it's complexity. We are applying parameter testing over the parameters on dataset to over come this limitation.. While apply the parameter tuning over the features in dataset, it will chart the importance of each parameters in the dataset for the accurate prediction of results. So we can make use of important features from the plot for making the prediction model which will be created after training data.

4.5 TRAINING

The data set that we acquired after the parameter tuning use for training the machine. We train a machine with historical data to give an experience to the system on previous events and make a system capable to predict an future event accurately. Our proposed system is using the Random forest machine learning algorithm for training. The below steps showing the working of the random forest machine learning algorithm as a trainer in the system that we are implementing

Step 1 – Selecting random samples from given dataset.
Step 2 – Constructing decision tree for every sample.
Step 3 – Get prediction result from every decision tree.
Step 4 – Voting performed for every predicted result.
Step 5 – Most voted prediction result set as final result.

4.6 TESTING

In this stage, with the random forest classifier created, we will make the prediction. The random forest prediction pseudocode is shown below:

1. Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome.
2. Calculate the votes for each predicted target.
3. Consider the high voted predicted target as the final prediction from the random forest algorithm.

Our system developed for the credit card fraud detection. When we gave an test data to this model, it will predict that this transaction is fraud or not. The test data that we give is the customer id. When we give a customer id to the system, it will checks all of its previous transactions and prevent that the current transaction is fraud or not. For this application, Random Forest algorithm is used to find loyal customers, which means customers who can take out plenty of money using credit cards and pay money back to the bank properly, and fraud customers, which means customers who have bad records like failure to pay back a money on time or have dangerous action.

4.7 SECURITY ENHANCEMENT

This module deals with the ,user level security enhancement for the credit card fraud detection. Our proposed system capable of detecting the fraud event in an credit card transactions. The prediction model can analyze the test data that we give through the random trees and detect whether the transaction is normal or not. When the system identifies an abnormal or fraud transaction, it is necessary to give an alert to the user who using the credit card were the fraud event occur. We are designed the UI in our system in a manner that any fraud transaction encountered, first security breach alert sent to the administrator. The administrator will send an alert to the user through the email to inform the customer that the one fraud event occurred, so user can change the password immediately.

4.8 COMPARISON

The random forest algorithm after parameter tuning is using as the classifier in this credit card fraud detection model. When we are comparing it with other machine learning algorithms we can see that the classification using random forest algorithm is more accurate than the classification using other algorithms. Here we are using support vector machine algorithm and random forest algorithm for the classification and finally we analyses which classifier is more accurate for the result prediction.

5. CONCLUSIONS

Popularization of e-commerce becomes increasing day by day. As along with this e-commerce popularization in our daily life, there is also a chance to occur the fraud events frequently. One of the major frauds occurring in this area is the credit card fraud events. So it is important to identify these fraud events. The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Here in this project we are proposing a method to detect the credit card frauds using the supervised machine learning technique called Random Forest with parameter tuning. The parameter tuning before using the random forest classifier can will capable for predicting the result more accurately with less execution time. This machine learning algorithm helps to classify the normal/abnormal credit card transaction and test the input data and then predict its behavior. After the prediction we are enabling an user interface to the change the password of the credit card when there any fraud events identified. We can improve the performance of the random forest classifier by tuning the parameter more efficiently and increasing the depth of the trees using in the random trees. There is still a huge scope for improvement in this model. Cross validation accuracy is generally more optimistic than true test accuracy. To make a prediction on the test set, minimal data preprocessing on categorical variables is required.

6. REFERENCES

[1] Behdad, M., Barone, L., Bennamoun, M., and French, T. (2012). Nature- inspired techniques in the context of fraud detection. IEEE Transactions on Systems Man and Cybernetics Part C, 42(6), 1273-1290.

[2] Quah, J. T. S., and Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. Expert Systems with Appli- cations, 35(4), 1721-1732.

[3] Srivastava, A., Kundu, A., Sural, S., and Majumdar, A. (2008). Credit card fraud detection using hidden markov model. IEEE Transactions on Dependable and Secure Computing, 5(1), 37-48.

[4] Kundu, A., Panigrahi, S., Sural, S., and Majumdar, A. K. (2009). Blast- ssaha hybridization for credit card fraud detection. IEEE Transactions on Dependable and Secure Computing, 6(4), 309-315.

[5] Shi, E., Niu, Y., Jakobsson, M., and Chow, R. (2010). Implicit Authen- tication through Learning User Behavior. International Conference on Information Security (Vol.6531, pp.99-113). Springer-Verlag.

[6] Mota, G., Fernandes, J., and Belo, O. (2014). Usage signatures analysis an alternative method for preventing fraud in E-Commerce applications. International Conference on Data Science and Advanced Analytics (pp.203-208). IEEE.