A Case Study for Credit Card Fraud Detection
Credit Card Fraud Detection Using MachineLearning

B.S.Swapnashanti[1], Itla Venkatasai[2], Marupa Akash[3], K.VenkatLaxmi Narisimha[4], M.A.OmerFarooq[5]

[1]Assistant Professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

[2,3,4,5] IVthBtech Student, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

ABSTRACT

In recent years credit card became one of the essential parts of the people. Sudden increase in E-commerce, customer started using credit card for online purchasing therefore risk of fraud also increases. Instead of carrying a huge amount in hand it is easier to keep credit cards. But nowadays that too becomes unsafe. Now a days we are facing a big problem on credit card fraud which is increasing in a good percentage. The main purpose is the survey on the various methods applied to detect credit card frauds. From the abnormalities, in the transaction, the fraudulent one is identified. We address this issue in order to implement some machine learning algorithm like random forest, logistic regression in order to detect this kind of fraud. In this paper we increase the efficiency in finding the fraud. However, we discussed and evaluated employee criteria. Currently, the issues of credit card fraud detection have become a big problem for new researchers. We implement an intelligent algorithm which will detect all kind of fraud in a credit card transaction. We handled the problem by finding a pattern of each customer in between fraud and legal transaction. Isolation Forest Algorithm and Local Outlier Factor are used to predict the pattern of transaction for each customer and a decision is made according to them. In order to prevent data from mismatching, all attribute are marked equally.

1. INTRODUCTION

1.1 GENERAL

Nowadays as we can see that there is a huge increase online payment and the payment is mostly done with the help of credit cards. It becomes a big problem for marketing company to overcome with the credit card fraudulent activities. Fraudulent can be done in many ways such as tax return in any other account, taking loans with wrong information etc. Therefore, we need an efficient fraudulent detection model to minimize fraudulent activity and to minimize their losses. There are a huge number of new techniques which provide different algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction by proposing a machine learning algorithms. Credit Card Fraudulent detection comes under machine learning, and the objective is to reduce such type of fraudulent activity. This type of fraud is happening from past, and till now not much research has done here in this particular area. The types of credit these features play an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. For the better performance of model, we need a better classifier. Different classifier can be combined together with help of meta-learning.

1.2 OBJECTIVES

To run a suitable business, vendors need to make a profit, which can be calculated by subtracting the cost of doing business from the total sell price. Therefore, fraudulent become a business's tolerance among online payment, among financing company, gross margin is calculated by (sell price - cost of goods sold). The lower the margin, there will be low risk for fraudulent payment. In practice, whenever fraudulent occurs, the cardholder have to complain to the financing company and the debit from card is usually cancelled, which means there is a loss for either cardholder's bank or the finance company. Fraudulent turns as a financial risk to the financial company and the cardholder's bank. To overcome with fraudulent, fraudulent detection techniques should be used. The main objective is to prevent the customer from fraud because if this kind of things keep happening then people will not show there interest in taking credit card and using there facility which is given by the banks and other financial company. Therefore, it's become an essential thing nowadays. People should also takes care of their personal information by keeping it to the limited source. The fraudulent activity start with the leaking of the someone personal information like credit number, one time password, registered mobile number and many more. The sharing of someone personal information should be reduced because fraudulent activity begin with the help of someone personal information like credit card number and many more.

1.3 EXISTING SYSTEM

The previous detecting technique takes a long time to catch fraud which is basically depend on the database, not that much accurate and not give the result in-time. After that algorithm which is used for the detection of credit card fraudulent is generally on basis of analysis, fraudulent detection based on credit card transaction made by cardholder and the credit rate for cardholders. There are certain limits of meta-learning. There are two features which is introduced here in our report is True Positive and False alarm. Both these features play an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. For the better performance of model, we need a better classifier. Different classifier can be combined together with help of meta-learning. Previously attempts have been made to work out Credit Card Fraud Detection system using SVM (Select Vector Machine). SVM makes use of hyperplane to classify the data points in a collection. A good hyperplane associates greater number of data points within its margin [2] . This is not efficient for a large amount of data sets. As, in large amount of data sets there is a probability of redundant data which will take more time to process. Therefore, it usually delayed in calculating the fraud or there might be probability to not calculate in time.

13.1 DISADVANTAGES OF EXISTINGSYSTEM

• In case of fraud there is a high amount losses and thus because of this loss, card limit should be reduced. • The fraudulent should be detected in real time and omission in false transactions is mandatory. • Reasons of fraudulent should be identified from data available. • System should be capable in identifying the trend of fraud transaction. • Credit card fraudulent transaction should be based on web service scheme.

1.4 THE SYSTEM PROPOSED

• In this model we overcome with the issues in a significant way. Using Isolation random forest and local outlier factor algorithm we can detect the fraud in actual time and find out the way to minimize the fraud to produces an optimized result so that it will perform a better prediction. On the basis of customer's behavior, we can detect fraudulent. Here the local outlier factor is used. • We have used logistic regression and random forest. We can get more accuracy like 0.99 etc… • We are taking the dataset with help of simple GUI from our local directory where we downloaded the dataset. • With the help of random forest algorithm and local outlier factor we are finding the data point which is different from its neighbor and can be a fraudulent transaction with its outlier behavior. • We have two classification class which is named as class 0 and class 1. • If there is legal transaction then the result will store in class 0 and if there is a fraudulent transaction then the result will store in class 1.

SYSTEM IMPLEMENTAION

2.1 GENERAL

Implementation phase brings out the design tweaked out into a operational system. Hence this can be deliberated to be most precarious juncture in accomplishing the efficacious system and in convincing the user faith that system will operate and be effective. This phase encompasses vigilant planning & design, examination of prevailing system and constraints on execution, design & scheming of methods to change over.

2.2 PROCEDURE FOLLOWED DURING IMPLEMENTATION

The application – Credit Card Fraud Detection which is in itself the complete & full-fledged GUI enabled application to envisage/foresee the authenticity & legitimacy of a transaction has been implemented, as per the following steps: • Install Anaconda from an reliable source. • Import packages: pandas, Scipy, Matplotlib, Seaborn • Load the dataset, a dataset is the pool of data for analytical/critical purpose, a (.CSV) file. • Reconnoiter and get through the dataset through data. shape, data. describe. • Split the dataset into training dataset and testing dataset.

• Plot histogram of the dataset to epitomize/depict numerical data. 31 • Determine the count of fraud cases by checking if class is 0 or 1. • In the similar procedure, get the correlation matrix. • Next, there is a need to determine the local outlier factor. • This is followed by use of random forest algorithm to find accurate results. • The GUI is developed using PyQt library. • The PyQt library, provides tools to achieve a complete GUI enabled application, similar to swings in java environment. • Define the constructor in the file. • Write down the entire implementation inside, thus encapsulating everything inside a GUI-enabled python file.

2.3 DATASET DESIGN

The dataset holds information about credit card transactions which has been made in a span of two days. The number of frauds have been calculated as 492 out of 284,807 transactions. The details have been given in form of positive and non-positive numerical values. The dataset contains 31 features which has been labelled as V1-V28 due to confidential reasons. The feature which has been reveled are Time and Amount of transaction. Here time denotes the number of seconds elapsed from the first transaction of Day 1. Amount of transaction consists of positive value denoting deposit and non-positive value denoting withdrawal.

3. RELATED WORK

Sharayu Pradeep and Nitin [2], have implemented credit card fraud detection system wherein the goal of their implementation is to minimize false alarms for authentic transactions. ML algorithms are used for detecting local outlier factor and even hidden Markov model has been used for fraud detection. Thirunavukkarasu M et.al [3], have proposed a credit card fraud detection system for real world scenarios. This system provides the required properties needed to determine the felonious and unauthorized transactions. Credit card data set is collected by the users and they are classified as trained and tested data set using random forest algorithm and decision tree. Based on the performance metrics used like accuracy, sensitivity, specificity and precision, random forest algorithm is found to have highest accuracy. A credit card fraud detection system which is found to be upgraded using machine learning has been proposed by Ramya M et.al[4]. They have put forward an original real time system by depicting 4 different patterns of fraudulent transactions using suitable algorithms. In order to address the issues faced like detecting frauds in a credit card transaction, predictive analytics is used for this purpose. Finally, with the help of GUI, end user is notified about it. Work carried out by Maniraj SP et.al [5] showed a credit card fraud detection system where their template or framework is used to acknowledge or admit whether a new transaction which has been carried out is found to be fraudulent or not. Their intention is to discover all transactions which are found to be fraudulent and then reduce the incorrect fraud classifications. Here several anomaly awareness algorithms have been deployed such a local outlier factor and isolation forest. This paper has explained how machine learning can be applied to get accurate results in fraud detection. An ensemble classifier has been used for fraud detection in the work carried out by Masoumeh et.al [6]. The basic machine learning algorithms like Naïve Bayes (NB), support vector machines (SVM), K-nearest neighbor (KNN) have been used here. Among all, ensemble method has been recognized as popular and common method due to its predictive performance on real world problems. Performance evaluation has been carried out lastly to justify the advantages of bagging ensemble algorithm.

4.LITERATURE SURVEY

4.1 GENERAL

In our paper we referred to various papers for improving the performance of routing, reduce delay of information, reduce packet loss rate, reduce link failure, to improve packet delivery rate, to reduce energy consumption. There are a huge number of new techniques which provide different algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction by proposing a machine learning algorithms. There are two features which is introduced here in our report is True Positive and False alarm. Both these features plays an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. As per today's Network plays an important role therefore it is mandatory for our models to be up to date to perform better detection capabilities. Whenever new fraudulent activity are detected then our model should be that much better to perform real time analysis. Other than traditional machine learning methods Fraudulent Detection System has been achieved through using Neural Networks [5] . To prevent personal information has become a huge task for financial company because there are a lot of attack on the system to steal someone personal information to perform fraudulent. Our model has two essential feature which will help in finding abnormal behavior in form of charts for different column such as time, amount etc.

4.2 CREDIT CARD FRAUDULENT DETECTION

We publish a Credit Card fraudulent detection model whose performance is evaluated on basis of anonymized data sets and found that detection model performance is good for this dataset. This is incorporated that this model creates two separate patterns for databases, one for fraud and other for legal transactions. The fraudulent detection model should be more accurate in order to detect the changing behavior of consumer and his behavior. We can predict this fraudulent by running our model after every fixed amount of transaction or after a fixed interval 6 of time. AI provides procedure for various types of calculations which can be performed independently. If there is any outlier value in our dataset, then our model can detect it. Outlier value means the value which deviates by a long margin from their neighbor can perform abnormal behavior. That outlier behavior is the fraudulent transaction in dataset. We have also reduced redundancy of datasets by removing some of the redundant data from our dataset. Because our main aim is achieving the real time analysis and for that we need to reduce the datasets so that we can speed up our algorithm performance.

4.3 DATA SAMPLING

Since, Random forest algorithm is a machine learning algorithm therefore we need trained dataset to perform our mechanism. These trained datasets are then loaded to the main memory of the system. Our dataset has almost 300,000 value so it's a difficult task to load trained dataset in main memory. For that purpose we have removed the redundant datasets. We have trained our dataset from previous data, we did like this because our model should be trained on previous data and should be able detect fraudulent transaction of the current month, which will help in real world.

4.4 CREDIT CARD FRAUDULENT DETECTION USING HIDDEN MARKOV MODEL

In our paper we utilized HMM to identify fraudulent. We demonstrated the exchanges of MasterCard by utilizing HMM. For swiping reason, we have utilized the RFID gadget to demonstrate the shopping exchanges. We identified the misbehaviors by observing the conduct of the client. We include High security addresses page additionally, in case card is stolen, we have given another profile ID to the consumer and gave ONE TIME PASSWORD for security reasons. We have given right to the admin to block the card from obstructing in case card is lost. As our aim is to achieve the better accuracy but our dataset we could achieve up to 99.97%. As for fraudulent detection, the false alarm plays an important role, as whenever there is a fraud transaction it shows an outlier transaction which will differ from its neighbor or we can say that deviate from the given data point. We give more priority to fraudulent catching algorithm then the false alarm because our aim is to catch the fraudulent at the very first moment. 7 2.5 CREDIT CARD FRADULENT DETECTIONUSING DECISION TREE INDUCTION ALGORITHM In Snehal Patiletal, describes the "Decision Tree Induction Algorithm" which is used for Credit Card Fraud Detection [1]. In this paper it discusses about the method, decision tree approach is a new cost sensitive technique compared with well-known traditional classification models on a real-world credit card fraud data set, which reduces the sum of misclassification cost, in selecting the splitting attribute at each of the non-terminal node become advance. Credit card fraud detection is to reduce the bank risks, also used to equalize the transaction information with credit card fraud transaction of historical profile pattern to predict the probability of being fraud on a new online transaction. In this model use of "Credit Card Fraud Detection Using Decision Tree for tracing Email and IP Address. By using this technique, we can able to find out the fraudulent customer/merchant through tracing the fake mail and IP address. If the mail is fake, the customer/merchant is suspicious and information about the owner/sender is traced through IP address. As prediction of score is much important task according to our model therefore we are predicting the score on the basis of the given formula: Score = 0.5 * TP + 0.5 * Deviation Where, TP is True Positive value and Deviation is the deviation of outlier data from the standard data point. On the basis of these score we made two classes 0 and 1. If the score is 1 it will move to class 1 and termed as legal transaction and if the score is 0 it will move to class 0 and termed as fraudulent transaction. At last, the accuracy is calculated on the basis of how many fraud transactions are there in our dataset and how many we predicted with the help of our model.

5.APPLICATION AND FUTURE ENHANCEMENT

5.1 GENERAL

Implementation is the most critical phase to attain a fruitful system and providing the users assurance that the new system is feasible and operative. Each module is tried and tested discretely using the data and substantiated in the mode indicated as per program specification, system and the environment is tested as per user requirement. The frequently techniques for fraud detection are Nave Bayes, support vector machine and the k - nearest neighbor algorithm. Here, this document has trained various machine learning practices and techniques used in detection of fraud in credit card and assess each methodology based on certain design measures and criteria. Nevertheless, if there is a need to contrivance a platform that performs real-time credit card fraud detection, it is imperious to reach precisions of 95%, as the odds of false positives along with false negatives is else quite elevated to be used for business application. Impending task must subsequently be focused at exploring further relevant features to enhance, execution of a thorough optimization, and doing real-time tests. Other than the major fraud practices some other types of frauds are done such as through phishing, skimming, credit card generator etc.[6] . Also the possibility regrettably not pursued for timing issues is to refine the metrics in form of commercial forfeiture resolution system, the tenacity of model wouldn't be to capitalize on the count of transactions precisely organized, but instead minimize the costs associated with following up on fraudulent transactions based on the confidence of the model and the associated financial loss. Finally, approaches for dealing with the 'refused' examples are to be explored. 40

5.2 FUTURE ENHANCEMENTS

There is a very strong possibility of the system being adopted as a norm for the major banking and financial services applications as fraud detection and prevention is the major checkpoint in financial and banking sector. The above system is also likely to be embedded in other applications based, modified as per platform-specific/application specific environment. The banks, financial and retail institutes have faced huge losses owing to cause of a robust and accurate system to predict and prevent the fraudulent transactions going on in an institution. This in-turn affects the business capabilities and consumer trust of the company. Thus, the organizations have moved their focus onto implementing a system which can depict inconsistent transactions, providing banks a privilege to act upon it take necessary measures.

6.CONCLUSION

In this model, we discussed about credit card fraud detection using machine learning. The proposed model has been extensively tested on different types of transactions. The results were promising, almost all the fraudulent transactions could be detected successfully, and the proposed methodology has been compared with existing method and the results shows that proposed method performs superior than existing methods. In this model, we detected the fraudulent transactions and recognized which illustrates the robustness of the proposed system. This proposed model took the trained dataset and performed classification on basis of them, if the transaction was legal then it moved to class 0 and if the transaction was fraud then it moved to class 1, and significantly improve the detection accuracy. The proposed method works efficiently in various platform, vivid environment and is a full□fledged cross platform application. The system has depicted robust, scalable and accurate performance to the degree that efficiency is taken into consideration in the Credit Card Fraud Detection System. The system takes into consideration various factors and has been fulfilling or meeting all the project specifications documented.

7.REFERENCES

[1] V. Bhusari S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011.

[2] Priya Ravindra Shimpi, Prof. Vijayalaxmi Kadroli Angrish, "Survey on Credit Card Fraud Detection Techniques", International Journal Of Engineering And Computer Science ISSN: 2319-7242 [3] Salvatore J. Stolfo, Wei Fan, WenkeLee, "Cost-based Modeling for Fraud and Intrusion Detection Results from the JAM Project", In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 207– 216, 2014.

[3] Delamaire. L. Abdou, HAH and Pointon. J,"Credit card f raud and detection techniques", Banks and Bank Systems, Volume 4, Issue 2, 2009.

[4] Suman, Nutan, "Review Paper on Credit Card Fraud Detection", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013.

[5] Renu, Suman, "Analysis on Credit Card Fraud Detection Methods", International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1 – Feb 2014. [7] Sushmito Ghosh and Douglas L. Reilly, "Credit Card Fraud Detection with a Neural□Network" Proc. IEEE First Int. Conf. on Neural Networks, 2014.

[6] Deepak Pawar, SwapnilRabse, Sameer Paradkar, NainaKaushi, "Detection of Fraud in Online Credit Card Transactions", International Journal of Technical Research and Applications e-ISSN: 2320-8163.