International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Fraud Detection in Mobile Payment System using Machine Learning: A Comprehensive Survey

Bhakti G. Gawas[1], Prof. Shruti Pednekar[2]
[1, 2]*Computer Engineering Department, Goa College of Engineering*

Abstract: Due to ever increasing bulk of digital transactions nowadays there is also increase in fraudulent transactions. There is need to detect this fraud transactions in online digital payment system. This paper is aimed at providing an expensive literature review of journal article produced between 2009 to 2018 in the selected domain. We have classified the research papers based on the machine learning methods, accuracy level of the technology, year of publishing, etc. and analysed the results.
Keywords: Classification; ensemble; fraud; fraudulent; transaction;

I. INTRODUCTION

Nowadays it's not only groceries and coffee that have used the mobile payments system but, consumers expect to pay for everything in digital mode. When choosing a product or service, consumers want to pay for it with ease and convenience. This resulted in expansion of e-commerce; digital payment technology has experienced significant improvements, especially in mobile payments. This has led to a growth of mobile commerce and an entire infrastructure of businesses has been born out of this ubiquitous mobile payment environment. From online banking to budgeting apps that record mobile payments and to digital loyalty programs and mobile-only sales, mobile payment system has become big business. Fraudsters have now become very smart, and adapted and developed new methods to carry out various crimes. Most common fraud that can be seen is Credit card frauds, Identity theft, and Mobile wallet frauds and so on. In this survey comparative study of recently used methods for fraud detection is briefly described. Fraud in online payments is a major and a serious problem growing nowadays. Several machine learning techniques have been used to detect fraudulent transactions. So the proposed work is to use ensemble methods and combine machine learning techniques.

II. LITERATURE REVIEW

Different types of studies have done in detecting fraud transactions. Several data mining techniques are utilized for fraud detection and have achieved different accuracy levels for different methods. In [13] survey was presented on various data mining and machine learning methods which are widely used for credit card fraud detections. In this paper author investigates about different methods of fraud detection in credit card. Survey was done on methods the are commonly used for detecting fraud those are genetic algorithm, decision tree, Artificial neural network, Convolution neural network(CNN), Outlier detection, clustering techniques, logistic regression, Deep learning, Rule based method, Hidden Markov model. From the previous survey of detection of fraud in credit card, few of the challenges were also identified in this paper. Reference [14] shows research on fraud detection on credit card data using machine learning methods. Advanced data mining techniques namely decision tree classifier, support vector machines, Logistic regression and random forest classifiers were used to detect the fraud transaction and then comparison was made to evaluate the best model. Credit card transactions dataset was used from European cardholders which had 284,786 transactions. Machine learning methods were applied on pre-processed data. The performance of the techniques was checked based the parameters such as accuracy, sensitivity, specificity, precision. The accuracies of techniques were found and listed. The result indicated that random forest was the best technique which gave accuracy of 98.6%. The accuracy level of the other classifiers which were used namely logistic regression, decision tree, SVM were 97.7%, 95.5%, 97.5% respectively. Hence it concluded Random forest was most precise classifier for detecting fraudulent transactions with dataset provided by ULB machine learning. In [10], authors used four different classifiers namely naïve Bayes classifier, Decision trees classifier, support vector machines, K-nearest neighbour. These four techniques were trained on real life data set of financial transactions for the purpose of fraud detection. The dataset used was pre-processed, which had 20 features and after removing some of the features it was made to 16 features. Confusion matrix was calculated. The performance of classifiers was evaluated on relevant metrics namely True Positive value, False Positive value, Balance Classification Rate (BCR), Matthews Correlation Coefficient (MCC). Results were calculated using weka tool. The accuracy level of the classifier used were namely SVM, K-nearest neighbour, Decision trees, Naïve Bayes were 69%, 70%, 65.3%, 73.8% respectively. They concluded that performance improvement could be achieved through developing a fraud detection model using an ensemble of different machine learning techniques.

In [1], authors studied on fraud detection based on financial data using machine learning methodology in small amount mobile payment system. This paper analyses the trend of detecting abnormal transactions in the financial sector and checks for the abnormal transaction patterns of mobile-based small-amount payment system. They studied several classification techniques for abnormal transaction detection such as Random forest classifier, neural networks, K-means, K-nearest neighbour and so on. This paper proposes abnormal transaction detection technique on decision tree methodology. They did test performance of the system with 5,000 actual mobile small-amount payment transactions. The results showed 86% sensitivity in detecting abnormal transactions using decision tree methodology. Another research was presented in [9] in which authors suggested bagging ensemble method using decision tree for detecting fraud transaction using credit card transactions. This paper also reviews several machine learning techniques namely naïve Bayes method, support vector machines, and K-NN. The dataset used was real time credit card data which was acquired from UCSD-FICO competition. The dataset had 100,000 credit card transactions and 20 attributes. The data was already labelled by bank, as legitimate and fraudulent. The performance of the various classification techniques were evaluated based on Fraud Catching Rate value, False Alarm Rate value, Balanced Classification and Matthews Correlation Coefficient. While the performance of other classifiers is lower than bagging classifier. Bagging ensemble method takes very less time. In [8] authors suggested an intelligent approach for the problem class imbalance in which frequent item set mining methodology was used. The proposed model was evaluated on data from UCSD Contest 2009. The testing set which was used was unlabelled data. The dataset had 100000 transactions of 73729 customers of up to 98 days. This dataset had 20 fields which after pre-processing made into 16 fields. Also the performance of the proposed fraud detection model which was Fraud Miner was measured with 4 other classifiers namely support vector machines , K-nearest neighbour classifier, naive Bayes , and Random forest used for detection of fraudulent transactions . In [7] comparative study was done on classifiers namely naive Bayes classifier and J48 by using bank dataset to increase true positive value and decrease false positive value rather than only achieving high classification accuracy. In this paper experimentation results was done on factors such as accuracy, sensitivity value and specificity value. J48 gave high classification accuracy than naive Bayes classifier.

III.OBSERVATION

The table below lists the different machine learning techniques used by various researchers used in their studies along with the accuracies obtained by them for each technique.

TABLE I OBSERVATION TABLE

Year	Reference Paper	Machine Learning Technique/Algorithm	Accuracy
2007	[5]	Decision tree, Neural networks, logistic regression	70%, 68%, 62%
2010	[2]	Naïve Bayes classifier	83.56%
2011	[3]	SVM, Random forest, logistic regression	96%, 93.2%, 92.1%
2013	[7]	Naïve Bayes J48	61.33%, 67.7%
2015	[9]	Bagging ensemble based on decision trees	96.77%
2016	[1]	Decision trees	86%
2016	[10]	Support vector machine, KNN, Decision trees, Naïve Bayes	69%, 70%, 65.3%, 73.8%
2017	[11]	Random forest, J48 tree	94.32%, 93.50%
2017	[12]	Naïve Bayes, KNN, Logistic regression	97.92%, 97.69%, 54.86%
2018	[14]	Logistic Regression, Decision tree, Random forest, SVM	97.7%, 95.5%, 98.6%, 97.5%
2018	[13]	Random Forest	96.97%

IV.CONCLUSIONS

In this paper, from the survey conducted, it gives the idea of different models available and the various machine learning techniques used in previous papers for detecting fraudulent transaction. From the analysis mode it is seen that accuracy of the classifiers varies in every paper that is because of the different number of attributes used. Hence different technologies give different results. So for further implementation, ensemble methods can be used to combine several machine learning techniques namely Naïve Bayes, Decision trees, KNN and so on.

REFERENCES

[1] EunYoung Choi, Woong Go, Taejin Lee, "A study on financial fraud detection method using machine learning in mobile small-amount payment service," International Journal of Innovative Research in Technology & Science, ISSN 2321-1156, Vol. 4, No. 5, September 2016.
[2] Qingshan Deng, "Detection of Fraudulent Financial Statements Based on Naïve Bayes Classifier," IEEE, The 5th International Conference on Computer Science & Education, pp. 24–27, 2010.
[3] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: a comparative study," Decision Support Systems, pp. 602–613, 2011.
[4] Shing-Han Li, David C. Yen, Wen-Hui Lu, Chiang Wang, "Identifying the signs of fraudulent accounts using data mining techniques," Computers in Human Behavior, Elsevier, pp. 1002–1013 , 2012.
[5] Aihua Shen1, Rencheng Tong1, Yaochen Deng2, "Application of Classification Models on Credit Card Fraud Detection," IEEE, pp.4244-0885, 2007.
[6] C. Whitrow, D. J. Hand , P. Juszczak, D. Weston , N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," Data mining knowledge discovery, Springer, pp. 30-55, DOI 10.1007/s10618-008-0116-z, 2008.
[7] Tina R. Patil, Mrs. S. S. Sherekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification," International Journal of Computer Science and Applications, Vol. 6, No.2, ISSN: 0974-1011, 2013.
[8] K. R. Seeja and Masoumeh Zareapoor, "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining," The Scientific World Journal, Volume 2014, Article ID 252797, doi.org/10.1155/2014/252797, 2014.
[9] Masoumeh Zareapoora, Pourya Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," International Conference on Intelligent Computing, Communication & Convergence, Elsevier, pp. 679 – 686, 2015.
[10] Marwan Fahmi, Abeer Hamdy, Khaled Nagati, "Data Mining Techniques for Credit Card Fraud Detection: Empirical Study," Sustainable vital technologies in engineering & informatics, pp. 8-10, Nov 2016.
[11] John Richard D. Kho, Larry A. Vea, "Credit Card Fraud Detection Based on Transaction Behavior," IEEE, 978-1-5090-1134-6/17, pp. 5-8, 2017.
[12] John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis," IEEE, 2017.
[13]] Vipul Patil1, Dr. Umesh Kumar Lilhore, "A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 03, ISSN : 2456-3307,pp. 320-325, 2018.
[14] Navanshu Khare and Saad Yunus Sait, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models," International Journal of Pure and Applied Mathematics, ISSN: 1314-3395, pp.825-838, 2018.