IRJEAS
NURTURING RESEARCH

INTERNATIONAL RESEARCH JOURNAL
OF
ENGINEERING & APPLIED SCIENCES

ISSN: 2322-0821(O)
ISSN: 2394-9910(P)
VOLUME 6 ISSUE 3
Jul 2018 - Sep2018
www.irjeas.org

ENSEMBLE PREDICTIVE MODEL FOR DETECTING CREDIT CARD FRAUD IN E-COMMERCE TRANSACTION

Swati Patel
MTech Scholar
NIRT, Bhopal

Anurag Shrivastava
AP & Head, Dept of CSE
NIRT, Bhopal

Anand Motwani
AP, CSE
NIRT, Bhopal

ABSTRACT

The models or techniques to assist fraud investigators, for efficient Credit Card Fraud Detection (CCFD), rely on machine learning algorithms. Proposing a predictive model for Credit Card Fraud determination is however mainly exigent due to the highly distributed and imbalanced data and the availability of only few transactions labelled as fraud in overall transactions. To seek out whether the transaction is fraud on E-commerce websites, is role of prediction models. To find out such transaction can be treated as a sort of machine learning (ML) problem. It is confirmed through several researches that use of Ensemble methods in ML certainly improves performance of prediction and classification tasks. This paper surveys fine classification and ensemble methods that are helpful in building model for CCFD. Further in this paper, a predictive model for CCFD based on ensemble method is proposed. The dataset from UCSD-FICO Data Mining competition is used for building and testing the model. The results obtained shows that the predictive model has potential in determining fraud and minimizing the risk in e-commerce transactions. The paper directs about the future research in the field.

General Terms
Credit Card Fraud Detection, E-commerce

Keywords
Credit Card Fraud Detection, E-commerce, Predictive Model, Machine Learning, Financial Organization, Classification, Ensemble Method.

1. INTRODUCTION

The Credit Card is used as a mode of payment provided to the customer of bank or such financial organization. It allows buying goods or services to its holder. It is generally made up ofPlastic with some secret numbers and Cardholder's Promise to pay for these goods and services availed [1, 2]. In Figure 1 'Clearing and Settlement under Credit Card System' is depicted.

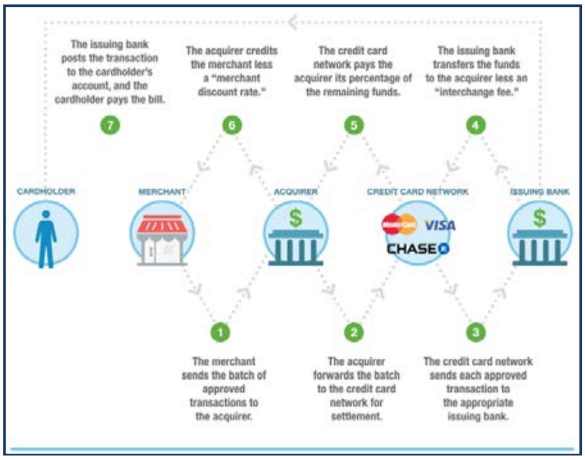
The issuing bank posts the transaction to the cardholder's account, and the cardholder pays the bill.
The acquirer credits the merchant less a "merchant discount rate."
The credit card network pays the acquirer its percentage of the remaining funds.
The issuing bank transfers the funds to the acquirer less an "interchange fee."
7
6
5
4
CARDHOLDER
MERCHANT
ACQUIRER
CREDIT CARD NETWORK
ISSUING BANK
$
VISA
CHASE
$
1
2
3
The merchant sends the batch of approved transactions to the acquirer.
The acquirer forwards the batch to the credit card network for settlement.
The credit card network sends each approved transaction to the appropriate issuing bank.

Figure 1: Clearing and Settlement under Credit Card System

2. LITERATURE REVIEW

Machine Learning (ML) through Ensembles is an important technique that combines outputs from multiple individual classifiers for improving classification accuracy [19, 20]. Theoretical and experimental results suggest that combining classifiers can give effective improvement in accuracy if classifiers within an ensemble are not correlated with each other [21, 22].

3. PROPOSED WORK

The framework proposed in this work is depicted in Figure 2. The proposed framework for prediction works for each transaction and separates the transaction with high or low risk using the method proposed. The proposed predictive model can be further used to generate alerts for transaction with high risks. Investigators check these alerts and provide a feedback for each alert, i.e. true positive (fraud) or false positive (genuine). The proposed model uses suitable pre-processing, attributes selection techniques along with proposed Bagging EM. K-fold cross validation is used as 'test split' method.

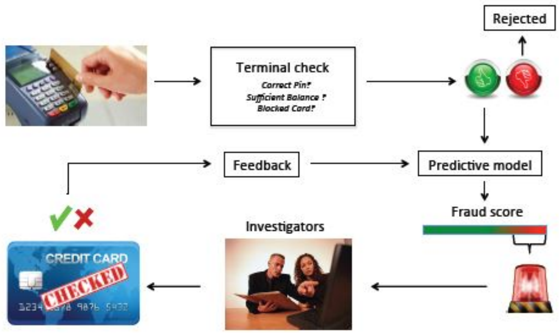

Figure 2: Proposed Model for Credit Card Fraud Detection

4. EXPERIMENTAL SETUP, METHODOLOGY AND PERFORMANCE ANALYSIS

4.1 Experimental Setup

Weka 3.8.1 is used as DM tool for simulation purpose. Weka is installed over Windows 10 Operating System. For this research a state of art research dataset from USCD-FICO competition [7, 8] is used. The competition is hosted by: FICO a leading provider of technologies and University of San Diego. Dataset description is presented in Table 1.

Dataset	No. of Features	Total Instances	No. of Instances (Yes)	No. of Instances (No)
USCD-FICO	20	10,0000	97346 (97.35%)	2654 (2.65%)

Table 1: Details of Dataset

4.2 Methodology

The experiment methodology involves following steps:

1. Implementing and integrating the proposed method with base classifier and building the proposed Classifier / model using the "training" dataset.
2. Applying base classifier to build the model with same training dataset.
3. The evaluation of proposed and base classifier on various metrics is done.
4. The proposed classifier is then compared with benchmark classifiers.

4.3 Performance Analysis

The performance analysis is done on the basis of following metrics:

Prediction Rate: Prediction rate refers to the percentage of correct predictions among all test data.

$$\text{Prediction Rate} = \frac{TP}{TP + TN} * 100$$

False Alarm Rate (FAR): The percentage of normal data which is wrongly recognized as of different class is FAR, and is defined as follows:

$$\text{False Alarm Rate} = \frac{FP}{FP + TN} * 100$$

The performance analysis is shown in Table 2.

Metrics	Naive Bayes	K-Nearest Neighbour	Proposed Method
Prediction Rate	96.9%	96.1%	97.8%
False Alarm Rate	0.68	0.72	0.65

Table 2: Performance Analysis

5. CONCLUSION & FUTURE WORK

Credit Card Frauds are common as attackers gather information through transactions and registered accounts. This opens new confronts in the field of fraud detection and prevention, but prevention is of course better than detection. The simple techniques like database comparison and pattern matching are not enough for detecting such frauds because fraudulent transactions are rare within huge number of genuine transactions. So, Predictive models are of prime importance for banks to detect CCFs. The proposed ensemble based CCFD predictive model is compared with base learner and state-of-art model. The proposed work is compared on basis of two functional metrics: Prediction Rate and FPR proved to be better. The efforts shown that, Ensemble ML methods are more suitable for detecting frauds with credit cards. In future, more efforts methods will be worked out to improve the Fraud Catching Rate. At the same time proposed predictive model would be integrated with live stream to find the online fraudulent transaction instantly. In future we intend to build up a cloud based ML application for detecting frauds in financial transactions done with cards.

6. REFERENCES

[1] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Systems with Applications, vol. 37, no. 8, pp. 6070–6076, Aug. 2010.

[2] Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", IEEE International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 978-1-4244-9394-4/11, 2011 IEEE.

[3] David Opitz and Richard Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of artificial intelligence research 169-198, 1999.

[4] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.

[5] Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm. In Proceedings of the thirteenth international conference on machine learning, Bari, Italy (pp. 148–156).

[6] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259.

[7] Masoumeh Zareapoor, Pourya Shamsolmolia, “Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier”, International Conference on Intelligent Computing, Communication & Convergence, (ICCC 2015), Elsevier, Procedia Computer Science 48 (2015) 679 – 685.

[8]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4180893/

[9] http://www.cs.waikato.ac.nz/~ml/weka/index.html

[10] Weka, University of Waikato, Hamilton,New Zealand.

[11] V.Mareeswari, Dr G. Gunasekaran, “Prevention of Credit Card Fraud Detection based on HSVM”, IEEE, International Conference On Information Communication And Embedded System (ICICES 2016), 978-1-5090-2552-7.

[12] Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic and Bj¨orn Ottersten, “Detecting Credit Card Fraud using Periodic Features”, IEEE 14th International Conference on Machine Learning and Applications, 978-1-5090-0287-0/15, 2015 IEEE.

[13] European Central Bank, “Third report on card fraud,” European Central Bank, Tech. Rep., 2014.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[15] Benson Edwin Raj, A. Annie Portia, “Analysis on Credit Card Fraud Detection Methods”, IEEE International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 978-1-4244-9394-4/11, 2011 IEEE.

[16] Alejandro Correa Bahnsen, Aleksandar Stojanovic, Djamila Aouada and Bj¨orn Ottersten, “Cost Sensitive Credit Card Fraud Detection using Bayes Minimum Risk”, 12th International Conference on Machine Learning and Applications 2013, 978-0-7695-5144-9/13, 2013 IEEE.

[17] Marwan Fahmi, Abeer Hamdy, Khaled Nagati, “Data Mining Techniques for Credit Card Fraud Detection: Empirical Study”, Sustainable Vital Technologies in Engineering & Informatics 2016, Published by Elsevier Ltd.

[18] Wen-Fang YU, Na Wang, “Research on Credit Card Fraud Detection Model Based on Distance Sum”, International Joint Conference on Artificial Intelligence 2009, 978-0-7695-3615-6/09, 2009 IEEE.

[19] T. G. Dietterich, “Machine-learning research: four current directions,” AI Magazine, vol. 18, no. 4, pp. 97–136, 1997.

[20] R. O. Duda, P. H. Hart, and D. G. Stork, Pattern Classification, Wiley-Interscience, New York, NY, USA, 2000.

[21] R. Bryll, R. Gutierrez-Osuna, and F. Quek, “Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets,” Pattern Recognition, vol. 36, no. 6, pp. 1291–1302, 2003.

[22] K. Tumer and N. C. Oza, “Decimated input ensembles for improved generalization,” in Proceedings of the International Joint Conference on Neural Networks (IJCNN '99), pp. 3069–3074, Washington, DC, USA, July 1999.

[23] T. Hastie and R. Tibshirani, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2009.