**RANDOM FOREST FOR CREDIT CARD FRAUD DETECTION**

**ABSTRACT**

Credit card fraud events take place frequently and then result in huge financial losses. Criminals can use some technologies such as Trojan or Phishing to steal the information of other people’s credit cards. Therefore, an effective fraud detection method is important since it can identify a fraud in time whena criminal uses a stolen card to consume. One method is tomake full use of the historical transaction data including normaltransactions and fraud ones to obtain normal/fraud behaviorfeatures based on machine learning techniques, and then utilizethese features to check if a transaction is fraud or not. Inthis paper, random forest is used to train thebehavior features of normal and abnormal transactions. We analyze the performance on creditfraud detection. The data used in our experiments come froman e-commerce company in China.

***Index Terms*—**Random forest, decision tree, credit card fraud

**INTRODUCTION**

Credit cards are widely used due to the popularization of ecommerce and the development of mobile intelligent devices. Card-not-present transactions (i.e., online transaction without a physical card) is more popular, especially all credit card operations are performed by web payment gateways, e.g., PayPal and Alipay. Credit card has made an online transaction easier and more convenient. However, there is a growing trend of transaction frauds resulting in a great losses of money every year. It is estimated that losses are increased yearly at double digit rates by 2020. Since the physical card is not needed in the online transaction environment and the card’s information is enough to complete a payment, it is easier to conduct a fraud than before. Transaction fraud has become a top barrier to the development of e-commerce and has a dramatic influence on the economy. Hence, fraud detection is essential and necessary. Fraud detection is a process of monitoring the transaction behavior of a cardholder in order to detect whether an incoming transaction is done by the cardholder or others. Generally, there are two kinds of methods for fraud detection: misuse detection and anomaly detection. Misuse detection uses classification methods to determine whether an incoming transaction is fraud or not. Usually, such an approach has to know about the existing types of fraud to make models by learning the various fraud patterns. Anomaly detection is to build the profile of normal transaction behavior of a cardholder based on his/her historical transaction data, and decide a newly transaction as a potential fraud if it deviates from the normal transaction behavior. However, an anomaly detection method needs enough successive sample data to characterize the normal transaction behavior of a cardholder. This paper is about misuse method. We use random forest to train the normal and fraud behavior features. Random forest is a classification algorithm based on the votes of all base classifiers.

The major contributions of this paper are summarized as follows.

1) To deal with normal/fraud detection problem, the Random Forest algorithm is used to train the normal/fraud behavior features.

2) From the result of experiments, some conclusions are made which would be helpful for future work.

**HARDWARE & SOFTWARE REQUIRMENT**

# H/W System Configuration:-

# Processor : Dual Core

# Speed : 1.1 G Hz

# RAM : 4 GB (min)

# Hard Disk : 20 GB

# Key Board : Standard Windows Keyboard

# Mouse : Two or Three Button Mouse

**Monitor** : SVGA

# S/W System Configuration:-

# Operating System : Windows XP,7,8,10

# Technology : Python

**Front End** : Tkinter

**IDLE**  : Python 2.7 or 3.7

**Database**  : MySQL

**EXHISTING SYSTEM**

Credit cards are widely used due to the popularization of ecommerce and the development of mobile intelligent devices. Card-not-present transactions (i.e., online transaction without a physical card) is more popular, especially all creditcard operations are performed by web payment gateways, e.g., PayPal and Alipay. Credit card has made an online transactioneasier and more convenient. However, there is a growing trend of transaction frauds resulting in a great losses of money every year.

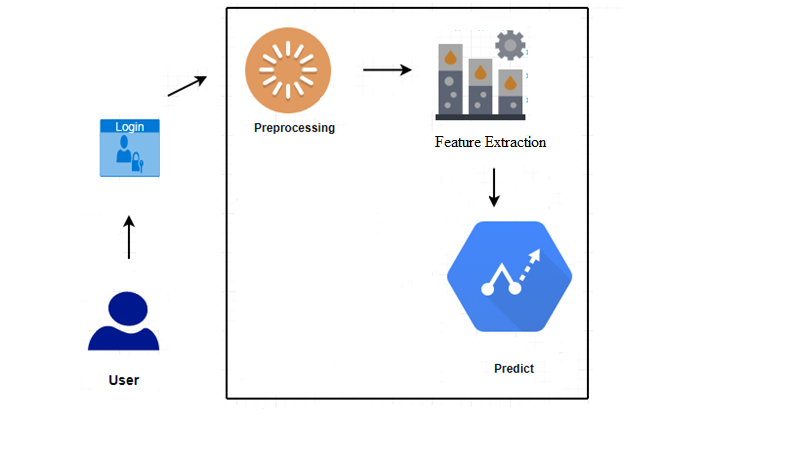
**PROPOSED SYSTEM**

In proposed system, we use misuse method which can ask the computer to find out whether it’s credit card fraud or not. In this story, we used Random Forest algorithm that analyzes and predicts the fraud and non-fraud/valid transactions.

Advantages:

* Performance is good.
* Reduces the time required to predict the output.
* Used for real time predictions of fraud transactions.

**SYSTEM ARCHITECTURE**

****

**CONCLUSION**

This paper has examined the performance of Random Forest model. A real-life B2C dataset on credit card transactions is used in our experiment. Although Random Forest obtains good results on small set data, there are still some problems such as imbalanced data. Our future work will focus on solving these problems. The algorithm of random forest itself should be improved. For example, the voting mechanism assumes that each of base classifiers has equal weight, but some of them may be more important than others. Therefore, we also try to make some improvement for this algorithm.

**REFERENCES**

[1] Gupta, Shalini, and R. Johari. ”A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant.” International Conference on Communication Systems and Network Technologies IEEE, 2011:22-26.

[2] Y. Gmbh and K. G. Co, “Global online payment methods: Full year 2016,” Tech. Rep., 3 2016.

[3] Bolton, Richard J., and J. H. David. ”Unsupervised Profiling Methods for Fraud Detection.” Proc Credit Scoring and Credit Control VII (2001):5–7.

[4] Seyedhossein, Leila, and M. R. Hashemi. ”Mining information from credit card time series for timelier fraud detection.” International Symposium on Telecommunications IEEE, 2011:619-624.

[5] 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.

[6] Drummond, C., and Holte, R. C. (2003). C4.5, class imbalance, and cost sensitivity: why under-sampling beats oversampling. Proc of the Icml Workshop on Learning from Imbalanced Datasets II, 1–8.

[7] Quah, J. T. S., and Sriganesh, M. (2008). Real-time credit card fraud

detection using computational intelligence. Expert Systems with Applications,

35(4), 1721-1732.

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

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

[10] Duman, E., and Ozcelik, M. H. (2011). Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications, 38(10), 13057-13063.

[11] Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C. (2011). Data mining for credit card fraud: a comparative study. Decision Support Systems, 50(3), 602-613.

[12] Sahin, Y., and Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines. Lecture Notes in Engineering and Computer Science, 2188(1).

[13] 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.

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

[15] Ju, W. H., and Vardi, Y. (2001). A hybrid high-order markov chain model for computer intrusion detection. Journal of Computational and Graphical Statistics, 10(2), 277-295.

[16] Bolton, R. J., and Hand, D. J. (2002). Statistical fraud detection: a review. Statistical Science, 17(3), 235-249.

[17] Vlasselaer, V. V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., and Snoeck, M., et al. (2015). Apate : a novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75, 38-48.

[18] Chan, P. K., Fan, W., Prodromidis, A. L., and Stolfo, S. J. (2002).Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and Their Applications, 14(6), 67-74.

[19] RONG-CHANG CHEN, TUNG-SHOU CHEN, and CHIH-CHIANG LIN. (2006). A new binary support vector system for increasing detectionrate of credit card fraud. International Journal of Pattern Recognitionand Artificial Intelligence, 20(02), 227-239.

[20] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[21] Dietterich, T. G. (2000). Ensemble methods in machine learning. , 1857(1), 1-15.

[22] Abeel, T., de Peer, Y. V. and Saeys, Y. Java-ML: A Machine LearningLibrary, Journal of Machine Learning Research, 2009, 10, 931-934

[23] Quinlan, J. R. (1986). Induction on decision tree. Machine Learning, 1(1), 81-106.

[24] Breiman, L., Friedman, J. H., Olshen, R., and Stone, C. J. (1984). Classification and regression trees. Biometrics, 40(3), 358.