**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.

**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.

**LITERATURE SURVEY**

# 1. A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant

Electronic Commerce (e-Commerce) and ease in the onsite transactions have led to the exponential growth in the acceptance of credit cards among consumers of all the sections. But despite their remarkable advantages, consumers are still reluctant in their use, especially for online transactions and reason being the increasing credit card fraud rate. A number of security models have been proposed and deployed for secure online transactions but the sharing of sensitive credit card data over the Internet has made online transactions vulnerable to threats. In this paper, we discuss and analyze the current developments in online authentication procedures including biometrics, one-time-password systems and use of mobile device and Public Switched Telephone Network for cardholder authentication. Then we propose a complete new framework for both onsite and online (Internet shopping) credit card transactions. This framework is more secure, robust, enhances user privacy and does not involve the deployment of special hardware systems at the customer's site.

**2. Random forest for credit card fraud detection**

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 effictive fraud detection method is important since it can identify a fraud in time when a criminal uses a stolen card to consume. One method is to make full use of the historical transaction data including normal transactions and fraud ones to obtain normal/fraud behavior features based on machine learning techniques, and then utilize these features to check if a transaction is fraud or not. In this paper, two kinds of random forests are used to train the behavior features of normal and abnormal transactions. We make a comparison of the two random forests which are different in their base classifiers, and analyze their performance on credit fraud detection. The data used in our experiments come from an e-commerce company in China.

## 3.Unsupervised Profiling Methods for Fraud Detection (2001)

Credit card fraud falls broadly into two categories: behavioural fraud and application fraud. Application fraud occurs when individuals obtain new credit cards from issuing companies using false personal information and then spend as much as possible in a short space of time. However, most credit card fraud is behavioural and occurs when details of legitimate cards have been obtained fraudulently and sales are made on a 'Cardholder Not Present' basis. These sales include telephone sales and e-commerce transactions where only the card details are required. In this paper, we are concerned with detecting behavioural fraud through the analysis of longitudinal data. These data usually consist of credit card transactions over time, but can include other variables, both static and longitudinal. Statistical methods for fraud detection are often classification (supervised) methods that discriminate between known fraudulent and non-fraudulent transactions; however, these methods rely on accurate identification of fraudulent transactions in historical databases -- information that is often in short supply or non-existent. We are particularly interested in unsupervised methods that do not use this information but instead detect changes in behaviour or unusual transactions. We discuss two methods for unsupervised fraud detection in credit data in this paper and apply them to some real data sets. Peer group analysis is a new tool for monitoring behaviour over time in data mining situations. In particular, the tool detects individual accounts that begin to behave in a way distinct from accounts to which they had previously been similar. Each account is selected as a target account and is compared with all other accounts in the database, using either external comparison criteria.

**4. Credit Card Fraud Detection using Time Series Analysis**

Credit card usage has been increased tremendously because of the popularity of E-commerce. As the usage of credit card grows the occurrence of fraudulent transactions also increases, thus comes the stipulation of fraud detection. Detection of fraudulent transaction using credit card plays a vital role in financial institutions. In the proposed work, fraud detection is done with data mining approaches. The parameters considered are transaction amount and transaction time. For every cardholder there is always a robust periodic pattern in the spending behaviour, centered on this fact the anomalies in the transaction are detected by analyzing the past history of transactions belonging to an individual cardholder. In this work two levels of detection methods are used. At the first level the fraud is detected by analyzing whether the new incoming transaction is fraud or not by using distance-based method. At the second level the next transaction is predicted by means of label-prediction methodology and compared with the actual transaction, if there is deviation then it is detected to be a fraudulent transaction. If the particular transaction is considered as a fraud then the cardholder is asked to continue the transaction by asking a secret question, if the cardholder does not give correct answer then the transaction will not be allowed to continue further. The approach used in the proposed work has also decreased the false positive situation and hence it is ensured that genuine transaction is not rejected.

**5.Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering**

Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. Fraudsters are so expert that they engender new ways for committing fraudulent transactions each day which demands constant innovation for its detection techniques as well. Many techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, decision tree, neural network, logistic regression, naïve Bayesian, Bayesian network, metalearning, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A steady indulgent on all these approaches will positively lead to an efficient credit card fraud detection system. This paper presents a survey of various techniques used in credit card fraud detection mechanisms and Hidden Markov Model (HMM) in detail. HMM categorizes card holder’s profile as low, medium and high spending based on their spending behaviour in terms of amount. A set of probabilities for amount of transaction is being assigned to each cardholder. Amount of each incoming transaction is then matched with card owner’s category, if it justifies a predefined threshold value then the transaction is decided to be legitimate else declared as fraudulent. Existing fraud detection system may not be so much capable to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system. In this paper Hidden Markov Model (HMM) is used to model the sequence of operation in credit card transaction processing. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent.

**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) are 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 great losses of money every year.

**Disadvantage:**

1. The main disadvantage of the existing system is the detection occurs only after gets a written complaint.

**PROBLEM STATEMENT:**

Credit card fraud detection is challenging task for the users online payment does not require physical card. And if anyone who knows the details of card can make the transactions currently, card holder comes to know only after the fraud transaction is carried out to proper mechanism are there to track the fraud transaction.

**OBJECTIVES:**

The aim of the project is to find out the normal transactions and fraud ones based on machine learning techniques, and then utilize these features to check if a transaction is fraud or not.

**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.

**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

**METHODOLOGY**

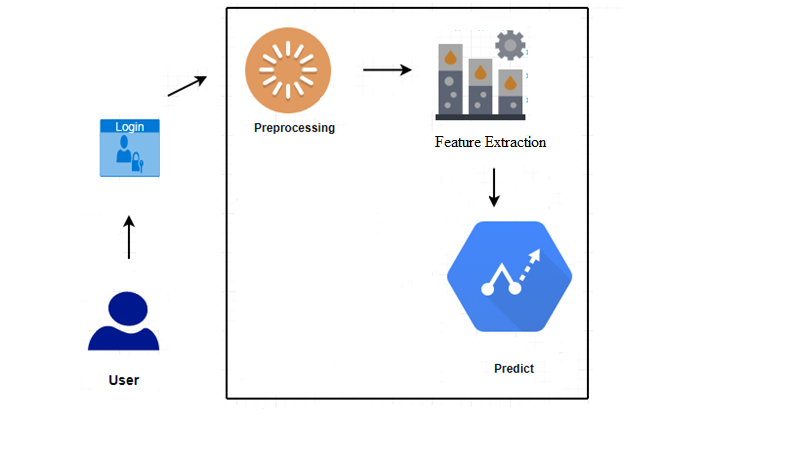
**List of Algorithms:**

1. KNN
2. SVM
3. Random Forest
4. Naïve Bayes

**List of Modules:**

1. Registration Module
2. Login Module
3. Prediction Module

**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.