Real Time Fraud Detection System using Data mining Techniques

K. Nithya #1, E. Lavanya *1

Mailam Engineering College, Mailam #1, *1

Nithyaknov14@gmail.com #1, lavanyakarthikeyanvpm@gmail.com *1

Abstract - Intelligent fraud detection system using random forest algorithm, detects the fraudulent card during transactions and alerts the customer regarding the fraud. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the system efficiency.

I. Introduction

Fraud refereed as to gaining goods/services and money by banned way. Fraud determines with events which involve criminal motives that, mostly, are hard to determine. Credit cards are one of the most popular aim of fraud but not the only one. Credit card fraud, a high range term for theft and fraud devoted or any similar payment mechanism as a fraudulent resource of funds in a transaction. Credit card fraud has been increasing issue in the credit card industry. Finding credit card fraud is a difficult task when using normal process, so the development of the credit card fraud detection models has become of importance whether in the academic or business organizations currently. In recent years, the prevailing data mining concerns people with credit card fraud detection model based on data mining. Since our problem is approached as a classification problem, classical data mining algorithms are not directly applicable. Intelligent fraud detection system using random forest algorithm, detects the fraudulent card during transactions and alerts the customer regarding the fraud. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the

system efficiency. It is documented in such way that, it is convenient to the user. Each section is divided into sub-sections. This chapter gives the information regarding analysis done for the proposed system. Here the goal of the project is explained, and also the cost and performance factors which will affect the feasibility of the project is explained, gets through the functional and nonfunctional requirement phase of the proposed system. This chapter illustrates the overall structure and responsibility of the project using UML. This chapter gets through the requirement phase of the proposed system and studies the requirements of the system in detail. It presents a formal document that crystallizes the user's requirements. The result of this study is being used in all the future steps of development of the project. In this chapter the detailed system design explores architecture of the system. It deals with the modules and their relationship in building the whole system. Design at this level explains about sub systems which are building blocks of the whole system. These sub systems have their well-defined functionality. The coding logic of the tools is explained with the code and syntax. We present how the code is organized with comments on code for understanding in future reference. We also discuss the Naming conventions that were followed during the Implementation phase of the

project and also the descriptions of the methods of all the modules used by the system. This chapter gives the conclusion of the report and also the possible enhancements that could be done in the future.

II.   Related Work

The Traditional detection method mainly depends on database system and the education of customers, which usually are delayed, inaccurate and not in-time. After that methods based on decimate analysis and regression analysis are widely used which can detect fraud by credit rate for cardholders and credit card transaction. For a large amount of data it is not efficient.

2.1 Proposed Method

The proposed system overcomes the above mentioned issue in an efficient way. This proposed system, aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the system efficiency

III.   Literature Review

[1] Describes about, Credit card business has developed quickly in the world in the past years. In order to prevent defaulting risk, some useful tools are used in credit card management. Application Scoring and Behavior Scoring are two important steps in this process, which data mining algorithms are used widely. In this paper, the application of data mining in credit card management are analyzed and the experiments show the difference of data mining in the application scoring and behavior scoring. The amount of issued credit cards has increased rapidly in Taiwan and is characterized as high risk business if comparing to that of the traditional banking loan. To minimize the operating risks and maximize business profits, the credit card issuing institutions need an intelligent system to support the process of the risk management after cards issued. The aim of this study is to construct an efficient risk prediction system to detect the possible defaults for the credit card holders. The system collects the personal and financial information about the credit card holders and then applies evolutional neural network which integrated with grey incidence analysis and Dempster-Shafer theory of evidence to predict the default cases. The experimental results show the integrated model has better prediction accuracy if compare to the model which applies evolutional neural network only and is capable of tracing and reducing the default risks.

[2] Describes about, Credit card fraud on the Internet is a serious and growing issue. Many criminals have hacked into merchant databases to obtain cardholder details enabling them to conduct fake transactions or to sell the details in the digital underground economy. The card brands have set up a standard called PCI DSS to secure credit card details when they are stored online. We investigate the standard and find significant flaws especially in its requirements on small businesses. Finally, we propose some general rules for the secure management of online data. [18] Describes about In this paper, we study how to switch the customers from an undesirable class to a desirable one in credit card churning management by post mining. Multiple Criteria Linear Programming (MCLP) classification model, an optimization-based data mining method, is firstly used to classify the samples. In post mining phase, we build a case base formed by a series of typical positive instances for the entire negative population as their "good examples". These positive instances are on or near the boundary between the two classes, and thus closest to negative objects to ensure lowest switching cost. Switching

plan for each negative object is then generated based on the case base, according to minimum cost principle. Real dataset from a large commercial bank of China is used to validate the method we proposed.

[16] Describes about First, we classify the selected customers into clusters using RFM model to identify high-profit, gold customers. Subsequently, we carry out data mining using association rules algorithm. We measure the similarity, difference and modified difference of mined association rules based on three rules, i.e. emerging pattern rule, unexpected change rule, and added/perished rule. In the meantime, we use rule matching threshold to derive all types of rules and explore the rules with significant change based on the degree of change measured. In this paper, we employ data mining tools and effectively discover the current spending pattern of customers and trends of behavioral change, which allow management to detect in a large database potential changes of customer preference, and provide as early as possible products and services desired by the customers to expand the clientele base and prevent customer attrition.

IV.   Random forest algorithm

We present pseudo-code for the basic algorithm only, without the bounded fringe technique. The addition of a bounded fringe is straightforward, but complicates the presentation significantly. Candidate split dimension A dimension along which a split may be made. Candidate split point One of the first m structure points to arrive in a leaf. Candidate split A combination of a candidate split dimension and a position along that dimension to split. These are formed by projecting each candidate split point into each candidate split dimension. Candidate children Each candidate split in a leaf induces two candidate children for that leaf. These are also referred to as the left and right child of that split.

V.   Modules

5.1 Registration of user details

In this module, user have to register their detail first, the details are user id, name, date of birth, account number, balance and transaction details, and address.

5.2 Login of user

In this module, user have login with their own id password and pin number, then only he or she will access the process. The user details are maintained in the database for future references.

5.3 Brute force attack for password guessing

In this module describes about, it will show alert when the unauthorized person access the account or something wrong about account.

VI.   Architecture

The above architecture describes the work structure of the system.The customer data in the data warehouse is subjected to the rules engine which consists of the fraud rule set. The filter and priority module sets the priority for the data and then sends it to the genetic algorithm which performs its functions and generates the output.
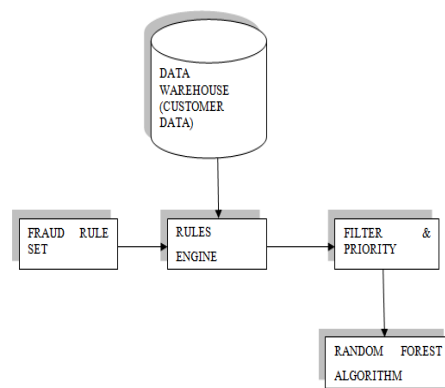
VII. Conclusion

From this, intelligent fraud detection system using random forest algorithm has been implemented. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. In future, the system will achieve more efficiency by improving the performances of the algorithm.

References

[1] Aihua Li , 1.Study on the Application of Data Mining Algorithms in Credit Card Management, Author(s) Aihua Li, 2009

[2]Blackwell.C,The management of online credit card data using the Payment Card Industry Data Security Standard, 2008

[3] Credit card fraud detection using hidden markov model – Abinav Srivastava,Amlan Kundu,Shamik Sural,Arun K.majumdar

[4] clifton phua, vincent lee1, kate smith & ross gayler, A Comprehensive Survey of Data Mining-based Fraud Detection Research,2005.

[5] Elio Lozano, Edgar Acu˜na, Parallel algorithms for distance-based and density-based outliers,2006.

[6] Leila Seyedhossein, Mahmoud Reza Hashemi Mining Information from Credit Card Time Series for Timelier Fraud Detection International Symposium on Telecommunications 2010.

[7] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), "Credit card fraud and detection techniques: a review"2009.

[8] Md Delwar Hussain Mahdi, Karim Mohammed Rezaul, Muhammad Azizur Rahman "Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business." 2010.

[9] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, Improving a credit

card fraud detection system using genetic algorithm, International conference on Networking and information technology 2010.

[10] Mirjana Pejic-Bach, "Profiling intelligent systems applications in fraud detection and prevention: survey of research articles", 2010.

[11] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection, 2008.

[12] Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network" 2011.

[13] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods" 2011.

[14] Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010).

[15] Tej Paul Bhatla, Vikram Prabhu & Amit Dua "Understanding Credit Card Frauds," 2003

[16]Ruey-Chyi Wu, Data mining application in customer relationship management of credit card business, 2005

[17] Wen-Fang YU, Na Wang, Research on Credit Card Fraud Detection Model Based on Distance Sum, IEEE International Joint Conference on Artificial Intelligence 2009.

[18]Yibing Chen , Post Mining of Multiple Criteria Linear Programming Classification Model for Actionable Knowledge in Credit Card Churning Management, 2011