ENGINEERING IN ADVANCED RESEARCHSCIENCE AND TECHNOLOGY

ISSN 2352-8648
Vol.03, Issue.01
May-2022
Pages: -273-277

A SUPERVISED LEARNING TECHNIQUE TO DETECT CREDIT CARD FRAUD DETECTION

[1]GANISETTI TULASI, [2]G TATAYYANAIDU

[1]M. Tech., Dept. of CSE, Prasiddha College of Engineering Technology, AMALAPURAM, AP, India.

[2]Associate Professor, Dept. of CSE, Prasiddha College of Engineering Technology, AP, India

Abstract:- Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data, the highly unbalanced classes distributions and the availability of few transactions labeled by fraud investigators. At the same time public data are scarcely available for confidentiality issues, leaving unanswered many questions about what is the best strategy. In this thesis we aim to provide some answers by focusing on crucial issues such as: i) why and how under sampling is useful in the presence of class imbalance (i.e. frauds are a small percentage of the transactions), ii) how to deal with unbalanced and evolving data streams (non-stationarity due to fraud evolution and change of spending behavior), iii) how to assess performances in a way which is relevant for detection and iv) how to use feedbacks provided by investigators on the fraud alerts generated. Finally, we design and assess a prototype of a Fraud Detection System able to meet real-world working conditions and that is able to integrate investigators' feedback to generate accurate alerts.

Keywords:- Random forest, decision tree, credit card fraud

INTRODUCTION:-

The online shopping growing day to day Credit cards are used for purchasing goods and services with the help of virtual card and physical card where as virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system Banks collect a lot of historical records corresponding to millions of customer's transactions. They are credit card and debit card operations, but unfortunately, only a small portion, if any, is open access. Fraud detection is a critical problem affecting large financial companies that have increased due to the growth in credit card transactions. The proposed method consists of the Predictive modeling and Logistic Regression. Now a day's bank transactions as well as credit card frauds increased. One of the most target frauds are credit card fraud, the fraud can occur any type of credit products, such products are retail, home loan and personal loan. During the last few decades as technology has changed, dramatically the face of fraud also changed. To detect credit card fraud, data mining techniques- Predictive modeling and Logistic Regression are used. In prediction model to predict the continuous valued functions. Credit card of CSV files will be analyzed to predict the outcome. In this paper, we propose to detect credit card transaction using available data set and data mining techniques of predictive modelling, Decision tree, and Logistic Regression. Predictive modeling splits the data into two partitions 70% of testing and 30%of training check output class distribution to predict the outcome. The decision tree to get the result as a tree with root node describes the best predictor in the data, the combination of two or more branches is denoted by decision node (non leaf nodes) and each branch represents a value for the attribute which is tested. The leaf node may be 1 in the case of fraud and 0 otherwise. Logistic regression or logistic model is a regression model, where the dependent variable is categorical of a linear generalized model

Related Work:-

A comprehensive understanding of fraud detection technologies can be helpful for us to solve the problem of credit card fraud. The work in [16] provides a comprehensive discussion on the challenges and problems of fraud detection research. Mohammad et.al., [14] review the most popular types of credit card fraud and the existing nature-inspired detection methods that are used in detection methods. Basically, there are two types of credit card fraud: application fraud and behavior fraud [3]. Application fraud is that criminals get new credit cards from issuing companies by forging false information or using

other legitimate cardholders information. Behavior fraud is that criminals steal the account and password of a card from the genuine cardholder and use them to spend. Recently, a kind of fraud detection method is popular in some commercial banks which is to check behaviors of the associated cardholder [7]. Almost all the existing work about detection of credit card fraud is to capture the behavior patterns of the cardholder and to detect the fraud transactions based on these patterns. Srivastava et.al. [5] model the sequence of transaction features in credit card transaction processing using a hidden markov model (HMM) and demonstrate its effectiveness on the detection of frauds. An HMM is initially trained with the normal behavior of the cardholder. If the current transaction is not accepted by the trained HMM with a high probability, it is considered to be fraudulent. However, they only consider the transaction amount as the feature in the transaction process. Amlan et.al [8] propose a method using two-stage sequence alignment which combines both misuse detection and anomaly detection [15]. In their method, a profile analyzer is used to determine the similarity of an incoming sequence of transaction on a given credit card with the legitimate cardholder's past spending sequence. Then, the unusual transactions traced by the profile analyzer are passed to a deviation analyzer for possible alignment with the past fraudulent behavior. The final decision about the nature of a transaction is taken on the basis of the observations by the two analyzers. However, this method cannot detect frauds in real time. Elaine et.al. [9] propose a user behavior model which treats the transaction features independently. Gabriel et.al [13] propose an alternative method to prevent fraud in E-commerce applications, using a signature-based method to establish a user's behavior deviations and consequently detect the potential fraud situations in time. However they only consider the click stream as the element of the signature. We believe that instead of using only one transaction feature for a fraud detection, it is better to consider multiple transaction features

Literature Survey:-

Automatic credit card fraud detection based on non-linear signal processing.

Fraud detection is a critical problem affecting large financial companies that has increased due to the growth in credit card transactions. This paper presents a new method for automatic detection of frauds in credit card transactions based on non-linear signal processing. The proposed method consists of the following stages: feature extraction, training and classification, decision fusion, and result presentation. Discriminate-based classifiers and an advanced non-Gaussian mixture classification method are employed to distinguish between legitimate and fraudulent transactions. The posterior probabilities produced by classifiers are fused by means of order statistical digital filters. Results from data mining of a large database of real transactions are presented. The feasibility of the proposed method is demonstrated for several datasets using parameters derived from receiver characteristic operating analysis and key performance indicators of the business.

Authors:- Delamaire, Linda, H. A. H. Abdou, and John Pointon

Analysis on credit card fraud detection methods

Due to the rise and rapid growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In real life, fraudulent transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A clear understanding on all these approaches will certainly 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 evaluates each methodology based on certain design criteria.

Authors: - S. B. E. Raj and A. A. Portia

On combining classifiers

We develop a common theoretical framework for combining classifiers which use distinct pattern representations and show that many existing schemes can be considered as special cases of compound classification where all the pattern representations are used jointly to make a decision. An experimental comparison of various classifier combination schemes demonstrates that the combination rule developed under the most restrictive assumptions-the sum rule-outperforms other classifier combinations schemes. A sensitivity analysis of the various schemes to estimation errors is carried out to show that this finding can be justified theoretically.

Authors: - J. K. Kittler, M. Hatef, R. P. W. Duin, and J. Matas,

Generalized Hammerstein–Wiener system estimation and a benchmark application

This paper examines the use of a so-called "generalized Hammerstein–Wiener" model structure that is formed as the concatenation of an arbitrary number of Hammerstein systems. The latter are taken here to be memory less non-linearity's followed by linear time invariant dynamics. Hammerstein, Wiener, Hammerstein–Wiener and Wiener–Hammerstein models are all special cases of this structure. The parameter estimation of this model is investigated using a standard prediction error criterion coupled with a robust gradient based search algorithm. This approach is profiled using a Wiener–Hammerstein Benchmark example, which illustrates it to be effective and, via Monte-Carlo simulation, relatively robust against capture in local minima.

Authors:- A. Wills and B. Ninness

SYSTEM ARCHITECTURE

Data description :-

The dataset was created combining two data sources; the fraud transactions log file and all transactions log file. The fraud transactions log file holds all the online credit card fraud occurrences while all transactions log file holds all transactions stored by the corresponding bank within a specified time period. Due to the confidential disclosure agreement made between the bank and the authors of the paper, some of the sensitive attributes such as card number were hashed. When evaluating the combined dataset, the shape of the data was much skewed due to the imbalanced numbers of legitimate transactions and fraudulent occurrences.

Data preparation:

Collected raw data were first divided into 4 data sets according to its fraud pattern. This process was done with the information gained by the bank. The four datasets are,

1. Transactions with Risky Merchant Category Code (MCC).
2. Transactions larger than $100.
3. Transactions with risky ISO Response code.
4. Transactions with unknown web addresses.

Those 4 datasets were used in two different ways.

1. By transforming raw data into a numerical form. (Type A)
2. By preparing raw data categorically without making any transformation. (Type B)

Data Cleaning –

Filling in missing values is an important task in the data cleaning process. There are many ways to overcome this issue, such as ignoring the whole tuple, but most of them are likely to bias the data. Since the source file which contained genuine transactions did not contain records with missing values, filling them was no more an issue. Tuples with meaningless value were removed from the files as they do not contribute to producing important data as well as they do not bias the data. Additionally, following changes such as removing unnecessary columns, separating the date time column into two x

Data Integration - Before the data were subjected to further change the two data sources were integrated together since fraudulent and genuine record files were in two separate files. Figure 1 shows how the mapping process was done.

Data Reduction –

The strategy used for this is Dimension reduction. We must prevent the risk of learning wrong patterns of data and the selected features should eliminate the irrelevant aspects and qualities of the fraud domain [10]. The principal component analysis which is well-known that PCA is a popular transform method Applying this method resolves the feature selection issue from the perspective of numerical analysis. PCA performed feature selection successfully by finding the suitable number of principal components. In type B, data cleaning and data integration were involved as same as in type A. Then those data were taken to the next step of the process.

Resampling Techniques.

The two data sources were characterized by a highly imbalanced distribution of examples among the classes. Fraudulent transactions contained a much smaller number of examples than the genuine transactions. To overcome this, we conducted under-sampling and over-sampling by reducing the majority occurrences and by raising the minority occurrences respectively. For over-sampling, Synthetic Minority Oversampling Techniques (SMOTE) and for under-sampling, condensed nearest neighbour (CNN) and random under-sampling (RUS) were used. The minority class is over-VDPSOHG E\ SURGXFLQJ ³V\QWKHWLF´ examples in SMOTE method [20]. Out of RUS and CNN, RUS is a non-heuristic method which balances the class distribution by using a method to eliminate random majority class examples [21] [22].

Modelling and testing:-

Our study analyses four different fraud patterns. For analyzing each pattern, we have reflected the following process as described in figure 2. Quite a few numbers of techniques were used in the data analysis. Four machine learning algorithms were prioritized in our analysis with the help of the literature. They are Support Vector Machine, Naive Bayes, K-Nearest Neighbor and Logistic Regression. We applied those selected supervised learning classifiers to our resampled data. When selecting machine learning models which can capture each fraud, the accuracy and performance of each model were taken into consideration. Optimal models were selected by filtering them out comparatively against an appropriate performance matrix

Real time Fraud Detection:-

In the past, fraud detection has been done by taking already happened transactions in bulk and applying machine learning models on them. Since the results can be seen after weeks or months, tracking down of detected frauds was found extremely difficult, and there have been many cases where the fraudsters were able to commit many more fraudulent purchases before being exposed. Real-time fraud detection is the execution of fraud detection models the second an online purchase is taken place. That way our system is capable of detecting frauds real-time. It gives an alert to the bank indicating its fraud pattern and accuracy rate, making it easy for fraud monitoring teams to move into their next action without having to waste their time and money.

Fraud Detection System.:- Real-time detection of credit card fraud can be stated as one of the main contributions of this project. The real-time fraud detection system consists of three main units; API MODULE, FRAUD DETECTION MODELS and DATA WAREHOUSE. All the components are involved in fraud detection simultaneously. Fraudulent transactions are being classified into four fraud types (Frauds occur due to Risky MCC, ISO-Response Code, Unknown web address, Transaction above 100$) using three supervised learning classifiers. API module is responsible for transferring real time transactions between the Fraud detection model, GUI, and Data warehouse. A Data Warehouse has been used for storing live transactions, the predicted results and other important data of the machine learning models. The user can interact with the fraud detection system with GUIs where it shows the real time transactions, alerts regarding frauds and historical data regarding frauds in a graphical representation. When a transaction is recognized as fraudulent by the fraud detection model, a message will be sent to the API module. Then the API module will notify the end user by sending a notification and the feedback.

Conclusion:-

This process is used to detect the credit card transaction, which are fraudulent or genuine. Data mining techniques of Predictive modeling, Decision trees and Logistic Regression are used to predict the fraudulent or genuine credit card transaction. In predictive modeling to detect and check output class distribution. The prediction model predicts continuous valued functions. We have to detect 148 may be fraud and other are genuine. In decision tree generate a tree with root node, decision node and leaf nodes. The leaf node may be 1 becomes fraud and 0 otherwise. Logistic Regression is same as linear regression but interpret curve is different. To generalize the linear regression model, when dependant variable is categorical and analyzes relationship between multiple independent variables

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 detection rate of credit card fraud. International Journal of Pattern Recognition
and 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 Learning Library, 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.