International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, Mar-Apr-2021, ISSN (Online): 2395-566X

Analysis of Credit Card Fraud Detection in Data Mining using Various Classifier Techniques

Research Scholar Sachin Jain, Professor Dr. Rohit Kumar Singhal
Institute of Engineering and Technology, Alwar

Abstract- The data mining is the technique which can mine useful information from the rough data. The prediction analysis is the technique of data mining which can predict new things from the current data. The classifications techniques are generally applied for the prediction analysis. This research work is based on the prediction of the credit card fraud detection. The various techniques are proposed by the authors for the credit card fraud detection. The technique which is proposed the study in the different paper is based on the conventional neural networks in which system learns from the previous experiences and drive new values.

Keywords- credit card fraud detection, prediction analysis etc.

I. INTRODUCTION

There is huge amount of data generated each day and the handling such large amount of data is very difficult. The generated data is stored in huge databases which can be retrieved as per the user. There are large sized repositories and databases generated in which the data can be stored. However, the retrieval of important data from such large databases is a major concern. There are numerous tools presented which can help in extracting useful information from the databases as per the requirement of users.

Researchers have proposed various machine learning, pattern recognition, neural networks, and various other techniques in this field. The mechanism through which the data can be stored and extracted efficiently as per the requirement is known as data mining. Knowledge discovery is the other name for data mining. The implicit data is extracted from huge databases which can be utilized further as important information through the data mining technique.

Data mining is referred to as the part of KDD even though they are considered to be very similar to each other. There are few steps followed in KDD in order to extract the raw data from the databases and covert into useful information [1]. Following are the steps followed:

- Data Cleaning: This step includes the removal of extra irrelevant data and any kinds of noise from the gathered data.
- Data integration: Within a common source, the multiple data sources are combined at this step.
- Data selection: The data that is in relevance to analysis is identified and the retrieved from the databases in this step.
- Data transformation: The transformation of selected data into a form that is required for mining purpose is known as data transformation or data consolidation.
- Data mining: The patterns that are useful can be extracted with the help of various techniques within this step.
- Pattern evaluation: On the basis of given measures, the interesting patterns are identified within this step.
- Knowledge representation: The knowledge is discovered are presented to the user in visual manner through this step. The data mining results can be understood and interpreted through the help of visualization techniques in this important step.

II.APPLICATIONS OF DATA MINING

Data mining has been utilized successfully in large number of companies. The companies that were involved here at first were mainly the information-intensive industries including the financial services as well as direct mail marketing. Now, a large data warehouse is managed as per the relationships with the users. For the success of data mining, there are two important factors which are; a huge, properly integrated data warehousing and the properly defined understanding of the business process as per which the data mining is applied.

III. CREDIT CARD FRAUD DETECTION

In order to purchase various goods and services in our day to day lives, the credit cards have been used. The cardholder can give his card to the merchant in order to make successful payment in case of physical-card based purchasing method. The credit card needs to be stolen by the attacker in order to conduct the fraudulent transactions. The credit card company can face huge losses in case when the cardholder does not realize the loss of card. In order to make any fraudulent transaction

© 2021 IJSRET

in online payment, the attacker only needs to know very little information.

The Internet or telephone is utilized in order to make transactions to buy something online. There are many times when the activity of stealing a card is not known by the user and the spending patterns are the only way through which this can be known. The rate of successful credit card frauds can be minimized by providing a fraud detection method that is made on the basis of analyzing the existing purchase data of the cardholder [14].

1. Credit Card Operation

The credit card operation consists of the steps as follows.

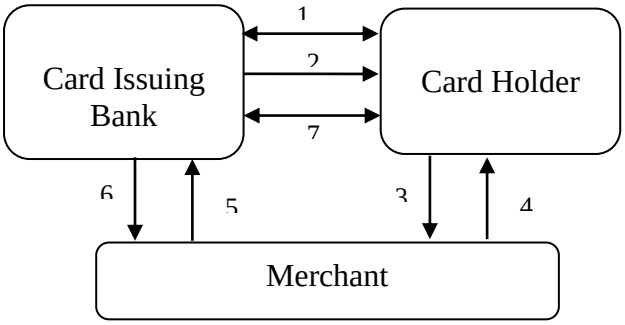

Figure 1. Credit Card Operation Process

- Contract for credit card: there is a contract between cardholder and the cared issuing bank regarding limit etc.
- Card issue: Once the contract is finished, the bank issues the credit card to their customer.
- Purchasing goods: A Cardholder purchases goods/services and offers the credit card.
- Deliver goods: A merchant establishment delivers goods once taking a valid credit card and noting the number and taking signatures.
- Raising of bill: The merchant establishment raises the bill for the purchase and sends it to the credit card issuing bank for payment.
- Payment for bill: The issuing bank pays the amount to the merchant establishment.
- Payment of Credit card: The issuing bank raises bill on the credit cardholder and sends it for payment. The credit cardholder then pays the amount to the issuing bank.

The credit card fraud detection is applied because of following reasons:

- In comparison to the existing system, there is high speed of identification of fraud utilization of card.
- Since, a virtual card is provided to the user till there is new card issued to the user; the transactions of the account holder cannot be halted.
- When the user realizes about the stealing of his card, he can block it on his own only.
- The virtual ID and password can be attained in secure manner only with the help of ONE TIMEPASSWORD type of security which is provided by this system.
- The identification mechanism of this technique is highly accurate.

2. Credit Card Risks

Credit Card Fraud is one of the biggest threats to business institutions these days. However, to combat the fraud effectively, it is vital to initial perceive the mechanisms of executing a fraud. Credit card fraudsters use a large variety of modus operandi to commit fraud [15]. In simple terms, credit card Fraud is outlined as: When a private uses another individuals' credit card for personal reasons whereas the owner of the card and also the card issuer are not aware of the actual fact that the card is getting used. Further, the individual mistreatment the card has no relation to the cardholder or establishment, and has no intention of either contacting the owner of the card or making repayments for the purchases created. Fraud detection systems are prune to several difficulties and challenges enumerated bellow. An effective fraud detection technique should have abilities to address these difficulties in order to achieve best performance.

- Imbalanced data: The credit card fraud detection data has imbalanced nature. It means that very small percentages of all credit card transactions are fraudulent. This cause the detection of fraud transactions very difficult and imprecise.
- Different misclassification importance: in fraud detection task, different misclassification errors have different importance. Misclassification of a normal transaction as fraud is not as harmful as detecting a fraud transaction as normal. Because in the first case the mistake in classification will be identified in further investigations.
- Overlapping data: many transactions may be considered fraudulent, while actually they are normal (false positive) and reversely, a fraudulent transaction may also seem to be legitimate (false negative). Hence obtaining low rate of false positive and false negative is a key challenge of fraud detection systems.
- Lack of adaptability: classification algorithms are usually faced with the problem of detecting new types of normal or fraudulent patterns. The supervised and unsupervised fraud detection systems are inefficient in detecting new patterns of normal and fraud behaviors, respectively.
- Fraud detection cost: The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it. For example, no revenue is obtained by stopping a fraudulent transaction of a few dollars.

III.LITERATURE REVIEW

© 2021 IJSRET

Sahil Dhankhad, et.al, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", 2018 The goal of data analytics is to delineate hidden patterns and use them to support informed decisions in a variety of situations. Credit card fraud is escalating significantly with the advancement of modernized technology and became an easy target for frauds [26]. Credit card fraud has highly imbalanced publicly available datasets. In this paper, many supervised machine learning algorithms are applied to detect credit card fraudulent transactions using a real-world dataset. Furthermore, these algorithms are applied to implement a super classifier using ensemble learning methods.

The most important variables that may lead to higher accuracy in credit card fraudulent transaction detection are identified. Additionally, the performance of various supervised machine learning algorithms that exist in literature was compared and discussed against the super classifier that was implemented in this paper.

Shiyang Xuan, et.al, "Random Forest for Credit Card Fraud Detection", 2018 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 when a criminal uses a stolen card to consume [27]. 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.

A comparison of the two random forests was made, which are different in their base classifiers and their performance was analyzed on credit fraud detection. The data used in our experiments come from an e-commerce company in China. This paper has examined the performance of two kinds of random forest models. 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.

Lutao zheng, et.al, "A New Credit Card Fraud Detecting Method Based on Behavior Certificate", 2018 In this paper, a novel credit card fraud detection system is proposed based on the cardholder's transaction records. What is different in our approach is that four attributes in a transaction and thirteen derived features are used as the behavior features, not just one unique attribute like the amount or time is considered [28]. Besides, we consider the influence of the special situation which will influence the transaction habit of cardholders such as festival and weekend. In addition, experiments show that accuracy of our method is above 90 percent over various input datasets with different fraud rate. Comparative experiments reveal that our method is better than the FDS with support vector machines. In the future we plan to consider more behavioral features such as the information of device or the relation of cardholders in order to make our method more effective.

Abhimanyu Roy, et.al, "Deep Learning Detecting Fraud in Credit Card Transactions", 2018 The analysis shows that there is a significant temporal component. The LSTM and GRU model significantly outperformed the baseline ANN which indicates that order of transactions for an account contains useful information in differentiating between fraud and non-fraudulent transactions. This suggests that that computational resources may be best spent training a larger network [29]. In our testing, performance improved whenever network size was increased. Further research is needed to determine when model performances cease to improve with network size increases. Further research could also be directed towards assessing model sensitivity to hyperparameters not included in our initial grid search such as momentum, batch size, number of epochs, and dropout rate. Our analysis held the number of neurons constant for each hidden layer. Varying the number of neurons in each layer may also reveal additional insight into the effect of network size on model performance.

Adrian Mead, et.al, "Detecting Fraud in Adversarial Environments: A Reinforcement Learning Approach", 2018 Credit card fraud is a costly problem for banks and a major frustration for consumers. As such, static models to detect fraud that rely on supervised training are exposed to the risk of being learned and circumvented. Previous adversarial learning work in fraud prevention showed increased effectiveness over static models that did not account for changing fraudster behavior Tyler Cody, et.al,

"A Utilitarian Approach to Adversarial Learning in Credit Card Fraud Detection", 2018 Credit card fraud detection can be modeled as an adversarial game between a fraudster and the fraud detection mechanism. Previous work uses a game theoretical adversarial learning approach to model the most successful strategy and preemptively adapt the fraud detection system. The game consists of the adversary's selection of strategy and the fraud detection system's decision to retrain. In the previous work, the detection system is retrained every round of the game. In application, there may be costs and risks associated with training and deploying a new model. Thus, it may be desirable to optimize the decision of whether to retrain the model based on the expected economic disutility [31].

© 2021 IJSRET

Lutao Zheng, et.al, "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity", 2018 In this paper, a logical graph of BP (LGBP) is proposed which is a total order-based model to represent the logical relation of attributes of transaction records. Based on LGBP and users' transaction records, one can compute a path-based transition probability from an attribute to another one [33]. At the same time, an information entropy-based diversity coefficient is defined in order to characterize the diversity of transaction behaviors of a user. In addition, a state transition probability matrix is defined to capture temporal features of transactions of a user. Consequently, a BP is constructed for each user and then used it to verify if an incoming transaction is a fraud or not. Our experiments over a real data set illustrate that our method is better than three state-of-the-art ones.

1. Credit Card Fraud Detection Techniques

The credit card fraud detection techniques are classified in two general categories: fraud analysis (misuse detection) and user behavior analysis (anomaly detection).

The first group of techniques deals with supervised classification task in transaction level. In these methods, transactions are labeled as fraudulent or normal based on previous historical data.

The second approach deals with unsupervised methodologies which are based on account behavior. In this method a transaction is detected fraudulent if it is in contrast with user's normal behavior. This is because we don't expect fraudsters behave the same as the account owner or be aware of the behavior model of the owner. To this aim, we need to extract the legitimate user behavioral model for each account and then detect fraudulent activities according to it. Comparing new behaviors with this model, different enough activities are distinguished as frauds. The profiles may contain the activity information of the account; such as merchant types, amount, location and time of transactions. This method is also known as anomaly detection.

2.Problem Formulation

The data mining is the approach which can predict the future from the existing datasets. This prediction analysis is the combination of the clustering and classification. The clustering is the approach which can group similar and dissimilar type of data. The classification is the approach which can assign the class to the grouped data. This research work is based on the prediction analysis of the risky credit information. The credit information is the type of information which is assigned to users by the bank and it is useful to take loans from the bank. In the base paper, performance of various classifiers like random forest, linear regression, SVM, adaptive boosting, decision tree and neural networks is compared for the credit score classification. The other classifier is required among thedefined for the credit information classification.

3. Objectives

Following are the various objectives of this research work:-

- Implement various classifiers for the credit score information classification in data mining
- To proposed naïve bayes classifier for the classification of credit score information in data mining .
- Implement proposed algorithm and compare with other classifiers in terms of accuracy, execution time.

IV.RESEARCH METHODOLOGY

Due to advancement in social web-sites, e-commerce, business modelling and various enormous technologies, various huge datasets are available which can be used for predictive modeling.. Advancement in communication and information, academics stakeholders are witnessing keen interest in predictive modelling of datasets. Predictive modelling analysis is performed to help academics organizations to drive innovation by acquiring new insight into their customers**.** Each application that is running today requires large size of databases such that huge amount of data that is being generated within them can be stored in them. There is a need to generate new and highly efficient tools which can manage such huge databases and the data present within them. For facilitating the large number of users as well as their applications several techniques have been proposed by researchers. Amongst such techniques is the data mining and knowledge discovery approach that is used in order to manage all such data. Data mining is known as the automated approach of identifying interesting patterns from huge databases**.** This technique helps in generating descriptive, understandable and predictive models from the existing data within the databases.

Naive bayes classifier depends on probabilistic classifier demonstrate that join solid autonomy suspicions which don't have affect on the truth that is the reason these are taken as Naive. This approach utilizes the ordinary circulation to show numeric properties. It likewise dealing with the numeric properties utilizing directed discretizatio **.** Naive bayes is a strategy for creating characterization display that appoints class names to issue occurrences, pointer as vectors of highlight esteems, where the class marks are venture from some limited set. This approach is not for preparing such classifier yet in addition gathering of calculation rely upon normal primary, all guileless bayes classifier continue that the estimation of any component is autonomous of estimation of some other element for given class marks. Leverage of naive bayes classifier is it require a little measure of preparing information to appraise the parameters basic for classification.

The credit card fraud detection algorithm and data classification methods were evaluated by conducting a detailed literature study in the first step. This was done to

© 2021 IJSRET

know their advantages and limitations. The limitation of the existing schemes is overcome by the design of proposed security solution that improves the structure of this method.

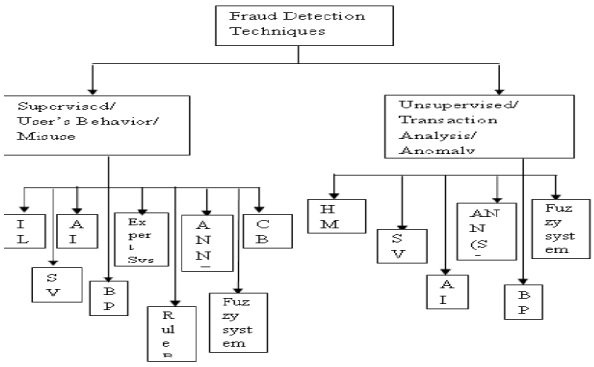

Figure2 : Complete Classification of Credit Card Fraud Detection Techniques.

It is important to highlight the key differences between user behavior analysis and fraud analysis approaches. The fraud analysis method can detect known fraud tricks, with a low false positive rate. These systems extract the signature and model of fraud tricks presented in oracle dataset and can then easily determine exactly which frauds, the system is currently experiencing [22]. If the test data does not contain any fraud signatures, no alarm is raised.

1. Artificial Neural Network

An artificial neural network (ANN) is a set of interconnected nodes designed to imitate the functioning of the human brain [9]. Each node has a weighted connection to several other nodes in adjacent layers. Individual nodes take the input received from connected nodes and use the weights together with a simple function to compute output values. Neural networks come in many shapes and architectures.

The Neural network architecture, including the number of hidden layers, the number of nodes within a specific hidden layer and their connectivity, most be specified by user based on the complexity of the problem. ANNs can be configured by supervised, unsupervised or hybrid learning methods.

2. Genetic Algorithm (GA)

Inspired from natural evolution, Genetic algorithms (GA), were originally introduced by John Holland. GA searches for optimum solution with a population of candidate solutions that are traditionally represented in the form of binary strings called chromosomes. The basic idea is that the stronger members of the population have more chance to survive and reproduce. The strength of a solution is its capability to solve the underlying problem which is indicated by fitness. New generation is selected in proportion to fitness among previous population and newly created offspring.

3. Hidden Markov Model (HMM)

A Hidden Markov Model is a double embedded stochastic process which is applied to model much more complicated stochastic processes as compared to a traditional Markov model. The underlying system is assumed to be a Markov process with unobserved states. In simpler Markov models like Markov chains, states are definite transition probabilities are only unknown parameters. In contrast, the states of a HMM are hidden, but state dependent outputs are visible. In credit card fraud detection a HMM is trained for modeling the normal behavior encoded in user profiles. According to this model, a new incoming transaction will be classified to fraud if it is not accepted by model with sufficiently high probability. Each user profile contains a set of information about last 10 transactions of that user liketime; category and amount of for each transaction. HMM produces high false positive rate.

V.CONCLUSION

In this work, it is concluded that data mining is the technique which can mine the useful information from the large amount of data. The credit card fraud detection is the application of prediction analysis. In the prediction analysis, future data is predicted from the current information. In the existing technique, neural networks is applied which can learn from the input data and drive the future values. In this work, classification technique is applied which can classify normal and fraud transactions. The naïve bayes classifier is applied in this work which can divided whole data into test and training sets. The test and training sets are taken as input and drive the future values. The performance of proposed modal is tested in terms of accuracy. The proposed technique shows high accuracy as compared to existing technique.

Future Work

1. In future proposed technique can be further compared with other technique of credit card fraud detection to test its reliability
2. The proposed technique can be further improved using the other classifiers like decision tree.

REFERENCES

[1] Ekrem Duman, M. Hamdi Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search", Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).

[2] S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011

[3] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", International Multiconference of Engineers and computer scientists March, 2011.

© 2021 IJSRET

[4] S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods," IEEE-International Conference on Computer, Communication and Electrical Technology; (2011). (152-156).

[5] Anshul Singh and Devesh Narayan, "A Survey on Hidden Markov Model for Credit Card Fraud Detection" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012

[6] S. Esakkiraj and S. Chidambaram, et al. "A Predictive Approach for Fraud Detection Using Hidden Markov Model", International Journal of Engineering Research & Technology (IJERT) ", Vol 2, January 2013

[7] Han, J., Kamber, M., & Pei, J. "Data mining: concepts and techniques: concepts and techniques. Elsevier (2011).

[8] Homayounfar, E., Sepehri, M. M., Hasheminejad, S. M. H., & Ghobakhloo, M. "Designing a chronological based framework for condition monitoring in heart disease patients-a data mining approach (DM-PTTD)". Iranian Journal of Medical Informatics, vol. 3, no. 3, pp. 1-6. (2014).

[9] Sebastiani, F. "Machine learning in automated text categorization ACM computing surveys (CSUR)", vol. 34, no. 1, pp. 1-47 (2002)

[10] Hasheminejad, S. M. H., & Jalili, S. "Selecting proper security patterns using text classification. In Computational Intelligence and Software Engineering," CiSE International Conference on IEEE, pp. 1-5 (2009)

[11] Hasheminejad, S. M. H., & Jalili, S. "Design patterns selection: An automatic two-phase method" Journal of Systems and Software, vol. 85, no. 2, pp. 408-424 (2012)

[12] West, J., & Bhattacharya, M. "Intelligent financial fraud detection: A comprehensive review. Computers & Security," vol. 57, pp. 47-66 (2016)

[13] Syeda, M., Zhang, Y. Q., & Pan, Y. "Parallel granular neural networks for fast credit card fraud detection. In Fuzzy Systems," 2002. FUZZIEEE'02. Proceedings of the 2002 IEEE International Conference on IEEE (Vol. 1, pp. 572-577)

[14] Phua, C., Lee, V., Smith, K., & Gayler, R "A comprehensive survey of data mining-based fraud detection research". arXiv preprint arXiv:1009.6119 (2010)

[15] Duman, E., & Ozcelik, M. H. "Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications," vol. 38, no. 10, pp. 13057-13063 (2011)

© 2021 IJSRET