ISSN:2229-6107

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY

E-mail :
editor.ijpast@gmail.com
editor@ijpast.in

www.ijpast.in

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

DATA MINING METHODS FOR BANKING SECTOR FRAUD DETECTION

P. Suresh,Dr. K. Sravan Kumar

Abstract:

Banking sector is having a great significance or value in our everyday life. Each and every person makes the use of banking sector in two ways, (i) physical and (ii) online. Physical fraud can take place like stealing the credit cards, sharing bank account details with corrupt bank employees, etc. Online fraud takes place by sharing the card details on the Internet or over the phone with a wrong person. It may also include spamming and phishing. While carrying out the transactions and all the relations with the bank policies, customers and the banks may face many problems due to fraudsters and criminals, and the chances of getting trapped are very higher. These kinds of frauds can be credit card fraud, insurance fraud, accounting fraud, etc. which may lead to the financial loss to the bank or the customers. Thus, detection of these kinds of frauds are very important. Fraud detection in banking sector is based on the data mining techniques and their collective analysis from the past experiences and the probability of how the fraudsters can steal from customers and banks. Therefore this paper addresses the analysis of data mining techniques of how to detect frauds and overcoming it in banking sector.

Keywords- Data Mining, Banking Sector, Spamming, FraudDetection, Collective Analysis.

I INTRODUCTION:

Financial fraud could be a growing concern with way reaching consequences within the government, company organizations, and finance trade. In Today's world high dependency on web technology has enjoyed magnified banking transactions. However, banking sector fraud had conjointly accelerated as online and offline transaction. As transactions become widespread mode of payment, focus has been given to recent procedure methodologies to handle the fraud downside. There are many fraud detections solutions and software system which prevents frauds in businesses like credit card, retail, e-commerce, insurance and industries. Data mining technique is one notable and common methods employed in determination banking sector fraud detection downside. It's not possible to be sheer sure concerning actuality intention associated right behind an application or transaction. In reality, to hunt out doable evidences of fraud from the accessible data using mathematical algorithms is the best effective possibility. Data mining is a method to extricate and analyse the entire and deducible knowledge and information which is applicable and serviceable for the future references [5]. The data used in data mining is extracted from several information like scrappy, noisy and unsystematic data my which the data mining process is executed. Banks are experiencing provocations in protective manner of online or internet banking challenge. The challenge keep customers account secure, where is avoiding complexness within the login process. But all the infinite number of passwords, different hardware token devices and tools that are out of the range used for communication made known by some banks has greatly discouraged some customers [1]. But we all know that security provided by the banks less attentive on customer comfort, but there will still be associate improvement. And all the fractals are having the use of subtle analytical models and rules

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

to correctly determine suspicious behaviour. However

customers are greatly annoyed if they are dealing declined-meaning the banks have to be compelled to choose a lot of complicated as different intimidating remark are found and also all the technological stuff must protect bank customers against all of them. Authentication accustomed be an easy watchword as all the growing intimidating remark over all these years it is fully capable from that to watchword with the numbers, then to information or data based mostly queries and lastly to this state in which all the token devices and different communication channels are there just to verify transactions. The solution made so is to combine all the safety strengths of associate authentication server with the logic and accuracy of a fraud detection system to spot intimidating behaviour and maximize authentication, else the opposite hand wherever traditional behaviour is found to lower the authentication so the client without much problem is disrupted – so resulting in low risk and a high value of customer's gratification. The descriptive data mining technique points out the relations or link in the data or information and proceeds to find all the properties and relations of the data and information examined. The below diagram tells about data mining technique/models and tasks carried out in it [4].

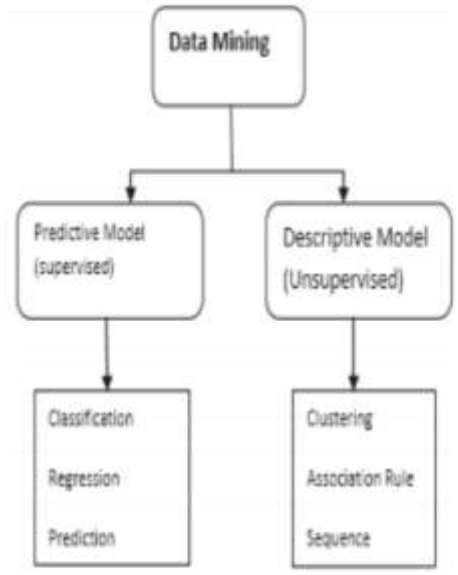

Fig.1. Data mining

The Reserve Bank of India-RBI maintains information on intimidating remarkable cases on the premise of operation beneath that all those cases have been accomplished. In step with this kind of information and data involving, the top most categories beneath that frauds are according by the deposits-saving a/c, housing loans, credit cards, term loans, ATM/debit cards, cash credit a/c, cash transactions, cheques, demand drafts, internet banking, etc.

II. APPLICATION AREAS OF DATA MINING IN BANKING

Fraud Detection:

Data Mining Techniques and Models can be found useful in the banking sector, mostly for the purpose of detection of fraud happening through the various techniques used by the fraudsters. By using all the different types of data mining models and techniques, the ever-increasing fraudulent activities, which are of a major concern for the business as well as the customers and banks, happening can be detected and also reported. Although, there are two processes through which the patterns of the frauds can be detected by the help of data mining [5]. The first process is the one in which the bank approaches different data warehouses which contain transaction information and implements its data mining codes to determine the frauds in it. Then, they can refer these patterns with their own collection of information of how frauds take place and determine the amount of trouble in it. Whereas, in the second procedure, the determination of the fraud pattern is done on the own personal information of the banks. The method which most of the banks use is the "Hybrid" approach to detect a fraud. It is used by majorly 9 banks and credit card issuers out of 10 in the US. Data Mining is not only the single factor which will facilitate the banking systems to achieve new customers, but also it can facilitate them to retain their existing customers. Customer accession and retentiveness are vital issues for any business, particularly the sector. Now-a-days, customers have such a lot of opinions with relevance where they will prefer to do their business. Executives within their banking system, therefore, should remember that if the bank employees are not giving every customer their full recognitions, the client will just search for the other bank that may notice. Data mining is also used to facilitate in marking new clients for merchandise and all the service and in finding or exploring a client's precious buying patterns in order that the banks are going to be ready to keep old clients by providing reason on individual basis customized to every client's requirements.

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

Risk Management:

Data Mining has also found use in the risk management within these systems. Bank employees want to introspect whether or not the clients they are dealing with are dependable or trustable or not. Providing new clients credit cards and a bank account, extending existing customer's, line of credits and giving loans maybe a dangerous and difficult decision for banking sector if they are not able to recognize any of the things regarding their clients and do not have their information in proper detail. Data Mining, moreover, can be used to cut back the danger and the risks of the banking sector that provide credit card by regulating those customers who are seemingly to highlight on their own bank account. One case was rumoured within media of one of the bank uncovering that all the cardholders who withdrew the cash at gaming house had a huge rated of law breaking and financial ruination. It is one of the practices in the banking sector to examine the clients dealing actions in their respective accounts to calculate their chance of predefined values in their respective loan account. Credit rating, was one in every of the old days financial risk managements implementations used which proved out to be a good technique during that time and this time period too [2]. Credit rating can be very useful to moneylenders within the banks when calculating to make choices to lend or not. Data mining may even obtain the praise behaviour of independent borrower having instalment, debt and credit card loans, having features like credit history, length of employment and length of residency. Thus the result is created permitting a loaner so that he/she can judge client further determining whether or not the individual could be better applicant for bank loan, or if there's a high risk of default. The extended period of time the customers who have been with the bank, stayed in better standing, having better income, results in receiving loan than a new client having no record with that particular bank or one whose income or wages is low. Some statistics associated with banking frauds The fraud criminal cases in banking sectors in nearly about 861 banks of Rs.1 lakh and more were reportedly within the half of 2015-16. Throughout 2014-15, one thousand six hundred and fifty one such cases were reportable that concerned associate quantity of Rs.11, 083.11 crore. Fraud Cases Related to Banking Sectors Fraudulent documentation involves sterilization, changing or modifying a document or papers involved in bank statements to deceive the bank. It can also involve approving half-truths provided in documents or the papers having the banking statements knowingly – cases of connivance of bank staff with fraudsters.

Siphoning of funds take place once funds borrowed from the banks are utilized for functions and reasons unrelated to the work or operations of the recipient. Absence of extant tips on the due diligence of bank professionals (like chartered accountants or money advisors) assisting borrowers at the time of disbursement of loans might end in valuation agencies or advocates facilitating the commission of frauds by colluding with the borrowers to inflate security valuation reports.

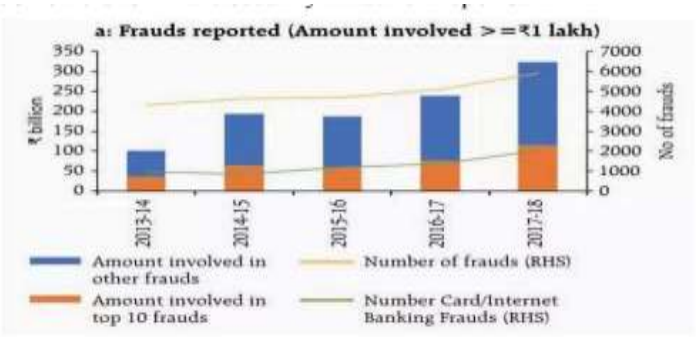

Fig.2. Frauds Reported

Fig. 3. Top Ten types of Online Fraud

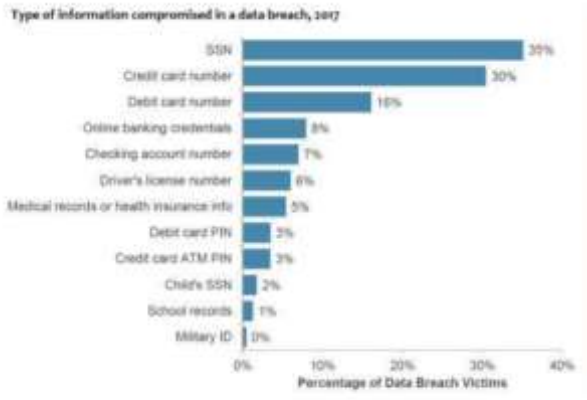

Fig. 4. Types of information compromised in a data breach

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

In 2014, around 65% of the whole fraud cases rumoured by banks were technology based frauds (covering frauds committed through/at a web banking channel, ATMs and different payment channels like credit/debit /prepaid cards). All the types of innovations in the business and the technological field that the banks are adopting in their seek growth are successively presenting increase in the levels of cybercrimes. These innovations have most likely introduced new vulnerabilities and complexities into the system [3].

Fig. 5. Bank Fraud in India

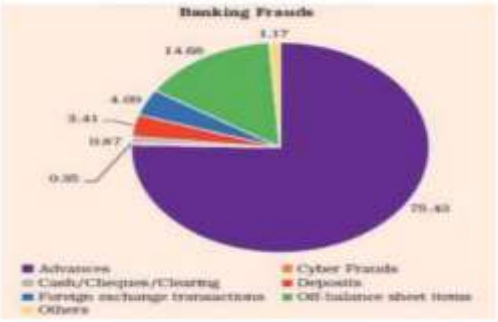

Fig. 6. Banking Frauds

III. DATA MINING TECHNIQUES:

Techniques applied for mining knowledge can be divided into various classes depending on the nature of knowledge that system is unearthing. We will now look into these important techniques.

3.1. Association This technique is used to unearth unsuspected data dependencies. In other words, it tries to detect data items that are associated or connected or correlated with each other which are not obvious previously. For example, if customers who are enquiring about a banking product, more often enquire about another unrelated product, then this technique can find this pattern out and inform the marketing team. More formally, the task is to uncover hidden associations from a large database. The idea is to derive a set of strong association rules in the form of "A1∧A2∧ … Am∧ B1∧ B2∧ … Bn" where Aj (for i∈{1… m}) and Bj (for j∈{1… n}) are set of attribute-values

from the relevant data sets in a database. For example, data recorded by a point of sales system would indicate that if customers buy certain items, they are most likely to buy certain other items. Such information can be used as decisions for marketing activities promotional pricing or product placements (Tiwari, 2010). In addition to this, association rules are employed in application areas including web usage mining, intrusion detection and bioinformatics. Typically all association rules are not interesting. From a large data set, a very large and a high proportion of the rules mined will be usually of little value. An associative relationship is considered to be useful if it satisfies a predefined support and confidence values (Geng and Hamilton, 2006). Hence, a rule is discarded if it does not satisfy this minimum support threshold and minimum confidence threshold. All these discovered strong association rules may not be interesting enough to present. Additional analysis need to be performed to uncover interesting statistical correlations between associated attribute-value pairs (Geng and Hamilton, 2006). Various types of association include (Ramageri, 2010): • Multilevel association rule • Multidimensional association rule • Quantitative association rule • Direct association rule • Indirect association rule

3.2. Classification and Prediction This is the most commonly applied data mining technique. It is employed when the classes of data in the population are known. For example, in the case of detecting fraudulent banking transactions from a bank's transactions database, there can only be two classes, namely fraudulent and non-fraudulent. It constructs a model from the sample data items with known class labels and use this model to predict the class of objects in the population whose classes are not known. Each tuple from the database contains one or more predicting attributes which determines the predicted class label of the tuple according to the constructed model. In the banking scene, classification technique is employed for Fraud detection (both corporate and credit fraud) (Ngai et al., 2011) These models are constructed usually using a decision tree model or a neural network model. A decision tree is a flow chart like recursive structure to express classification rules where each node specifies a test on an attribute value, each branch specifies a mutually exclusive outcome of the test together with a subsidiary decision tree for each outcome and tree leaves represent classes or class distributions. It can easily be converted to classification rules or can be used to compact description of data (Asghar and

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

Iqbal, 2009). Fuzzy sets are applied to the classification techniques when parameters to consider are of fuzzy in nature. For example, the length of URL parameter for detecting phishing sites can range from low to high with other values in between (Aburrous et al., 2010). Other commonly used classification technique involves application of neural networks. A neural network is essentially a network of processing nodes with weighted connections between the nodes where the weights are determined by a learning process using training data. Neural networks are computationally more expensive than their decision tree counterparts (Kumar et al. 2011). Classification works with discrete and unordered data and helps to identify class labels of the members of the population. But prediction models works with continuous-valued functions. That is, it is used to predict missing or unavailable numerical data values from the sample attribute values. Commonly used technique for prediction is regression analysis. It is a statistical methodology that is used to forecast values from existing numerical values. In predictive models for data mining, we have a set of independent variables whose values are already known and a set of dependent or response variables whose values we want to predict. Regression helps us to express the relationship between these variables as a linear or non-linear function. In many real world problems related to banking such as stock price predictions, or credit scoring follow complex models with many independent variables and requires multidimensional regression analysis and logistic regression (Li and Liao, 2011).

IV. DATA MINING TECHNIQUES USED FOR FRAUD DETECTION IN BANKING SECTOR:

There are many data mining techniques, and most are used in data mining analysis projects. Amongst these developed techniques include classification, clustering, association, prediction and serial patterns. A. Classification Classification can be defined as the most typically applied data mining technique, which deals with a collection of preclassified examples that helps us to develop a model that classifies the record of a population at a larger scale. Moreover, this retrieves vital information about data. It is

unlike clustering. An example is ‘Gmail’. The algorithm which is used by them helps them to classify whether the email received is Spam or a legit email Fraud detection and credit card risk are of a similar temperament for the purpose of this sort of research. The classification approach implies the use of decision tree or neural network-based classification algorithms in it. Data Classification involves the methods of learning and classifications in it. In learning method, the training data are analysed by classification of the algorithms of the data mining classification technique. Classification rules’ accuracy can be determined by employing a data check. New data tuples will be employed to the principles if the accuracy of the classification is appropriate as per the requirement. This is able to embrace complete records of each fraudulent and valid activity which are determined on a record-by-record basis. This classifier-training algorithm uses the pre-classified examples to work out on the set of parameters needed for correct discrimination. These parameters are then encoded into a model called as classifier by an algorithmic program. Classifier plays a major role in it. There are various types of classification models. They are as follows: Classification by decision tree Bayesian Classification Neural Networks Support Vector Machines Classification Based on Associations. B. Clustering The identification of indistinguishable kind of objects is often known to be clustering. In the process of clustering, all the different bank transactions are grouped together into one cluster. We can say that the approach of pre-processing for classification and attribute set choice is used in clustering. For example, one customer of a given geographical location and of a specific job profile request some particular set of services, like in banking sectors the purchasers from the service class always demands for the policy that ensures a lot of security as they’re not willing to take risks, likewise a similar set of service class people in countryside areas are having the preferences for a some particular brands only which can be demur from their counterparts in cities [2]. Because of this information, the organisation can facilitate their cross selling products, instead of casting and focusing on one of their particular product, the bank’s customer service representatives are often equipped with customer profiles that are so furnished by data mining that makes it easy for them to find that which of products and services are most applicable to the callers. This data mining technique can provide the management to find the answer of 80/20 theory of selling, which says: 20% of your clients can make your profits by 80%, then the difficulty is to identify those 20% and also the techniques of people that will help in accomplish the constant. The clustering methods are of the following methods: Partitioning Methods Hierarchical methods Density based methods Grid-based methods Model-based methods C. Predication As the name itself suggests, this technique helps us to determine the relation between dependent and independent variables. It can help us to understand the changes in the variables which are independent. The technique used is regression analysis for

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

determining the relationship between them. Predication analysis technique is mostly used in banks for the determination and prediction of the frauds and threats in the banking system and sector. The money in this case can be related to as an independent variable whereas the fraudster here can be the dependent variable as the money is independent for everyone. Then on the basis of the previous historical data mined from the past, a fitted regression curve can be drawn for the prediction of a fraud which is attempted. Regression analysis will help to get the connection between the dependent and the independent variables. In data mining, the response variables are the ones which we would like to predict firstly and the independent variables are already bestknown. But, unfortunately, several real-world issues do not seem to be predictable easily. They may be hard to predict. There are various types of regression method known to us, they are as follows: Linear Regression Multivariate Linear Regression Non-Linear Regression Multivariate Non-Linear Regression D. Association Rule In data mining techniques – association rule, the main focus is to find out the sets of binary variables that occurs so commonly during a transaction database, and the aim is to find different groups that are related to each other having a specific target variable. For simplicity we can say that in this data mining technique if/then statements are used, that helps to find the link between the independent knowledge during a comparative database or different data storage area. One example for this technique is 'if a customer buys bread, he/she is 80% likely to purchase milk as well.' So the two elements, an antecedent (if) and a consequent (then) are used in the association rule data mining technique. The association rule is of the following types: Multilevel association rule Multidimensional association rule Quantitative association rule

E. Neural Network The set of input/output units that are connected and each integration features a weight present in it is a neural network. There is a learning part in which network adjust the weights and learn resulting it to predict the proper category labels of the input tuples. These kind of networks have an amazing capability extract meaning from tough or general data, used to extract patterns and remark different courses that are too much interconnected or complex to be observed by either humans or alternative computer techniques. The compatibility of neural networks is continuous valued inputs and outputs.

F. Sequential Patterns In data mining technique, sequential pattern technique is also used in which it finds the similar or we can say homogeneous patterns in the data transactions over a business time. To acknowledge relationship among data, the uncovered patterns are used for any business analysis. They determine the user buying behaviours. This mining technique is a special case in structured mining technique.

V. Conclusion:

In the data mining technique all the important information from such a huge amount of knowledge and changes all the high level decision making in the banks and retail sectors. From different databases they mix the varied data and store the mixed data using data storage in acceptable format that the data mining can be done for it. Analysis of data is done further and thus the captured information is used in banking sector or in any organisation to support decision-making. In banking sector the data mining techniques are a huge help to them for targeting and exploit new clients, most useful in fraud interference, providing phase based mostly merchandise, fraud detection in real time, risk management, analysis of the customers. Data Mining is a most important tool for detecting the fraud activities happening in the banks related information and prevent the frauds happening in our daily life due to the fraudsters. Data mining operates to provide the security to database and to enhance and the choice creating power, taking right decisions at the right time, and selecting the correct options. It fetches the important pattern from the large database which helps us in improving the quality of the database. Hence, this research paper includes lots of problems associated with the banking information security and ways to overcome the problems of the banking system frauds easily through the techniques provided my data mining.

References:

[1] "Use of Data Mining in Banking" Kazi Imran Moin, Dr. Qazi Baseer Ahmed - International Journal of Engineering Research and Applications

[2] "Data Mining Techniques and its Applications in Banking Sector" Dr. K. Chitra, B. Subashini - International Journal of Emerging Technology and Advanced Engineering

[3] Fraud Detection in the Banking Sector,K. Chitra, B.Subashini - Proceedings of National Level Seminar on Globalization and its Emerging Trends.

[4] "Data Mining Techniques and Their Implementation in Blood Bank Sector" Ankit Bhardwaj, Arvind Ssharma, V. K. Shrivastava/International Journal of Engineering Research and Applications

[5] "Real-time Fraud Detection In The Banking Sector Using Data Mining Techniques/Algorithms"

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 3 Sep 2023

S. N. John Department of Electrical and Information Engineering Covenant University, Ota, Ogun State, Nigeria