Detection of Cyber Crime in Banking Sector

[1]Pranjali P. Nalavade, [2]Bhagyashri P. Dewoolkar, [3]Sampada S. Prabhu

[123]Student of Computer Engineering
[123]Department of Computer,
[123]S.S.P.M'S college of Engineering, Kankavli, India-416602

Abstract

Nowadays, the banking industry is facing an acute problem of fraud. The problem is international, and no country is fully protected. Fraudsters became specialists in hijacking on-line sessions: they steal shopper credentials and use malware to swindle funds from unaware account holders. Data mining applications ar employed in many banking sectors for shopper segmentation and productivity, credit scores and authorization, predicting payment default, advertising, and detecting fake transactions.. This paper presents a general plan concerning the model of information Mining techniques and numerous cyber-crimes in banking applications. It conjointly provides associate comprehensive survey of competent and valuable techniques on data processing for cyber-crime knowledge analysis. The objective of cyber-crime data processing is to acknowledge patterns in criminal manners so as to predict crime anticipate criminal activity and forestall it. This paper implements a completely unique data processing techniques like K-Means, Influenced Association Classifier and J48 Prediction tree for investigating the cyber crime data sets and sorts out the accessible problems. The K-Means algorithmic program is being used for unsupervised learning cluster at intervals influenced Association Classification. K-means selects the initial centroids so the classifier will mine the record and formulate predictions of cyber crimes with J48 algorithmic program. The collective knowledge of K-Means, Influenced Association Classifier and J48 Prediction tree tends certainly to afford a enhanced, incorporated, and precise result over the cyber crime prediction in the banking sectors Our enforcement organizations need to be adequately outfitted to defeat and forestall the cyber crime.

Index Terms - Cyber crime, Data mining,k mean algorithm, clustering, Influenced Association Classification, J48

I. INTRODUCTION

Data mining is the computer-assisted process to break through and analyzing large amount of data, then extracting the meaning of data. It is conjointly the method of analyzing knowledge from completely different views and summarizing it in to helpful data [1]. The Data mining prediction techniques ar capable to boost the accuracy, performance, speed of predicting the cyber-crime. Cyber-crime has been increasing in complexness and money prices since firms began to utilize computers within the course of doing business. Cyber criminals are becoming more sophisticated and are targeting consumers as well as public and private organizations [2]. Cyber-crime analysis includes a terribly big responsibility of enforcement system in any country. Cyber-crime involves the breakdown of privacy, or harm to the pc system properties like files, website pages or software [3].

Developing a high-quality cyber-crime tool to acknowledge crime prototypes apace and capably for future cyber-crime pattern exposure is crucial. Banking sector has been hotspot for cyber-crime is it natural or unnatural. And also the technology is becoming indispensible part of banks it has become easy for users so as for attackers as now they have more mode to exploit the vulnerabilities [4]. Banking sectors ar liable to several interruptions originated by associate assortment of classes of threats; various threats ar distinct underneath numerous teams that's cyber fraud, trade length development and information safety measures. Cyber-crimes that are committed in banks include hacking, Credit card fraud, money laundering, Does attacks, phishing, salami attacks, ATM card cloning etc. Cyber threats like pharming, phishing, tempted reveal of personal details like fraud ar the safety qualms that exist within the brains of shoppers in banking and money sectors. Perceiving cyber-crime is extraordinarily robust in addition, as of numerous online business transactions and hectic network traffic which generate enormous quantity of data and just a segment of which relay to prohibited actions.

Crime prediction uses past knowledge and when analyzing knowledge, predict the future crime with location and time. At present, serial criminal cases rapidly occur so it is a challenging task to predict future crime accurately with better performance [5]. Credit card and net primarily based crime ar more and as more technologies ar rising high. To deal and overcome fraud, clustering and classification techniques are implemented [6]. Fraud detection is one of the difficult processes not only technically, but also in crime investigations. The method of fraud detection relies on easy comparisons, and also based on association, clustering, prediction and outlier detection [7]. Association Rule mining as generates "n" best association rules based on n selected and Classification and Regression Tree (CART) that predict categorical class labels [8]. Clustering technique is devised as a multi purposive optimization crisis. The suitable clustering algorithm and parameter locations depend on the entity dataset and projected use of consequences. Clustering in and of itself isn't a routine task, but it is an iterative method of knowledge innovation or interactive multi purposive optimization that engrosses test and malfunction.

Cyber-crimes is reduced from the banking transactions by applying the updated technology and appointing reliable officers and devices [9]. Cyber-attacks in the banking sectors might be in the type of illegal access, devastation, bribery or amendment of data or any kind of malicious practice to source network malfunction, reboot or sling. Current security system have created cracking terribly monotonous but not impossible. So some very important intensity of security should be recognized before trade on the web is systematically performed. To preserve against cyber-crimes, intrusion recognition methods ought to be intended, executed and governed. The planned model is fancied to access large amount of criminal records so prediction is created as per the precedent performance of the criminal people. As the records within the system enhances, there's no got to insert the knowledge. In the current digital era, the probable risk to shield large amount of information with a various society of cyber criminals could be a vast dispute.

II. LITERATURE SURVEY

• Dr. Zachariah SulimanZubi and AymanAltaher Mahmoud had proposed a model for crime and criminal data that analyzes using simple K-means algorithm for clustering and Apriori algorithm for data Association rules. It conjointly tends to assist specialist in discovering patterns and trends, creating forecasts, finding relationships and potential explanations, mapping criminal networks and distinctive potential suspects. They showed the promising results of their model planned model from the attributes for crime, criminal and also the results of K-means algorithmic rule. They also gave the overall statistical knowledge about the criminal age versus crime type which provided the input to the K-means algorithm [10].

• RasoulKiani, SilamakMahdavi and Amin Keshavarzi (2015) had applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales among 1990 to 2011. They allotted weights to the options so as to enhance the standard of their model and removed low price from them. They employed Genetic algorithm for optimizing of Outlier detection operator parameters using Rapid miner tool .

• Dr. K. Chitra and B. Subhashini (2013) had analyzed the data mining techniques and its applications in banking sectors like fraud detection and prevention, customer retention, marketing and risk management. They mentioned the necessity of data mining techniques within the banking sectors for higher targeting and exploit new customers, most valuable customer retention, automatic credit approval which is used for fraud detection and interference in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing .

• Akshay Kumar Singh, Neha Prasad, NohilNarkhede and Siddharth Mehta (2016) had described a system for analyzing crime and also discussed the method of increasing the accuracy of crime prediction for the prevention of crimes. They used Apriori algorithm for the identification of trends and patterns in crime. They conjointly used call trees for crime prediction owing to its strong nature and conjointly it will be utilized with giant datasets. Their system provides the lead to the shape of crime category or class of crime by that effective measures is deployed by the native enforcement agencies to avoid crime and secure the neighborhood. It may also predict the timeframe for the crimes if their collected information is sort of made .

• TusharSonaqwanev, ShirinShaikh, ShaistaShaikh, Rahul Shinde and AsifSayyad (2015) had grouped crime data according to various types of crimes that had taken place against women in different states and cities of India. They used K-means algorithm for clustering, Pearson¨s correlation coefficient for correlating crimes between two variables and Linear regression for crime prediction .

• UttamMande, Y. Srinivas and J. V. R. Murthy (2012) had collected the crime dataset from Andhra Pradesh local department and probably aimed to spot a criminal supported the witness or clue at the crime spot. They used binary cluster and classification techniques to investigate the criminal information. They conjointly tried to spot the criminal by mapping criminal exploitation the strategy of car correlation and also the approach during which the incident had taken place and their options of the crime are thought of to validate the criminal .

• Lawrence McClenden and NatarajanMeghanathan (2015) had proved how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns. With the help of wood hen tool, they observed that Linear regression algorithm was very effective and accurate in predicting than Additive regression and Decision stamp algorithms when implemented them with same finite set of options on the Communities and Crime dataset .

• JavadHosseinkhani, Suhaimi Ibrahim, SuriyatiChuprat and javidHosseinkhaniNaniz (2014) had afforded a review for extracting useful information by means of Data mining, in order to find the crime hot spots out and predict crime trends for them using crime data mining techniques. They conjointly evaluated progressive approaches for extracting helpful info by suggests that of information mining, in order to find crime hot spots out and predict crime trends for them using crime data mining techniques .

• RaghavendraPatidar and Lokesh Sharma (2011) had tried to deduct fraudulent transaction through Neural network along with the Genetic algorithm. They used Genetic algorithms for creating the selections concerning configuration, number of hidden layers, number of nodes that would be used in the design of neural network for their problem of credit card fraud detection. For the training of purpose of artificial neural network they used supervised learning feed forward back propagation algorithmic rule .

• AnishaAgarwal, DhanashreeChougule, ArpitaAgarwal and DivyaChimote (2016) had utilized frequent pattern mining with association rule mining for analyzing the various crimes done by a criminal and predict the possibility of every crime that may once more be performed by that criminal. This prediction was supported attributes like list, education, occupation, friend circle, family background and other factors. They implemented Apriori algorithm for generating frequent item sets. They designed AN application that may be accessed and out there to the licensed users anytime and anyplace, in conjunction with the most practicality of prediction of the additional crimes by individual criminal .

• Linda Delamaire, Hussein Abdou and John Pointon (2009) had intended to identify the different types of credit card fraud and to review alternative techniques that have been used in fraud detection. They mentioned the manner however call trees, Genetic algorithms, Clustering techniques and Neural networks can be utilized as detection techniques in credit card fraud detection. They known the various forms of frauds like bankruptcy fraud, counterfeit fraud and additionally mentioned the measures to deduct those frauds mistreatment data processing techniques .

III. PROPOSED MODEL FOR CYBER CRIME PREDICTION

Discovering and exploring cyber crimes and inquiring their affiliations with virtual criminals area unit involved in evaluating cyber crime progression. The proposed work presents the model over cyber crime prediction with K-Means clustering technique, Influenced Associative classifier and J48 classifier. For the cyber crime prediction in banking sectors, the proposed model grants an enhanced prediction outcome. Influenced Associative Classifier affords a well-organized thanks to utilize the classification technique with Association Rule Mining, which enhances the prediction accuracy for classification. It additionally employs the influenced support and confidence structure for excavation out the Association rule from crime information. The incorporated implementation of J48 technique with K-Means and Influenced Association Classifier provides the improved prediction outcome over the cyber crime hazards in banking sectors.

A. Collection of cyber crime dataset

A diversity of cyber crime knowledge should be collected for the prediction of cyber crime category in banking sector by the analysis of crime pattern. So this knowledge has got to be collected from numerous news feeds, articles and blogs, police

department websites over the internet. The collected cyber crime knowledge is keep in crime information for additional handling of knowledge.

B. Pre-processing of cyber crime dataset

The cyber crime dataset stored in Crime database has to be preprocessed before applying data mining techniques on them. Because preprocessing removes noisy data, missing values etc.
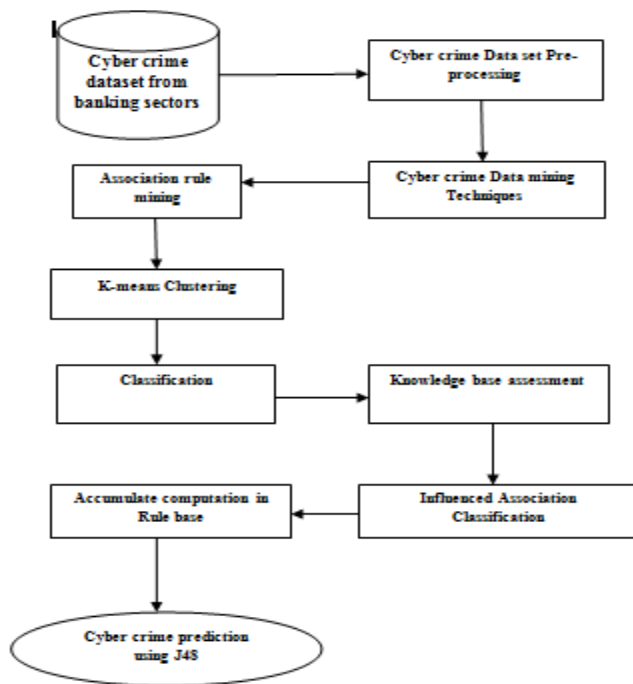

Fig. 1. Proposed model for Cyber crime prediction

C. Data mining Techniques

For Pre-processed data, Data mining techniques and algorithms are implemented to identify or forecast fraud through Knowledge innovation from abnormal patterns and also it achieves recognition in combating cyber credit-card fraud data processing aids by contributory in determination tribulations in banking sector by discovering patterns, relationships and links that are unseen in the business information accumulated in the crime databases.

1. Association Rule mining

Based on frequent occurrences of the crime patterns, Association rule mining produces rules for cyber crime dataset. These generated rules assist the assessment producer of defense society to require a hindrance action. The procedure comprises the subsequent measures:

•   The method of determining commonly occurring item sets in the cyber crime database.

•   The identification of patterns in program implementation and customer behaviors as association rules known as intrusion recognition.

2. Clustering

Splitting of a collection of records or things to variety of teams is named bunch. Clustering is tacit on discovering interactions linking cyber crime and criminal characteristics having some past mysterious general characteristics. For discovering frauds in banking sectors, clustering techniques are utilized. Clustering is phrased as unsupervised learning as a result of its categories don't seem to be definite and determined current and syndicate of knowledge is thru exclusive of superintendence. K-means partition algorithmic program is enforced in bunch cyber crime datasets as a result of its minimal art and fewer procedure complexity. At first, the quantity of data items are assembled and precised as „k" clusters. Between the mean distances of objects, the mean value is intended. The positioning repetitive methodology is employed to recover the partitions by transferring things from one cluster to different. Then until the union occurs, the number of iterations is carried out.

3. Classification

Classification is the most frequently used data mining technique, which executes a set of pre-classified examples to build up a model that can classify the instances of attributes at huge scale. The classification technique creates an association between a dependent variable and an independent variable by mapping the data points. Within the given dataset, Classification is used to bring out in which group each data occurrence is associated. Classification is employed to make many models of unknown patterns and prospect assessment on the idea of the previous higher cognitive process. Automatic credit authorization is that the nearly major procedure within the banking sector and monetary organizations. Frauds may be prohibited by building a superior assessment for the credit consents victimisation the classification illustration supported call trees like J48, CART etc.

4. Influenced Association Classification

For accomplishing additional exactness, the associative classification is extremely novel and improved method which assimilates the mining of association rule and classifications of the model prediction. This method is being implemented for ruling out the link and association over item sets. The associative categoryification comes beneath unattended learning since it will have interaction any class characteristic for rule extraction. Two steps employed to extract association rules are,

1. Through cyber crime knowledge set, classes are generated based on the association rule.

2. within the category labels, perform analysis on the dataset classification.

The Influenced Association classification is entirely novel perception for rule categorization. It conjointly intends weighted confidence and support structure for mining association rules over the cyber crime knowledge set. Various steps enforced in Influenced Association Classifier has been summarized below:

Initially, Pre-process the cyber crime dataset thus additional mining practices may be achieved on them.

• To replicate the assessment in the replica of prediction, every element is assigned within a range of weight

(0-1). Attributes having additional significance are allocated maximum weight(0.9) and having fewer significance are allocated minimum weight(0.1).

• Influenced Association Rule Mining algorithm is implemented on pre-processed cyber crime data set for obtaining fascinating pattern invention. Influenced Association Classification uses weighted support and confidence and also the rules spawned by this method square measure referred to as Classification Association Rule.

• The extracted Classification Association Rules are stored in the Rule base index.

• At any time if any new cyber crime record is updated, this CAR rule forecast the class label from the Rule base.

5. Cyber crime Prediction using J48

For the classification of problems and issues within the cyber crime prediction analysis, J48 algorithmic program is additional peaky and precise. Two steps in J48 are:

• Formation of tree.

• Validate the built tree over the cyber crime data set.

The J48 rule uses pruning methodology for construction of the tree. The pruning technique diminishes the scale if the tree by removing applicable information that guides the terrible concert in prediction. The anticipated J48 rule classifies the information until the complete categorization and affords utmost accuracy over the coaching of cyber crime data. It also stabilizes the precision and litheness. The J48 rule is that the in depth version of decision tree C4.5. The J48 rule produces the classifier output within the style of rule sets and decision tree. The rule sets square measure easy to acknowledge and too simple for using at intervals the applying.

IV. CONCLUSION

The planned model generates a superior conception over the cyber crime prediction by implementing the novel data processing techniques like K-Means, Influenced Association Classification with Prediction tree J48. The Influenced Association Classification is an improved model for classification and association with weighted support and confidence measures. From cyber crime datasets, K-Means rule bunches the item sets. The Classification concert and precision can be enhanced with K-Means, Influenced Association Classification with Prediction tree J48. In the banking sectors, the clients have to be aided through precise requirements in the application software to discover alert while a stern interruption is recognized. Intrusion tools need to be established where it's practicable and appraised on a typical b asis. To scrap beside cyber attacks, customer tutoring must be prepared in association with government and other confidential organizations. Awareness agenda ought to be place into follow to ensure that purchasers acknowledge information issues, intensity of privacy and the method to make the banking transactions secure.

REFERENCES

[1] Ms. H. N. Gangavane and Prof. Ms. M. C. Nikose, "A Survey on Document Clutering for identifying Criminal", IJRITCC, Vol. 2, Issue 2, February 2015, pp. 459-463.

[2] Ms M. Lakshmi Prasanthi and Tata A S K Ishwarya, "Cyber Crime Prevention & Detection", IJARCCE, Vol. 4,Issue 3, March 2015, pp. 45-48.

[3] Shubham Kumar, Dr. SantanuKoley and UdayKuamr, "Present Scenrio of Cyber Crime in INDIA and its Preventions", IJSER, Vol.6, Issue 4, April 2015,pp. 1972-1976..

[4] Manpreet Kumar, DivyaBansal and SanjeevSofat, "Study of Cyber Frauds and BCP Related Attacks in Financial Institutes", IJICT, Vol. 4, Issue 16, 2014, pp. 1647-1652.

[5] Nikhil Dubhey and SetuKumarChaturvedi, "A Survey of Crime Prediction Technique using Data mining",IJERA,Vol. 4,Issue 3, March 2014, pp. 396-400.

[6] Muhammad Arif, KhabaibAmjadAlam and Mehdi Hussain, "Crime Mining : A Comprehensive Survey", International Journal of u- and e-Service, Science and Technology, Vol. 8, Issue 2, 2015, pp. 357-364.

[7] Dr. R. Jayabrabu, Dr. V. Saravanan and Dr. J. JebamalarTamilselvi, "A Framework for Fraud Detection System in automatic data processing victimization Intelligent agent for higher deciding process".

[8] K. ChitraLekha and Dr. S. Prakasam, "An analysis of finding the Influencing Factors of supporting for the "GiveitUp" LPG Subsidy for the Government using Data mining Techniques", IJCA, Vol. 143, Issue 5, June 2016, pp.34-39.

[9] Dr. M. Imran Siddique and SamaRehman, "Impact of Eletronic Crime in Indian Banking Sector", IJBIT, Vol. 1, Issue 2, September 2011, pp. 159-164.

[10] Dr. ZakariaSulimanZubi and AymanAltaherMahmmud, "Crime Data Analysis using Data mining Techniques to Improve Crimes Prevention"