Webology (ISSN: 1735-188X)
Volume 18, Number 5, 2021

An Analysis Of Cyber Crime Prediction Model In Financial Sector Using Big Data Analytics

Mrs. P.Punithalakshmi[1] , Dr. M.Rajakumar[2]

[1]Research Scholar, Jamal Mohamed College, Trichy. (Affiliated to Bharathidasan University, Tiruchirappalli)

[2]Research Advisor, Assistant Professor, Jamal Mohamed College, Trichy.(Affiliated to Bharathidasan University, Tiruchirappalli)

Abstract

In advancement of Technology development in various fields of industry, has a major issue of handling data. Many of industry are facing a different kinds of problems to maintain the data about industry faculties, business, and financial-related information. It is very difficult to hold the information securely in the data warehouse. The data mining concepts are used to detect the cybercrime problems and analyze the problems of cybercrime in financial sectors. Biggest support of big data analytics belonging to Velocity, Veracity, Volume of data that can be handled with safely and processed with an efficient manner of any algorithm execution. With the response of online automation system facility provide the biggest challenges in financial sectors, like banking, online payment transaction and account holder information of banks must be preserved and the data must be secure. In this way major support for online banking system with information technology department, cybercrime attack is possible for the crackers, intruders, hackers, phishing. Online financial transactions of each and every step should be noted by the servers of the financial sector and they are maintaining the details securely. Accuracy and response time being the most important quality concerned for Big Data cyber security Analytic that are compromised with existing solution and time efficiency of various algorithms, that is a process with various attributes and provide the crime prevention models of maintaining the big data things.

Keywords: Big Data, Cyber Security, Random-Forest, ANN, Financial Sector scope.

1. INTRODUCTION

Cybersecurity deals with the security on the internet to avoid the unauthorized access of information about their customers. This is only for the customer trust and support business. They are expected to maintain their account with safe and securely. In the highest population country, many customers have a bank account, financial sector or altered. That is decided by the customer depends upon the simplification and sophistication of handling and maintain money on digital-based transactions. In this way major support for online banking system with information technology department, cybercrime attack is possible for the crackers, intruders, hackers, phishing. Online financial transactions of each and every step should be noted by the servers of the financial sector and they are maintaining the

details securely.

2. CYBERSECURITY ISSUE

Security is the most important aspect of computer systems, networking, and electronic data storage places. Many security- related issues are arrive at the time of handling data on network or offline mode of data repository. Each data organization is providing an authentication application for accessing the data[7]. There are many security-related threads are affect confidentiality information.It is one of the biggest challenges to prevent information from hackers, crackers, and intruders.

There are different ways of security-related issues like threads, attacks, hacking, eavesdropping, etc. These kinds of issues arise in all the networks. Information security for the customers ofthe bank, employees of the organization, a product of business warehouse, retrieving code about the licensed software, and many data related network of companies are preserved [21] [22] [23] [24] [25]. They are used to provide the hope and authentication accessing of information retrieval online or offline. Cybersecurity is a vast area of securing information on the web [26]. Especially, huge financial-related organizations are affected by cybercrime. That is information forhandling the data on their repository with less security.

3. CYBERCRIME IN THE FINANCIAL INDUSTRY

The financial industry and another large volume of organizations depend on big data analysis. That is a large volume of data collected from a different source which may be structured or unstructured. Big data analytics uses extensive techniques and tools for analyzing large, multidimensional data set. However, there are many challenges in dealing with a large volume of available data[6]. A new method and technologies need to be devised in order toanalyze the heterogeneous and multi-sourced data.

One of the fundamental techniques of Big data analytics (BDA), data mining is an interdisciplinary[8] [13] [14] [15] [16] [17] [18] [19] [20] and growing research area. Data mining is useful in not only the discovery of knowledge but also enhancing of known one. With support of such a technique, big data analytics can help us easily identify the crime patternswhich occur in particular area and how they are related with time. The implication of machine learning and statistical techniques on crime or other big data applications such as traffic accidents or time series data will enable the analysis, associated pattern and trends.

4. PROGRESS OF WORK REVIEW

In 2013, Mr. Jyoti Agarwal from Amity University introduced the concept of Crime analysis using K-means Clustering[5], The prediction of crime based on spatial distribution of existing data and anticipation of crime rate using different data mining technique.

In the year of 2016, Mr. Udhya Thupakula and Mr. Vijay varadharajan from Advanced Cybersecurity Research Center Faculty of Computer science, has proposed “Securing Big Data Environmentfrom Attacks” to revival the technique for big data environments such as public cloud with tenants using their virtual machine for different services such as utility and healthcare. In this

work makes use of Trusted Component Model(TC) in each physical servers for monitoring the usages of resources allocated to the tenant virtual machine. Here Entity deduction, Store and Restore, validation are important in TC sub component that are using for deducting service- specific attack on the tenant virtual machine. [1] Cybercrime investigations analysis in 2017 at Norwegian University of science and technology[2], the investigators are more than ever confronted with vast amount of heterogeneous data, increased complexity in distributed stored information. Constantly increasing network bandwidth it makes extremely challenging to process to store information. In this work cybercrime investigation is to use computational forensics based on advanced data analytics to prevent and combat cybercrime. It is applied on large volume of unstructureddata.

In 2018 SRM Institute of Science and Technology, from Department of Information technology by Aarathi srinivas Nadathur,Gayathri Narayanan and others [3], publish the crime analysis prediction using big data, that is The exponentially increasing population in our country leads to increase in crime and in turn generating massive amount of data which could be analyzed for the government to make critical and essential decisions as to maintain law and order. In this work implemented using joins, partitions and bucketing techniques in Hadoop tool. This schema could be extendedwith suggestions to recommend actions, corresponding measures andconstructive policies according to the offence type and the subsequent crime incident.

An Information security of Information banking system on Network security involves the protection of data at the time of transmission over communication line and protection from unauthorized remote access to the network. This article discuss aboutinternal and external attacks on the reliability position[4]. The protection should be divided into two directions. That is, the minimum sufficient awareness of system users and multilevel identification of user and control their rights. Anshu sharma, et al.,
[6] proposed k means clustering algorithm which was used for constructing patterns of data. Data were collected and circulated, twothird of true data and misrepresentation history information were utilized for preparing and remaining information were utilized for predict and web crime discovery. The precision of the proposedwork was 94.75 % and it productively recognized the false rate of 5.28%.

K. K. Sindhu et al., [7] explained scientific investigation ventures in the capacity media and hidden data investigation in the record framework, network forensic and cyber-crime mining. Devices was proposed by combining digital forensic investigation and mining of crime data intended for discovering motive and pattern of attacks and checks of assaults sorts occurred in that time period.

5. ANALYSIS OF ALGORITHM OF CYBER CRIME

Computer-enabled cyber-crimes include traditional crimes thatcan be enhanced in scale and reach using computers and networks For example, cyber-fraud and data theft using hacking, key-logging, and social engineering can be classified as computer-enabled crimes.

- Hacking is an unauthorised access or control over computernetwork security systems for some illegal or criminal purpose. (ET)

- Identity Theft is by gaining access to one’s personally identifiable information (PII) and to commit fraud (Norton).

- Cyberstalking - When online activities of a person are observed closely without his/her knowledge this is called cyberstalking. It is a breach of his/her privacy too.

- Credit/debit card theft over a phone call/e mail/sms occurs when an authorized user of credit/debit card shares its information over, email, sms or call with an unauthorized person with a belief that the caller is an authorized person from the bank.

The Data mining algorithms are finding towards the accuracy of analysis and pattern discovering for given training data set. Theprocess of clustering and classification are may fit into some special set of data that are evaluated in real time. There are two approach of Data mining technique for discovering analysis that is Supervised[9] and Unsupervised. The crime analysis data are already classified by supervised learning with class label prediction. The decision tree mechanism and navie bays theorem and support vector mechanisms are evaluate the analysis in crime[12]. The crime analyses are engaged by two ways. One form computer system level data securityand Network level security. Computer security was depends on Virus, warms, Threads are affected our authenticated files or data. But the network level, it is very complex type of attacks, like intruders, eavesdroppers, hackers, crackers they affect our data at thetime of network communication.

In this work to improve the security levels on network using high secure algorithms. Because of in financial sectors are highly affected by network level security. The data mining algorithm with analyze crime and give the best prediction of attack on the network. The big data analytics are the huge volume of data can be operated through online with a separate server. All the transaction and communication on financial sectors are fully handled in online. To analyse the crime factor and provide the security for the communication are discussed with data mining classification algorithms.

Random forest algorithm Random forest algorithm is a supervised classification algorithm and bootstrapping algorithm with decision tree model. It is a learning method of classification and regression. It is operated by constructing decision trees at training time and result comes as individual trees. Advantages of this algorithm are it uses classification and the regression task both, also handle the missing values and it won't over fit when there is more number of trees in the forest and it also classifies categorical values. This algorithm is mostly used in banking, medicine, stock market and e-commerce. Steps for random forest algorithm:

(i) k features are randomly selected from total m features,where k << m
(ii) Node d is calculated.
(iii) Nodes are splited into its branch nodes.
(iv) Steps from 1 to 3 are repeated still number of node becomel.
(v) Steps from 1 to 4 are repeated until a forest is created.

6. PROBLEM DEFINITION FOR CURRENTISSUES IN CYBER CRIME

The cyber crimes are increased from day-to-day in India. Because increasing the financial growth through online trade and marketing in this country, The NCBR Report where introduced in 2017 new crime heads such as Cyber blackmailing, Cyber stalking and dissemination of fake news. India recorded 21, 769 cyber crime in 2017, an increase of 77% from 2016. Country wide 1.7 cyber crimes were

committed per one Lakh population in 2017.

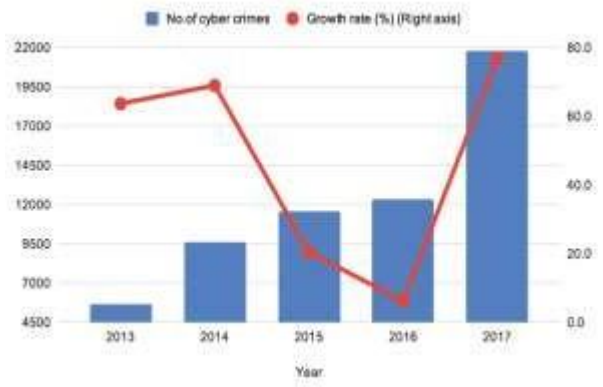

Fig.1. Cyber Crime Over the Year

In 2014 cyber crimes had grown by 69 percent compared to 2013. There are 11,592 cyber crime were recorded in 2015 in India. The growth rate of cyber crimes has gone down in 2015 and 2016 before registering a sharp spike in 2017. In 2016 growth rate of cyber crime fell further by 6.3 percent as 12,317 crimes were registered. The Figure.1. shows the year wise cyber crime growth rate.

In cyber crime has different way of fraud are classified by the survey[11]. There are seven type of crime are off the record. In the crime lied in high with ATM Fraud because it is highly sensitive of secure transaction through online. The figure.2. shows the different way of frauds in crime. There were a problem analyzed in cyber crime the maximum of fraud occur in cyber crime through ATM process. So It is the major issue for the cyber crime on the network.

Fig.2. Cyber crime Fraud in India

a) Analyse the Cyber Crime on financial sectors with Random Forest algorithm

Cybercrime is the crime done utilizing the computer and the internet. It is an illegal activity that includes illegal interception, data interferences, e-fraud and misuse of devices. Random forest is a supervised algorithm that is used for both classifications as well as regression. This algorithm creates a decision tree on data sample and then gets a prediction from each of them and finally selects the best solution by means of voting. It is an example method which is better than a single decision tree because it reduces the over-fitting by averaging the result. The dataset are collected from web resources that are already used in many researchers in different analyses and pattern and model derivation of various suitable algorithms. Workingwith random forest algorithm to analyse cybercrime data in financial sectors, it is used to generate the decision tree for each sample data start with the selection of random samples for the entire training dataset. Than every sample set of decision tree algorithm will get the

prediction result from every decision tree. In the next step of voting will perform for every predicted result. Finally select at most voted or the majority of voted prediction result.

7. PROPOSED METHODOLOGIES OF CYBERCRIME ANALYSIS IN CREDIT CARD FRAUD

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers arenot charged for items that they did not purchase. The dataset containstransactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days. It contains only numerical input variables which are the result of a PCA transformation. The only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

We have 492 frauds out of 284,807 transactions. Thus, the dataset is highly **unbalanced**, the positive class (frauds) account for 0.172% of all transactions. To deal with an unbalanced dataset, we can upsample minority class or downsample majority class. I choose to downsample majority class as upsampling will create duplicate or fake observations which can divert my model. Downsampling the majority class involves randomly removing observations from the majority class to prevent its signal from dominating the learning algorithm. For credit card fraud in transaction training dataset are applied to random forest algorithm that are produced a confusion matrix using legitimate transaction and fraud transaction yield the following table. This algorithm support 99.94 percent of the accuracy of given training dataset .

Actual/Predicted	**Legitimate**	**fraud**
Legitimate	85279	8
Fraud	38	118

Table 1. Confusion matrix

8. USING ANN FOR PREDICTION OF CREDITCARD FRAUD TRANSACTION

Artificial Neural Networks are used to provide the prediction pattern based on analyzed data or given data. Other soft computing techniques like fuzzy time series, Bayesian networks are also used for prediction purposes of ANN are more powerful as they provide more accuracy as compared to other techniques. A Perceptron network with one or more hidden layers is called a Multilayer perceptron network. A multi perceptron network is also a feed-forward network. It consists of a single input layer, one or morehidden layers and a single output layer. Due to the added layers,MLP networks extend the limitation of limited information processing of simple Perceptron Networks and are highly flexible in approximation ability. Multi-layer perceptron networks are networks with one or more hidden layers. The back propagation network is a type of MLP that has 2 phases i.e. Feed Forward Phase and Reverse Phase. In the Feedforward phase, the input neuron pattern is fed tothe network and the output gets

calculated when the input signals pass through the hidden input and output layer. In the Reverse Phase, the error is back propagated to the hidden and input layer for weightsadjustment. The error is calculated at the output layer when theactual output is compared with the target value. Some networks also calculate the error at the hidden layer which is propagated back to the input layer. This helps in more accuracy and convergence.

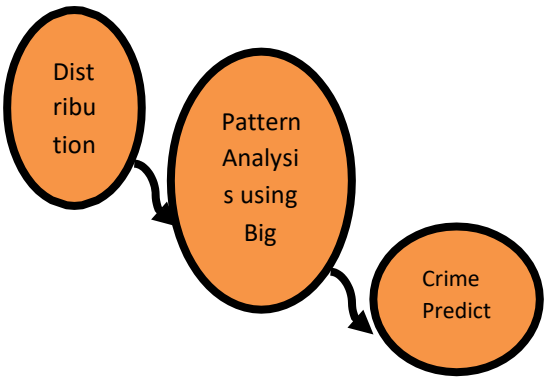

Fig.3. Crime prediction process model

	Pred: Legitimate	**Pred: fraud**
Act: legitimate	8527 2	15
Act: fraud	37	11 9

Table 2. Confusion Matrix

In ANN model for MLP process enable the training dataset with credit card transaction details of the attribute. After the pre- processing and data transformation step that training data reduced with values are inputted into ANN and provide the accuracy ofclassification and confusion matrix. That also provides the accurate and efficient result for large volume of data with optimal time. It also produced the 99.84 percent of accuracy for the given training dataset.

9. RESULT AND DISCUSSION

In this paper discussed with cybercrime dataset from the region. The data mining concept of analyzing the best pattern for cybercrime for given dataset. Most of the algorithms support the big data process. It is major process of prevent cyber crime in current digital world. The financial sectors must improve the security for vendor application and provide hope for the dependent people for that money and preserved by the financial sectors. Comparing these two algorithms for the prediction of credit card transaction fraud provide the better improvement result given by the random forest algorithm with ANN – multiple linear processing classifier. so our training dataset produces an output for better than

ann. So we can apply the test dataset to this algorithm provide the accuracy of resultsfor fraud detection in credit card for the cybercrime process. Table-3 denotes the ANN MLP algorithm and Random Forest algorithms efficiency and accuracy of the prediction of given big data attributes. In this paper analysesfs the big data analytics the data set are referred from UCI web source provide the highest accuracy for Random Forest algorithm also efficient for algorithm processing time is betterthan ANN. Figure -4 denotes the graphical representation of analysis.

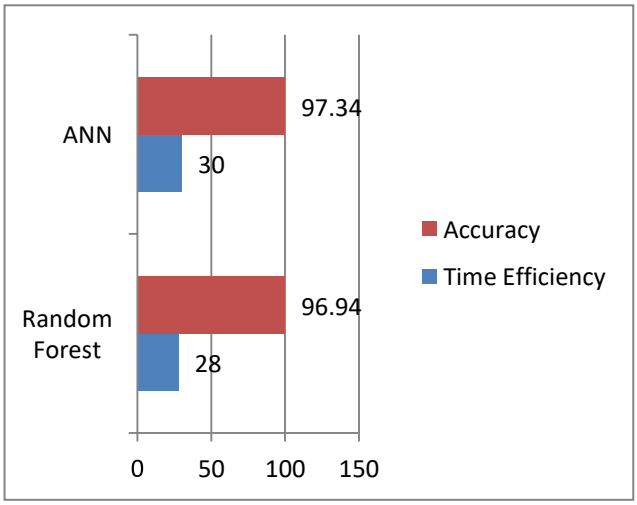

Fig.4. Time efficiency and Result Accuracy graph

	Accuracy	Time Efficiency(seconds)
Random Forest	96.94	28
Artificial NN	97.34	30

Table 3. Accuracy and Efficiency

10. CONCLUSION

In this research article for crime analysis system, supportan algorithm for large volume of dataset that fit into the two various kind of datamining process. Cyber crime analysis and protection wasthe most challenges aspect in finance sectors. The given dataset are collected from government website in real time. In this article conclude that the Maximum of fraud action in financial sectors in ATM card transaction. That are very complex to give protection for these transaction. Dataset are inputted into Random forest and ANN algorithm process for analyze the prediction and time efficiency in large volume of data. Both algorithms are produce well efficient of high true positive(legitimate) analysis. ANN algorithm process, fix the problem and predict the fraudulent transactions more quickly.

REFERENCES

[1]. “Securing Big Data Environment from Attacks” by Udhaya Tupakula and Vijay Varadharajan , Advanced Cyber Security Research Centre, Faculty of Science and Engineering, Macquarie

University, Australia in 2016 at IEEE publication[978-1-5090- 2403-2/16] DOI. 10.1109/Bigdatasecurity-HPSC-IDS-2016.

[2]. ”Cyber Crime Investigation in the Era of Big Data”, by Andrii Shalaginov, Jon William Johnsen, Katrin Franke, NTNU Digital forensics group, Faculty of information technology and electrical engineering, Norwegian University. 2017 IEEE International Conference on Big Data.978-1-5386-2715-0/17.

[3].“Crime Analysis and Predictin Using Big Data” by aarthi srinivasnadathur, gayathri narayannan, Indraja ravichandran, srividhya.S,kavalvizhi form department of information technology , SRM Institute of Science and Technology, Kattankulathur,Kancheepuram, Tamilnadu, India. Published in “International Journal of Pure and Applied Mathematics” Volume 119 No. 12, 2018. ISSN: 1314-3395.

[4].”Technologies of safety in the bank sphere from cyber attacks”by Nyrkov Anatoliy .P, Abramova Kritstina.V , Koroleva,Gaskarov from Admiral Kakarov State University of Maritime and Inland Shipping, Russia. 978-1-5386-4340-2/18 @IEEE in year 2018.

[5]. “Crime analysis using K-Means Clustering” by Jyothi Agarwal,Renuka Nagpal, Rajini sehgol form Amity University, Nodia, in the International Journal of Computer Application [0975-8887] volume 83,No 4 December 2013.

[6]. A Data Mining Framework To Analyze Road Accident Data Journal Of Big Data, Sachin Kumar and Durga Toshniwal 2015.

[7]. “Cyber Crime Analysis in Social Media using Data Mining Technique”, by M. Ganesan, P. Mayilvahanan, Department of Computer Science, Vels University. In International Journal of Pure and Applied Mathematics(IJPAM) volume 116 No. 22 [1311-8080] 2017.

[8]. “Survey of Analysis of crime detection techniques using data mining and Machine Learning”, by S. Prabhakaran, and silpa mitra , in National Conference on Mathematical Techniques and its applications [1742-6596].

[9]. “Predictive Modelling of Crime Dataset using Data mining”,by prajakta yerpude and Vaishnavi Gudur, Department of Compute science, in International Journal of Data Mining and Knowledge Process. vol-4 -2017.

[10]. “Using big data Analytics for developing crime predictive model” by Tirthraj chauhan, and Rajanikanth aluvalu in RK university International Conference on Research & Enterpreneurship 2016. (ISBN : 978-93-5254-061-7).

[11]. The Hindu “article” cyber crime reached a new high in 2017.
Published in November 2019.

[12]. “Cyber crime investigations in the Era of Big Data” ,by Andrii Shalaginov, Jan William Johnsen, Katrin Franke, , Department of Information Security and Communication Technology. DOI: 10.1109/BigData.2017.8258362

[13] Subhashini, M., & Gopinath, R., Mapreduce Methodology for Elliptical Curve Discrete Logarithmic Problems – Securing Telecom Networks, International Journal of Electrical Engineering and Technology, 11(9), 261-273 (2020).

[14] Upendran, V., & Gopinath, R., Feature Selection based on Multicriteria Decision Making for Intrusion Detection System, International Journal of Electrical Engineering and Technology, 11(5), 217-226 (2020).

[15] Upendran, V., & Gopinath, R., Optimization based Classification Technique for Intrusion Detection System, International Journal of Advanced Research in Engineering and Technology, 11(9), 1255-1262 (2020).

[16] Subhashini, M., & Gopinath, R., Employee Attrition Prediction in Industry using Machine Learning Techniques, International Journal of Advanced Research in Engineering and Technology, 11(12), 3329-3341 (2020).

[17] Rethinavalli, S., & Gopinath, R., Classification Approach based Sybil Node Detection in Mobile Ad Hoc Networks, International Journal of Advanced Research in Engineering and Technology, 11(12), 3348-3356 (2020).

[18] Rethinavalli, S., & Gopinath, R., Botnet Attack Detection in Internet of Things using Optimization Techniques, International Journal of Electrical Engineering and Technology, 11(10), 412-420 (2020).

[19] Priyadharshini, D., Poornappriya, T.S., & Gopinath, R., A fuzzy MCDM approach for measuring the business impact of employee selection, International Journal of Management (IJM), 11(7), 1769-1775 (2020).

[20] Poornappriya, T.S., Gopinath, R., Application of Machine Learning Techniques for Improving Learning Disabilities, International Journal of Electrical Engineering and Technology (IJEET), 11(10), 392-402 (2020).

[21] Karthikeyan, B., and Dr S. Hari Ganesh. "Encrypt-Security Improved Ad Hoc On Demand Distance Vector Routing Protocol (En-SIm AODV)." ARPN Journal of Engineering and Applied Sciences (ISSN: 1819-6608) 11.2 (2016): 1092-1096.

[22] Karthikeyan, B., N. Kanimozhi, and S. Hari Ganesh. "Analysis of reactive AODV routing protocol for MANET." 2014 World Congress on Computing and Communication Technologies. IEEE, 2014.

[23] Karthikeyan, B., Dr S. Hari Ganesh, and Dr JGR Sathiaseelan. "High Level Security with Optimal Time Bound Ad-Hoc On-demand Distance Vector Routing Protocol(HiLeSec-OpTiB AODV)." International Journal of Computer Science Engineering (E-ISSN: 2347-2693) 4.4 (2016): 156-164.

[24] Karthikeyan, B., N. Kanimozhi, and Dr S. Hari Ganesh. "Performance and analysis of ad-hoc network routing protocols in manet." NCAC (2013): 65-71.

[25] B. Karthikeyan, Detection of Selective Forwarding Attacks in Wireless Sensor Networks, International Journal of Electrical Engineering and Technology (IJEET), 11(9), 2020, pp. 376-392.

[26] B. Karthikeyan, Cluster based Malicious Node Detection for Mobile Ad Hoc Networks, International Journal of Advanced Research in Engineering and Technology (IJARET), 11(12), 2020, pp. 3501-3510.