IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 3, March 2024

Impact Factor: 5.731

Analyzing Impact of Cybercrimes on Society using Data Mining Techniques

Mrs. Sneha R. Paringe[1] and Mrs. Sonali M. Huddar[2]

Assistant Professor, Department of Information Technology[1,2]

Changu Kana Thakur Arts, Commerce and Science College, New Panvel (Autonomous), India

Abstract: *This study was conducted to review cybercrime detection and prevention techniques. Data mining techniques are used to recognize patterns in criminal manners to predict criminal activities and prevent them. Data mining techniques are also used to detect Denial of service attacks and to review alternative techniques that have been used in fraud detection. Cybercrime is defined as, a PC or hardware device used as a tool to dedicate the offense. It means criminal activity through the computer networks violating the rules & regulation and laws. Cybercrime is an illegal act committed using a computer network. It is a subset of computer crime and involves a computer and a network. It is ubiquitous and has become a major security issue. Common forms of cybercrimes are child pornography, cyber stalking, identity theft, cyber laundering, credit card theft, cyber terrorism, drug sale, data leakage, sexually explicit content, phishing, and other forms of cyber hacking. The computer may have been used in the commission of a crime, or it may be the target to identify the different types of fraud and to review alternative techniques that has been used in fraud detection. To reveal the varied cyber-attack strategies adopted by cybercriminals to target various data mining techniques can be used.*

Keywords: Data mining, Denial of service attacks, Cyber-attack, Cybercrime

I. INTRODUCTION

Today's generation is mostly addicted to the virtual world than to the real world because of advancement in emerging technology and its wide spread knowledge leads to security issues, cybercrime, internet hackers and intruders. Cybercrime is defined as, a PC or hardware device used as a tool to dedicate the offense. It means criminal activity through the computer networks violating the rules & regulation and laws. Our online presence is only increasing day by day. Be it shopping for clothes online or buying groceries at a store, with every transaction; we leave behind a digital footprint that online criminals are always looking to exploit. With increased internet usage, the number of cybercrimes has also increased globally.

Cybercrime is an illegal act committed using a computer network. It is a subset of computer crime and involves a computer and a network. It is ubiquitous and has become a major security issue. The computer may have been used in the commission of a crime, or it may be the target. They are offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm, or loss, to the victim directly or indirectly, using modern telecommunication networks such as Internet (Chat rooms, emails, notice boards and groups) and mobile phones (SMS/MMS)".

II. LITERATURE REVIEW

"Review On Cybercrime Prediction with Data Mining Techniques", Farhat Bibi1& Dr. Hamid Ghous2, INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR) ISSN: 2349-4689 Issue 179, Volume 79, Number 01, January 2021

It is observed from this paper that the systematic analysis addressed various methods of mdetecting cybercrimes and reviewed various studies concerning the detection rates achieved and some of the results. This paper guides for the implementation of a cybercrime classification algorithm in which, compared to current techniques, cybercrime can be efficiently detected.

"Role of Data Mining in solving Cyber Crime Problem in India", Shruti Bajaj, Dr. Rajesh Kumar Singh, IJRECE VOL. 5 ISSUE 1 JAN.-MAR. 2017

Data mining has many applications in security including in national security as well as in cyber security. The threats to national security include attacking buildings and destroying critical infrastructures such as power grids and telecommunication systems. Data mining techniques are being used to identify suspicious individuals and groups, and to discover which individuals and groups are capable of carrying out terrorist activities.

Hosseinkhani J., Ibrahim S., Chuprat S., Naniz J.H., Web Crime Mining by Means of Data Mining Techniques, Research Journal of Applied Sciences, Engineering & Technology 7(10) (2013), 2027-2032.

This paper provides advice on how to enhance the privacy of users on social networking sites. To avoid identification of hacking, they advise making the user profiles, personal information for best friends, if you want to reduce the facts about security dangers on Social Networking websites.

Objectives:

- To recognize patterns in criminal manners to predict criminal activities and prevent them.
- To detect Denial of service attacks and to review alternative techniques that have been used in fraud detection.
- To identify the different types of fraud and to review alternative techniques that have been used in fraud detection.
- To use mining techniques to analyse the impact of cybercrime

Purpose:

In today's global and digital world, safeguarding intellectual property, financial information and your company's reputation is a crucial part of business strategy. Securing vital resources and information in the network is the most challenging fear for system enterprises. Also, Communication through social sites leads the people to share the personal life in the public through likes, dislikes, share and status on various social platforms. This will lead to Cybercrime by misuse of Personal Data.

This Project Provides a comprehensive review of Cybercrimes and its impact on society. The study first presents the different types of cybercrimes and discusses their consequences against individuals, organizations, and societies. Second, it comprehensively reviews the data mining techniques for analysing the impact of cybercrime data on society.

III. METHODOLOGY

Data mining is the process of analyzing information, discovering new patterns and data, and predicting future trends. It's often used in scientific research, business development, customer relations, and other spheres. Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions. Data mining is the computer assisted process to break through and analyze large amounts of data and then extract the meaning of data. Data mining techniques like classification, Association Rules, and Clustering are capable of enhancing the accuracy, performance, and speed of predicting cybercrime.

In this project, we are testing 97 samples of people age from 16 to 35 so our sample size is 97. Classification techniques, Association rules and Clustering are used to analyse the data. The above-mentioned techniques classify the data to analyse its impact on people in the society.

Data Mining Techniques:

Classification

Classification is a data mining technique which assigns data instances into predefined target classes. The purpose of classification is to predict the outcome based on given input. In this project various classes have taken to do analysis of collected data. Several kinds of classification algorithms are available like C4.5, ID3, K-nearest neighbour classifier, Naive Bayes etc. that help in such a decision-making process.

IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 3, March 2024

Impact Factor: 5.731

In this project, C4.5 algorithm is used that builds decision trees from a set of training data using the concept of information entropy. The training data is a set S=s1,s2,..... of already classified samples. Each sample consists of a p-dimensional vector ….,(X1i, X2i,....Xpi) , Where Xi represent attribute values or features of the sample, as well as the class in which Si falls. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sublists.

Generally, a classification technique follows 3 approaches Statistical, Machine learning and Neural Network for Classification. While considering these approaches, Machine Learning approach is best suitable for this project. J48 is an open-source Java implementation of the C4.5 algorithm in the Weka data mining tool.

Association Rules:

Association Rules are used to show the relationships between data items. It is used for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. Association is a data mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules. Association rules are calculated from itemset. If rules are generated from all possible itemset, there may be a very high number of rules and rules may not be very meaningful.

The problem of association rule mining is defined as:

Given a set of items I={I1,I2,.....,Im} and a database of transactions D={t1,t2,.....,tn} where ti={Ii1, Ii2,....,Iik} and Iij ∈ I, an association rule is an implication of the form X⇒ Y where X,Y⊂ I are sets of items called itemsets and X∩ Y= ∅ [1]. The strength of the Association rule can be measured in terms of its support and confidence.

The support(s) for an association rule X⇒ Y is the percentage of transaction in the database that contains X∪ Y. It measures how often it should occur in the database.

The confidence (α) for an association rule X ⇒Y is the ratio of the number of transactions that contains X ∪ Y to the number of transactions that contain X, it measures the strength of rule.

Clustering (K-Means):

Clustering is the task of dividing the population or data points into a number of groups as data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is a collection of objects on the account of similarity and dissimilarity between them.

WEKA Tool:

Weka contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to these functions. The original non-Java version of Weka was a Tcl/Tk front-end to (mostly third-party) modelling algorithms implemented in other programming languages, plus data pre-processing utilities in C and a make file-based system for running machine learning experiments.

This original version was primarily designed as a tool for analyzing data from agricultural domains. Still, the more recent fully Java-based version (Weka 3), developed in 1997, is now used in many different application areas, particularly for educational purposes and research.

Data Preprocessing:

The main objective of the project is to evaluate the impact of cybercrime on society. The task to evaluate the performance is difficult because we need to consider all the parameters on which the final results depend. As per the weightage of the parameters and the collected data, attributes to be analysed for cybercrime and its impacts are created in pre-processing phase of the project. These attributes are as follows:

IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 3, March 2024

Impact Factor: 5.731

- Full Name - Name of the person
- Which age group do you fall into? - Age to fit the person in different age group
- Are you currently using your web browser or a password manager app to save or create passwords? - Whether person is using default system generated passwords.
- Have you enabled Two-factor authentication (2FA) for any of your online accounts?
- Do you regularly backup your devices? E.g., Phones, laptops, tablets?
- Do you install software updates on your devices when they become available?
- How safe do you feel about your information, when you are online?
- Have you ever lost money due to Cyber Crime?
- Have you ever experienced any of these situations?
- Have your ever heard denial of service attack?
- Have you stopped shopping online due to this issue?
- How many times have you been a victim of a cybercrime?
- Do you think that the laws in effect are able to control cyber criminals?
- Have you ever experienced any type of fraud?
- Have you ever experienced any type of fraud? If Yes, Mentioned it.
- For statistical parameters, answers given by people for different attributes according to the experience of the person. Here, parameters like difficulties faced by people while using the internet, fraud, experience with spam or phishing mails are analysed. for every person the above-mentioned parameters have to analyze so that we can check the impact of cybercrime on society.

Data Analysis:

We collected data from the survey method via Google form questionnaire.

IV. RESULT AND DISCUSSION

The main objective of the project is to evaluate the impact of cybercrime on society. The task to evaluate the performance is difficult because we need to consider all the parameters on which the final results depend. As per the weightage of the parameters and the collected data, attributes to be analysed for cybercrime and its impacts are created in pre-processing phase of the project. These attributes are as follows:

1. Full Name - Name of the person
2. Which age group do you fall into? - Age to fit the person in different age group
3. Are you currently using your web browser or a password manager app to save or create passwords? - Whether person is using default system generated passwords.
4. Have you enabled Two-factor authentication (2FA) for any of your online accounts? -
5. Do you regularly backup your devices? E.g., Phones, laptops, tablets?
6. Do you install software updates on your devices when they become available?
7. How safe do you feel about your information, when you are online?
8. Have you ever lost money due to Cyber Crime?
9. Have you ever experienced any of these situations?
10. Have your ever heard denial of service attack?
11. Have you stopped shopping online due to this issue?
12. How many times have you been a victim of a cybercrime?
13. Do you think that the laws in effect are able to control cyber criminals?
14. Have you ever experienced any type of fraud?
15. Have you ever experienced any type of fraud? If Yes, Mentioned it.

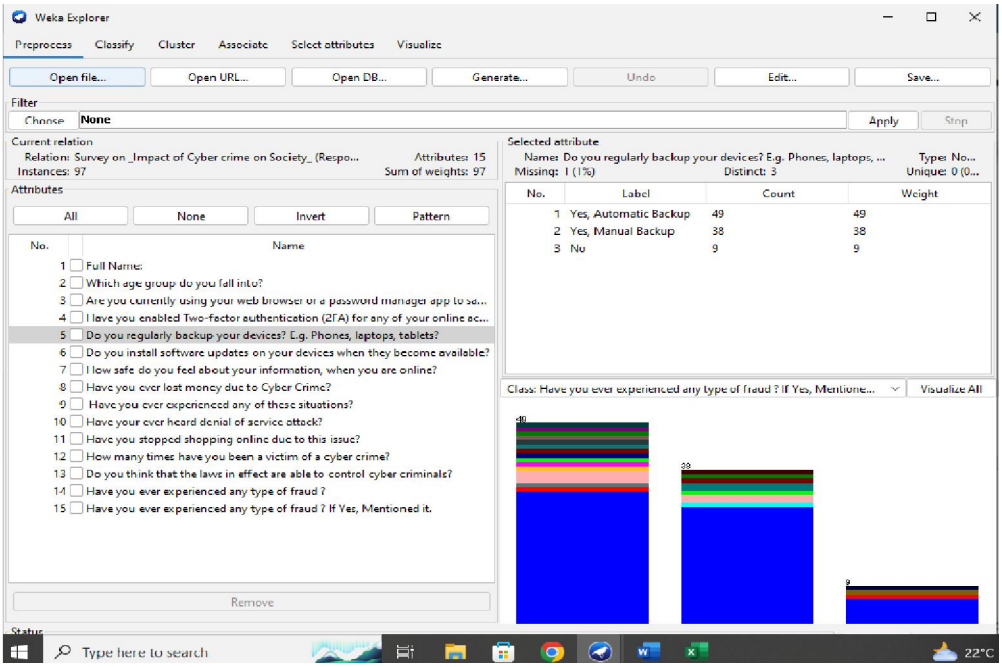
IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 3, March 2024

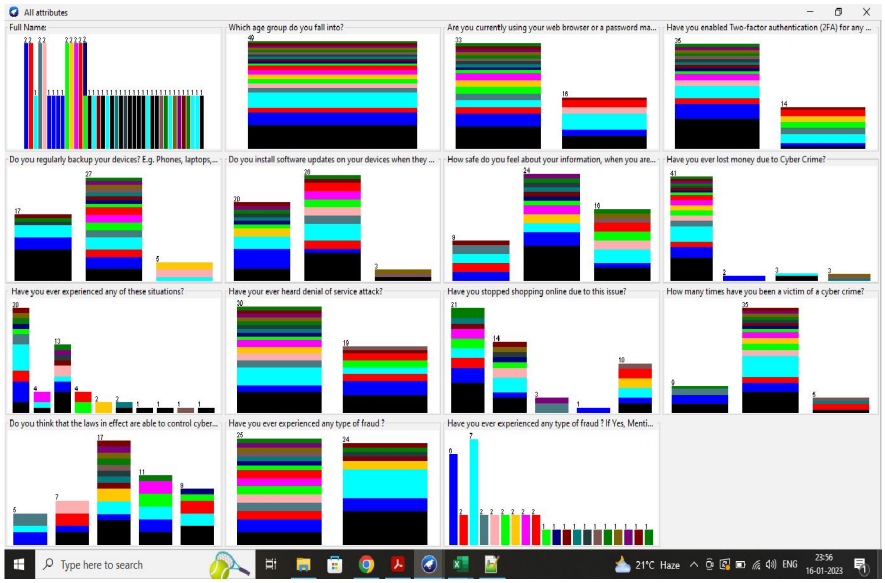
Impact Factor: 5.731



The above window shows that the data is ready to analyze. It shows all the attributes from our file in the attribute box. We can keep only those attributes which we want to analyze. The attributes which are not necessary for the classification process, we can remove those attributes from here.

Following window shows the description of all the instance details related to each attributes. For example, we have 35 instances who knows Two Factor Authentication and 14 instances who didn't know about Two- Factor Authentication. These details gives information that 35 people are aware about this security feature.

Thus, when we look for the Experience to cybercrime attribute, we can say that there are 97 instances who have experienced cybercrime while using internet.



IJETIR

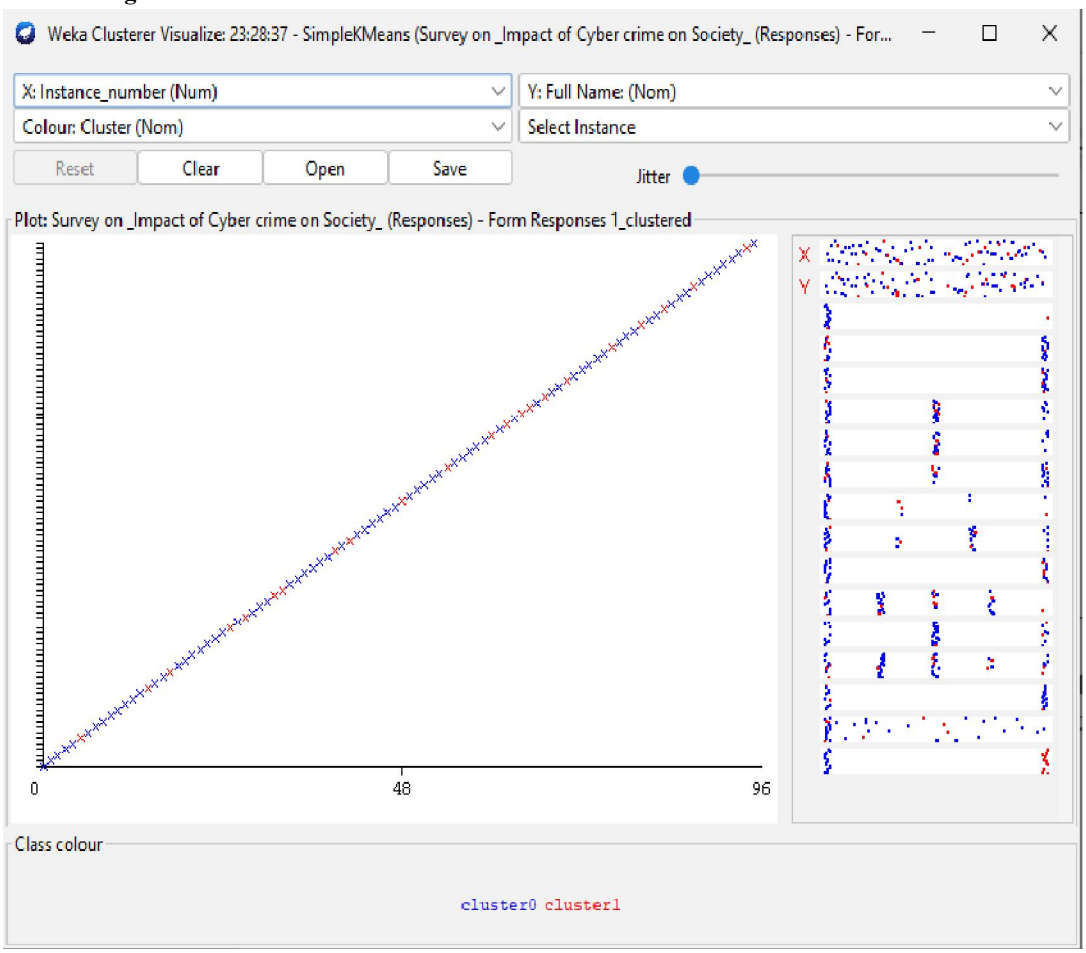
ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 3, March 2024

Impact Factor: 5.731

Attribute Ranking:



IV. CONCLUSION

Data Mining Techniques are efficiently used in this project for analyzing impact of Cybercrime on society using data mining techniques such as classification, Association rules and Clustering. Collected data is analysed based on various attributes selected by using classification, association rules and clustering with the help of Weka Machine Learning Tool. Analyzed data from different aspects are as follows:

Data mining techniques mentioned in this project are very useful for society people to be aware of cybercrimes. These techniques can also identify the areas where this cybercriminals mainly attack. People can take precautions so that they cannot face any difficulties while using various online platforms, telecommunications fraud etc.

REFERENCES

[1]. OnlineAvailable:https://www.idsa.in/system/files/book/book_indiacybersecurity.pdf.

[2]. Online Available:https://en.wikipedia.org/wiki/Cybercrime.

[3]. Shruti Bajaj, Dr. Rajesh Kumar Singh ―Role of Data Mining in solving Cyber Crime Problem in India, IJRECE VOL. 5 ISSUE 1 JAN.-MAR. 2017, ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE)

[4]. Https://www.statista.com/statistics/309435/india-cyber-crime-it-act/https://www.moneycontrol.com/news/india/cyber-crimes-in-india-rise-6-a-year-in-2021-telangana-tops-list-ncrb-data-9115161.html

[5]. [6]https://zeenews.india.com/technology/two-months-of-2022-saw-more-cyber-crimes-than-entire-2018-why-e-fraud-is-a-ticking-time-bomb-2458733.html

[6]. Https://factly.in/fraudulent-website-collecting-user-information-in-the-name-of-oppo-phone-giveaway

[7]. Https://www.geeksforgeeks.org/k-means-clustering-using-weka/

[8]. V. Jain, A. Bhatia, Y. Sharma, and V. Arora, ―Crime Prediction using K-means Algorithm,‖ vol. 2, no. 5, pp. 206–209, 2017.

[9]. Journal, M. Vol, I. Factor, and J. Homepage, ―IMPLEMENTATION OF DATA MINING TECHNIQUES FOR CYBER CRIME DETECTION K. Chitra Lekha □ Dr. S. Prakasam □□,‖ vol. 7, no. 4, pp 607–613.

[10]. Chewae M., Hayikader S., Hasan M.H., Ibrahim J., How Much Privacy We Still Have on Social Network?, International Journal of Scientific and Research Publications 5(1) (2015).

[11]. "Review On Cybercrime Prediction with Data Mining Techniques", Farhat Bibi1& Dr. Hamid Ghous2, INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR) ISSN: 2349-4689 Issue 179, Volume 79, Number 01, January 2021

[12]. "Role of Data Mining in solving Cyber Crime Problem in India", Shruti Bajaj , Dr. Rajesh Kumar Singh, IJRECE VOL. 5 ISSUE 1 JAN.-MAR. 2017

[13]. [13] Hosseinkhani J., Ibrahim S., Chuprat S., Naniz J.H., Web Crime Mining by Means of Data Mining Techniques, Research Journal of Applied Sciences, Engineering & Technology 7(10) (2013), 2027