International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.6, June 2021 E-ISSN: 2347-2693

Crime Analysis and Prediction Model Using Data Mining and Machine Learning Techniques: Comparative Analysis

Neetu Singh[1]*, Tripti Ajariya[2], Shailesh Raghuvanshi[3], Neha Singh[4]

[1,2]Dept. of Computer Science, Bhabha University, Bhopal-462047 (M. P) India
[3,4]Department of Physics, Bhabha University, Bhopal-462047 (M.P) India

Corresponding Author: dangineetu1991@gmail.com

DOI: https://doi.org/10.26438/ijcse/v9i6.97104 | Available online at: www.ijcseonline.org

Received: 17/Jun/2021, Accepted: 20/Jun/2021, Published: 30/Jun/2021

***Abstract*—** Crime is the illegal task perform by human system for perform unauthorise3d access of various services without legal permission. in computer system there are many categories of crime related with cybercrime. the crime is basically relating to offence which is happen by illegal task and permission. N this computer system the many crime is describing various cybercrime. Like malware, hacking, cyber-attacks, cyber illegal access, IOT hacking, phishing scams, ransom ware, etc.

These are described various types of cybercrime in computer technology. for determine the any one we have used various algorithm and simulation tool performance. Thyis is destroy our human society transaction system. It is also created by some human; many companies also do this illegal work for our business growth. This cyber various crime is also impact of our country growth and development also decreased.

In this research work we have determinate cybercrime attacks detection and prediction model of determine the cyber-attacks using data mining and machine learning techniques. For implementation of this proposed work we are using WEKA simulation tool for execute data mining algorithms and Jupiter anaconda navigator simulate tool for machine learning algorithms for determinate accuracy of data mining and design production model using machine learning algorithms.

The attacks categories by cybercrime are ddos attacks, U2r, R2L, probe, ICMP attacks, UDP attacks, TCP attacks, FAR high false alarm rate, Dos attacks, these are cyber-attacks detection by data mining and machine learning algorithms based determinate accuracy and efficiency of data mining algorithms and prediction model by machine learning algorithms

Keywords—cybercrime, cyber-attacks, data mining, KDD cup data set, machine learning, weak tool, Jupiter anaconda navigator simulate tool.

I. INTRODUCTION

Cybercrime is primary example of cross-border crime. Computer networks connect all countries of the world, and evil-doers can cause significant harm anywhere in the world without leaving home desk. The potential harm is remarkably varied, ranging from individuals not being able to access them personal computer for few hours or stumbling across racist or obscene material on the Internet, to a company's internal network being inaccessible for 24 hours or trade secrets being stolen, and to government's public websites being blocked or seeing state secrets appear on the web. Financial losses range from a few hundred dollars being extorted to multi-million dollar losses caused by cyber fraud or cyber sabotage. Increasingly, as the internet penetrates ever more into world's core activities, Cybercrime also involves the risk of terrorist attacks bringing down major part of the internet, and therewith causing an economic and social disaster on global scale.

The criminal abuse of information technology and the necessary legal response are issues that have been discussed ever since the technology was introduced. Over the last 50 years, various solutions have been implemented at the national and regional levels. One of the reasons why the topic remains challenging is the constant technical development, as well as the changing methods and ways in which the offences are committed.

The significance of the study lies in the fact that internet, which is decentralized mode of communication has given birth to new form of crime called Cybercrime. The researcher via his work tries to analyse what kind of jurisdictional problems the court have to face while resolving disputes relating to Cybercrimes. Society is facing crimes like hacking, launching virus, password sniffing, cyber stalking, internet gambling, spamming, software piracy etc. what has become worst night mare is the then law enforcing machinery fails in this cyber jungle. The moot question that has troubled all most all the nations is-"which court has jurisdiction to resolve the Cybercrime case." Use of cyberspace is a unique and entirely new means of mass communications that is located in no specific geographical location and yet is accessible to anyone, anywhere, who has availability to it through computer link to the internet. Consequently, there is no single organization or nation that control membership in

his virtual land, nor is there a centralized location from which access is regulated.

Any study is futile without a reasonable objective. The researcher believe that the increasing dominance and wide spread scope of cyber space and Cybercrimes represents a major breakdown of international law and poses jurisdictional challenges to the courts in combating Cybercrimes. Following are the main study points with which this study will be conducted:

- To develop a conceptual understanding of cyberspace and Cybercrime.
- To understand Freedom of Speech and Human Right issue in cyberspace.
- To examine the need for regulating the cyberspace.
- To point out the legal implication of Cybercrime and cyberspace.
- To understand the global character of Cybercrime and different types of Cybercrime.
- To highlight the challenges posed by the Cybercrime with special emphasis on jurisdictional issues.
- To develop understanding of jurisdictional theories and there applicability to transnational Cybercrime.
- To analyse, are the International laws are strict enough to enforce an arrest in cases involving extra territorial jurisdiction in connection with Cybercrime.
- To present an analysis of jurisdictional provisions of Convention on Cybercrime and to highlight the existing flows in the convention.
- To portray the existing state of Cybercrime in India. To examine the extraterritorial provisions of IT Act, 2000 and present a critical analysis of its effectiveness.
- To suggest the future solution to the jurisdictional issues posed by trans-border nature of Cybercrime.

II. TYPES OF CYBER CRIME

The cybercrime may be broadly classified into three groups:

- Crime against the Individuals a. Person b. Property of an individual.
- Crime against Organization a. Government b. Firm, Company and Group of Individuals.
- Crime against Society The following are the crimes that have been committed against the followings group: Harassment via electronic mails b. Dissemination of obscene material. c. Cyber-stalking. d. Defamation. e. Indecent exposure. f. Cheating. g. Unauthorized control/access over computer system. h. Email spoofing. i. Fraud.
- AGAINST INDIVIDUAL PROPERTY a. Computer vandalism b. Transmitting virus. c. Net repass. d. Unauthorized access / control over computer system e. Intellectual Property crimes f. Internet thefts.
- AGAINST ORGANIZATION a. Unauthorized access / control over computer system. b. Cyber terrorism against the government organization. c. Possession of unauthorized information. d. Distribution of Pirate software.
- AGAINST SOCIETY a. Child pornography c. Indecent exposure of polluting the youth financial crimes. d. Sale of illegal articles. e. Trafficking. f. Forgery. g. Online gambling.

III. CATEOGRIES OF CYBER CRIMINALS ASPECT OF ITCRIMES:

Coders They are the comparative veterans of the hacking community. With a few years' experience at the art and a list of established contacts, ‘coders’ produce ready-to-use tools (Trojans, mailers, custom bots) or services (such as making a binary code undetectable to AV engines) to the cybercrime labour force – the ‘kids’. Coders can make a few hundred dollars for every criminal activity they engage in.

Drops These individuals convert the ‘virtual money’ obtained in cyber crime into real cash. Usually located in countries with lax e-crime laws (Bolivia, Indonesia and Malaysia are currently very popular), they represent ‘safe’ addresses for goods purchased with stolen financial details to be sent, or else ‘safe’ legitimate bank accounts for money to be transferred illegally, and paid out of legitimately.

Mobs These are professionally operating criminal organization which combines all of the above covered functions. Organized crime makes particularly good use of safe ‘drops’, as well as recruiting accomplished ‘coders’ onto their payrolls.

White hat hacker: A white hat hacker is an ethical hacker who ethically oppose to the abuse of computer systems. A white hat generally focuses on securing IT systems, whereas a black hat (the opposite) would like to break into them. The term white hat hacker is also often used to describe those who attempt to break into systems or networks in order to help the owners of the system by making them aware of security flaws, or to perform some other altruistic activity.

Black hat hackers: A black hat is a person who compromises the security of a computer system without the permission of authorized party, typically with malicious intent. The somewhat similar activity of defeating copy prevention devices in software which may or may not be legal in a country's laws is actually software cracking. The primary difference between white and black hat hackers is that a white hat hacker claims to observe ethical principles. Like black hats, white hats are often intimately familiar with the internal details of security systems, and can delve into obscure machine code when needed to find a solution to a tricky problem. Some use the term grey hat and fewer use brown hat to describe someone's activities that crosses between black and white.

Grey Hat Hackers: A Grey Hat in the computer security community, refers to a skilled hacker who sometimes acts legally, sometimes in good will, and sometimes not. They are a hybrid between white and black hat hackers. They

usually do not hack for personal gain or for malicious intentions, but may or may not occasionally commit crimes during the course of their technological exploits.

Internet Crime Hackers: Internet crime hackers commit crime on the internet, using the Internet and by means of the Internet. Internet crime is a general term that includes crimes such as phishing, credit card frauds, bank robbery, illegal downloading, industrial espionage, child pornography, kidnapping children via chat rooms, scams, cyber terrorism, creation and/or distribution of viruses, Spam and so on. All such crimes are computer related and facilitate crimes. The different types of Internet crime vary in their design and easy accessibility to be committed. It can be separated into two different categories.

Blackmail Hackers: Blackmail is a long-established illegal act that has been given a new twist in the modern age. The blackmailer may threaten to release embarrassing or other harmful information via the Internet or a private network if the victim does not comply with the demands of the criminal. A cybercrime of this type may go as far as having the victim transfer funds to an untraceable bank account using some type of online payment program, thus making full use of modern technology to commit the crime Blackmail Hackers hack.

Official government websites. ii) On line credit card scam iii) Work on cyber-crime operations and they make money.

IV. TYPES OF CYBER CRIME:

The meaning of cyber-crime may be any crime that is committed by means of the special knowledge or expert use of computer technology; harmful acts committed from or against a computer or network; an unlawful act wherein the computer is either a tool or a target or both. Hence the types of cyber-crimes, in the opinion of scholars are as following:

- **Cyber-crime can be categorized as: Software related crimes, Data related crimes. , Physical crimes. , Internet and other computer related crimes.**

SOFTWARE RELATED CRIMES:

- Unauthorized Access.
- Salami Attack.
- Logic Bomb.
- Virus/Vorm Attack.
- Trojan Attack.
- Intellectual Property Crime.

DATA RELATED CRIMES:

- Data Diddling.
- Data Leakage.
- Data Spying.
- Scavenging.

PHYSICAL CYBER CRIMES:

- Theft.
- Breakage.
- Destroying Data, Output or Media.
- Inter-Processing Manipulations:

INTERNET RELATED CYBER CRIMES:

- E-mail Bombing.
- Internet Time Theft.
- Cyber Pornography.
- E-mail Spoofing.
- Cyber Stalking.
- Password Attacks.
- Brute-Force Attack.
- Main-in-the-middle.
- Hacking.
- Black mailing.
- Fraud creation.

Important Security Issues:

- Cyber-attacks is the cloud main issues right now.
- Security issues.
- Restricted access for users.
- External shairing of data.
- Lack of visibility.
- Hijacking of acoounts.
- Insecure interfaces/APIrules and interface.

threats:

The cloud computing has many advantageous like speed, efficiency, dynamic scalling of product application on runtime sevieces on various enviorment and infrastructure based scalling. But so many harm and attcks also occurring with cloud computing as repect to service model representation of cloud based appplications.

The various threats are given below for cloud computing:

1. DDOS attacks.
2. Malware threats detecton.
3. Hackers attcks.
4. Cross scripting attcks .
5. Volinearble attacks.
6. Malware based cloud computing attacks,
7. Data breaches.
8. Human error .
9. Systax based errors.
10. Security issues and challenges consist varioys threats based application.

V. OBJECTIVE OF THE PROPOSED WORK

Aims:

The main is to determine the cyber-attacks by using data mining algorithms and machine learning algorithms based prediction model.

Objectives:

1. To determine the U2r, R2l attacks by suing data mining techniques.
2. To design prediction model by machine learning algorithms.

3. To determine the accuracy & efficiency of data mining algorithms for detection of cyber-attacks.
4. To determinate predictor of cybercrime based attacks detection using Jupiter anaconda navigator simulate tool.

VI. LITERATURE SURVEY: SLR.

The DATA MINING AND MACHINE LEARNIING based literature review consist various international and national authors papers which is involved for specific challenges, services and security problem solving siisues and resolved techniques representation.

01.Juels et al. [07] this paper presents paper on cloud computing and scope of challenegs and security issues for cloud computing on the basisi of various deployment models.

Shacham et al. [22] in this authors review on various threats and infrastrcuure on the basis of various fiels measurement on the critics based analysis.

02.Arbors et al. [23] The adequacy of their plans lays on an exceptionally fundamental level on the preprocessing steps that the customer directs prior to re-appropriating the information report F. Any change to the substance of F, several pieces, should spread through the blunder adjusting code, henceforth introducing critical computation and correspondence intricacy.

03.Ateniese et al. [24] in this author reprsesnt the clous attcks security iissues is basically belongs to storage and data based system . the access of cloud service is represent the main feature of cloud based services.

04.In [28] this paper, we execute DAA convention for the security of the individual devicesunder the cloudcomputing climate. It moreover describes TPM module using JAVA. Additionally, we have attempted affirmation and correspondence of each hub through DAA convention. Additionally, the proposed model chooses effectively security even out and will control for the ordinary resources.Therefore, it should give fitting security administrations as per the unique changes of the normal assets. We can anticipate that our proposed model should tackle the security issues in the static climate utilizing MAUT and straightforward heuristics 05.In [11] The proposed distributed storage security model gives an exceptionally secure cloud climate by acquainting the three areas with store client information base on the security boundaries specifically validation, classification, uprightness, accessibility, non-renouncement, security, and protection. It confines unapproved elements to deal with the client's information by carrying out twofold confirmation components. It likewise gives insurance against different security breaks, for example, beast power assault, disguise assault, information altering, and cryptanalysis of respectability key.

06.In [12], maker presented the five pivotal qualities of distributed computing, three cloud administration models, and four cloud sending models. Exploration in the protected distributed storage is disturbed by the way that user'sdata may be kept at a couple of regions for one or the other repetition/adaptation to internal failure to non-basic disappointment or because the organization is given through a chain of organization providers. Examined the safety efforts got by the majorcloud specialist organization (Amazon webservices or AWS) including their system security and security best practices took after by AWS.

VII. PROPOSED KF MODEL

The proposed IDS has three major functionalities which implemented using WEKA Java API for machine learning. Firstly, the IDS consists with a dataset pre-processing technique such as, attribute selection, attribute filtering, and instance filtering. Secondly, the IDS consists with a Bayesian network classification model, which is the key component in the system, which does the classification of the network data. Thirdly, Inference analyser which has designed as the prediction module for incoming testing network traffic. These modules named as data pre-processing, decision tree and KF model learning and inference algorithm module to classify the incoming new data respectively.

VIII. ISSUES WITH THE PRESENT CYBER ATTAKS DETECTION SYSTEM.

The following issues have been identified in the presently available CYBER ATTACKS Detection System:

High False Alarm Rate (FAR)
Not completely adaptable
Low detection rate of u2r type of attacks
Low detection rate of r2l type of attacks

PERFORMANCE ANALYSIS

Table 1

Parameters	Definition
True Positive (TP) or Detection Rate (DR)	Attack occur and alarm raised
False Positive (FP)	No attack but alarm raised
True Negative (TN)	No attack and no alarm
False Negative (FN)	Attack occur but no alarm

Detection rate (DR) and false positive (FP) are used to estimate the performance of intrusion detection system [17], which are given as below:

$$DC = \frac{Total\ Detected\ Attacks}{Total\ Attacks} \times 100$$

$$FP = \frac{Total\ misclassified\ process}{Total\ Normal\ Process} \times 100$$

IX. KDDCUP'99 DATASET

The classes in KDD99 dataset can be categorized into 5 main classes (one normal class and four main intrusion classes: probe, DOS, U2R, and R2L).

1) *Normal* connections are generated by simulated daily user behaviour such as downloading files, visiting web pages [4].
2) *Denial of Service (DoS)* attack causes the computing power or memory of a victim machine too busy or too full to handle legitimate requests. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users like apache2, land, mail bomb, back, etc.
3) *Remote to Local (R2L)* is an attack that a remote user gains access of a local user/account by sending packets to a machine over a network communication, which include send-mail, and Xlock.
4) *User to Root (U2R)* is an attack that an intruder begins with the access of a normal user account and then becomes a root-user by exploiting various vulnerabilities of the system. Most common exploits of U2R attacks are regular buffer-overflows, load module, Fd-format, and Ffb-confige.
5) *Probing (Probe)* is an attack that scans a network to gather information or find known vulnerabilities. An intruder with a map of machines and services that are available on a network can use the information to look for exploits.

X. RESEARCH METHODOLOGY

Decision tree: algorithm
DT induction is one of the classification algorithms in data mining. The classification algorithm is inductively learned to construct a model from the pre-classified data set. Inductive learning means making general assumptions from the specific examples in order to use those assumptions to classify unseen data. The inductively learned model of classification algorithm is known as classifier. Classifier may be viewed as mapping from a set of attributes to a particular class. Data items are defined by the values of their attributes and *X* is the vector of their values **$\{x_1, x_2,\ldots\ldots,x_n\}$,** where the value is either numeric or nominal. Attribute space is defined as the set containing all possible attribute vectors and is denoted by *Z*. Thus X is an element **of *Z* ($X \in Z$). The set of all classes is denoted by C = $\{c_1, c_2,\ldots\ldots,c_n\}$. A classifier assigns a class $c \in C$ to every attribute of the vector *X*** $\in$ Z..

J48: algorithms:
The decision tree learning is one of the machine learning approaches for generating classification models. In this paper, C4.5, a later version of the ID3 algorithm [49], will be used to construct the decision trees for classification. In ID3, a decision tree is built where each internal node denotes a test on an attribute and each branch represents an outcome of the test. The leaf nodes represent classes or class distributions. The top-most node in a tree is the root node with the highest information gain. After the root node, one of the remaining attribute with the highest information gain is then chosen as the test for the next node. This process continues until all the attributes are compared or when all the samples are all of the same class or there are no remaining attributes on which the samples may be further partitioned.

XI. SIMULATION AND RESULT ANALYSIS:

Implementation file the implementation performs following points:

- **Jupiter notebook python and machine learning tool**
- **weka simulation tool.**

1. implementation start on Jupiter notebook 1.6.3 simulation tool.
2. anaconda navigator are used to run python and machine learning algorithms.
3. weka simulation tool are used for design kf model like knowledge flow model and network file access for create virtual network for detection of cyber-attacks on networks.
4. after that we are simultaneously using Jupiter notebook 1.6.3 python and machine learning implementati0on simulation tool.

- **weka simulator:**
- **Jupiter notebook 1.6.3 using anaconda.**
- **kf model file.**
- **network file – virtual network system for create cyber-attacks detection.**

Table 2

Measurement	Single classifier Naïve Bayes	Decision Tree Based classification
Accuracy	83.19	99.60
Detection rate	94.70	99.80
False Alarm	19	0.50

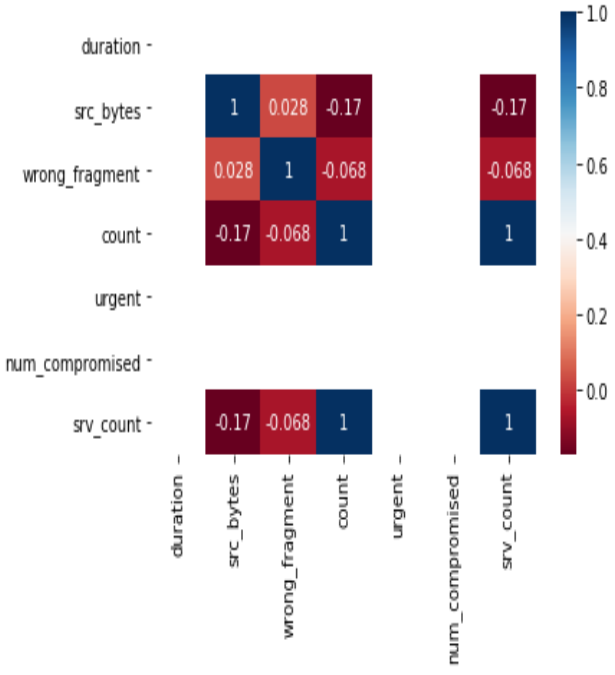

Fig. 1

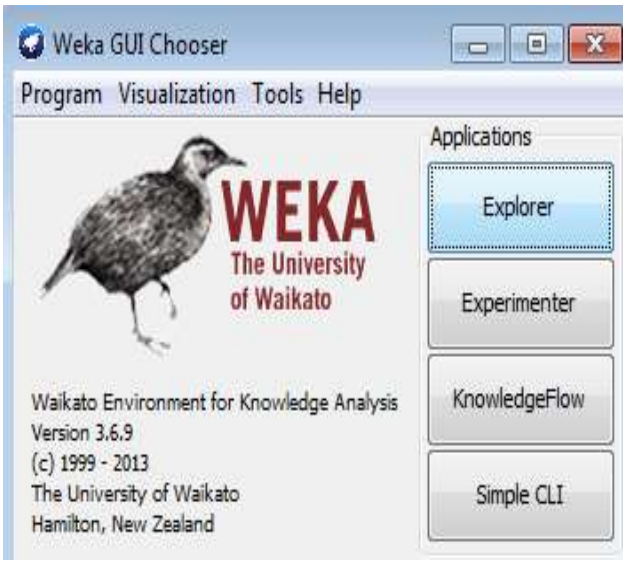

Fig. 2 WEKA SIMULATE

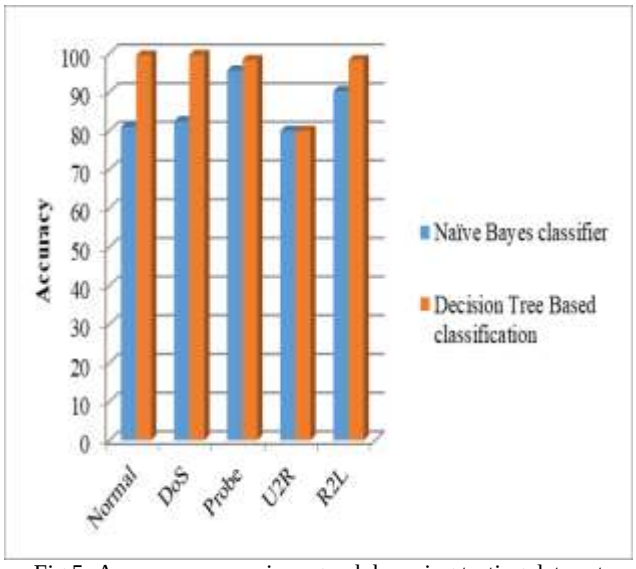

Fig.5: Accuracy comparison graph by using testing data set

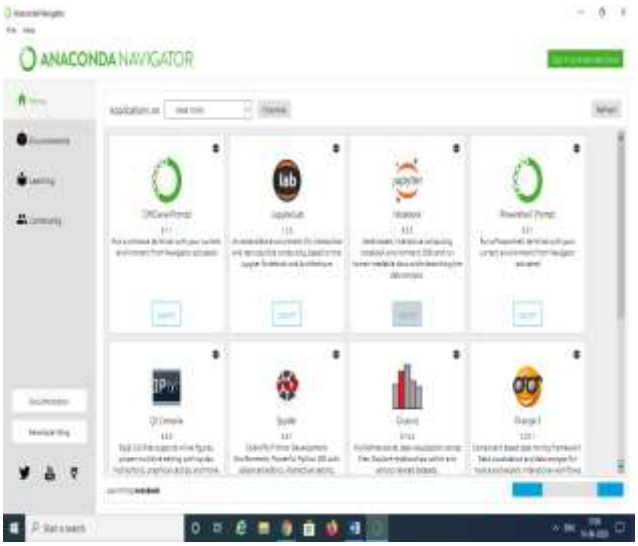

Fig.3: JUPITER ANACONDA NAVIGATER.

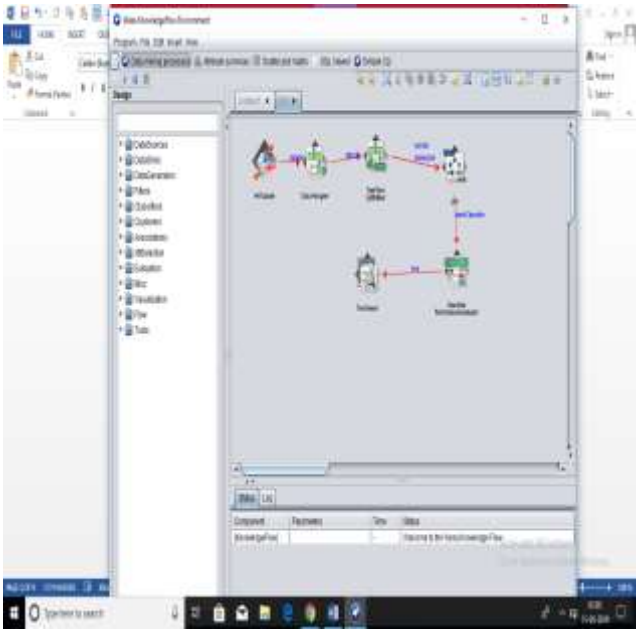

Fig.6

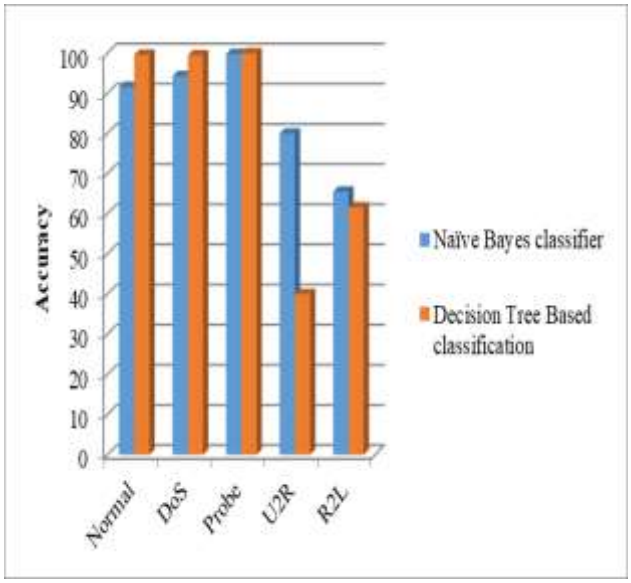

Fig.4: Accuracy comparison graph by using training data set

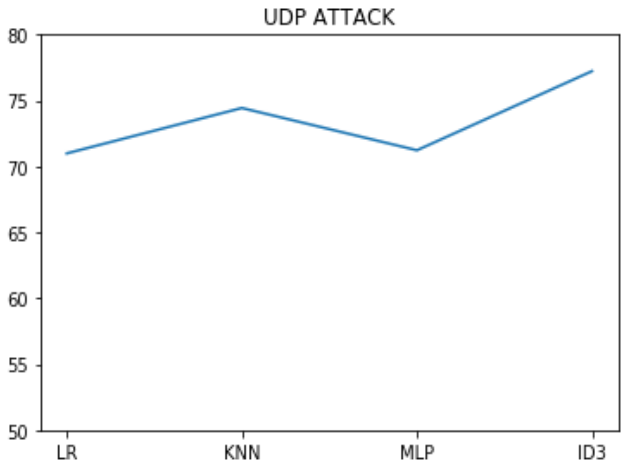

Fig.7

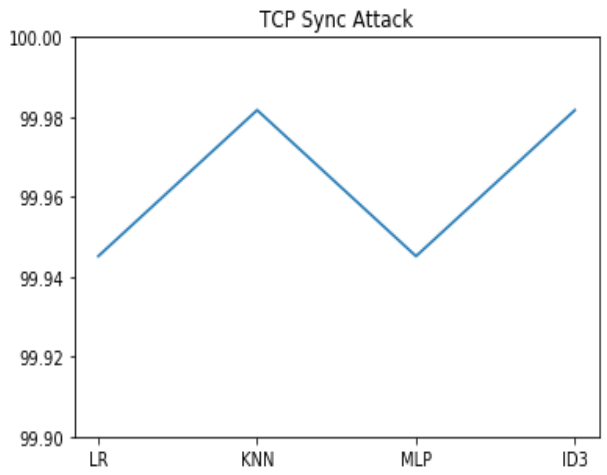

Fig.8

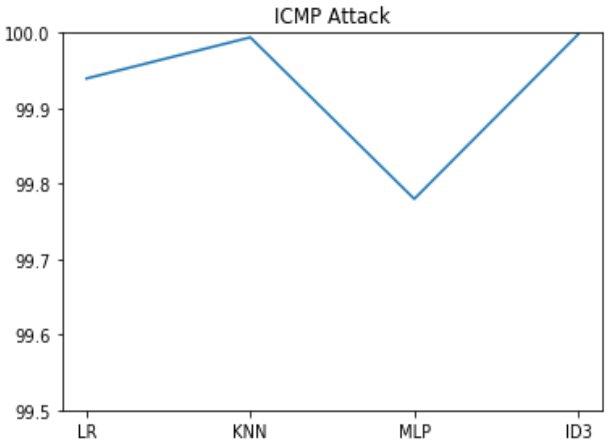

Fig.9

XI.CONCLUSION

Authentication is needed for giving upgraded security in cyber-crime prevention by cyber-attacks. In this PAPER introducing cyber-attacks security and prediction model by machine learning approach.

The Denial of administration assault is one among the sort of dynamic assault. The Denial of administration assaults that income that the assailants will send bound messages that is in danger of the framework. Regularly they send bundles to the objective framework which can prompt failure [2].

This DDoS assault become more perilous with regards to cloud as a result of its heap appropriating nature. This anyway start from a solitary actual machine yet may defile the total cloud inside no time. Consequently, a proactive method to shield from DDoS assault is proposed. Result shows that when this plan is utilized with cloud DDoS assaults have been limited without bringing about any extensive overhead.

The recommended approach called Decision Tree Based arrangement is assessed and contrasted and the single Naïve Bayes classifier utilizing KDD Cup '99 informational collection. The test results show that the k Decision Tree Based order approach accomplishes better exactness and discovery rates while lessening the bogus caution by identifying novel interruptions precisely. The exhibition of Naïve Bayes classifier has been improved by applying Decision Tree Based characterization. Notwithstanding, Decision Tree Based grouping has limit to recognize interruptions that are basically the same with one another like U2R and R2L.

FUTURE WORK

In future we apply optimizatyptoion based cryptographic procedure, on the grounds that advance based crypto method are light weighted. The following future work depends on IOT with distributed computing application representation simultaneously.

REFERENCES

[1]. Halder, D., &Jaishankar, K. (2011) Cybercrime and the Victimization of Women: Laws, Rights, and Regulations. Hershey, PA, USA: IGI Global. ISBN 978-1-60960-830-9, **2011.**

[2]. Lambert, Glenn M. II, "Security Analytics: Using Deep Learning to Detect Cyber Attacks" (2017). UNF Graduate Theses and Dissertations. 728, https://digitalcommons.unf.edu/etd/728

[3]. ManjeetRege& Raymond Blanch K. Mbah, Machine Learning for Cyber Defense and Attack , DATA ANALYTICS 2018 : The Seventh International Conference on Data Analytics, Copyright (c) IARIA, 2018. ISBN: 978-1-61208-681-1 , **pp.73–78., 2018.**

[4]. Dmitri Koteshov, How Can Ai Change The State Of Cybersecurity, **March 7, 2018,** https://www.elinext.com/industries/financial/trends/aiand-security/.

[5]. Anti-Phishing Working Group, “Phishing and Fraud solutions”. [Online], [Accesses: March 18, 2019] http://www.antiphishing.org/.

[6]. S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, “A Comparison of Machine Learning Techniques for Phishing Detection”, APWG eCrime Researchers Summit, Pittsburg, PA. **October 4-5, 2007,**

[7]. N. Lu, S. Mabu, T. Wang, and K. Hirasawa, “An Efficient Class Association Rule-Pruning Method for Unified Intrusion Detection System using Genetic Algorithm”, in IEEJ Transactions on Electrical and Electronic Engineering, **Vol. 8, Issue 2, pp. 164 – 172, January 2, 2013.**

[8]. Knowledge Discovery and Data Mining group, “KDD cup 1999” [Online], [Accessed:March 18, 2019], http://www.kdd.org/kddcup/index.php.

[9]. K. Lee, J. Caverlee, and S. Webb, “Uncovering social spammers: social honeypots + machine learning”, SIGIR’10, Geneva, Switzerland, **July 19-23, 2010.**

[10]. Nilaykumar Kiran Sangani & HarootZarger, Machine Learning in Application Security, [Accessed: March 18, 2019] http://dx.doi.org/10.5772/intechopen.68796.

[11]. Security Week Network. Symantec Adds Machine Learning to Endpoint Security Lineup [Internet]. 2016. Available from: http://www.securityweek.com/symantec-addsmachine-learning-endpoint-security-lineup .

[12]. Ozlem Yavanoglu & Murat Aydos, A Review on Cyber Security Datasets for Machine Learning Algorithms, 11-14 Dec. 2017, 2017 IEEE International Conference on Big Data (Big Data), INSPEC Accession Number: 17504859.

[13]. Md. Zeeshan Siddiqui &Sonali Yadav, Application Of Artificial Intelligence In Fighting Against Cyber Crimes: A Review, International Journal of Advanced Research in Computer Science April 2018 , (ISSN: 0976-5697), ISBN: 978-93-5311-643-9, page[118-121].

[14]. Nilaykumar Kiran Sangani&HarootZarger, Machine Learning in Application Security, [Accessed: March 18, 2019] http://dx.doi.org/10.5772/intechopen.68796 [15] Dmitri Koteshov, How Can Ai Change The State Of Cybersecurity, March 7, 2018, https://www.elinext.com/industries/financial/trends/aiand-security/.

[15]. Richard Power, "1999 CSI/FBI Computer Crime and Security Survey," Computer Security Issues & Trends, Computer Security Institute, winter **1999.**

[16]. Denning D E, "An Intrusion-Detection Model," In IEEE Transaction on Software Engineering, **Vol. Se-13, No. 2, pp. 222-232, February 1987.**

[17]. Lee, W, Stolfo S and Mok K , "Adaptive Intrusion Detection: A Data Mining Approach," In Artificial Intelligence Review, Kluwer Academic Publishers, **14(6), pp. 533 - 567, December 2000.**

[18]. Satinder Singh, Guljeet Kaur, **"**Unsupervised Anomaly Detection In Network Intrusion Detection Using Clusters," Proceedings of National Conference on Challenges & Opportunities in Information Technology RIMT-IET, Mandi Gobindgarh. **March 23, 2007.**

[19]. Eric Bloedorn , Alan D. Christiansen , William Hill , Clement Skorupka , Lisa M. Talbot , Jonathan Tivel, "Data Mining for Network Intrusion Detection: How to Get Started," CiteSeer, **2001**

[20]. L. Portnoy, "Intrusion Detection with Unlabeled Data Using Clustering," Undergraduate Thesis, Columbia University, 2000.

[21]. Theodoros Lappas and Konstantinos Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems," http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.2533&rep=rep1&type=pdf.

[22]. Dewan Md. Farid, Nouria Harbi, Suman Ahmmed, Md. Zahidur Rahman, and Chowdhury Mofizur Rahman, "Mining Network Data for Intrusion Detection through Naïve Bayesian with Clustering", World Academy of Science, Engineering and Technology, **2010.**

[23]. The KDD Archive. KDD99 cup dataset, 1999. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[24]. X. Li and N. Ye., "A supervised clustering algorithm for computer intrusion detection," Knowledge and Information Systems, 8, pp498-509, ISSN 0219-1377, 2005

[25]. Kruegel C., Mutz D., Robertson W., Valeur F., "Bayesian event classification for intrusion detection," In: Proceedings of the 19th Annual Computer Security Applications Conference; **2003.**

[26]. Portnoy L., Eskin E., Stolfo S.J., "Intrusion detection with unlabeled data using clustering," In: Proceedings of The ACM Workshop on Data Mining Applied to Security; **2001.**

[27]. Paxson V., "Bro: A System for Detecting Network Intruders in Real-Time", Computer Networks, **31(23-24), pp. 2435-2463, 14 Dec. 1999.**

[28]. D.Barbara, J.Couto, S.Jajodia, and N.Wu, "ADAM: A test bed for exploring the use of data mining in intrusion detection", SIGMOD, **vol30, no.4, pp 15-24, 2001.**

[29]. P.Domingos, and M.J. Pizzani, "On the optimality of the simple Bayesian classifier under zero-one loss", m/c learning, **Vol.29, no2-3, pp 103-130, 1997.**

[30]. F. Provost, and T. Fawcett, "Robust classification for imprecise environment," Machine Learning, **vol. 42/3, pp. 203-231. 2001.**

[31]. Athanasios Papoulis and S. Unnikrishna Pillai,. "Probability, Random Variables and stochastic Processes ", McGraw-Hill, Fourth Edition, ISBN 0073660116, **2002.**

[32]. P. Kabiri and A.A. Ghorbani, "Research on Intrusion Detection and Response: A Survey," International Journal of Network Security, 1, 84-102, **September 2005.**

[33]. A. Patcha and J-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Comput. Netw., 51, 3448-3470. ISSN 1389-1286. **2007.**

[34]. T.M. Mitchell. Machine Learning. McGraw-Hill. ISBN: 0-07-115467-1, **1997.**

[35]. N. Ben Amor, S. Benferhat and Z. Elouedi, "Naive Bayes vs Decision Trees in Intrusion Detection Systems," Proceedings of the ACM symposium on Applied computing, ISBN 1-58113-812-1, pages 420-424, New York, USA, **2004.**

[36]. M. Panda and M.R. Patra, "Network intrusion detection using naive bayes," IJCSNS International Journal of Computer Science and Network Security, **7, 258-263, 2007**

[37]. F. Gharibian and A.A. Ghorbani, "Comparative Study of Supervised Machine Learning Techniques for Intrusion Detection," In CNSR '07: Proceedings of the Fifth Annual Conference on Communication Networks and Services Research, Pages 350-358, Washington, DC, USA, **2007**

[38]. L. Portnoy, E. Eskin and S. Stolfo, "Intrusion Detection With Unlabeled Data Using Clustering," In Proceedings of the ACM Workshop on Data Mining Applied to Security, **2001.**

[39]. K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," Proceedings of the 28th Australasian conference on Computer Science, ISBN 1-920-68220-1, pages 333-342, Darlinghurst, Australia, Australia, **2005.**

[40]. W. Wang, X. Guan and X. Zhang, "Processing of massive audit data streams for real-time anomaly intrusion detection," Comput. Commun., 31, 58- 72. ISSN 0140-3664, **2008**

[41]. J. Song, K. Ohira, H. Takakura, Y. Okabe and Y. Kwon, "A Clustering Method for Improving Performance of Anomaly-Based Intrusion Detection System," IEICE Transactions on Information and Systems, E91-D, 1282-1291. ISSN 0916-8532, **2008**

[42]. E.J. Spinosa, A.P. de Leon F. de Carvalho and J. Gama, "Cluster-based novel concept detection in data streams applied to intrusion detection in computer networks," Proceedings of the ACM symposium on Applied computing, pages 97.

