ZKG INTERNATIONAL ISSN: 2366-1313

Machine Learning Techniques for Analyze and prediction of Cyber Attack Detection

[1] Koppada Kumar Sai, [2] M. Rama Bhadra Rao

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

kumarsai72880@gmail.com

[2,] Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: *One of the most crucial concerns within the global is the problem of cybercrime, which causes widespread economic losses to international locations and their residents each day. The frequency of cyber attacks has extended, highlighting the need to become aware of the individuals behind those crook activities and recognize their strategies. Detecting and stopping cyber attacks poses a big problem, but recent advances have added AI-based protection fashions and predictive tools to clear up these problems. Although there is a wealth of records on crime prediction techniques, they need to be adapted to better fit cybercrime and cyber attack expectations. One technique to this problem is to apply actual statistics to determine the final results of the assault and discover the position of the party. This data consists of details on crime, perpetrators, belongings harm, and attack vectors. Forensic teams can acquire information about sufferers of cyber attacks via the utility system. This studies look at makes use of machine learning techniques to discover cybercrime using two fashions and expect how behaviour can assist pick out the cyber attack manner and the crime.*

KEY WORDS- Machine learning algorithm, Cybercrime, SVM, Cyber-attacks.

I. INTRODUCTION

Machine mastering aims to make forecasts for the future based on the past data. The process of mastering machines (ML) is a branch of artificial intelligence which allows devices to "research" new skills with the aid of explicit programming. The purpose of the device that it gains information about is to improve

ISSN: 2366-1313

computers with software which can adapt to changing information. Prediction and training processes depend extensively on specialized algorithms.

The statistics on the schooling process are fed into an algorithm which utilizes this knowledge to predict regarding the fresh test records. Three distinct parts that make up system-learning. Three types of classes, supervised, less supervised and more desired - could be used to categorize research. Before software is able to use records to master supervised the data must first be classified. Unlabeled studies do not use labels. This is a way to get familiar set of guidelines. The method is used to determine how to group the input data. In the end, reinforcement learning is developed when you receive a lot of low-quality feedback. It is also attractive to watch interactive interactions with the environment.

Python lets statisticians employ system-learning techniques to find patterns that offer relevant statistics. Based on the way they "analyze" facts to make predictions, these algorithms could generally be classified into two categories: supervised learning or unsupervised. "Classification" is the term used to define what category a particular set of facts is part of.

The term "class" can be used to refer to as objectives, indicators or even businesses. Predictive modeling within the realm of classes involves attempting to approximate the map function between continuous in-put variables (X) as well as variable outputs that are discrete (y). The device gains knowledge about the information and category, it is an instance of supervised mastering. It is where the computer program learns to sort out new information by reference to the previously-established patterns. The statistics component may contain binary information (for example if it's a female or male, or whether it is spam, or is no longer) It might also contain other directions in the field of statistics. Document categorization and biometric identity as well as voice recognition and writing studies are just a few examples of situations that require a lot of attention in class.

II LITERATURE REVIEW

Bilen, Abdulkadir & Ozer, Ahmet. (2021). E475. 10.7717/peerj-cs.475.

Cyber-attacks are now among the major problems of the globe. They can cause significant financial losses to both individuals and nations daily. In addition, the increase in cyber attacks creates cybercrime. The most crucial factors to fight criminals and crimes is to identify those who commit cybercrimes and becoming aware of the tactics used by those who attacking. The detection and prevention of cyber-attacks is challenging task. The problem is solved by deciphering an attack's source and the person responsible for it by using real information. These include details about the nature of crime, gender attacker, and the extent of damage and the method of attack. They can be gathered through the application of people who have been exposed to cyber-attacks and forensic agencies. In this article we examine cyber-crimes through two models employing algorithms for machine learning and forecast the effects of the identified attributes on the identification of the method of cyber-attack and its person who is responsible for the attack. We employed eight methods of machine learning to conclude that their accuracy ratios were comparable. It was the Support Vector Machine Linear was identified as the most effective cyber-attack strategy, with the accuracy of 95.02 percent. With the initial model, we were able to predict the kinds of attacks victims are likely to be subjected to, with high precision. It was found that the Logistic Regression was the leading technique for identifying attackers, with a rate of 65.42 percent. The second model we determined whether criminals were able to be identified through comparing their traits. The results show that the likelihood of a cyber attack diminishes when the educational and amount of income a victim earns. We are convinced that the cyber crime organizations will adopt the strategy. It can also aid in investigation of cyber attacks and makes fighting these threats much easier and effective.

Al-majed, Rasha & Ibrahim, Amer & Abualkishik, Abedallah & Mourad, Nahia & Almansour, Faris. (2022). 10. 261.10.21533/pen.v10i3.3035.

Integration of networks is prevalent for cyber-physical systems (CPS) in

order to provide remote access, surveillance and data analysis. The systems have also been vulnerable to cyber-attacks due to their connection to an unsecure network. If there was an infraction to internet security it was possible for an attacker to alter the functions of the system, and this can have devastating results. Therefore, the detection of attacks on critical CPS is of paramount importance. Recognizing attacks against CPSs which are becoming increasingly attacked by cybercriminals as well as cyber-attacks are becoming more challenging. Machine Learning (ML) and Artificial Intelligence (AI) have the capability of making these some of the most stressful times but they could also make it the best of moments. There are many methods that AI technology could aid increasing the profitability and expansion across a range of sectors. This data can be processed by using ML as well as AI methods that are designed to detect attacks on CPSs. In this paper we present a new cyber attack detection method that combines AI as well as ML (ML) techniques. In the beginning, we gather data of the CPS database, and then we process the information using normalization to ensure elimination of redundant and error-prone data. The data is extracted with Linear Discriminate Analysis (LDA). The proposed method is Self-tuned Fuzzy logic-based Hidden Markov Model (SFL-HMM) that incorporates a Heuristic Multicar Optimization (HMS-ACO) method to aid in the identification of cyber-attacks. The method proposed is tested by using the MATLAB simulation software and it is compared against the existing methods. The outcomes of the tests demonstrate that the framework has more success than conventional methods in attaining high levels of security. In addition, when it comes to the detection rate and false positive rate and time to compute the framework is superior to traditional detection techniques.

Mazhar, T.; Irfan, H.M.; Khan, S.; Haq, I.; Ullah, I.; Iqbal, M.;

Smart grids are quickly taking over conventional systems on a global scale. Smart grids have their own drawbacks as with any new technology. The cyber attack on the smart grid is one of the most difficult problems to prevent. Most of the problems are due to the millions of

sensors continuously communicating and receiving data messages through the network. Cyber attacks may compromise the grid's reliability, availability, as well as privacy. Network administrators, users that include sensors and smart devices as well as network administrators are three of the layers that make up a modern grid system that is at risk of cyber attacks. This study will examine the numerous vulnerabilities and risks that could compromise the security of vital elements of a grid network that is innovative. To protect ourselves from the risks, we provide various security options using various strategies. Additionally, we offer suggestions for lessening the risk that these three kinds of cyber attacks can happen.

Starker, I.H. Current and Future Prospects. Ann. Data. Sci. (2022).

Thanks to the digital revolution of the world and Internet of Things revolutions, our world today is an abundance of data on cyber security. Finding a way to effectively address cyber security issues and threats is an increasing concern for the cyber security industry across the globe. The traditional security tools are not sufficient to deal with the modern security concerns due to the increasing number of various cyber-attacks and dangers. Making use of the artificial intelligence expertise particularly machine learning technologies are crucial to provide the most dynamically upgraded, automated and current security system by analyzing data from security. In this report we offer a thorough analysis of the machine learning algorithms and the way they can be utilized for data analysis that is intelligent as well as automation in cyber security, due to their ability to gain useful insights from cyber information. Also, we look at a range scenarios that use data-driven insights, automation and a decision-making process can provide next-generation cyber security that are more proactive than the traditional methods. Future prospects for machine learning and cyber security is the main focus of the findings of our study and the relevant research areas. Our goal is to study not just the state of play in machine learning, but also the relevant methods as well as their potential to future security breakthroughs.

III System Analysis

The detection systems that are primarily system-based systems employ regarded styles as well as the signatures of attacks to recognize their targets. They're effective when it comes to known attacks, but they are vulnerable to new attacks.

An anomaly-based total detection system the machines detect deviations of normal behavior patterns in order to identify potential threats. They may detect new attacks however, many create excessively fakes incredible cost

Rule-based total detection systems use rules that are predefined to identify and stop attacks. They're easy to put in place; however they require constant replacement as attacks develop.

Machine mastering-based detectors: These devices employ various device learning algorithms to study sources and determine patterns that suggest threats. They are effective against novel and well-known threats; however they require huge amounts of information about schooling from professionals.

Proposed device:

Hybrid strategies: combining techniques based on signatures, anomaly-based and rules-based systems and machine learning algorithms that can draw on the advantages of each and improve the accuracy of normal detection.

Techniques for deep study deep learning algorithms that include convolution neural networks as well as recurrent neural networks can be utilized to analyze complex information about network traffic, and to pick the patterns of subtitles that indicate threats.

Generative Adverbial Network A community that is generative could be utilized to create fake attack data to instruct methods of machine learning and increase their performance overall towards new attacks

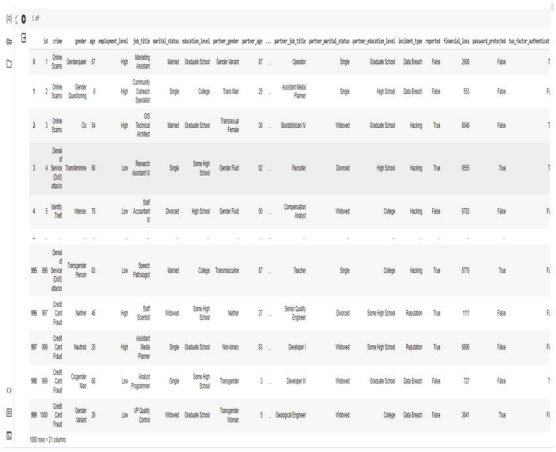
Explainable AI (XAI): XAI techniques can help explain the choices made using a gadget to gain knowledge about models which can improve the agreement and clarity in the detection of cyber attacks.

Federated learning: This allows organizations from multiple institutions to teach students using their gadgets on their private records and not share confidential information

IV DATA SET DESCRIPTION

ISSN: 2366-1313

To investigate and anticipate the technique of cyber attack detection that uses gadget-learning and gadget learning, we require a database that includes attributes associated with cyber attacks, community member's devices, logs of device usage and other relevant data. The data collected will be used to train the systems to learn about fashions. It is important to be aware of and prepare for cyber attacks. By analyzing and forecasting methods our goal is to enhance the effectiveness of cyber security detection tools and assist in the creation of proactive security methods.



Data Set Size: 1000 rows& 21 columns

V Design

INPUT AND OUTPUT DESIGN

INPUT DESIGN:

The design of the input provides the link between the information system and the individual. It is a growing set of specifications and techniques for guidance in records and these steps are essential for converting transaction data an acceptable format for processing. It can be accomplished by utilizing the information on a computer from a printed or written document or through the use of humans who input details at the same time to the device. The input layout is specialized in controlling the quantity of input needed and preventing errors. What records have to be organized or codified?

The dialogue to instruct employees in providing the opportunity to enter.

Methods to prepare input validation and the ways to be followed when errors occurs.

OBJECTIVES:

1. it's important in order to prevent error in the statistical entry procedure and to show the correct route to control in receiving accurate statistics of the computerized gadget.

2. It can be accomplished by means creating user-friendly monitors that allow information entry, which can manage a huge amount of data. The

ISSN: 2366-1313

reason for designing the enter screen in order to make the process of entering data simpler and free of mistakes. The screen for information access can be designed in such a fashion that all of the statistical manipulators can be achieved. Additionally, it provides access to document-viewing features.

3. Properly formatted messages are displayed they are desired in order to ensure consumers will not feel like a shopper in a jumble of information.

OUTPUT DESIGN:

A good output fulfills the expectations of the user who is not using it and provides the data in a clear manner. When output layout is established how records are moved to allow for immediate need and the final replication output. It's the most important and straight supply of information for the individual. Effective and efficient layout of output improves process's courting and assists the user in to make a better decision.

The output of a record gadget should meet any or all of these goals.

Convey information about past games, the most cutting-edge name or future projections of

* Future.

Alert people to important events such as opportunities, problems or issues.

The trigger can be used to initiate a movement.

Verify a motion.

VI MACHINE LEARNING ALGORITHMS

MODULES:

User

Admin

Machine Learning

MODULES DESCRIPTION:

User:

Users can register at the beginning. In order to sign up, the user must provide an email address for the person and a mobile phone for further messages. When the user has signed up, and admin prompts for the individual to sign up. When admin has enabled the account, the user can log in to our device. The user can then upload their data in complete accordance with our data column that is matched. In order to execute the set of rules data should be entered in an into format or use the flow formatting. This is the format we used. Adage Technologies Limited dataset for testing motive. Users can also

ISSN: 2366-1313

upload the newly created data in the current data set based on our Django tool. Users can select Data Preparations inside the internet page and the system for cleaning of data begins. The clean statistics as well as its needed graph will be shown.

Admin:

The administrator can sign in using his login details. Administrators can activate the authentic users. After activation, user can sign in on our device. Administrators are able to view users as well as view the universal documents in the browser. Additionally, it will load the records. Administrators can access the list of school statistics and the list of test records. Administrators can download the data and view forecasted outcomes.

A machine that gains knowledge of:

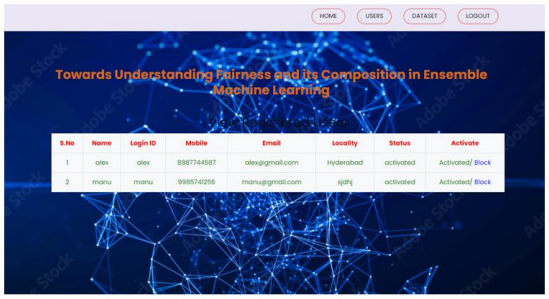
As per the cut up criteria, the clean data are cut into the eighty% education and 20 percent test after which the data is tested on one machine to gain understanding of the classifier along with Natural language Process (NLP). The evaluation of sentiment is done through auto encoding methods that are quality-tuned such as ALBERT and BERT obtain a comprehensive understanding of the public's sentiment. So, we've analyzed the findings of our study and the method using the context information, and have confirmed the results.

Output Screens

Home page:

Registration page:

View Users page:



Dataset View:

View Dataset

id	crime	gender	age	employment_level	job_title	marital_status	education_level	partner_gender	partner_age	partner_employment_level	partner_job_title	par
1	Online Scams	Genderqueer	67	High	Marketing Assistant	Married	Graduate School	Gender Variant	87	Medium	Operator	Sing
2	Online Scams	Gender Questioning	0	High	Community Outreach Specialist	Single	College	Trans Man	25	High	Assistant Media Planner	Sing
3	Online Scams	Cis	54	High	GIS Technical Architect	Married	Graduate School	Transsexual Female	30	Low	Biostatistician IV	Wid
4	Denial of Service (DoS) attacks	Transfeminine	68	Low	Research Assistant IV	Single	Some High School	Gender Fluid	82	High	Recruiter	Div
5	Identity Theft	Intersex	76	Low	Staff Accountant III	Divorced	High School	Gender Fluid	50	High	Compensation Analyst	Wid
6	Denial of Service (DoS) attacks	Transfeminine	65	Low	Legal Assistant	Widowed	High School	Transsexual	87	High	Financial Analyst	Mar
7	Credit Card Fraud	Transgender Male	9	Low	Software Test Engineer I	Single	High School	Cis Man	92	Low	Senior Financial Analyst	Mar
8	Online Scams	Transsexual Person	40	Medium	VP Product Management	Divorced	High School	MTF	78	Medium	Biostatistician IV	Wid
9	Online Scams	Transsexual Man	61	Low	Help Desk Operator	Divorced	Some High School	MTF	60	Medium	Senior Developer	Sing
10	Credit Card Fraud	Transgender	61	Low	Recruiter	Single	Graduate School	Cisgender Female	69	Low	Physical Therapy Assistant	Mar

ISSN: 2366-1313

Output values:

VII CONCLUSION

The study aims to come at and prevent cyber attacks the help of combing devices mastering techniques with previous information on similar tries. The model predicts the characteristics of a victim's ability as well as the kinds of threats they could encounter. The methods of machine learning have enough convincing proof. A good approach is to utilize SVMs that are linear.

It can effectively detect the threat and execute the cyber attack approximately 61% of the times. I would like to increase this choice through using a variety of AI techniques. It is essential to improve the awareness about malware-related attacks as well as social engineering. Cyber security is a risk that has attack has changed to be inversely associated to the educational and financial level of the person who is victimized. The primary goal of the study is to provide regulation enforcement with efficient and effective tools so that they can be even more effective in stopping cybercriminal. Through analyzing the characteristics of those who have been victims through our investigation, the development of new education and alarming methods for other users who have similar characteristics will be technologically advanced.

REFERENCES

1. Bilen, Abdulkadir & Özer, Ahmet. (2021).Cyber-attack method and perpetrator prediction using machine learning algorithms. PeerJ Computer Science. 7. E475. 10.7717/peerj- cs.475.

2. Al-majed, Rasha & Ibrahim, Amer& Abualkishik, Abedallah & Mourad, Nahia & Almansour, Faris. (2022). Usingmachine learning algorithm for detection of cyber-attacks in cyber physical systems. Periodicals of Engineering and Natural Sciences (PEN). 10. 261.10.21533/pen.v10i3.3035.

3. Mazhar,T. ;Irfan, H.M.; Khan,S. ;Haq,I. ;Ullah,I. ;Iqbal, M.; Hamam, H. Analysis of Cyber Security Attacks and Its Solutions for the Smart grid Using Machine Learning and Block

chain Methods. *Future Internet***2023**,*15*, 83. https://doi.org/10.3390/fi15020083

4. Sarker, I.H. Machine Learning for Intelligent Data Analysis and Automation in Cyber security: Current and Future Prospects. *Ann. Data.*

A.Alshehri, N. Khan, A.Alowayr and M. Yahya Alghamdi, "Cyber attack detection framework using machine learning and user behaviour analytics, "*Computer Systems Science and Engineering*, vol.44, no.2,pp. 1679–1689,2023.

5. Amjad Rehman,Tanzila Saba, Muhammad Zeeshan Khan, Robertas Damaševičius, Saeed Ali Bahaj,"Internet-of-Things-Based Suspicious Activity Recognition Using Multimodalities of Computer Vision for Smart City Security", *Security and Communication Networks*, vol. 2022,Article ID8383461,12 pages, 2022. https://doi.org/10.1155/2022/8383461

6. Liu Qiang, Qu Xiaoli, Wang Dake, Abbas Jaffar, Mubeen Riaqa, Product Market Competition and Firm Performance: Business Survival Through Innovation and Entrepreneurial.

7 Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.