WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning and Cyber Security: A Review

Ajeesha M I[1],Muhamed Jamshir M[2]

[1]Research Scholar, School of Computer Studies, Rathnavel Subramaniam College of Arts and Science, Coimbatore, Tamilnadu, India.
[2]Adhoc Faculty, NSS College of Engineering, Palakkad, Kerala, India.
Email: ajeesha1393@gmail.com, jamshir20@gmail.com

ABSTRACT

Machine Learning is the evaluation of computer algorithms that improve automatically by way of experience. It has become a vital technology for cybersecurity. Machine learning is one of the leading tools for handling information security. Machine learning is also called Computational learning theory. Cybersecurity is the discipline of technologies and practices intended to protect networks and data from attack, theft, or unauthorized access. It encompasses everything that pertains to protect our data from theft and damage attempted. This survey paper describes a literature survey of machine learning in cybersecurity.

KEYWORDS: Machine learning, Cybersecurity.

1. INTRODUCTION

The information superhighway known as the internet has revolutionized technology. From 1983 it makes a start for computer crimes. Morris Worm is the first denial of service attack(DoS). Prediction is the easiest way for securing data. ML system is able to process data to make predictions, and do this in near real-time, in turn providing intelligence to administrators who look and monitor to identify anomalies. Machine learning means solving certain tasks with the use of an approach and particular methods based on data you have and cybersecurity is the protection of internet connected systems including software, hardware, and data from cyber attacks[1]. Each module will have a certain focus, but the overall goal is to protect people's data and privacy. It monitors systems and mitigates threats when they happen[10]. A strong security infrastructure includes multiple layers of protection dispersed throughout a company's computers, programs, and networks[9].

Machine learning develops the patterns and manipulates the patterns with algorithms using data from everywhere is needed. Machine learning algorithms essentially build models of behaviors and use those models for making future predictions based on new input. Machine learning can help businesses to analyze threats and respond to attacks. It also helps to automate more menial tasks previously carried out by under-skilled security teams[8]. Machine learning is to formulate algorithms that can receive input data and use statistical analysis to predict an output.

2. MACHINE LEARNING

Machine learning is remolding the world both in and out of the digital realms. Machine learning is a potentially transformative technology for cybersecurity. Machine learning can make cybersecurity simpler, more proactive, less expensive and more effective[7]. Machine learning is analyzing data pulled from contacts, chat, and voice operations. It will help businesses to detect malicious activity faster and stop attacks before they get started.

Machine learning algorithms allow computers to train input data and use statistical analysis to output values that fall in a specific range. It is to understand the structure of data and apply that data into models that can be realized and used by people. Machine learning accelerates computers in building models from sample data to automate decision-making processes based on data inputs. The generic machine learning methods of supervised and unsupervised learning, and the algorithmic approaches in machine learning, including the k-nearest neighbor algorithm, decision tree learning, and deep learning. Supervised learning organizes algorithms based on example input and output data that is labeled by humans and unsupervised learning provides the algorithm with no labeled to allow it to find structure within its input data. Every machine learning algorithm has three components:

Representation: It is how to represent knowledge. Examples are decision trees, sets of rules, instances, graphical models, neural networks, support vector machines, model ensembles, etc.

Evaluation: It is the way to evaluate candidate programs (hypotheses). Examples include accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence, etc.

Optimization: It is the way candidate programs are generated and is the search process. For example, combinatorial optimization, convex optimization, constrained optimization.

2.1. TYPES OF MACHINE LEARNING ALGORITHM

2.1.1 Supervised Learning

In supervised learning, the computer gives example inputs that are labeled their desired outputs. The algorithm "learns" by correlating output with the "taught" output to find errors and refine the model accordingly. This model uses patterns to predict label values on additional unlabeled data. The generic use case of supervised learning is to use historical data to predict statistically likely future events. Practical machine learning mostly uses supervised learning. It has input variables as (x) and an output variable as (Y) and the algorithm for the mapping function from the input to the output Y = f(X).

The objective is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data. Supervised learning problems are regression and classification problems.

- Classification: A classification problem is when the output variable is a category, such as "red" or "blue".
- Regression: A regression problem is when the output variable is a real value, such as "dollars" or "weight"[5].

2.1.2 Unsupervised Learning

In unsupervised learning, the data is unlabeled, so the learning algorithm is left to find commonalities among its input data. As unlabeled data are more abundant than labeled data, machine learning methods that facilitate unsupervised learning are valuable. Unsupervised learning is for transactional data. Unsupervised learning problems are clustering and association problems.

- Clustering: A clustering problem is to find out the inherent groupings in the data, such as grouping customers by purchasing behavior.
- Association: An association rule learning problem is where you want to find rules that describe large portions of your data, such as people that buy X also tend to buy Y.

2.1.3 Reinforcement Learning

Reinforcement learning directly takes influence from how human beings learn from data in their lives. It features an algorithm that favours upon itself and learns from new situations using a trial-and-error method. Favorable outputs are reinforced and non-favorable outputs are discouraged.

3. MACHINE LEARNING TECHNIQUES

Regression:

Regression methods fall within the group of supervised ML. They help to determine or to explain a particular numerical value based on a set of prior data, for example predicting the price of a property based on previous pricing data for similar properties. Regression techniques run from simple to complex. Simple like linear regression and complex like regularized linear regression, polynomial regression, decision trees and, random forest regressions, neural nets etc.

Classification:

Classification is the Supervised machine learning method. It can predict or describes a class value. The output can be yes or no values. The simplest classification method is Logistic regression. Logistic regression evaluates the probability of an occurrence of an event based on one or more inputs.

Clustering:

Clustering comes under the category of unsupervised learning method. The goal is to group or cluster observations that have similar characteristics. Clustering methods do not use output information for training. The most popular clustering method is K-means. K represents the number of clusters that the user chooses to create.

APPLICATIONS OF MACHINE LEARNING

Machine learning algorithms use in circumstances where the solution is required to continue improving post-deployment. The dynamic nature of adaptable machine learning solutions is one of the main selling points for its assumption by companies and organizations across verticals.

Machine learning algorithms and solutions are versatile and can use as a substitute for medium-skilled human labor given the right circumstances. For example, the customer service executives in large B2C companies have now been replaced by natural language processing machine learning algorithms known as chatbots. The chatbots evaluates customer queries and provide support for human customer support executives or deal with the customers directly.

3. CYBERSECURITY

Today it is impossible to employ effective cybersecurity technology without relying heavily on machine learning. With machine learning, cybersecurity systems can analyze patterns and learn from them to help prevent similar attacks and respond to varying behavior. It can help cybersecurity teams be more proactive in preventing attacks and responding to active threats in real time. It can reduce the amount of time spent on routine tasks and enable organizations to use their resources more strategically. Cybersecurity refers to a set of techniques used to protect the integrity of networks, devices, and data from attack, damage, or unauthorized access. Cyberattacks take many forms, such as social engineering, malware, and ransomware. Current technological advancements have opened up new possibilities for cybersecurity. But adversely, adversaries have benefited from these advances as well.

Cyber-attack is now a global concern and has given many reviews that hacks and other security attacks could risk the global economy. Organizations convey sensitive information across networks and to other devices in the course of doing business, and cybersecurity describes protecting that data and the systems used to process or to store. It consists of technologies, processes, and practices designed to protect networks, devices, programs, and data from attack, theft, damage, modification, or unauthorized access. It may also be referred to as information technology security.

The rapid growth of the Internet, cybersecurity has become a major concern to organizations all over the world. The reality that the information and tools & technologies needed to penetrate the security of corporate organization networks are widely available has increased that security concern. Cybersecurity is to protect information from being hijacked, compromised, or attacked. Cybersecurity can be calculated by at least one of three goals-

- Protect the confidentiality of data.
- Preserve the integrity of data.
- Advances the availability of data for authorized users.

These goals are the basis of all security programs form the confidentiality, integrity, availability (CIA) triad. The CIA triad is a security model that is planned to guide policies for information security enclosed by the premises of an organization or company. It is considered as the three most important components of security.

3.1. TYPES OF ATTACKS

Malware

A malware attack is a common cyberattack. Malware or malicious software executes unauthorized actions on the victim's system. The malicious software encompasses many specific types of attacks such as ransomware, spyware, command and control, and more.

Phishing

Phishing is a type of attack which tries to steal perceptive information like user login credentials and credit card number. It results when an attacker is masquerading as a trustworthy entity in electronic communication.

Man-in-the-middle attack

It is a type of attack that allows an attacker to obstruct the connection between client and server and acts as a bridge between them. As a result, an attacker will be able to read, insert and modify the data in the obstructed connection.

Zero day attack

A zero-day attack is an advanced attack. It is an undetermined attack that exposes a vulnerability in software or hardware and can produce complicated problems well before anyone realizes something is wrong. A zero-day attack leaves NO chance for detection at first.

Cybersecurity is the practice of protecting systems, networks, and programs from malicious attacks. These cyberattacks are usually aimed at accessing, changing, or destroying sensitive information; extorting money from users; or interrupting normal business processes.

3.2. CHALLENGES OF CYBERSECURITY

Cybersecurity is taking preventive action before any threat exploits the system. Information theft is the most expensive and fastest growing sector of cybercrime. The challenges of cybersecurity are:

- Network Security: The process of defending the network from unwanted users, threats and intrusions.
- Application Security: Apps require constant updates and testing to ensure these programs are secure from thefts.
- End point Security: Endpoint security is the process of protecting remote access to a company's network.
- Data Security: Data is within the networks and applications. Protecting company and customer information is an isolated layer of security.
- Database and infrastructure Security: Everything in a network involves databases and physical equipment. Protecting these devices is also important.
- Cloud Security: Many files are in digital environments or "the cloud". Protecting data in a 100% online environment presents a large amount of challenges.
- Mobile Security: Cell phones and tablets include virtually every type of security challenge in and of themselves.

4. CONCLUSION AND FUTURE SCOPE

In this paper, a survey about machine learning for cybersecurity is explained. Cybersecurity is one of the most fertile and fast-moving areas of technology. Cybersecurity protects the system and data from malicious attacks. Different machine learning algorithms are used for intrusion detection. Cyber security is to protect our valuable data and information from malicious attacks. Cybersecurity is a vast field of science, for detecting host based intrusions and network based intrusion detection system makes future scope of this work.

CONFLICT OF INTEREST

The authors reports no conflicts of interest. They have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors are alone responsible for the content and writing of this article.

REFERENCES

[1] Anna L. Buczak, Erhan Guven "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection" IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 18, NO. 2, SECOND QUARTER 2016.

[2] Idan Amit, John Matherly, William Hewlett, ZhiXu,Yinnon Meshi, Yigal Weinberger,"MachineLearning in Cyber-Security-Problems, Challenges and Data Sets", http://arxiv.org.

[3] Iqbal H. Sarker, A.S.M. Kayes, Shahriar Badsha Hamed Alqahtani, Paul Watters and Alex Ng, "Cybersecurity data science: an overview from machine learning perspective", Sarker et al. J Big Data (2020) 7:41.

[4] Rishabh Das, Thomas H. Morris,"Machine Learning and Cyber Security",DOI: 10.1109/ICCECE.2017.8526232

[5] Supervised and Unsupervised Machine LearningAlgorithms (machinelearningmastery.com)

[6] Tahir Mehmood and Helmi B Md Rais, "Machine Learning Algorithms In Context Of Intrusion Detection", 2016 3rd International Conference On Computer And Information Sciences(ICCOINS),978-1-5090-2549-7/16/$31.00 ©2016 IEEE.

[7] The Growing Role of Machine Learning in Cybersecurity (securityroundtable.org)

[8] Top machine learning use cases for security CSO Online

[9] What Is Cybersecurity? Why Is It Important? Built In

[10] Why You Should Study a Cyber Security Degree in 2021 - MastersPortal.com

AUTHORS PROFILE

Ms. Ajeesha M I currently pursuing PhD in the area of Data mining/Machine Learning from Rathnavel Subramaniam College of Arts & Science affiliated to Bharathiar University Coimbatore. She obtained M. Phil in the area of Data Mining from Rathnavel Subramaniam College of Arts & Science affiliated to Bharathiar University Coimbatore in 2018. Her research interests are Data Mining and Machine Learning.

Mr. Muhamed Jamshir M received his M. Tech in the area of Electronics Design and Technology from National Institute of Technology, Calicut in 2012. He obtained his B. Tech, in Electronics and communication engineering from Government Engineering College, Trissur, in 2010 from Calicut University. At present he is working as an Ad hoc Faculty, NSS College of Engineering Akathethara, Palakkad. His research interest lies in the area of Networking and Machine Learning. He served as a key note speaker for various seminars country wide.