ISSN: 2096-3246
Volume 54, Issue 02, April, 2022

Comparative Analysis of Cyber Physical System Classifications Utilizing Machine and Deep Learning

J. Sagar Babu[1] and Prof. Dr. Ashok Kumar P.S[2]
Department of CSE, St. Peter's Engineering College, Hyderabad[1]
Department of CSE, Don Bosco Institute of Technology, Banglore[2]

Keywords:
Big Data, Cyber-attacks, Machine Learning, Deep Learning, and Cyber-Physical Systems.

ABSTRACT

When it comes to resolving the vast range of problems that crop up in regular life, the cyber physical system (CPS) is by far the most popular infrastructure option. However, making the right decision quickly remains challenging in the big data era. Transforming the manufacturing sector and other applications will rely on Internet of Things (IoT) or CPS. These advantages of CPS come at a price, however, as companies struggle to deal with the massive amounts of data being produced by Internet-connected gadgets, resulting in a slew of difficulties for individuals. These infrastructures are too complex for even the most knowledgeable individuals to manage, monitor, or evaluate. As a result, there is a pressing need for the convergence of Machine Learning (ML) and cyber security in CPS, which equips experts with the tools they need to monitor the internet for potential threats in record time. Since the proposed study examines the various frameworks used for Cyber-attack detection using a learning method, it demonstrates the significance of ML and Deep Learning (DL) in a CPS for more accurately identifying potential dangers. Many academics rely on security analytics, and the tool may also be used to prioritize alarms and signals. It has been brought to the attention of the researchers that the suggested study on various assaults has also emphasized the need to be mindful of rare attacks that may become highly harmful. Additionally, the pros and disadvantages of various methodologies and datasets used in the study of various works in assessing different assaults are presented to aid in selecting the appropriate strategy according to demand.

1. INTRODUCTION

The proliferation of ICT applications has made it more difficult to keep private information secure in the modern world. Numerous dangerous applications designed to steal or destroy sensitive information have been launched by cybercriminals through the Internet. Malware and other forms of harmful software compromise data security, leading to the disclosure of sensitive or illegal material [1]. High-quality and many malware attacks are quite costly. Due to the proliferation of cyberattacks, researchers now prioritise the safety of their most sensitive data. As a result, the business, IT, and industrial user communities now place a premium on protecting sensitive information [2]. As a result, ML has flourished in this area, enhancing previously established approaches and their capacity to grasp and address the underlying problems at hand. This has led to an uptick in the use of ML in several fields, including healthcare, advertising, gaming, and computer vision [3]. ML approaches are superior than rule-based algorithms and human operators in some settings. This development is having an impact on the many cyber security domains, with certain detection systems being improved by integrating ML components [4]. However, ML based analysis and detection technologies are helpful in the long run goal of creating a fully automated cyber security system.

Our research was based on a comprehensive evaluation of trials conducted on genuine, massive, and network traffic. Existing works evaluate cyber security ML solutions by taking into account the particular application in addition to AI for security operators [5, 6]. Considering various problems with the current cyber security systems, this research article first provides a rationale for this endeavour. Later, the relevance of ML and DL for a cyber physical system is discussed, and the foundation for a cyber-attack detection system that uses this technique is described. Both the attack model and the ML/DL technique used to counter it are described in the same place. This study has so focused on the ways in which learning methods might be used to cyber security. In addition, concise explanations of each method's advantages and disadvantages are provided. Cyberattack detection work is also highlighted in our paper, as is the use of several ML algorithms.

2. CHALLENGES

When it comes to resolving problems that arise in the course of our daily lives, the cyber physical system is the most often used infrastructure. Taking the right choice quickly in a large data setting remains challenging. In order to take manufacturing, industry, and other applications to the next level, the integration of Internet of Things (IoT) or cyber physical systems (CPS) is employed [7]. With all of CPS's advantages, however, comes a slew of difficulties caused by a dearth of insightful analytical tools, which has an impact on businesses because to the difficulty of dealing with the massive amounts of data produced by Internet-connected gadgets. No one, not even experts, knows how to keep tabs on, manage, or evaluate these systems. As a result, there is an urgent need to combine machine learning with cyber security in CPS, so that experts can monitor online dangers more efficiently. Cybersecurity systems that use ML will be able to evaluate trends, learn from them, and use that knowledge to help avoid future assaults and adapt to new situations. It aids cybersecurity teams in anticipating and stopping threats, as well as dealing with assaults when they happen. It helps businesses save time and effort in their resource allocation by streamlining the routing of jobs. It saves time on routine chores and frees up personnel for more strategic endeavours in businesses. Incorporating ML into the cybersecurity industry has reduced costs, increased effectiveness, simplified operations, and facilitated preventive measures. However, this can only be achieved if the data used to power ML provides an accurate depiction of the surrounding environment.

3. RELATED WORKS

The economy, the privacy of its residents, and the physical infrastructure of the nation are all at risk from cyber attacks, which have grown more common and dangerous. This part outlines the overall structure of cyber-attack detection through learning model, which entails a number of processes, such as data gathering,

pre-processing, model training and solution, and final-output prediction. Then, we go on to discuss threat models and real-world applications of learning methods in Cyber security, as well as the significance of ML and DL for CPS.

A. Using a Model Learning Framework to Identify Cyber Attacks

Data gathering is the starting point for every learning model used to identify cyber attacks. One may get datasets from a variety of places, and then choose the sort of dataset needed for their operational model which is shown in Figure 1.

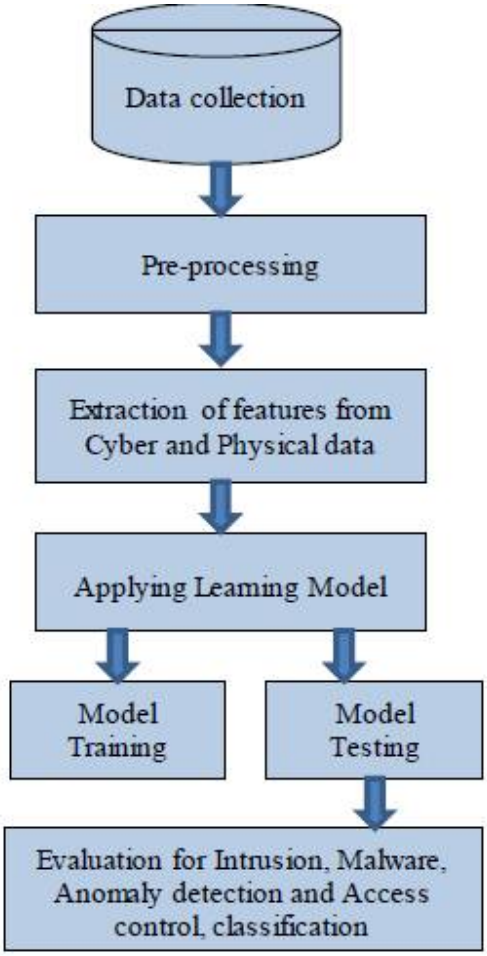

Fig.1 The use of a learning framework for the detection of cyber attacks

Some researchers collect or use publicly available information on network or workplace-based attacks. Following collection, data undergoes several forms of pre-processing to enhance the quality of the final product. While some threats travel in packets, others may exploit the TCP or UDP protocols to launch attacks [8]. Together, the attacking IP, the victim IP, and the targeted IP port form a unique identification for the TCP flow. When a cyber-attack is launched against a port that has not been designated a specific service, the accompanying TCP connection is closed as part of the pre-processing phase. Many other components are used throughout the data preparation phase. In addition, when the data has been pre-processed, the relevant properties are extracted for further study [9]. Next, the features are forwarded on to the model-building stage, when the employed learning model is put through its paces in both training and testing. Training data is tailored to the methodology used for evaluation, since different learning strategies need different forms of input and output. Training iterations for a network are selected according to the needs of the job at hand.

Final findings are obtained in the same or following rounds of testing the model. The goals of intrusion detection, malware detection, access control, anomaly detection, and classification are finally achieved. Predictions of cyber-attack rates may be made using out-of-sample data in a similar fashion, with the latter used to evaluate the system's or network's prediction accuracy.

B. Insights on the value of ML and DL for CP systems

Different types of researchers employ security analytics since it helps to prioritise the signals and alerts that are created in response to possible risks. As a result, we found a viable answer to the problem in a very short length of time. It's possible that ML may be used to make cyber defences even more robust. Data science methods are used, for instance, to analyse and manage cyber security companies' enormous dataset. These documents might go back years, connecting to established threat intelligence [7]. Regression, clustering, classification, and dimensionality reduction are just some of the areas where ML is put to use in F-score. Also, ML may be used to test the efficiency of authentication systems, smart metre data profiling, human interaction proofs, and protocol implementation. Cybersecurity is a crucial application area for ML and DL because of the many threats it faces that cannot be recognised by conventional mathematical models. According to ABI's estimates, cyber security spending on intelligence, data, and analytics will reach $96 billion by 2021. Security software is essential for healthcare organisations with larger networks to prevent unauthorised access by hackers who are after sensitive patient information [10]. Because they allow companies to access the information, they are known as "endpoints." In order to detect cyber security threats, machine learning models must be deployed within the client healthcare company's network, and real-time network activity analysis must be enabled [24]. Machine learning and deep learning are trained on typical network behaviour to determine the likelihood of anomalous activity. For instance, dark-trace is a machine learning service that may be used to spot abnormalities in healthcare organisations by flashing warning lights anytime a user's behaviour looks to drastically deviate from the usual. Cyber security systems that use ML are better equipped to see patterns and devise defences against future attacks. This shows that ML improves cyber security teams in both preventing risks and reacting to active attacks in real time. Here, the information is essential to the accomplishment of ML and DL in cyber defence. Multiple algorithms are employed to generate and analyse new patterns in ML and DL

C. Model of Threat

A cyber-attack is any malicious attempt to gain unauthorised access to a computer system, network, or other technologically dependent company or organisation. Malicious code is used in cyberattacks, which may have far-reaching effects like data breaches and lead to cybercrimes including theft and information identification if left unchecked. There are many different kinds of attack, each of which is defined in the following references [11-25]:

Computer viruses and worms are malicious software that secretly replicates itself over a network or on a user's hard drive.

Unwanted electronic mail, sometimes known as spam, is a kind of newsgroup. As a result, if spam filters aren't in place, unsolicited emails might be delivered to the recipient without their knowledge.

To put it simply, a Trojan programme masquerades as a legitimate one. However, once it is executed, it either changes the password location or makes the system more open to future intrusion attempts. Otherwise, the information and/or software on the hard drive will be lost.

Denial of service (DoS) attacks happen when hackers try to crash servers, IP addresses, or whole networks by sending an overwhelming number of messages.

For a computer virus to spread to other gadgets or users' profiles on social media websites, it needs to be controlled by a human being. Also, by using this software, hackers can create a botnet and exert remote control over a network of computers.

Scareware: Some users are tricked into downloading software by cybercriminals who use scare tactics. Though initially harmless, such programmes eventually become malicious and launch attacks on the user's computer.

Phishing attacks are created to steal users' login credentials.

Cybercriminals may disrupt tax collection by attacking legitimate payment systems.

Man in the middle attack happens when an attacker sets up shop at the interface between two ends of a communication channel, allowing him to read the message before it reaches its intended recipients. Furthermore, the attacker can gain access to private user data.

By using brute force, an attacker tries numerous times but ultimately fails to gain access to the system or data being guarded.

When an attacker uses a DDoS attack, they flood the target with commands, preventing the user from being able to access any data or use the system.

D. Cyber security and the Use of Machine Learning

At now, there are three main cyber applications that rely heavily on ML algorithms. Here you can find some introductions to malware analysis, spam detection, and intrusion detection.

The purpose of intrusion detection is to identify malicious activity on a computer system or network (IDS). Modern corporate networks arrange the older networks depending on threat patterns. Although ML-based categorization is useful, there are additional methods for detecting threats and anomalies in a contemporary deployment [12]. Two particular issues that fall under the broader umbrella of intrusion detection are the generation of domains and the detection of bonnets.

Modern malware automatically generates new variations with the same destructive effects but looks as various executable files, making malware analysis a challenging task. These metamorphic and polymorphic characteristics of malware are a challenge for rule-based malware detection systems. Through the use of ML methods, malware variants can be categorised correctly within their family tree.

Spam and phishing detection include a wide range of strategies with the goal of minimising the disruption and possible danger caused by unsolicited electronic messages. Hackers' favourite method of gaining first access to a corporate network now involves phishing and unwanted emails. Phishing emails often include links to malicious websites or other forms of malware. Attackers have been using a wide variety of cutting-edge evasion tactics, which makes it very challenging to identify phishing and spam. The use of ML techniques in the spam detection process may help enhance accuracy.

4. BACKGROUND WORK

Researchers in several fields rely on security analytics because it allows them to prioritise signals and warnings and so identify risks more accurately. This helps in locating the problem's resolution as soon as possible. Using ML, we can further strengthen cyber defences. However, there are still constraints on the full scope of its usefulness. We've discussed the benefits and drawbacks of several methods used in previous research. The primary purpose of this analysis is to compare and contrast the merits and shortcomings of various scholars' published works. Researchers employed many techniques like Support Vector Machine (SVM), K Nearest Neighbors (KNN), Multi Layer Perceptron (MLP), Long Short Term Memory, Naïve Bayes (NB), Random Decision Forest (RDF), and Replicator Neural Networks (RNN).

There was use of RDF, SVM, and NN by Youness Arjoune et al. (2020). A significant quantity of data may be processed quickly using their technology. There is a rise in productivity and a decrease in processing time. In other words, it can only detect a single jamming attempt.

SVM and RNN were used by Andrea Pincet I et al. (2018). The closest neighbour method is the most effective detector and it has a high degree of accuracy. With a low percentage of false positives when detecting at tacks. Scalability issues arise during the training phase of bigger systems.

SVM, NB, J.48, and a decision table were used by Tahir Mehmood et al. (2016). Compared to other algorithms, J.48 has a lower misclassification rate and higher overall accuracy. Algorithms for various

classes provide varying results. The TRP cannot be increased by an algorithm.
In 2019, Yong Jin and coworkers used Machine learning with human oversight. For each part, an in-house experimental network is built so it may be put to use in the appropriate contexts. The actual network environment has to be deployed. DNS information is not properly prepared.
The LSTM method was used by Nathezhtha. T, et al. (2019), and it can detect insider threats. A lower false alarm rate is achieved by the discovery of both malfunctioning and fresh user nodes. SVM and RDF were used by Yaokai Feng et al. (2018).
Honeypots are used to gather information, which is then used to improve performance and provide new capabilities. Detect ion maintains a steady performance when 40 features are added.
SVM, C4.5, KNN, and MLP were used by Doyeon Kim et al. (2018). In this way, both legitimate and malicious APs may be identified and prevented. In order to ensure the safety of sensitive information stored on smart devices for the duration of a person's life, a proper security system must be developed. It is also necessary to use smart algorithms to examine these.
When it came to RDF algorithm, Kinam Park et al. (2018) used it. Provides superior detection efficiency, especially on the initial training data. They are unable to function well in a variety of training scenarios.
Recently, SVM was used by V. Deepa et al. (2018). High detection rates, less false alarms, and improved accuracy are all results. It allows for effective DDoS attack detection on the data plane.

5. COMPARATIVE ANALYSIS OF MANY ASSAULTS

Several publications are discussed herein, each of which employs a unique algorithm or method for identifying and stopping an assault. Researchers have discovered that using k-means and support vector machines (SVM) on synthetic datasets yields superior outcomes.
In the below paragraph, you'll find the results of a comparative analysis of assaults and the various learning strategies employed by researchers on various datasets to eliminate these attacks and identify them in a timely manner. Individual outcomes highlight the value of tailoring one's strategy to a particular collection of facts and method of assault.
Black hole attacks were analysed by S.Vidhya et al. (2014). They created their own communication system from scratch. They use a method called the Message Digest 5 Algorithm. Their stated figures include a throughput of 83% and a packet delivery ratio of 95%.
Network assaults were analysed by Dheeraj Pal et al. (2014). They are use the KDD'99 dataset in their analysis. The Fuzzy-GA Fuzzy-Genetic Algorithm The detection rate they reported was 97.5 percent.
In 2016, researchers M. Elif Karslig El, et al. Specifically, they use NSLKDD as their data source. Their method of choice is the k-means algorithm. Accuracy= 80.119% is what they reported.
According to research by Xi Qin et al. (2015), DDoS attacks (DDoS). DARPA IDeval is the dataset they utilise. They use an approach based on an algorithm for entropy cluster analysis. According to the findings, their strategy is more effective than the standard one, especially when the DDoS traffic scale is more than 20%.
Recent research on DDoS assaults was conducted by S.Sumathi et al. (2018). KDDcup99 and DARPF are the data sets used by them. They employed data mining and a number of different ML approaches. Fuzzy c-mean = 98.7 percent accuracy in categorization, researchers say.
Unknown and novel variant assaults in a network setting were investigated by Ze-Hong Chen et al. (2018). They used the NSLKDD, CIDDS -001, GPRSWPA2, and synthetic c datasets. They used k-means clustering and support vector machine analysis. Accuracy was calculated to be 88.45 for NSL-KDD, 93.28% for CIDDS-001, 90.64% for GPRSWPA2, and 99.16% for a synthetic dataset.
Data fabrication attacks were investigated by Bhavya Kailkhura et al. (2017). It was their own network that they built. They used a learning strategy and the Robust distributed weighted average consensus method. Increased efficiency in detecting was one of the results.
DoS and DDoS assaults were investigated by Seyyed Meysam Tabatabai e Nezhad et al. (2016). They

tapped into the Darknet dataset. They use the Chaotic System and ARIMA Time Series Model to analyse the data. The team calculated a detection rate of 99.5%.

Clone assaults were analysed by P. Sherubha et al. (2019). They're using the KDD cup data set. They use a method called the Randomized Forest Multiobjective Cuckoo Search (RFMOCS) algorithm, which is based on adaptive random forests. Results showed an accuracy of 96%.

To further understand Enhanced Resilient Sensor Attack, Kang Yang et al. (2018) analysed sensor data from a 20-run unattacked EV3 and LEGO device. They used a whole new approach to detecting sensor attacks called EPI. It was determined that the rate of detection and identification was equal to 92%.

Attack detection in network traffic was investigated by Liangchen Chen et al. (2020). They used the KDD99 and CICIDS -2017 datasets. They used a technique called fuzzy entropy weighted natural nearest neighbour (FEWNNN). Results revealed that space was decreased by 33.3% for the KDD99 dataset and 25% for the CICIDS-2017 dataset.

6. CONCLUSION

Researchers have concluded that security analytics is critical for enhanced threat detection because it allows one to more accurately prioritise signals and alerts. That helps get us closer to a solution that will actually work. The ML technique can be used to further fortify cyber defences. For instance, the massive dataset collected by the cyber security industry is being analysed with data science methods. These documents have been kept for a long time so that they can be used to put threat intelligence in context. Regression, clustering, classification, and dimensionality reduction are all areas where ML in F-score is put to use. In addition, ML can be utilised to assess the efficiency of authentication systems, the profiling of data collected by smart metres, the safety of proofs involving human interactions, and the implementation of protocols. This review emphasises the significance of employing ML and DL during the detection phase and highlights the framework for doing so. The paper also includes a detailed analysis of the many varieties of cyberattacks that target computer networks and telephone systems. Some of the ways that ML can be used to enhance the effectiveness of attack detection are briefly discussed. These include malware analysis, spam detection, and intrusion detection. The report highlights the benefits and drawbacks of the various attack detection and prevention techniques. In the not-too-distant future, one of the most promising approaches will be implemented, and it will be used to detect numerous types of assault.

3. REFERENCES

[1] M. I. Jordan, T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” Science, pp. 255-260, 2015.

[2] Y. LeCun, Y. Bengio, G. Hinton, “Deep learning,” Nature, pp. 436-444, 2015.

[3] A. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, pp. 1-5, 2015.

[4] E. Blanzieri and A. Bryl, “A survey of learning - based techniques of email spam filtering,” Artificial Intelligence Review, pp. 3-30, 2008.

[5] J. Gardiner and S. Nagaraja, “On the Security of Machine Learning in Malware C8C Detection,” ACM Computing Surveys , pp. 441-446, 2016.

[6] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar, “Adversarial Machine Learning,” in ACM workshop on Security and artificial intelligence, pp. 43-58, 2011.

[7] K. Sravanthi, M. Shamila, Amit Kumar Tyagi, “Cyber Physical Systems: The Role of Machine Learning and Cyber Security in Present and Future”, Computer Reviews Journal, pp. 66-80, 2019

[8] Xing Fang, Maochao Xu, Shouhuai Xu, Peng Zhao, “A deep learning framework for predicting cyber-attacks rates”, EURASIP Journal on Information Security, pp. 1-11, 2019.

[9] Z. Zhan, M. Xu, S. Xu, “Characterizing honeypot-captured cyber-attacks: Statistical framework and case study” IEEE Trans. Inf. Forensic Secur. pp. 1775–1789 , 2013.

[10] M. Shamila, K. Vinuthna and T. Amit Kumar.” A Review on Several Critical Issues and Challenges in

IoT based e-Healthcare System" International Conference on Intelligent Computing and Control Systems (ICCS), pp. 23- 30, 2019.

[11] Junaidu Bello Marshall, Mua'zu Abdullahi Saulawa, "CYBER-ATTACKS: THE LEGAL RESPONSE", International Journal of International Law, pp. 1-18, 2015.
[12] Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Alessandro Guido, Mirco Marchetti, "On the Effectiveness of Machine and Deep Learning for Cyber Security", 10th International Conference on Cyber Conflict, pp. 371-390, 2018
[13] Youness Arjoune, Fatima Salahdine, Md. Shoriful Islam, Elias Ghribi, Naima Kaabouch,"A Novel Jamming Attacks Detection Approach Based on Machine Learning for Wireless Communication", International Conference on Information Networking (ICOIN), pp. 459-464, 2020.
[14] Andrea Pinceti, Lalitha Sankar, Oliver Kosut, "Load Redistribution Attack Detection using Machine Learning: A Data-Driven Approach", IEEE Power & Energy SocietyGeneral Meeting (PESGM), pp. 1-5, 2018.
[15] Tahir mehmood, helmi b md rais, "machine learning algorithms in context of intrusiondetection", 3rd international conference on computer and information sciences (iccoins), pp. 369-373, 2016.
[16] yong jin, masahiko tomoishi, satoshi matsuura, "a detection method against dns cache poisoning attacks using machine learning techniques", ieee 18th international symposium on network computing and applications (nca), pp. 1-3, 2019.
[17] T. Nathezhtha, V. Vaidehi, "Cloud Insider Attack Detection Using Machine Learning",International Conference on Recent Trends in Advanced Computing (ICRTAC-CPS), pp. 60-65, 2018.
[18] Yaokai Feng, Hitoshi Akiyama, Liang Lu, "Feature Selection For Machine Learning-Based Early Detection of Distributed Cyber Attacks", IEEE 16th Int. Conf. on Dependable, Autonomic & Secure Comp., pp. 173-180, 2018.
[19] Doyeon Kim, Dongil Shin, Dongkyoo Shin, "17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications", pp. 1876-1878, 2018.
[20] Kinam Park, Youngrok Song, Yun-Gyung Cheong, "Classification of Attack Types for Intrusion Detection Systems using a Machine Learning Algorithm", IEEE Fourth International Conference on Big Data Computing Service and Applications, pp. 282-286, 2018.
[21] V. Deepa, K. Muthamil Sudar, P. Deepalakshmi, "Detection of DDoS Attack on SDN Control plane using Hybrid Machine Learning Techniques", International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 299-303, 2018.
[22] S. Vidhya, T.Sasilatha, "Performance Analysis of Black Hole Attack Detection Scheme using MD5 Algorithm in WSN", International Conference on Smart Structures & Systems (ICSSS-2014), pp. 51-54, 2014.
[23] Dheeraj Pal, Amrita Parashar, "Improved Genetic Algorithm for Intrusion Detection System", Sixth International Conference on Computational Intelligence and Communication Networks, pp. 835-839, 2014.
[24] Raj, Jennifer S. "Machine Learning Based Resourceful Clustering With Load Optimization for Wireless Sensor Networks." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 01, pp.29-38, 2020.
[25] Duraipandian, M. ,"Performance Evaluation Of Routing Algorithm For Manet Based On The Machine Learning Techniques." Journal of trends in Computer Science and Smart technology (TCSST), pp.25-38, 2019.