Artificial Neural Networks for Enhanced Cybersecurity Through Event Profile Analysis

Veera Venkata Gayatri Alamanda[1], S. Surya Godha Devi[2]
#1. M.Tech Scholar in Department of Artificial Intelligence,
#2. Assistant Professor, Department of Artificial Intelligence, Kakinada Institute Of Engineering & Technology
or Women, AP, India.

ABSTRACT

The modern world is completely reliant on the internet for all aspects of daily life. With each passing day, the amount of time spent in the virtual world is increasing. Everyone in the world has money to spend more time than ever before is spent on the Internet. [7] Consequently, the dangers the number and severity of cyber-threats and -crimes are both rising.

"Cyber" is a phrase used to describe "threat" refers to criminal behaviour that is carried out through the use of the Internet. The methods used by cybercriminals are evolving as a result of passing through the barrier is now possible.

Conventional there is no method that can identify zero-day attacks. Attacks with a high degree of sophistication. [5] There has been a lot of machine learning so far. Detection methods for cybercrimes have been created a fight against cyber attacks. The purpose of this investigation presents an assessment of many commonly used machines. Gaining knowledge of detection methods for some very dangerous risks to cyberspace from cyber attacks. A basic machine learning framework consisting of three components the main focus of the research is on approaches, especially strong religious belief. a network, a decision tree, and an SVM As of right now, we don't have any examined the effectiveness of these for a brief period of time spam detection, intrusion detection, and other applications of machine learning Based on commonly used and known malware detection and prevention techniques datasets for comparison purposes.

Index Terms— Intrusion Detection System, Malware Detection, Spam Classification, Performance Evaluation, Machine Learning Application.

I. INTRODUCTION

All digital technologies from all over the world can be exchanged in cyberspace, a global environment that enables this exchange. An electronic document, music, video, image, and tweet are all examples of resources. The Internet, technical expertise, system resources, data, and unskilled users are all part of the cyberspace. Information and resources may be easily accessed around the world because to the internet's worldwide reach. Cyberspace is now the primary means of transferring data and exchanging information, despite the fact that the costs and benefits of doing so are rapidly increasing. In the year after 2017, the internet grew in popularity. There has been an 81% increase in internet use in industrialised countries, and it's continuing expanding worldwide [1]. Cybercrime and cyber dangers have increased as a result of the rise in cyberspace.

Cyber-security has also achieved a significant number of improvements to compete against cybercrimes as the range of threats grows. There are a number of technologies, specialists, and methods that are utilised to safeguard the cyberspace from cybercriminals [2]. [1] Traditional information technology as well as automated cyber security are two of the most common ways to cyber security. Unqualified individuals, poorly configured system resources, and limiting access to clean data are a few of many drawbacks of conventional cyber security that amplify cybercrime [3]. Automated cyber security is the future of the field. The requirement for advanced and automatic cyber security measures is critical. '

To keep up with the ever-changing nature of cybercrime, they have the ability to learn from previous experiences and detect new polymorphic cyber-attacks [4]. A cyber-attack is an attempt to steal data, violate integrity standards, or otherwise damage a computer system or network. Phishing, malware, IOT device attacks, denial of service attacks, spam and intrusions into networks and mobile devices are just a few examples of cyber threats. This paper discusses malware detection, intrusion detection, and spam detection.

In the world of email, spam is a term used to describe an unwanted or unsolicited email. Spam e-mails are typically used to promote products or spread bogus information. It consumes computer and network resources like network bandwidth, memory, and time [7]. Malware is a further cyber-threat. Antivirus software, also known

as anti-malware and antispyware, are two types of anti-malware that are used to protect computers from malware. Malware includes a wide range of threats, including viruses, worms, ransomware, spywares, malware, malvertising, and Trojan horses [8]. Another cyberattack to cyberspace is malicious intrusions on computer networks and devices. The purpose of these intrusions is to identify and scan a computer system's vulnerabilities. For this reason, intrusion detection (IDS) is employed to protect the network. Signature/misuse, anomaly and hybrid intrusions are all types of intrusions. [9–10]. By using machine learning (ML) as a primary defence against cyber threats, it is able to overcome the shortcomings of conventional security measures [11]. Despite its many advantages, machine learning has its limitations and drawbacks. A classification of artificial intelligence (AI) is machine learning [12]. One of the most fascinating aspects of machine learning is that it does not require explicit programming because it can learn from its own experience to produce results [13].

Aside from cyber security, medical science and education, machine learning techniques are being used in practically every other sector of our lives as well. This is due to the numerous advantages of machine learning techniques. There are numerous machine learning techniques that have been used to identify and categorise various cyber threats. Decision trees, random forests, naive Bayes, support vector machines, K-nearest neighbours, deep belief networks, artificial neural networks, and K-means are only a few of the machine learning approaches commonly utilised [14, 15]. The decision tree, max pooling layer, and support vector machine algorithms have all been considered in this article, however On the basis of commonly used and benchmark datasets, we've presented a comparative study of machine learning algorithms.

II. METHODOLOGY

Deep learning techniques including FCNN, CNN, and LSTM (long short term memory) are used to identify assaults in this paper's AI-SIEM (Artificial Intelligence-Security Information and Systems Involving) method. [6] The effectiveness of the suggested work is assessed using SVM, Decision Tree, Random Forest, KNN, as well as Naive Bayes. I'm implementing CNN and LSTM techniques here.

The following modules make up a proposed algorithm.

Data Parsing: This module creates an original data ability to establish by parsing an input dataset.

TF-IDF: In order to create an event vector that contains both normal and attack signatures, we will use this module.

Event Profiling Stage: Based on profile events, the information generated would be split into training and testing model. Deep Learning Neural Network Model: This module offers a framework for future analysis based on data from the training and testing sets [14]. Recall, precision, and FMeasure will be assessed on test data using the resulting training set. So, an approach that learns appropriately would be more accurate, and that notion would be picked for use in an actual system to detect attacks.

To execute all the datasets takes five to 10 minutes if all techniques can be executed, however KDD training analysis [9] of the research runs flawlessly. Additional datasets could be examined by lowering their length or executing them on a more powerful system.

III. RESULT AND DISCUSSION

To run project to get below screen

It's easy to upload your train dataset by clicking the Upload Train Dataset tab on the upper right of the result, then selecting KDD train.csv.

Now that we've seen that the dataset comprises 9999 entries, we can click on the 'Run Pre-processing TF-IDF Algorithm' tab to transform the raw dataset into TF-IDF values.
If you want to produce an event vector from TF-IDF, you can click on the "Generate Event Vector" tab.


In the above result, we can see the total number of unique event names, as well as the overall size of the dataset and the percentage of the dataset used for training and testing. To generate an LSTM and CNN model, click the "Neural Network Profiling" button after the data has been prepared and the training and testing events models have been completed.


We can see the accuracy, precision, recall, and FMeasure values in the results shown above. Afterwards, Run all Algorithms and to see the accuracy of all algorithms, click the 'Accuracy Comparison Graph' button.

We can see from the graph above that LSTM and CNN perform well on the y-axis, which depicts algorithm name and the x-axis, which depicts accuracy. To get the graph below, click on 'Precision Comparison Graph'.
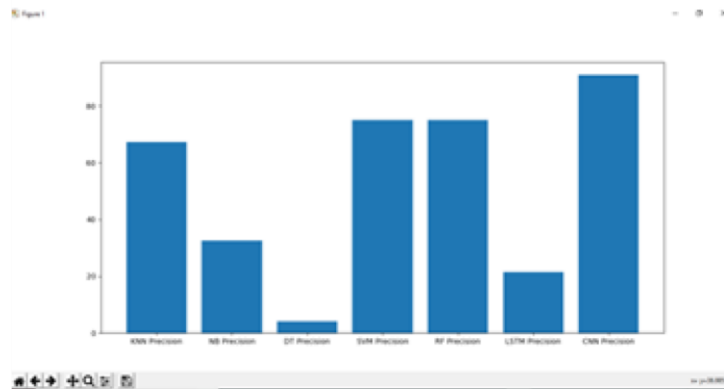
The graph above shows how well CNN performs; now select 'Recall Comparison Graph' to see how well it compares to the competition.
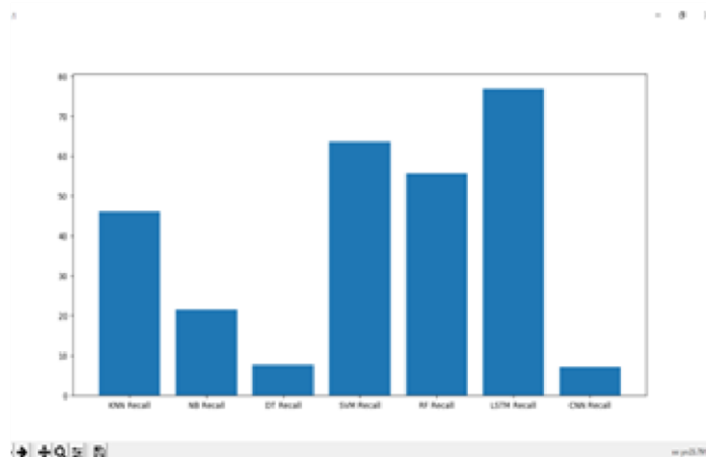
The LSTM is showing good results in the graph above; to see the comparison graph shown below, select the FMeasure Comparison Graph button.
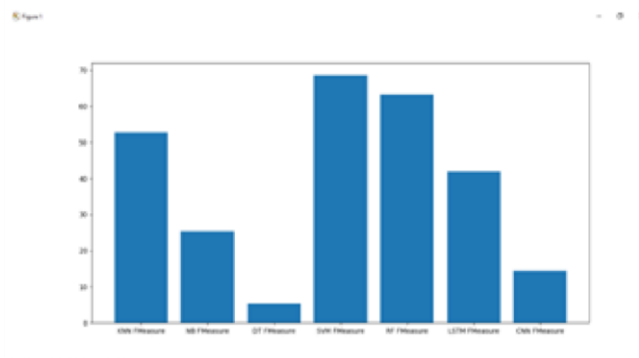
We can observe from the comparison graphs that LSTM and CNN perform well in terms of accuracy, recall, and precision.

IV. CONCLUSION

At an ever-increasing rate, cyber risks are on the rise. In the face of such dangers, traditional security measures are woefully inadequate. In order to overcome the limits of conventional security systems, machine learning algorithms are being used. At both the defender's and the attacker's end, machine learning approaches are making an impact. [10] For the detection and classification of incursion, spam, and malware, we have evaluated the effectiveness of three machine learning models (MLP). According to our findings, we have compared the results of the evaluations using regularly used and benchmark datasets. According to the information presented before, we cannot suggest a specific training method for every type of cyber threat. A variety of training methods are being utilised to combat a variety of cyberattacks [15].

Many authors have sought to emphasise the limitations of machine learning approaches, which is a good thing. We have observed and indicated that a new benchmark dataset is urgently needed to test the most recent advancements in machine learning for cybercrime identification. [12] Missing values and a lack of diversity characterise the datasets that are currently available. There is a pressing demand for specialised and tailored training programmes for security-related purposes. Detection of cyber threats will be the focus of our future research.

REFERENCES

[1]. I. Firdausi, A. Erwin, and A. S. Nugroho, "Analysis of machine learning techniques used in behavior-based malware detection," in 2010 second international conference on advances in computing, control, and telecommunication technologies, 2010: IEEE, pp. 201-203.
[2]. G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," in 2018 10th International Conference on Cyber Conflict (CyCon), 2018: IEEE, pp. 371-390.
[3]. F. Mercaldo and A. Santone, "Deep learning for image-based mobile malware detection," Journal of Computer Virology and Hacking Techniques, pp. 1-15, 2020.
[4]. J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, S. J. Shepherd, and R. A. Abd-Alhameed, "An autonomous host-based intrusion detection system for android mobile devices," Mobile Networks and Applications, vol. 25, no. 1, pp. 164-172, 2020.
[5]. C. Chen et al., "A performance evaluation of machine learning-based streaming spam tweets detection," IEEE Transactions on Computational social systems, vol. 2, no. 3, pp. 65-76, 2015.
[6]. Z. Chen, S. Liu, K. Jiang, H. Xu, and X. Cheng, "A data imputation method based on deep belief network," in
[7]. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015: IEEE, pp. 1238-1243.
[8]. D. M. Farid, N. Harbi, and M. Z. Rahman, "Combining naive bayes and decision tree for adaptive intrusion detection," arXiv preprint arXiv:1005.4496, 2010.
[9]. Q. J. Ross, "C4. 5: programs for machine learning," San Mateo, CA, 1993.
[10]. P. S. Oliveto, J. He, and X. Yao, "Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results," International Journal of Automation and Computing, vol. 4, no. 3, pp. 281-293, 2007.
[11]. C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data mining and knowledge discovery, vol. 2, no. 2, pp. 121-167, 1998. [39] G. D. Forney, "The viterbi algorithm," Proceedings of the IEEE, vol. 61, no. 3, pp. 268-278, 1973.
[12]. M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," in The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), 2014: IEEE, pp. 1-6.
[13]. D. Gunawan, R. F. Rahmat, A. Putra, and M. F. Pasha, "Filtering Spam Text Messages by Using Twitter-LDA Algorithm," in 2018 IEEE International Conference on Communication, Networks and Satellite (Comnetsat), 2018: IEEE, pp. 1-6.
[14]. B. Klimt and Y. Yang, "Introducing the Enron corpus," in CEAS, 2004. [45] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in 2015 International Conference on Signal Processing and Communication Engineering Systems, 2015: IEEE, pp. 92-96.
[15]. A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A novel hierarchical intrusion detection system based on decision tree and rules-based models," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2019: IEEE, pp. 228-233.
[16]. R. Kokila, S. T. Selvi, and K. Govindarajan, "DDoS detection and analysis in SDN-based environment using support vector machine classifier," in 2014 Sixth International Conference on Advanced Computing (ICoAC), 2014: IEEE, pp. 205-210.