International Journal for Modern Trends in Science and Technology, *9*(04): 178-181, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0904029
Available online at: http://www.ijmtst.com/vol9issue04.html

Detection of Intrusions in Network using Machine Learning Algorithms

M. Ravi Kanth[1] , S. Jahnavi[2] , P. Yuva Raju[3] , R. Chanakya[4] , M. Sujith Bhanu[5]

[1]Assistant Professor, Department of CSE, Dhanekula Institute of Engineering & Technology, A.P, India
[2]Department of CSE, Dhanekula Institute of Engineering & Technology, A.P, India

To Cite this Article
M. Ravi Kanth, S. Jahnavi, P. Yuva Raju, R. Chanakya, M. Sujith Bhanu. Detection of Intrusions in Network using Machine Learning Algorithms, International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 178-181. https://doi.org/10.46501/IJMTST0904029

Article Info
Received: 08 March 2023; Accepted: 02 April 2023; Published: 05 April 2023.

ABSTRACT

Cyber assaults are a sensitive subject in the area of Internet security. Businesses and governments everywhere are working very hard to protect sensitive data. Competitors employ a number of techniques and strategies to sustain the business while attempting to breach security and spread harmful software like botnets, viruses, Trojans, etc. to access crucial data. Due to the vast volume of data being transmitted over the Internet, it is necessary to secure the data. An intrusion detection system (IDS) can aid in this situation by assisting in the detection of any virtual security risks. In order to find any intrusions into the system or network, intrusion detection systems (IDS) monitor and analyse data. Several methods are used by intruders to enter a network. In order to classify assaults, detect them whenever an attack occurs, and determine which machine learning technique is most appropriate for recognising the attack, the suggested IDS is being built utilising cutting-edge technology like machine learning algorithms. IDS protects against unauthorised user access and monitors networks and systems for harmful activity. This study's goal is to assess current IDS research using a Machine Learning (ML) strategy with a concentration on datasets.
KEY WORDS: Machine Learning, Cyber Security, Decision Tree, SVM, Random Forest, Multilayer Perceptron.

1. INTRODUCTION

The strong demand for internet use is expanding quickly, as are the dangers to the network. According to Symantec research from 2016, the company only found over 430 million new malware programmes in 2015, a 36 percent rise from the previous year. There are many other types of attacks that can be used, including Brute Force Attacks, Heartbleed Attacks, DoS Attacks, DDoS Attacks, Web Attacks, etc. The network's bandwidth is expanding quickly in tandem with the growth of internet users. Nowadays, the normal speed ranges from 1 Gbps to 10 Gbps for a typical data centre. For large corporations or giant tech firms like Google, Facebook, etc., the download and upload speeds range from 40 Gbps to 100 Gbps. A security tool known as a network-based intrusion detection system (NIDS) guards against internal and external attacks as well as unwanted access to networks that are created by both software and hardware. The most well-known idea is that of a firewall, which is designed to guard against unwanted access by IP address and port number throughout the whole network and to manage these actions using NIDS. Its many and varied working applications include counting the number of intrusion attempts on the network, such as denial of service attacks and hacking activities that could jeopardise the

security of any one computer or the entire network. NIDS are typically installed outside the firewall, where they can monitor all external traffic by sensing and detecting anomalous activity.

2. LITERATURE SURVEY

Khalil Ibrahimi and Mostafa Ouaddane published Management of Intrusion Detection Systems based-KDD99: Analysis with LDA and PCA in 2017. Recently, the computer and network security sectors have devoted a lot of time to researching the intrusion detection problem. The Intrusion Detection System (IDS) then becomes a hot research issue, especially in data mining and machine learning. Principal component analysis (PCA) and linear discriminant analysis (LDA) classification techniques are used to detect intrusions and classification anomalies in order to increase classification accuracy and lower the high false alarm rate from the traditional data base such as KDD99 or others. The NSL-KDD data set is used in the IDS studies to enhance the current classification strategies.

Saqr Mohammed Almansob and Santosh Shivajirao Lomte wrote Addressing Challenges for Intrusion Detection System Using Naive Bayes and PCA Algorithm. The largest problem that networks today are facing is defending against outside threats. Hence, an intrusion detection system is a tool for monitoring and analysing every computer or network activity. Two solutions to issues with intrusion detection systems were suggested in this research. One of these methods is called Principal Component Analysis (PCA) for feature extraction and uses the Naive Bayes method to solve classification problems. The model was therefore applied to the KDD99 dataset. The results collected indicate an improvement in detection and accuracy rates as well as a reduction in false positive rates.

Doukan Aksu, Serpil Üstebay, Muhammed Ali Aydin, and Tülin Atmaca published Intrusion Detection with Comparative Analysis of Supervised Learning Techniques and Fisher Score Feature Selection Algorithm (2018). Rapid technological advancement not only simplifies daily living but also highlights several security flaws. Attack kinds that are developing and evolving have an impact on several individuals, groups, and businesses. In order to prevent these kinds of losses, intrusion detection systems have been developed. In this study, we used the CICIDS2017 dataset, which includes both typical and cutting-edge benign assaults. The Fisher Score algorithm is used to choose the best features. With the help of the algorithms for Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Decision Tree (DT), real-world data that was extracted from the dataset is categorised as benign or DDoS. The study's findings led to success rates of 0,9997%, 0,5776%, and 0,99%, respectively.

Rashmi TV published Forecasting the System Failures Using Machine Learning Algorithms in 2020. The ability to recover quickly continues to be a major challenge for the designers and managers of large organised frameworks. But, in order to recover from a failure, one must first recognise and understand the setback. Without a set setback expectation, this makes recovering from a failure difficult. Any component failure causes the system to slow down. Failure Prediction aids in the early discovery of system failure, allowing us to prevent the damage brought on by that specific system's failure. We can prevent the loss of data and application services by anticipating the breakdown. By making a prediction, we may also determine whether the framework will fail, as suggested by the analysis of the evaluation of the framework's current conduct and authentic investigation.

3. RELATED WORK

This section highlights several recent accomplishments in this area. You should be aware that we only look at research that used the NSL-KDD dataset for performance benchmarking. Every dataset mentioned after this point should be thought of being NSL- KDD. This strategy enables a more thorough comparison of the work with other materials discovered in the writing. Another limitation is that most work uses information preparation for both planning and testing. Finally, we look at a few deep learning-based approaches that have been used in the past for work of a similar nature.

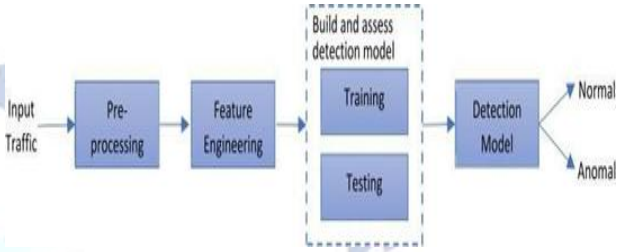
One of the most timely pieces of literature used an ANN with a strengthened back-spread for the design of such an IDS. Just the preparation dataset was used in this study for planning (70%), approval (15%), and testing (15%). Use of unlabeled data for testing resulted in a decrease in execution, as was to be expected. Further research tested the J48 decision tree classifier on the preparation dataset using 10-overlay cross-approval. Instead of using the entire arrangement of 41

capabilities, this work used a smaller list of 22 capabilities. A similar study compared other well-known regulated tree-based classifiers and discovered that the Random Tree model performed best with the highest degree of accuracy and the lowest false alarm rate. Moreover, a number of 2-level characterization techniques have also master presented. One such study used 10-crease cross approval for testing, Nominal-to-Binary directed separation at the second level, and Discriminative Multinomial Naive Bayes (DMNB) as the basis classifier. Ensembles of Balanced Nested Dichotomies (END) at the primary level and Random Forest at the secondary level were used in this work. In line with expectations, this update led to a higher location rate and a reduced fake positive rate. Another 2-level execution, which used PCA to reduce the list of capabilities and SVM (using Radial Basis Function) for final classification, produced good recognition precision using only the training dataset and the entire collection of 41 features. A drop in features set to 23 improved location accuracy in several of the attack classes, but the overall performance was lowered. By using data to rank the highlights and then a conduct-based element determination to narrow the list of capabilities to 20, the authors enhanced their job. Using the preparation dataset, this resulted to an increase in detailed precision. The test and preparation datasets were both used in the next class to be examined. This classification's underlying effort combined fluffy characterisation with hereditary computation, which resulted in detection accuracy of 80%+ and a low false-positive rate.

Another important piece of research used unassisted grouping algorithms to discover that the exhibition using only preparation information was drastically reduced when test information was also included. Using both training and test datasets, a comparable execution using the k-point computation resulted in a slightly higher recognition accuracy and a reduced false positive rate. As compared to SVM RBF methodology, a different, less well-known method called OPF (optimal way woodlands), which employs chart apportioning for include classification, was discovered to have a good identification accuracy within 33% of the time.

4. PROPOSED SYSTEM

Generation better accuracies than the previous implementation is the target. It is observed that the existing system is lacking efficiency an accuracy for certain algorithms. So, various datasets are implemented on ML algorithms and intrusion detection accuracies are measured.



Implementation:

1) Every dataset should be normalised.
2) Create training and testing datasets using that dataset.
3) Use ML algorithms to create IDS models.
4) Assess the effectiveness of each model.

Advantages of Proposed System:

- Defense against harmful assaults on your network.
- Eliminating or ensuring the presence of harmful components within an already- established network.
- Prevents people from using the network in an unlawful way.
- Block access to resources that could be contaminated for programmes.
- Protecting sensitive data.

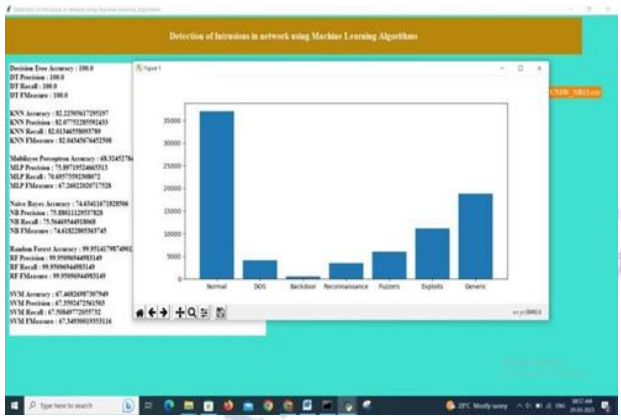
5. RESULTS

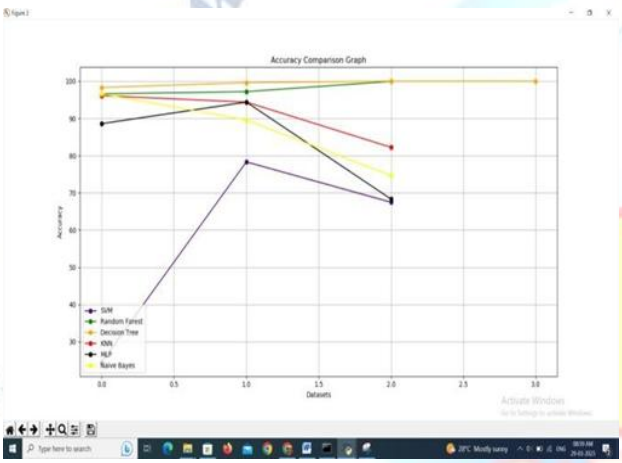
The used ML algorithms produce superior accuracy results. The performance of different machine learning algorithms, including SVM, Random Forest, Decision Tree, Naive Bayes, KNN, and Multilayer Perceptron, is discussed in this work. Decision Tree and KNN perform better than the other methods. Using a filter- based features selection technique, which will choose the most significant attributes from the training dataset and eliminate any that have correlation values below a predetermined threshold, will shorten training times.

We employed 3 datasets, including KDDCUP, NSLKDD, and CICIDS2017, to carry out this study. These datasets feature numerous attacks, including Denial of Service, R2L, U2R, Probe, and many others.

In order to conduct the study, libraries like numpy, pandas, and scikit-learn were employed. Using the

Python programming language and the Jupyter notebook integrated development environment (IDE), the application is being created.





6. CONCLUSIONS:

The used ML algorithms (Support Vector Machine, Random Forest, Decision Tree, Naive Bayes, and K-Nearest Neighbor) produce better accuracies with three different datasets than the existing system, which is deficient in efficiency and accuracy for some algorithms when detecting the intrusions of web pages. KNN and Decision Tree outperform the competition. We want to use Apache Hadoop, key learning algorithms, AI, and other technology advancements in the future to defend against attacks utilising this information. Port scope operations and other types of assaults will both be met with our defences.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Rashmi TV. "Predicting the System Failures Using Machine Learning Algorithms". International Journal of Advanced Scientific Innovation, vol. 1, no.1, Dec.2020, doi:10.5281/zenodo.4641686.

[2] K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp.1–6.

[3] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141– 149.

[4] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[5] Yalei Ding and Yuqing Zhai," Intrusion Detection System for NSL- KDD Dataset Using Convolutional Neural Networks," doi.org/10.1145 /3297156.329720.

[6] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[7] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.