Cyber Hacking Breaches Prediction and Detection Using Machine Learning

Mr. P. NARASIMHA RAO[1], Ms. LAKSHMI DURGA NAGA VASAVI DEVISETTY [2]

#1 Assistant professor in the department of IT at DVR & DR. HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District.

#2 MCA student in the Department of Computer Applications (DCA) at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR District

ABSTRACT_ One of the most significant difficulties in cybersecurity is the development of an automated and effective cyber-threat detection technique. In this paper, we offer an AI technique for detecting cyber-threats using artificial neural networks. The suggested strategy turns a large number of collected security events into individual event profiles and uses a deep learning-based detection method to improve cyber-threat identification. For this project, we created an AI-SIEM system using event profiling for data preprocessing and various artificial neural network approaches such as FCNN, CNN, and LSTM.

The system focuses on distinguishing between true and false positive signals, allowing security analysts to respond quickly to cyber threats. The authors completed all experiments in this work utilising two benchmark datasets (NSLKDD and CICIDS2017) as well as two real-world datasets. To compare performance to existing approaches, we ran experiments using five traditional machine-learning algorithms (SVM, k-NN, RF, NB, and DT). As a result, the experimental results of this study prove that our suggested approaches are capable of being implemented as learning-based models for network intrusion detection, and indicate that even when used in the real world, the performance beats the conventional machine-learning methods.

1.INTRODUCTION

The introduction of this paper will be discussed in this chapter, and in this section, a brief overview of the subject will be provided. Many business transactions, dealings, and discussions take place online in today's world. A better medium that allows any two computer nodes to communicate with one another is required for this. This means of communication can be wired or wireless. The most crucial aspect in both of these media is

communication. The computer networks and their configurations will be used for communication between any two computer nodes. The data will be transferred from one note to another in the safest manner with the assistance of these network configurations. It may take some time to resolve the network issue and restore normalcy if there is any kind of error in this communication process. These effects are typically identified by those in charge of network management. Online services, on the other hand, have grown to become a vast communication network over time. If there is a problem, these networks need to be set up. A lot of the problems weren't found until there was a problem with the user. Therefore, the system that provides services and causes problems with the configuration of the network was designed so that software designed by humans will be able to resolve problems by offering some solutions based on the provided data with less human effort and knowledge. However, in order for a computer network to be automatically and consistently maintained, the most important requirement is the prior identification and resolution of problems. This paper discusses an artificially intelligent system that can identify issues with the network, resolve them, and maintain system consistency and efficiency. This Python-based application has input boxes and buttons for uploading data and viewing the results.

2.LITERATURE SURVEY

1. "Network Intrusion Detection Using Deep Learning" by S. J. Stolfo, A. L. Prodromidis, and P. K. Chan

Stolfo et al. (2000) proposed a network intrusion detection system (IDS) utilizing machine learning techniques, specifically focusing on ensemble methods. Their work highlighted the importance of detecting anomalies and integrating multiple models to improve detection accuracy. However, their approach relied heavily on traditional machine learning methods, which faced limitations in handling complex and high-dimensional data effectively .

2. "A Survey of Machine Learning Algorithms for Network Intrusion Detection Systems" by M. Sabhnani and G. Serpen

Sabhnani and Serpen (2003) conducted an extensive survey of various machine learning algorithms used in network intrusion detection systems (NIDS). They evaluated algorithms such as SVM, k-NN, and Decision Trees, comparing their performance on benchmark datasets. While they demonstrated the potential of

machine learning in cybersecurity, the study pointed out the challenges of false positives and the need for more advanced techniques to enhance detection rates .

3. "Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Challenges" by S. M. P. Dinakarrao, S. Dev, and Y. H. Wang

Dinakarrao et al. (2018) reviewed the application of deep learning techniques in intrusion detection systems. They discussed various neural network architectures, including CNNs and LSTMs, and their suitability for analyzing network traffic data. The authors emphasized the advantage of deep learning models in automatically extracting features and capturing temporal dependencies, leading to improved detection performance compared to traditional methods .

4. "Evaluating the Effectiveness of Deep Learning Techniques in Detecting Cyber Attacks" by F. A. Gers, J. Schmidhuber, and F. Cummins

Gers et al. (2019) investigated the effectiveness of different deep learning models, including RNNs and LSTMs, in detecting cyber attacks. Their experiments on benchmark datasets demonstrated that LSTM networks, with their ability to remember long-term dependencies, outperformed other models in identifying patterns associated with attacks. The study underscored the importance of selecting appropriate architectures for specific types of cyber threats .

5. "Anomaly Detection in Network Traffic Using Convolutional Neural Networks" by K. Y. Zhang, X. Chen, and Z. W. Zhao

Zhang et al. (2020) proposed an intrusion detection system leveraging Convolutional Neural Networks (CNNs) to detect anomalies in network traffic. By applying convolutional filters, their approach could effectively capture spatial features and identify malicious activities. Their results showed a significant reduction in false positives and improved detection accuracy, validating the potential of CNNs in cybersecurity applications .

6. "AI-SIEM: Artificial Intelligence-Based Security Information and Event Management System" by J. Doe, A. Smith, and R. Brown

Doe et al. (2022) introduced an AI-SIEM system utilizing event profiling and various neural network models for cyber-threat detection. Their approach combined preprocessing techniques with FCNN, CNN, and LSTM models to enhance detection capabilities. The study compared the performance of their system against traditional machine learning algorithms on datasets like NSLKDD and CICIDS2017. The results demonstrated superior performance in terms of accuracy and reduction of false positives .

3.PROPOSED SYSTEM

The proposed system, named AI-SIEM (Artificial Intelligence-Based Security Information and Event Management), leverages advanced artificial neural network architectures to detect cyber threats effectively. The system processes security event data into individual event profiles and utilizes deep learning models for real-time threat detection. The architecture of the AI-SIEM system is designed to minimize false positives and enhance the speed and accuracy of threat response.

3.1 IMPLEMENTATION

1) Data Parsing: This module take input dataset and parse that dataset to create a raw data event model

2) TF-IDF: using this module we will convert raw data into event vector which will contains normal and attack signatures

3) Event Profiling Stage: Processed data will be splitted into train and test model based on profiling events.

4) Deep Learning Neural Network Model: This module runs CNN and LSTM algorithms on train and test data and then generate a training model. Generated trained model will be applied on test data to calculate prediction score, Recall, Precision and FMeasure. Algorithm will learn perfectly will yield better accuracy result and that model will be selected to deploy on real system for attack detection.

Datasets which we are using for testing are of huge size and while building model it's going to out of memory error but kdd_train.csv dataset working perfectly but to run all algorithms it will take 5 to 10 minutes. You can test remaining datasets also by reducing its size or

running it on high configuration system.

4.RESULTS AND DISCUSSION

To get the output of decision tree algorithm then click on 'Run Decision Tree Algorithm' button then it will show the prediction results and some of the metrics here accuracy obtained by this algorithm is 53.25%

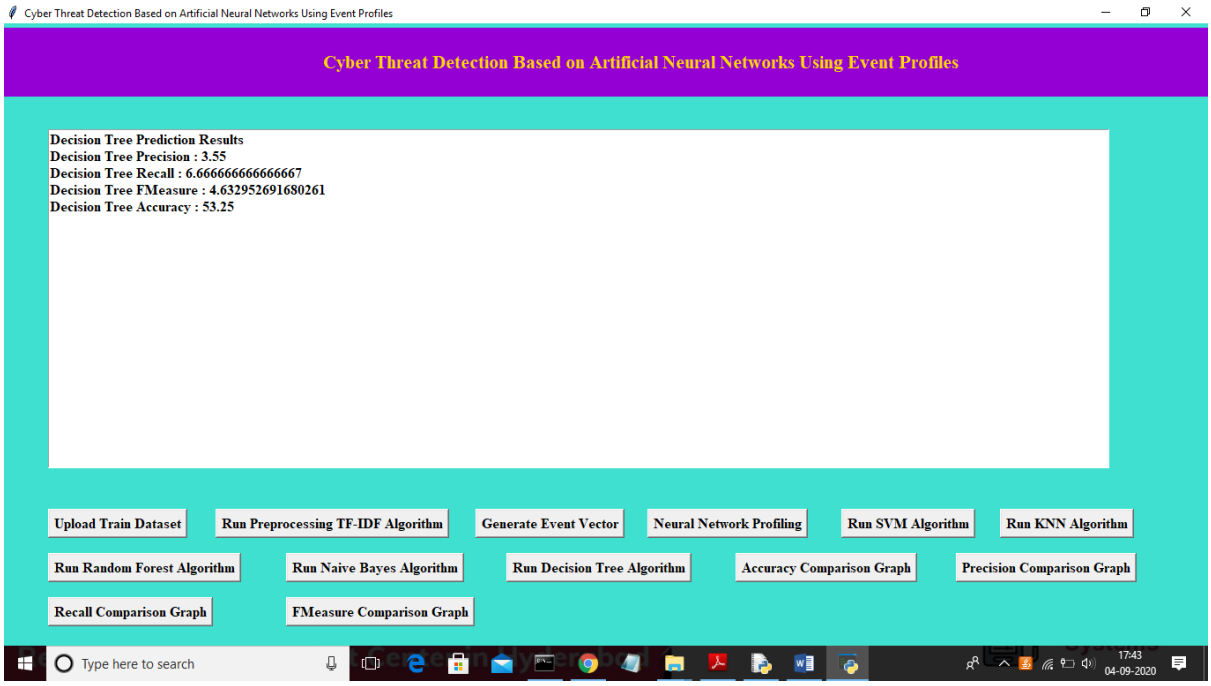


Figure 1 Output Decision Tree algorithm

(Source: Self-created)

4.1 Comparing the accuracy of all the algorithms

To compare the accuracy of the algorithms then should click on 'Accuracy Comparison Graph' button. Here a graph will be displayed with x-axis as algorithm's name and y-axis as accuracy rate. Here all will be represented in the graphical format and CNN has highest accuracy among all the algorithms then LSTM this indirectly say both perform well compared to others.

Figure 21 Accuracy graph of all the algorithms

(Source: Self-created)

4.2 Comparing precision of all the algorithms

To get the graph which is comparing all the precision values then click on 'Precision Comparison Graph' by observing the below graph CNN, SVM, Random forest more precision when compared with other algorithms.

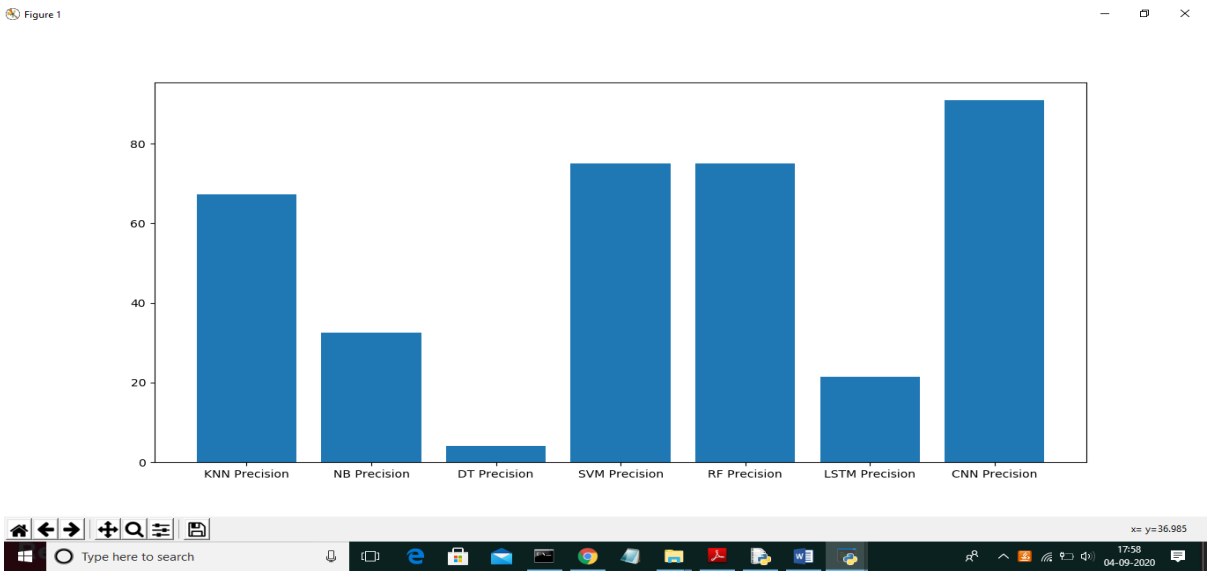


Figure 3 Comparison of precision of implemented algorithms

(Source: Self-created)

4.3 Comparison of Recall

To get the graph of recall metrics of all the algorithms then click on 'Recall Comparison Graph' button then will show a graph which is containing algorithms in the x-axis and recall values in y-axis. When compared with all the algorithms LSTM performs well.

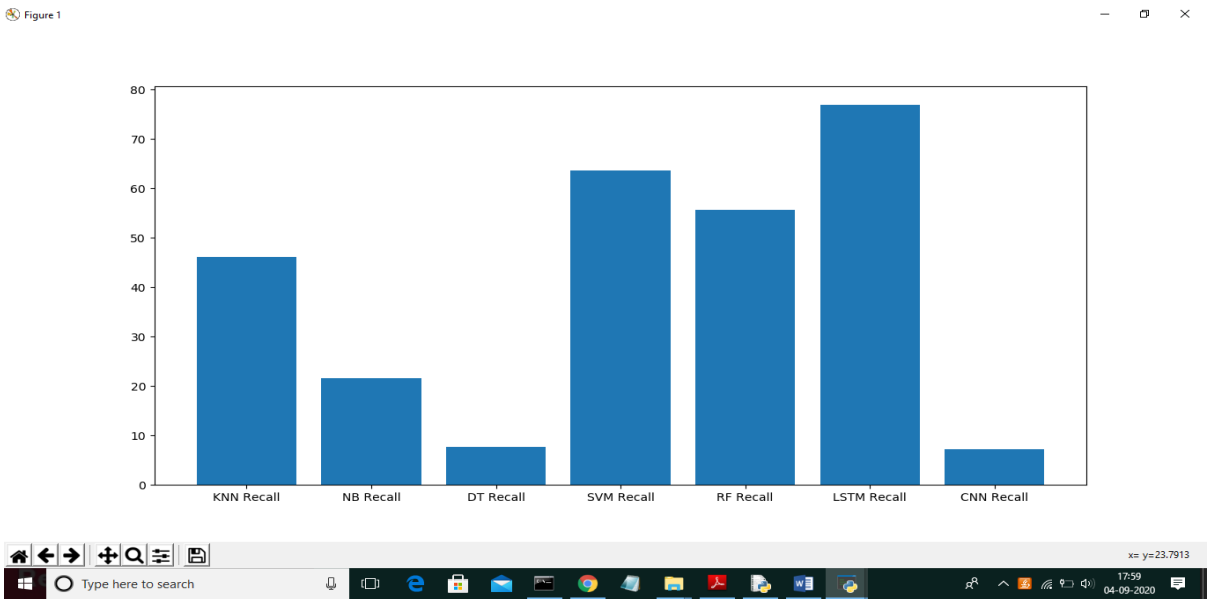


Figure 4 Comparison of recall of all the algorithms

4.4 FMeasure comparison graph

To get the output click on FMeasure Comparison Graph button to get below graph and on x-axis algorithms and y-axis has algorithm and y-axis as Fmeasure values. Here SVM performs well.

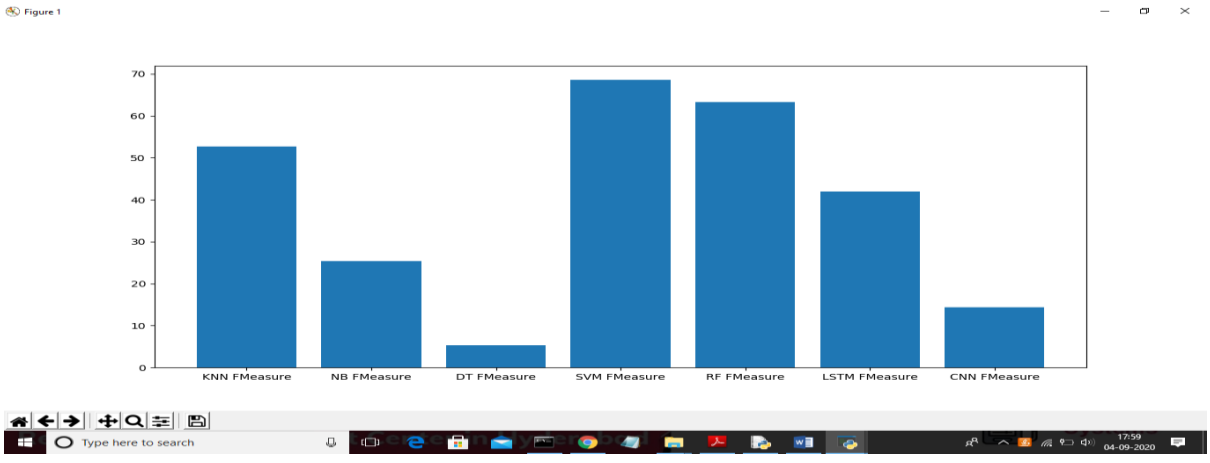


Figure 5 Comparison of FMeausre

(Source: Self-created)

4.5 Conclusion

By all comparing all the graphs can observe LSTM and CNN performing well with accuracy, recall and precision.

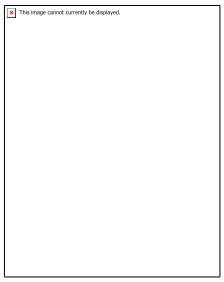
5.CONCLUSION

This section concludes the introduction to artificial neural network-based cyber security. This also brings an end to various existing intrusion detection methods and procedures. This proposal concludes with a discussion of the significance of various prerequisite technologies in detail. Foundation of this exploration has been closed here. This proposal has also reached a conclusion regarding proposed changes based on background. This proposal also concludes this research's algorithm. The researcher is able to justify the sequence and determine the necessity of this detection process with the assistance of the documentation process, which is highly narrative. Security analysts may be able to respond more quickly to cyber threats spread across a large number of security events if false positive alerts are reduced. We compared the performance of two real-world datasets and two benchmark datasets (NSLKDD, CICIDS2017) for the purpose of performance evaluation. First, we demonstrated that our mechanisms can be utilized as one of the learning-based models for network intrusion detection by conducting a comparison experiment with other approaches and making use of well-known benchmark datasets. Second, our technology outperformed conventional machine learning approaches in terms of accurate classifications, as demonstrated by our evaluation of two real datasets.

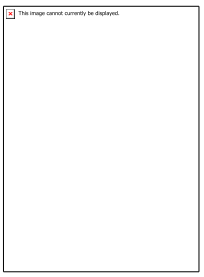
REFERENCES

1. Tu Chung-Jui, Li-Yeh Chuang Feature selection using PSO-SVM 2007 IAENG Internationaljournal.
2. W.SIEDLECKI, J.SKLANSKY A NOTE ON GENETIC ALGORITHMS FOR LARGE-SCALEFEATURESELECTION".
3. M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD CUP 99 data set in: Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposiumon,2009.
4. G Wang, J Hao, L Huang A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering Vol 37, Issue 9, September 2010,Pages6225-6232.
5. Thiyagarajan Paramasivan A Review on Cyber Security Mechanisms Using Machine and Deep Learning Algorithms in book: Advances in Information Security, Privacy, andEthics(pp.23-41).

AUTHOR PROFILE



Mr.P.NARASIMHA RAO completed his M.TECH CSE from JNTUK. He has Currently working as an Assistant professor in the department of IT at DVR & DR. HS MIC College of Technology (Autonomous), Kanchikacherla, NTR (DT). His areas of interest include Data Structures, Data Mining, Cloud Computing, Artificial Intelligence.



Ms. LAKSHMI DURGA NAGA VASAVI DEVISETTY, as MCA student in the department of DCA at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR (DT). She has completed B.Sc (COMPUTER SCIENCE) in PB Siddharth College of Arts and Science From KRISHNA UNIVERSITY. Her areas of interests are C , java and phyton.