International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

Cyber Threat Detection Based On Artificial Neural Networks Using Event Profiles

Chakala Navya[1], G Upendra reddy[2]
[1]*M.Tech Student, Department of CSE, Golden Valley Integrated Campus, Madanapalli*
[2]*Associate Professor, Department of CSE, Golden Valley Integrated Campus, Madanapalli*

Abstract:
One of the major challenges in cybersecurity is the provision of an automated and effective cyber-threats detection technique. In this paper, we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helping security analysts to rapidly respond to cyber threats. All experiments in this study are performed by authors using two benchmark datasets (NSLKDD and CICIDS2017) and two datasets collected in the real world. To evaluate the performance comparison with existing methods, we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection, and show that although it is employed in the real world, the performance outperforms the conventional machine-learning methods.
Keywords: Cyber security, intrusion detection, network security, artificial intelligence, deep neural networks

INTRODUCTION

With the emergence of artificial intelligence (AI) techniques, learning-based approaches for detecting cyber attacks, have become further improved, and they have achieved significant results in many studies. However, owing to constantly evolving cyber attacks, it is still highly challenging to protect IT systems against threats and malicious behaviors in networks. Because of various network intrusions and malicious activities, effective defenses and security considerations were given high priority for finding reliable solutions [1], [2], [3], [4].

Traditionally, there are two primary systems for detecting cyber-threats and network intrusions. An intrusion prevention system (IPS) is installed in the enterprise network, and can examine the network protocols and flows with signature-based methods primarily. It generates appropriate intrusion alerts, called the security events, and reports the generating alerts to another system, such as SIEM. The security information and event management (SIEM) has been focusing on collecting and managing the alerts of IPSs. The SIEM is the most common and dependable solution among various security operations solutions to analyze the collected security events and logs [5]. Moreover, security analysts make an effort to investigate suspicious alerts by policies and threshold, and to discover malicious behavior by analyzing correlations among events, using knowledge related to attacks.

Nevertheless, it is still difficult to recognize and detect intrusions against intelligent network attacks owing to their high false alerts and the huge amount of security data [6], [7]. Hence, the most recent studies in the field of intrusion detection have given increased focus to machine learning and artificial intelligence techniques for detecting attacks. Advancement in AI fields can facilitate the investigation of network intrusions by security analysts in a timely and automated manner. These learning-based approaches require to learn the attack model from historical threat data and use the trained models to detect intrusions for unknown cyber threats [8], [9].

A learning-based method geared toward determining whether an attack occurred in a large amount of data can be useful to analysts who need to instantly analyze numerous events. According to [10], information security solutions generally fall into two categories: analyst-driven and machine learning-driven solutions. Analyst-driven solutions rely on rules determined by security experts called analysts. Meanwhile, machine learning-driven solutions used to detect rare or anomalous patterns can improve detection of new cyber threats [10]. Nevertheless, while learning-based approaches are useful in detecting cyber attacks in systems and networks, we observed that existing learning-based approaches have four main limitations.

First, learning-based detection methods require labeled data, which enable the training of the model and evaluation of generated learning models. Furthermore, it is not straightforward to obtain such labeled data at a scale that allow accurate training of a model. Despite the need for labeled data, many commercial SIEM solutions do not maintain labeled data that can be applied to supervised learning models [10].

Second, most of the learning features that are theoretically used in each study are not generalized features in the real world, because they are not contained in common network security systems [3]. Hence, it makes difficult to utilize to practical cases. Recent efforts on intrusion detection research have considered an automation approach with deep learning technologies, and performance has been evaluated using wellknown datasets like NSLKDD [11], CICIDS2017 [12], and Kyoto-Honeypot [13]. However, many previous studies used benchmark dataset, which, though accurate, are not generalizable to the real world because of the insufficient features. To overcome these limitations, an employed learning model requires to evaluate with datasets that are collected in the real world.

Third, using an anomaly-based method to detect network intrusion can help detect unknown cyber threats; whereas it can also cause a high false alert rate [6]. Triggering many false positive alerts is extremely costly and requires a substantially large amount of effort from personnel to investigate them.

Fourth, some hackers can deliberately cover their malicious activities by slowly changing their behavior patterns [10], [14]. Even when appropriate learning-based models are possible, attackers constantly change their behaviors, making the detection models unsuitable. Moreover, almost all security systems have been focused on analyzing short-term network security events. To defend consistently evolving attacks, we assume that over long-term periods, analyzing the security event history associated with the generation of events can be one way of detecting the malicious behavior of cyber attacks.

These challenges form the primary motivation for this work. To address these challenges, we present an AI-SIEM system which is able to discriminate between true alerts and false alerts based on deep learning techniques.

Our proposed system can help security analysts rapidly to respond cyber threats, dispersed across a large amount of security events. For this, the proposed the AI-SIEM system particularly includes an event pattern extraction method by aggregating together events with a concurrency feature and correlating between event sets in collected data. Our event profiles have the potential to provide concise input data for various deep neural networks. Moreover, it enables the analyst to handle all the data promptly and efficiently by comparison with longterm history data.

EXISTING SYSTEM

1. A learning-based method geared toward determining whether an attack occurred in a large amount of data can be useful to analysts who need to instantly analyze numerous events. According to [10], information security solutions generally fall into two categories: analyst-driven and machine learning-driven solutions. Analyst-driven solutions rely on rules determined by security

DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

experts called analysts. Meanwhile, machine learning-driven solutions used to detect rare or anomalous patterns can improve detection of new cyber threats [10]. Nevertheless, while learning-based approaches are useful in detecting cyber attacks in systems and networks, we observed that existing learning-based approaches have four main limitations.

2. First, learning-based detection methods require labeled data, which enable the training of the model and evaluation of generated learning models. Furthermore, it is not straightforward to obtain such labeled data at a scale that allow accurate training of a model. Despite the need for labeled data, many commercial SIEM solutions do not maintain labeled data that can be applied to supervised learning models [10].
3. Second, most of the learning features that are theoretically used in each study are not generalized features in the real world, because they are not contained in common network security systems [3]. Hence, it makes difficult to utilize to practical cases. Recent efforts on intrusion detection research have considered an automation approach with deep learning technologies, and performance has been evaluated using wellknown datasets like NSLKDD [11], CICIDS2017 [12], and Kyoto-Honeypot [13]. However, many previous studies used benchmark dataset, which, though accurate, are not generalizable to the real world because of the insufficient features. To overcome these limitations, an employed learning model requires to evaluate with datasets that are collected in the real world.
4. Third, using an anomaly-based method to detect network intrusion can help detect unknown cyber threats; whereas it can also cause a high false alert rate [6]. Triggering many false positive alerts is extremely costly and requires a substantially large amount of effort from personnel to investigate them.
5. Fourth, some hackers can deliberately cover their malicious activities by slowly changing their behavior patterns [10], [14]. Even when appropriate learning-based models are possible, attackers constantly change their behaviors, making the detection models unsuitable. Moreover, almost all security systems have been focused on analyzing short-term network security events. To defend consistently evolving attacks, we assume that over long-term periods, analyzing the security event history associated with the generation of events can be one way of detecting the malicious behavior of cyber attacks.
6. These challenges form the primary motivation for this work.

Disadvantages :
- Can predict outcome only from one aspect.
- The output is not reliable

PROPOSED SYSTEM

Our proposed system can help security analysts rapidly to respond cyber threats, dispersed across a large amount of security events. For this, the proposed the AI-SIEM system particularly includes an event pattern extraction method by aggregating together events with a concurrency feature and correlating between event sets in collected data. Our event profiles have the potential to provide concise input data for various deep neural networks. Moreover, it enables the analyst to handle all the data promptly and efficiently by comparison with longterm history data.

Advantages :
- Predicts outcomes from multiple aspects.
- Reliable outcome based on the different alogorithms.

RESULTS

In this paper author is describing concept to detect threats using AI-SIEM (Artificial Intelligence-Security Information and Event Management) technique which is a combination of deep learning algorithms such as FCNN, CNN (Convolution Neural Networks) and LSTM (long short term memory) and this technique works based on events profiling such as attack signatures. Author

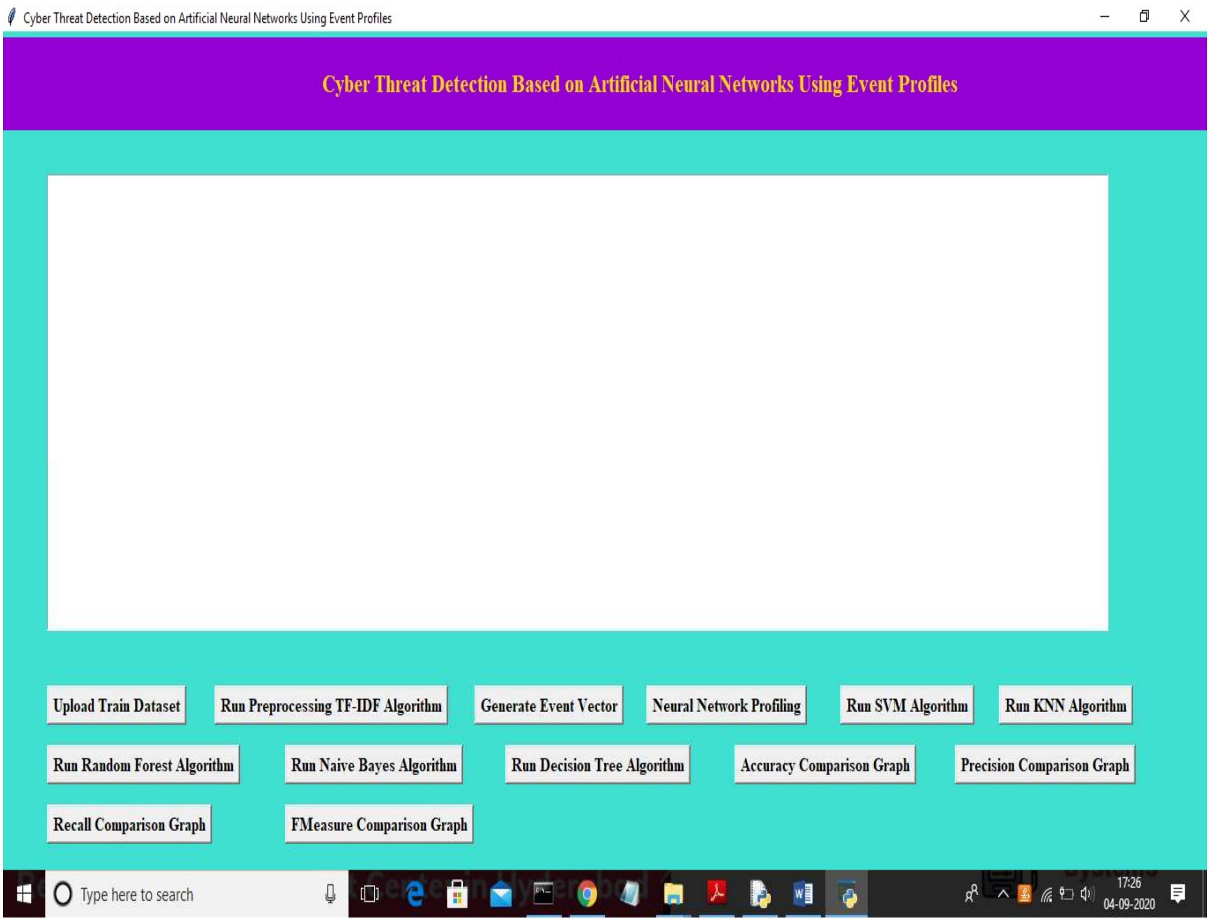
International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

evaluating propose work performance with conventional algorithms such as SVM, Decision Tree, Random Forest, KNN and Naïve Bayes. Here I am implementing CNN and LSTM algorithms.

Propose algorithms consists of following module

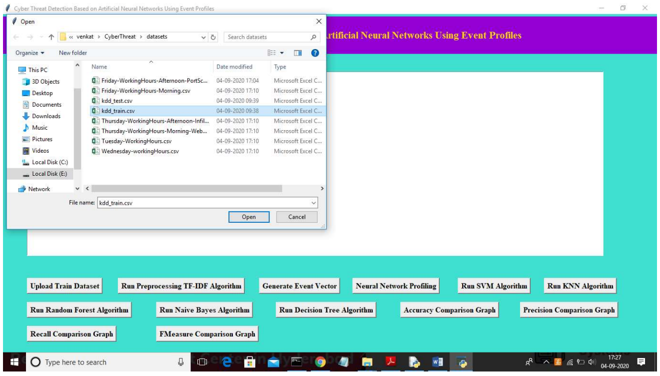
1)Data Parsing: This module take input dataset and parse that dataset to create a raw data event model
2)TF-IDF: using this module we will convert raw data into event vector which will contains normal and attack signatures
3)Event Profiling Stage: Processed data will be splitted into train and test model 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.

To run project double click on 'run.bat' file to get below screen

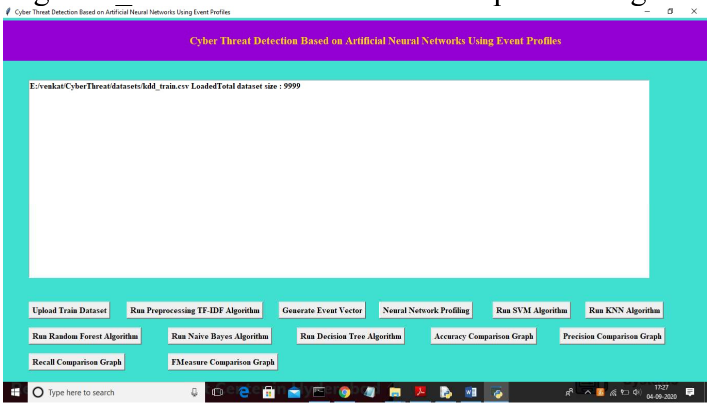


In above screen click on 'Upload Train Dataset' button and upload dataset

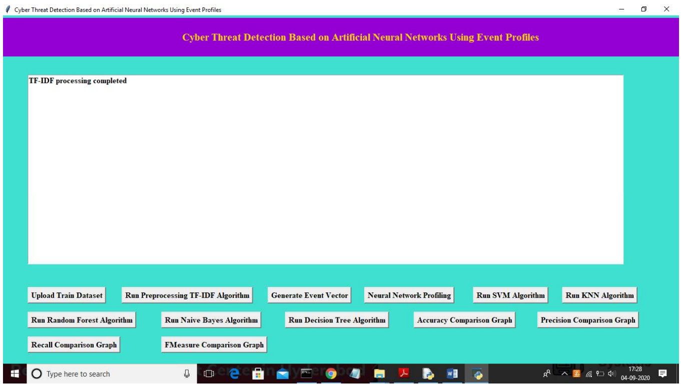
International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621



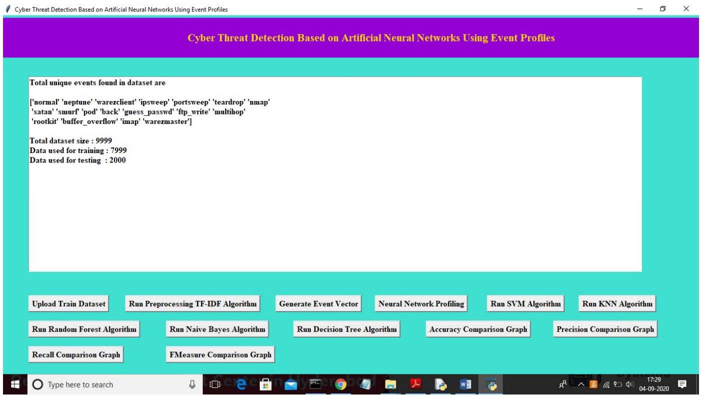
In above screen uploading ‘kdd_train.csv’ dataset and after upload will get below screen



In above screen we can see dataset contains 9999 records and now click on ‘Run Preprocessing TF-IDF Algorithm’ button to convert raw dataset into TF-IDF values



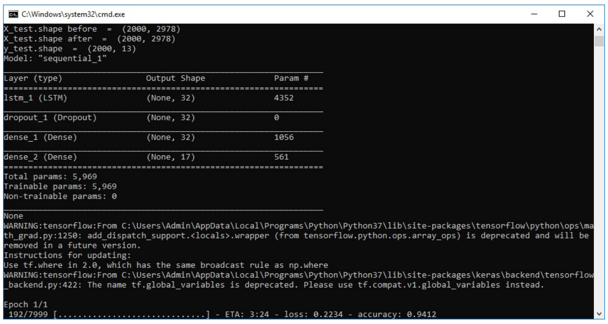
In above screen TF-IDF processing completed and now click on ‘Generate Event Vector’ button to create vector from TF-IDF with different events



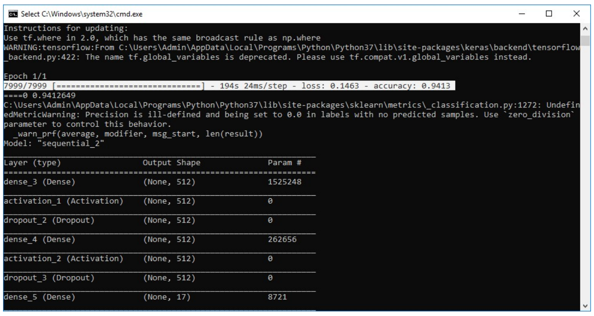
In above screen we can see total different unique events names and in below we can see dataset total size and application using 80% dataset (7999 records) for training and using 20% dataset

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

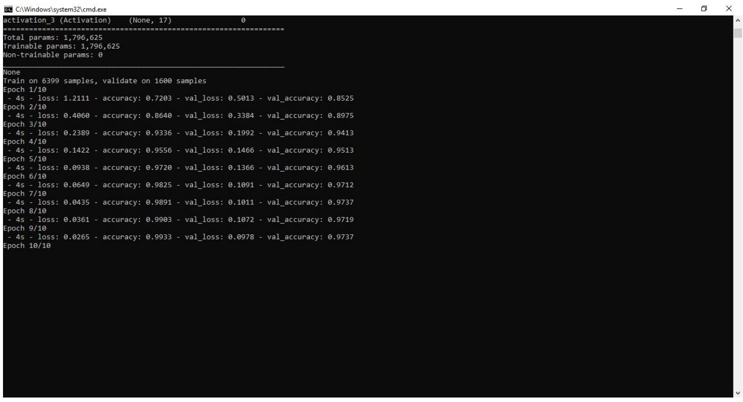
(2000 records) for testing. Now dataset train and test events model ready and now click on 'Neural Network Profiling' button to create LSTM and CNN model



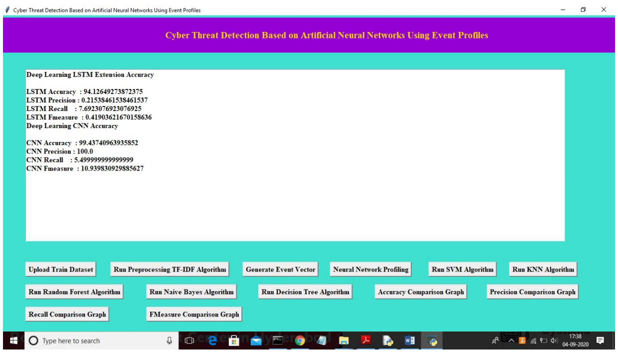
In above screen LSTM model is generated and its epoch running also started and its starting accuracy is 0.94. Running for entire dataset may take time so wait till LSTM and CNN training process completed. Here dataset contains 7999 records and LSTM will iterate all records to filter and build model.



In above selected text we can see LSTM complete all iterations and in below lines we can see CNN model also starts execution

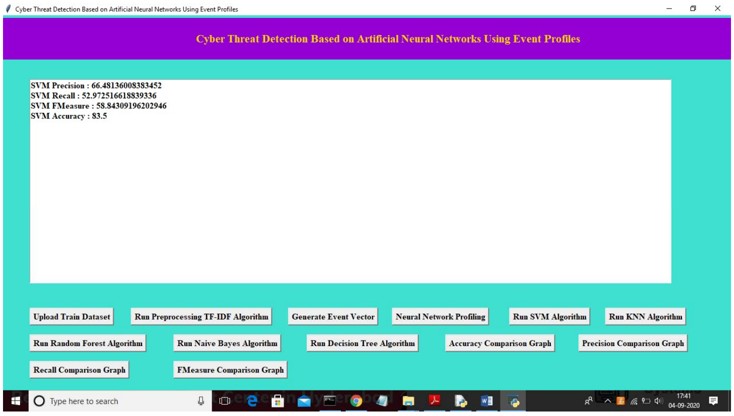


In above screen CNN also starts first iteration with accuracy as 0.72 and after completing all iterations 10 we got filtered improved accuracy as 0.99 and multiply by 100 will give us 99% accuracy. So CNN is giving better accuracy compare to LSTM and now see below GUI screen with all details

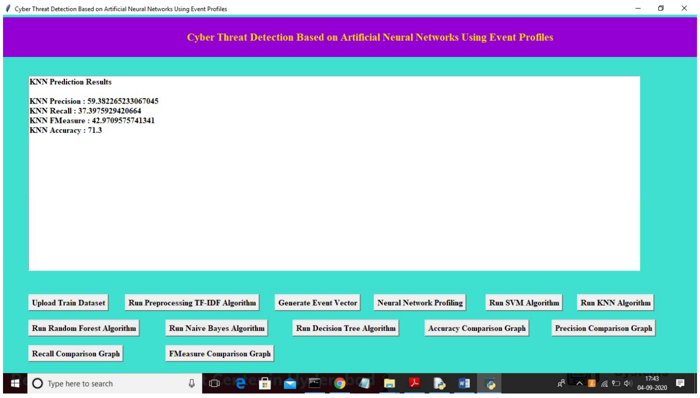


In above screen we can see both algorithms accuracy, precision, recall and FMeasure values. Now click on 'Run SVM Algorithm' button to run existing SVM algorithm

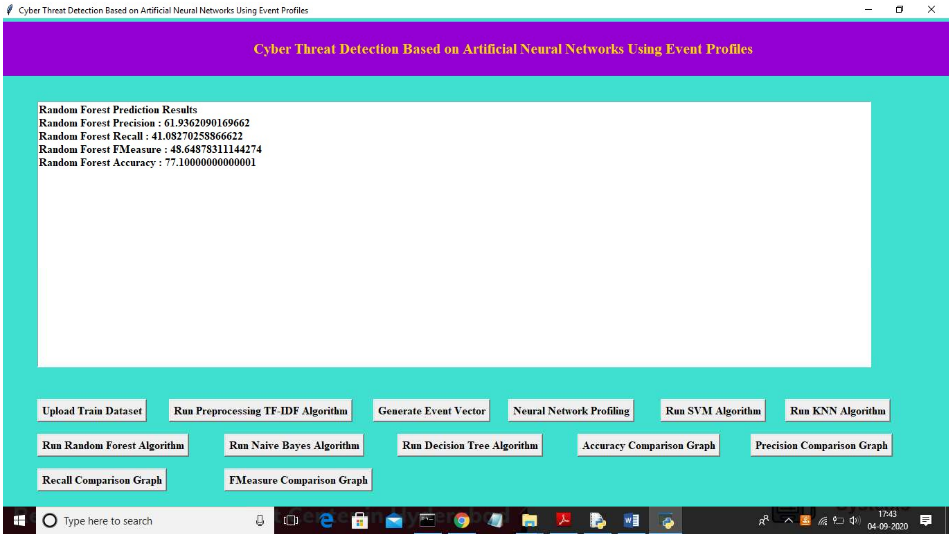
International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621



In above screen we can see SVM algorithm output values and now click on ‘Run KNN Algorithm’ to run KNN algorithm

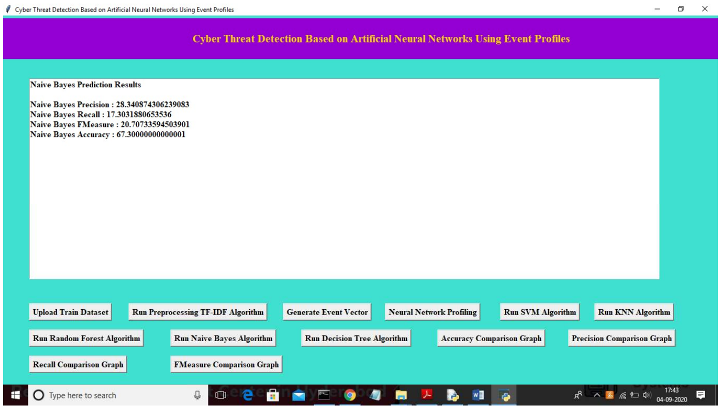


In above screen we can see KNN algorithm output values and now click on ‘Run Random Forest Algorithm’ to run Random Forest algorithm

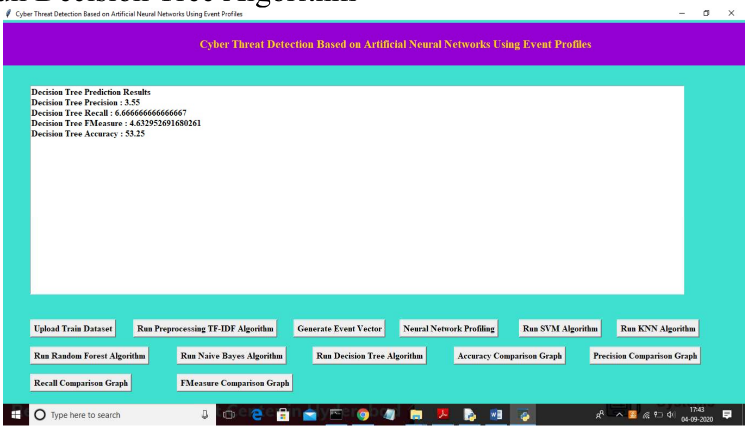


In above screen we can see Random Forest algorithm output values and now click on ‘Run Naïve Bayes Algorithm’ to run Naïve Bayes algorithm

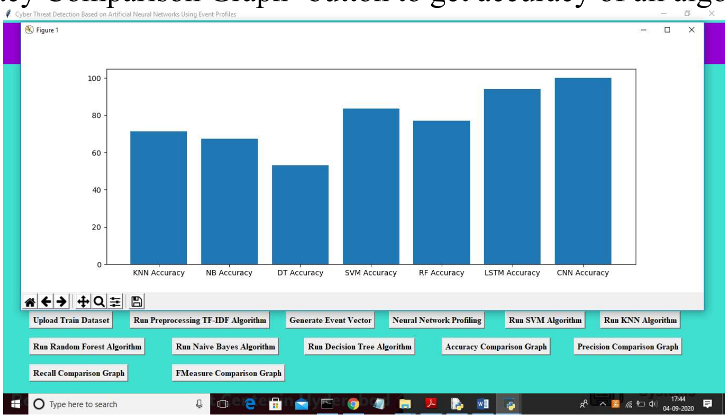
International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621



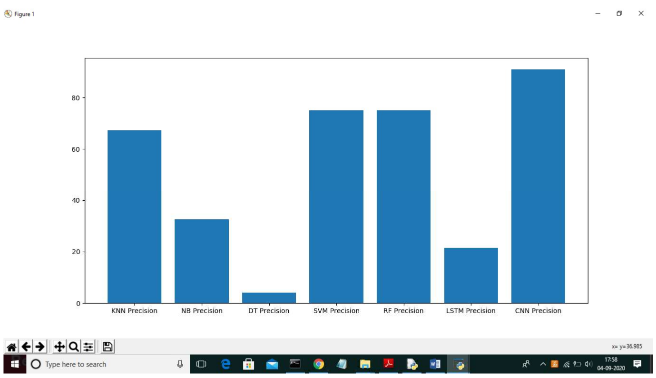
In above screen we can see Naïve Bayes algorithm output values and now click on 'Run Decision Tree Algorithm' to run Decision Tree Algorithm



Now click on 'Accuracy Comparison Graph' button to get accuracy of all algorithms



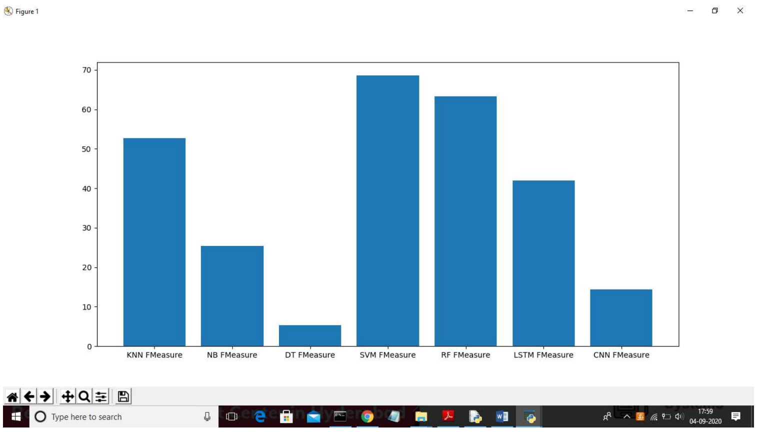
In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that LSTM and CNN perform well. Now click on Precision Comparison Graph' to get below graph



International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.7 September - October – 2023
DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

In above graph CNN is performing well and now click on 'Recall Comparison Graph'

In above graph LSTM is performing well and now click on FMeasure Comparison Graph button to get below graph



From all comparison graph we can see LSTM and CNN performing well with accuracy, recall and precision.

CONCLUSION

In this paper, we have proposed the AI-SIEM system using event profiles and artificial neural networks. The novelty of our work lies in condensing very large-scale data into event profiles and using the deep learning-based detection methods for enhanced cyber-threat detection ability. The AI-SIEM system enables the security analysts to deal with significant security alerts promptly and efficiently by comparing longterm security data. By reducing false positive alerts, it can also help the security analysts to rapidly respond to cyber threats dispersed across a large number of security events.

For the evaluation of performance, we performed a performance comparison using two benchmark datasets (NSLKDD, CICIDS2017) and two datasets collected in the real world. First, based on the comparison experiment with other methods, using widely known benchmark datasets, we showed that our mechanisms can be applied as one of the learning-based models for network intrusion detection. Second, through the evaluation using two real datasets, we presented promising results that our technology also outperformed conventional machine learning methods in terms of accurate classifications.

REFERENCES

[1]S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal,

DOI:10.46647/ijetms.2023.v07i05.001 ISSN: 2581-4621

K. Han, "Enhanced Network Anomaly Detection Based on Deep Neural Networks," *IEEE Access*, vol. 6, pp. 48231-48246, 2018.

[2]B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qiao, L. Chang, "Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base", *ETRI Journal*, vol. 39, no. 4, pp. 592-604, Aug. 2017

[3]W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," *IEEE Access*, vol. 6, no. 99, pp. 1792-1806, 2018.

[4]M. K. Hussein, N. Bin Zainal and A. N. Jaber, "Data security analysis for DDoS defense of cloud based networks," *2015 IEEE Student Conference on Research and Development (SCOReD)*, Kuala Lumpur, 2015, pp. 305-310.

[5]S. Sandeep Sekharan, K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics," *In Proc. Int. Conf. Wireless Com., Signal Proce. and Net.(WiSPNET)*, 2017, pp. 717-721.

[6]N.Hubballiand V.Suryanarayanan,‘‘False alarm minimization techniques in signature-based intrusion detection systems: A survey,’’ *Comput. Commun.*, vol. 49, pp. 1-17, Aug. 2014.

[7]A. Naser, M. A. Majid, M. F. Zolkipli and S. Anwar, "Trusting cloud computing for personal files," *2014 International Conference on Information and Communication Technology Convergence (ICTC)*, Busan, 2014, pp. 488-489.

[8]Y. Shen, E. Mariconti, P. Vervier, and Gianluca Stringhini, "Tiresias: Predicting Security Events Through Deep Learning," *In Proc. ACM CCS 18*, Toronto, Canada, 2018, pp. 592-605.

[9]Kyle Soska and Nicolas Christin, "Automatically detecting vulnerable websites before they turn malicious,", *In Proc. USENIX Security Symposium.*, San Diego, CA, USA, 2014, pp.625-640.

[10] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, K. Li, "AI2: training a big data machine to defend," *In Proc. IEEE BigDataSecurity HPSC IDS*, New York, NY, USA, 2016, pp. 49-54

[11] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu and Ali A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," *In Proc. of the Second IEEE Int. Conf. Comp. Int. for Sec. and Def. App.*, pp. 53-58, 2009.

[12] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization", *Proc. Int. Conf. Inf. Syst. Secur. Privacy*, pp. 108-116, 2018.

[13] [online] Available: http://www.takakura.com/Kyoto_data/