Science and Technology Publishing (SCI & TECH)
ISSN: 2632-1017
Vol. 9 Issue 3, March - 2025

Decision Tree-Based Network Threat Classification For Deep Packet Inspection Framework

[1] Okafor, Joyce Odu
Department of Electrical and Electronic Engineering,
Federal University Otuoke, Bayelsa State
ORCID: 0009-0001-3983-6994
Odujp@fuotuoke.edu.ng

[2]Ezema D.C.
Department of Electrical and Electronic Engineering
[3]State University of Medical and Applied Sciences(SUMAS),Igbo-Eno,Enugu State, Nigeria
donatus.ezema@sumas.edu.ng.

[3]Ifeagwu E.N.
Department of Electrical and Electronic Engineering.
Federal University Otuoke, Bayelsa State.
ORCID:0009-0005-7448-0187
ifeagwuen@fuotuoke.edu.ng

Abstract*— This paper presents decision tree-based network threat classification for deep packet inspection framework. Specifically, Decision Tree (DT) algorithm trained for network threat classification using Software Defined Network (SDN) intrusion. The DT algorithm is designed to be integrated into Deep Packet Inspection (DPI) for application in intrusion detection on Software SDN. The case study SDN intrusion dataset consists of 79 SDN threat attributes with the following six different attack classes; brute force, SQL injections, Distributive Denial of Service (DDoS), normal packet, Cross-Site Scripting (XSS) intrusions, and benign traffic. The dataset has 1,188,333 rows of network intrusions and white-listed traffic related data items. The DT model performance was assessed using the following metrics: Receiver Operating Characteristic (ROC), Accuracy, False Discovery Rate (FDR), False Negative Rate (FNR), Positive Predictive Value (PPV), and True Positive Rate (TPR). The results show that across the different treat classes, benign attack recorded PPV of 99.9% and FDR of 0.1%, DDoS recorded PPV of 97.9% and of 2.1% . Web-based brute force attack was classified with PPV of 55.4% and FDR of 44.6%, while the SQL injection attack recorded PPV of 100% and FDR of 0%. For web-based XSS attack, it reported PPV of 57.1% and FDR of 42.9% while normal packets was excellent classified with PPV of 100% and FDR of 0%. In all, it was observed that while the model was able to perform excellently in four out of the six attack classes, it recorded fair results for web based brute force and XSS attack respectively.

Keywords — Threat Classification, Decision Tree, Deep Packet Inspection, Network Threat, Software Defined Network

1. INTRODUCTION

In the last few decades, Machine Learning (ML) algorithms have evolved into a very versatile tool for many application [1,2,3]. They are now used in so many smart applications and complex systems and autonomous systems implementation [4,5]. They are also extensively employed in detecting and preventing cyber threats such as malware, phishing attacks, and intrusion attempts [6,7].

Basically, different ML approaches are used to address different aspect of the issues associated with the protection of the cyber infrastructure [8,9]. Ml are used in the detection of cyber threats from the incoming packets. In addition, the ML models are also used to determine the type of attack and even to coordinate the response of the system towards such attack. In this paper, decision tree-based network threat classification for deep packet inspection framework is presented [10,11]. Essentially, Decision Tree (DT) machine learning approach is used to detect and classify cyber-attack; in this case, the DT algorithm is integrated in the deep packet inspection framework so that as the deep packet inspection of the network traffic is implemented, the DT will be used to detect and classify the threat in the incoming packets. This is very useful for network access control and cyber-attack mitigation. In this work, the DT algorithm is trained and validated using cyber threat dataset with different attack records. The model

performance is expressed in terms of precision, accuracy, recall and F1-score.

2. METHODOLOGY

In this work, Decision tree (DT) model is employed for network threat detection. Notably, the DT model is a popular machine learning algorithm commonly applied for deep packet inspection (DPI) modeling (Rahul et al., 2021). In order to accurately predict the target variable for new data points, this predictive modeling tool iteratively separates a Software Defined Network (SDN) attack dataset into homogeneous subsets by posing a series of feature-related queries. It begins with a root node and divides the data iteratively according to the feature that yields the best separation, generating branches until a stopping condition is satisfied. A prediction is represented by each leaf node according to the dominant class that exists there. The sequence for training DT to generate the packet inspection model is presented in Figure 1.

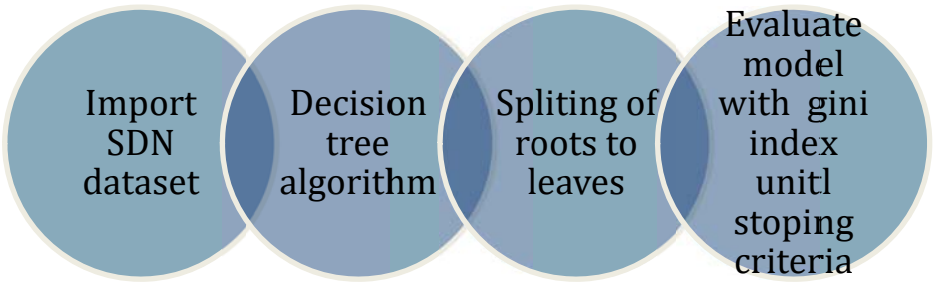

Figure 1: Sequence of the DT training process for the packet inspection generation

In order to effectively forecast the target variable, the DT is trained by recursively splitting the dataset based on the values of the features. Starting with the entire dataset at the root node, the algorithm selects the best feature and split point to maximize information gain or decrease impurity at each node. This procedure continues recursively, branching out, until a stopping condition is met, such as reaching a maximum tree depth or having nodes with a minimum amount of samples. By traversing the tree from the root to a leaf node and assigning the prediction to the majority class, the tree can be trained to predict more data points. Finally, the model is improved and evaluated by validation techniques through the adjustment of hyper-parameters, before the deep packet inspection model is generated. The procedure for the DT-based DPI is presented Algorithm 1, as well as Algorithm 2.

Algorithm 1: The procedure for the DT-based DPI packet generation

Ste 1: Start
Ste 2: Load training dataset of SDN attack
Ste 3: Split data into training, test and validation sets respectively
Ste 4: Load decision tree algorithm
Ste 5: Generate tree roots with the dataset
Ste 6: Initialize recursive splitting of trees into data sub-sets
Ste 7: Evaluate model with gini index
Ste 8: Generate model for packet inspection

Algorithm 2: The procedure for the DT-based DPI model packet inspection operation

Ste 1: Start
Ste 2: Load incoming packet from network
Ste 3: Initialize DT packet inspection model
Ste 4: Extract packet features
Ste 5: Classify packet features with DT model
Ste 6: Identify malicious packets
Ste 7: Classify as intrusion
Ste 8: Return

The Software Defined Network (SDN) intrusion dataset was used in the study to train the model as well as to

validate the model. The SDN intrusion dataset consists of 79 SDN threat attributes with the following six different attack classes; brute force, SQL injections, Distributive Denial of Service (DDoS), normal packet, Cross-Site Scripting (XSS) intrusions, and benign traffic. The dataset has 1,188,333 rows of network intrusions and white-listed traffic related data items. The breakdown of each the SDN intrusion dataset is presented in Table 1. The DT model performance was assessed using the following metrics: Receiver Operating Characteristic (ROC), Accuracy, False Discovery Rate (FDR), False Negative Rate (FNR), Positive Predictive Value (PPV), and True Positive Rate (TPR).

Table 1: The breakdown of each the SDN intrusion dataset

Attack class	Attack feature size
Benign Traffic	798,322
Web Attack XSS Traffic	1,962
Web Attack SQL Injection Traffic	60
Web Attack Brute Force	4,550
DDoS Traffic	338,139
Normal packet	45,345

3. RESULTS AND DISCUSSIONS

The model was trained and validated on Python platform. The TPR and FNR performance evaluation results for the DT-based DPI model is presented in Figure 2. According to the results in Figure 2 , for benign attack, the model reported a 98.6% and 1.4% for the FNR, for the DDoS attack the model reported 99.9% for TPR and 0.1% for FNR, for the web-based brute force attack the model reported TRP of 99.9% and FNR of 0.1%,, for the web-based SQL injection attack the model recorded 100% for TPR classification and 0% for FNR, while for the web-based XSS attack, it reported 0.1% and normal packet TPR it recorded 100% with 0% for FNR.

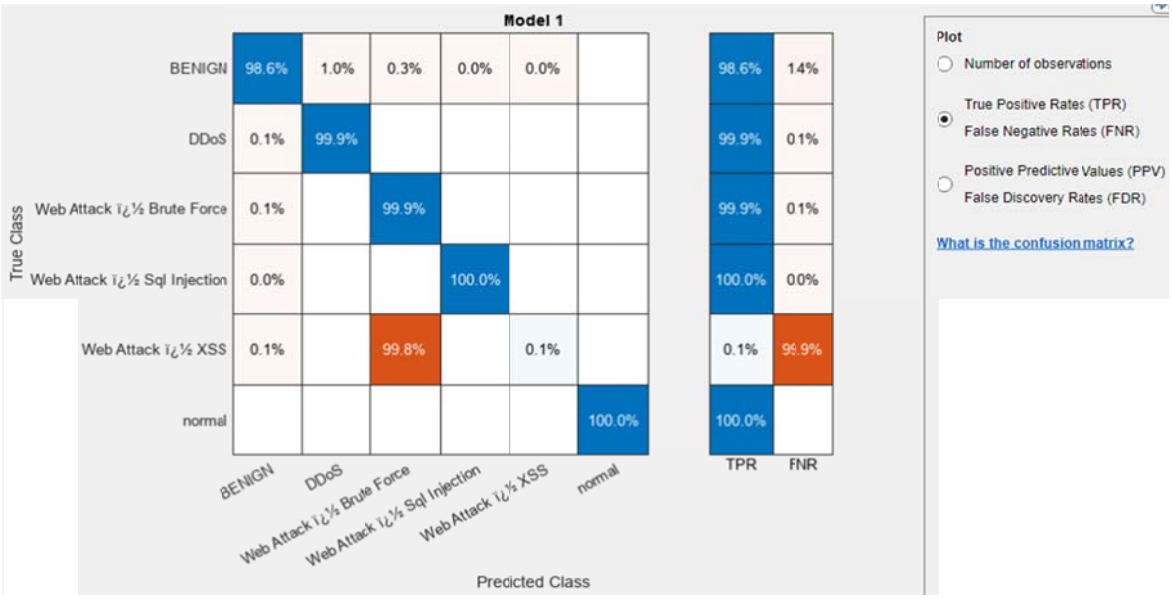

Figure 2: DT-based DPI result considering the TPR and FNR

The PPV and FDR performance evaluation results for the DT-based DPI model is presented in Figure 3. According to the results in Figure 3 it was observed that across the different treat classes, benign attack recorded 99.9% PPV and 0.1% FDR, DDoS recorded 97.9% PPV while FDR recorded 2.1% FDR. Web-based brute force attack was classified with 55.4% PPV and 44.6% FDR, while the SQL injection attack recorded 100% PPV and 0% FDR. For web-based XSS attack, it reported 57.1% PPV and 42.9% FDR while normal packets was excellent classified with 100% PPV and 0% FDR.

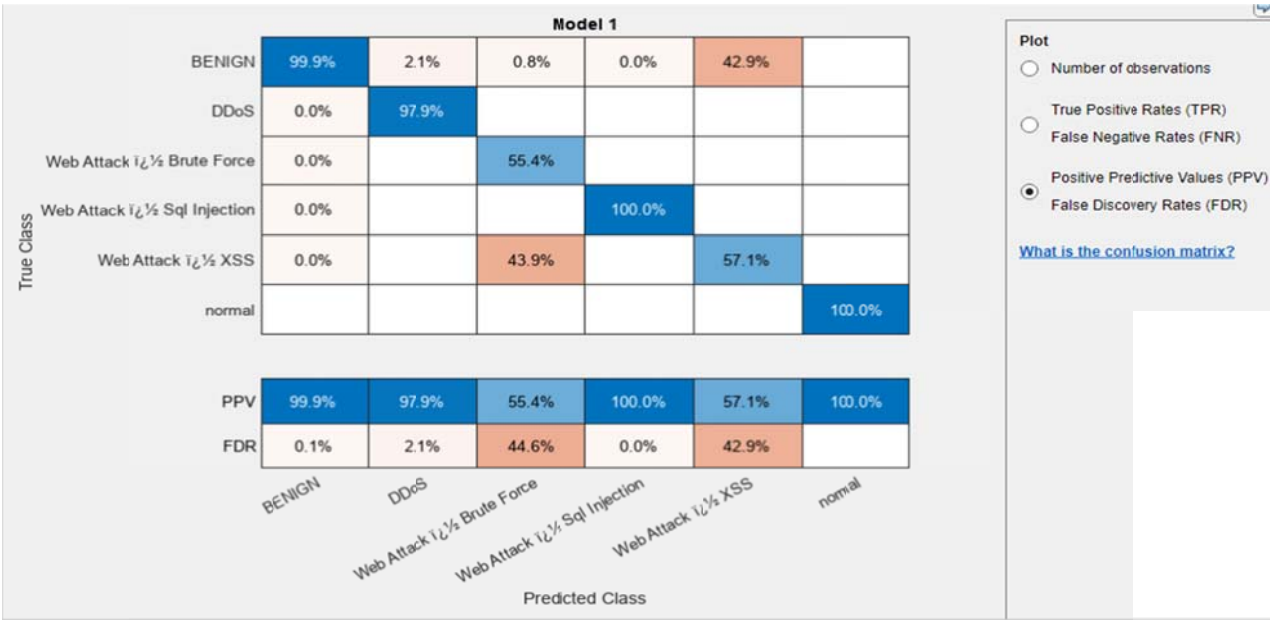

Figure 3: PPV and FDR for DT based DPI model

The Receiver operating characteristic (ROC) results for the DT-based DPI model are presented in Figure 4. The ROC reported the Area under the curve (AUC) of the mode against each of the threat classes and normal packet. The AUC for benign reported 0.9984, DDoS recorded 0.9904, we based brute force recorded 0.9567, web-based SQL injection attack reported 0.9999, web-based XSS attack recorded 0.9471 and normal packet AUC recorded 1. The model accuracy reported 91.8%.

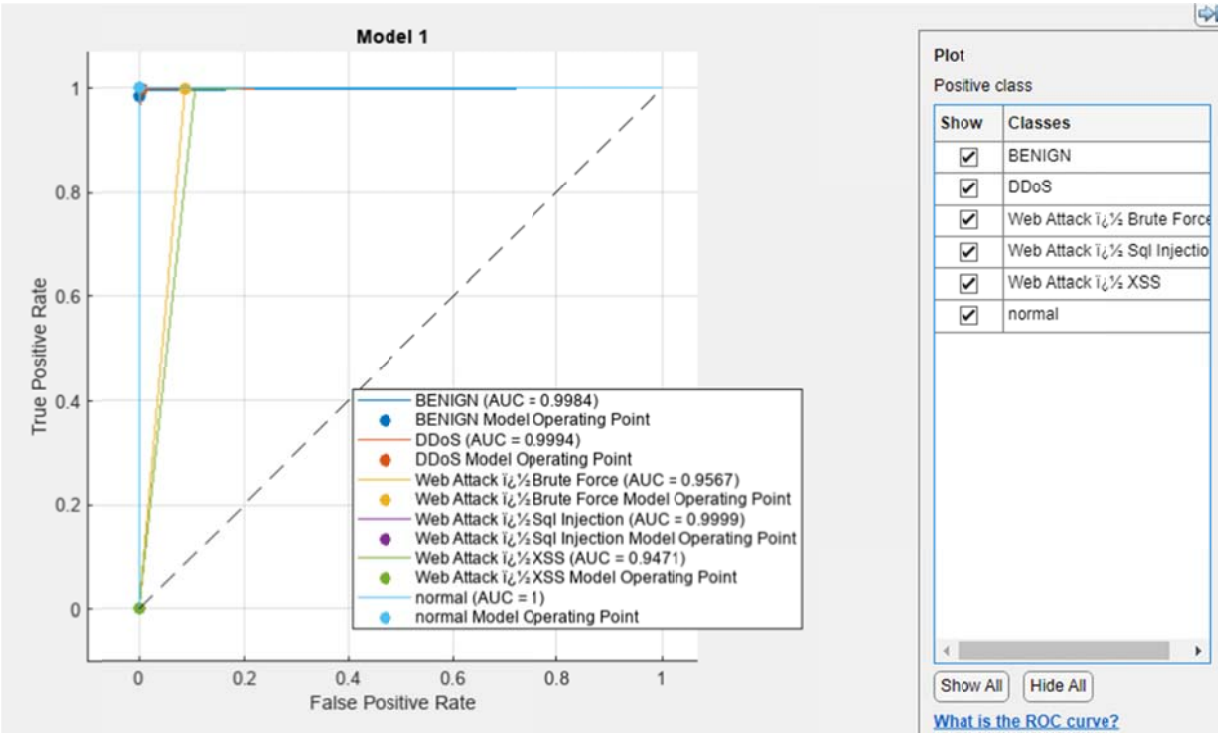

Figure 4: ROC analysis of the DT based DPI model

The summarized results of the DT-based DPI model performance for all the performance parameters and attack categories is presented in Table 2. Overall the results have demonstrated the effectiveness of the DT model for the classification of incoming threats to the network infrastructure. In the results, it was observed that while the model was able to perform excellently, it recorded fair results for web based brute force and XSS attack respectively.

Table 4.2: The summarized results of the DT-based DPI model performance for all the performance parameters and attack categories

Metrics	Benign	DDoS	Brute force	SQL injection	XSS	Normal
PPV	99.9	97.9	55.4	100	57.1	100
FDR	0.1	2.1	44.6	0	42.9	0
TPR	98.6	99.9	99.9	100	0.1	100
FNR	1.4	0.1	0.1	0	99.9	0
ACC	91.8	91.8	91.8	91.8	91.8	91.8
ROC	0.9984	0.9904	0.9567	0.9999	0.9472	1

4. CONCLUSION

The study presented network traffic detection scheme based on Decision Tree (DT) algorithm. The DT algorithm is integrated into Deep Packet Inspection (DPI) for application in intrusion detection on Software Defined Network (SDN). The DT algorithm was trained using SDN intrusion dataset consists with six different attack classes.

The overall results demonstrated the effectiveness of the DT algorithm for the classification of incoming threats to the network infrastructure. In the results, it was observed that while the model was able to perform excellently in four out of the six attack classes, it recorded fair results for web based brute force and XSS attack respectively.

REFERENCES

1. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 160.
2. Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, *3*, 58-73.]
3. Pramod, A., Naicker, H. S., & Tyagi, A. K. (2021). Machine learning and deep learning: Open issues and future research directions for the next 10 years. *Computational analysis and deep learning for medical care: Principles, methods, and applications*, 463-490.
4. Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, *3*, 54-70.
5. Adeyeye, O. J., & Akanbi, I. (2024). Artificial intelligence for systems engineering complexity: a review on the use of AI and machine learning algorithms. *Computer Science & IT Research Journal*, *5*(4), 787-808.
6. Marengo, A., & Pagano, A. (2024). Machine learning for cybersecurity for detecting and preventing cyber attacks. *Machine Intelligence Research*, *18*(1), 672-689.
7. Geetha, R., & Thilagam, T. (2021). A review on the effectiveness of machine learning and deep learning algorithms for cyber security. *Archives of Computational Methods in Engineering*, *28*(4), 2861-2879.
8. Pinto, A., Herrera, L. C., Donoso, Y., & Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, *23*(5), 2415.
9. Berghout, T., Benbouzid, M., & Muyeen, S. M. (2022). Machine learning for cybersecurity in smart grids: A comprehensive review-based study on methods, solutions, and prospects. *International Journal of Critical Infrastructure Protection*, *38*, 100547.
10. Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. *IEEE Access*, *11*, 80348-80391.
11. Polat, O., Türkoğlu, M., Polat, H., Oyucu, S., Üzen, H., Yardımcı, F., & Aksöz, A. (2024). Multi-stage learning framework using convolutional neural network and decision tree-based classification for detection of DDoS pandemic attacks in SDN-based SCADA systems. *Sensors*, *24*(3), 1040.