https://doi.org/10.33271/nvngu/2025-1/104

**I. Syamsuddin**[*1],
orcid.org/0000-0002-6017-7364,
**D. Al-Dabass**[2],
orcid.org/0009-0001-7312-4712

1 – Politeknik Negeri Ujung Pandang, Makassar, Indonesia
2 – Nottingham Trent University, Nottingham, the United Kingdom
[*] Corresponding author e-mail: irfans@poliupg.ac.id

# FOSDET: A NEW HYBRID MACHINE LEARNING MODEL FOR ACCURATE AND FAST DETECTION OF IOT BOTNET

**Purpose.** This study is aimed at introducing a new hybrid machine learning model to enhance the accuracy and speed in detecting botnet attacks in Internet of Things networks. The new model is derived from an integration of decision tree algorithm and feature selection algorithms to produce a novel hybrid machine learning for better performance in IoT botnet detection.

**Methodology.** The study adopts a six steps research methodology. It consists of dataset collection, dataset preprocessing, applying machine learning, comparing feature selection algorithms, combining both machine learning and feature selection algorithms, and finally comparing the results.

**Findings.** A novel hybrid machine learning (ML) model called FoSDeT has been obtained as a result of combination of decision tree algorithm and feature selection algorithm called Forward Selection which shows a significant improvement in IoT botnet detection in comparison to standard decision tree model.

**Originality.** The paper proposes a simple yet powerful hybrid approach which integrates Decision Tree algorithm with two predefined feature selection algorithms namely, Forward Selection and Backward Elimination. The new hybrid model called FoSDeT shows a significant enhancement in terms of IoT botnet detection.

**Practical value.** The hybrid model obtained from this study might be used by IT security practitioners in developing real intrusion detection system for defending IoT networks from botnet attacks.

**Keywords:** *IoT, botnet, cyber attack, machine learning, detection accuracy, detection speed*

**Introduction.** Internet of Things (IoT) is a system of interconnected devices that serves to facilitate the exchange of information between physical devices. These devices can be driverless vehicles, SmartTVs, Smart City infrastructure and various other devices that can be monitored and controlled remotely [1]. A great development of IoT implementation always accompanied by security risks which also increase. Weaknesses of IoT networks makes them vulnerable to various cyber attacks as mentioned in many studies [2]. Security risks that can occur include Man in the middle attacks, Eavesdropping, drive-by, and malware.

Malicious software (maware) is one of the problems that continues to grow and become a serious threat. Malware itself is software that was created to infiltrate or damage a computer system or computer network without the permission of the device owner. There are several types of malware commonly used by criminals, one of which is botnets [3, 4].

Botnet is a rapidly growing problem which has been raising lots of concerns by research nowadays. In short, botnet is a collection of computers running malware, controlled by hackers (usually called botmasters). Botnets turn computers into an army of cyber attacks, usually for spam, fake websites, DoS (Denial of Service) attacks, viruses, as well as gathering information through phishing and scams [4].

Many studies discuss classification using machine learning methods. Machine learning allows machines to know and learn the types of data so as to produce information. The application of machine learning has been carried out by analyzing several machine learning techniques to detect P2P (peer-to-peer) botnets [5]. Experiment with different machine learning algorithms to compare their ability to classify botnet traffic from normal traffic by selecting distinguishing features from network traffic. Among many machine learning algorithms for classification is Decision Tree (DT). In the case of detecting intrusion in computer networks, the algorithm of J48 and Naive Bayes were compared. Both algorithms provide good results which can detect zero-day threats with high precision [6].

Other research that uses supervised learning algorithms on data network traffic for accurate identification of IoT devices is connected to the network. The accuracy obtained for the IoT classification is 99.28 % [7]. The authors argue the application of machine learning as a solution for better classifying botnet attacks.

Feature selection is one of the main factors that affect the performance of the results of machine learning algorithms [8]. If the data contains a number of features, the processed data will be time consuming, which is ineffective. An efficient feature selection method helps in reducing parts of data that are not significantly needed in the classification process so that the results obtained will be more accurate and faster [8, 9].

Therefore, the research aims to assess the application of feature selection algorithms in order to establish a new hybrid machine learning model with improved accuracy and speed in detecting anomaly within IoT caused by botnet attacks. The study employs Decision Tree algorithm with two different feature selection methods using the BoT-IoT dataset to see the effect on the performance of both combinations.

The organization of the paper is as follows. In the second section, literature review is presented. Section 3 provides the research methodology to carry out the study. It is then followed by section 4 that presents the results and analysis. Finally, section 5 concludes the research.

**Literature review.** Among the earliest study in terms of machine learning potential to improve the detection of botnet was a study by Beigi, et al. [10]. The paper underlines the importance role of effective feature selection in machine learning-based botnet detection. They developed and used a diverse data set (16 botnets) to fully test the effectiveness of the feature for accurate detection.

Later, Singh, et al. [11] report research on machine learning as a big data analytics framework for peer-to-peer botnet detection. They introduce a scalable implementation of a quasi-real-time intrusion detection system with machine learning to improve the detection rate of peer-to-peer botnet attacks.

Survey research by Alejandre, et al. [12] deals with the challenge of using feature selection to detect botnets using machine learning. It implemented the Genetic Algorithm (GA) to select features and the C4.5 algorithm to perform the calcification process between connections that have and do not have botnets.

Miller and Busby-Earle [13] provide a brief overview of the different machine learning (ML) methods and the role they play in botnet detection. A clear understanding of this role is essential for developing an effective real-time online detection

approach and efficient and more powerful models. Similarly, Pektaş and Acarman [14] also suggest the importance of dataset engineering by effective feature selection in the approach analyzing the most distinguishing features for the purpose of building an efficient and effective botnet detection system.

Gadelrab, et al. [15] in his research entitled "*BotCap: Machine learning approach for botnet detection based on statistical features*" describe a detailed approach to developing a botnet detection system using machine learning techniques. They have identified a set of statistical features that can help differentiate between harmless traffic and malicious botnets. Then, they have conducted several machine learning experiments to test the suitability of machine learning techniques and also to select a minimal subset of identified features that provide the best detection.

Likewise, Hoang and Nguyen [16] conduct a study about botnet detection based on machine learning techniques using Domain Name Service (DNS) query data. They found that machine learning using DNS are effective in making botnet detection more accurate up to 90 %.

In addition, Mathur, et al. [17] conduct research on mining network flow using machine learning. According to their analysis, botnet detection via mining of network traffic flow is able to train classification by specific network flow datasets which in turn it is able to distinguish between normal traffic and bot traffic with high accuracy and low false positive rate.

Another survey related paper written by Khraisat, et al. [18] presents the results of a contemporary IDS classification survey, a comprehensive overview of the most recent popular work, and an overview of datasets commonly used for evaluation purposes. The paper also presents evasion techniques used by attackers to evade detection and discusses future research challenges to counter these techniques thereby making computer systems more secure.

Nomm and Bahsi [19] made another approach by tackling unsupervised anomaly based botnet detection in IoT networks. They suggest that it is possible to induce high-accuracy unsupervised learning with a reduced feature set size, allowing for reduced computational resources required.

Likewise, Shafiq, et al. [20] introduce a unique feature selection approach known as 'CorrAUC' and applied it to the Bot-IoT dataset. The novel approach chose five characteristics that accurately characterized the dataset and could be utilized for training. In this study, they employ four machine learning algorithms (Decision Tree, SVM, Naive Bayes, and Random Forest) and make a systematic comparison of their performance on a test set.

Furthermore, Baig, et al. [21] published a paper entitled "*Averaged dependence estimators for DoS attack detection in IoT networks*", which presents a DoS detection framework consisting of modules for data generation, feature ranking, training and testing. The algorithms used are C4.5, MLP, Naïve Bayes, Bayesian Network, A1DE, and A2DE. The application used is Weka. A2DE shows the highest accuracy of 99 %.

Bovenzi, et al. [22] use a hierarchical Network Intrusion Detection technique to detect attacks across several situations. H2ID performs (i) anomaly detection utilizing a unique lightweight approach based on a MultiModal Deep AutoEncoder (M2-DAE), and (ii) attack classification with soft-output classifiers. We validate our approach with the recently released Bot-IoT dataset, inferring between four key attack categories (DDoS, DoS, Scan, and Theft) and unknown assaults.

Machine learning techniques are being used to detect malware in [23]. They compare Support Vector Machine, Decision Tree and Deep Belief Networks on malware dataset. It is finally concluded that SVM has better results than decision tree and Deep Belief Networks in terms of accuracy and recommend SVM for future applications.

Soe, et al. [24] in their research paper entitled "*Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features*" implements machine learning-based IDS using a new feature selection algorithm on the Raspbery Pi system. The Weka software is used to check and compare the performance of all machine learning.

Ullah and Mahmoud [25] in their research paper entitled "*A two-level flow-based anomaly activity detection system for IoT networks*" justify the use of Random Forest algorithm with two levels of anomaly activity detection nodes for intrusion detection systems in IoT networks.

A new approach to use of balanced network traffic to effectively identify IoT botnet is offered by Shobana and Poonkuzhali [26]. The study highlights the issue of class imbalance within the dataset which was handled using the random oversampling approach. Further machine learning analysis was performed using Support Vector Machine, Naive Bayes, Decision Tree, and Deep Neural Networks.

In 2022, Syamsuddin and Barukab proposed Sukry, a novel machine learning approach applied in Raspberry Pi hardware using Enhanced kNN algorithm. The results show a significant improvement of IoT detection in comparison to traditional kNN [27]. Their approach shows improvement results, since kNN algorithm is commonly considered inefficient to deal with large dataset.

To deal with specific attack flow in industrial control network bases on IoT, a new solution is proposed by Qian [28]. They introduce a hierarchical interval-based belief rule base (HIBRB). At the end, it is shown that HIBRB model can improve the detection rate of attack flow while maintaining high accuracy.

Alhaddad, et al. [29] developed a Convolutional Neural Network (CNN) based IoT cyber attack monitoring system. It is also equipped with an intuitive Kafka based real-time monitoring in order to streamline network attack surveillance and resilience. Their CNN based cyber detection has achieved a high accuracy rate of 99.86 %.

Then, Karmous et al. address the problem IoT cyber attacks by proposing an enhanced IDS using Software Defined Network (SDN). This new approach shows a high accuracy machine learning model for real-time prediction [30].

The evaluated research shows several developments in IoT network security and intrusion detection systems (IDS). To improve accuracy while reducing time consumption in IDS for IoT contexts, gaps are evident when it comes to combining decision tree methods with strong feature selection approaches. Despite the widespread use of decision tree algorithms (e.g., Shafiq, et al. [20], Baig, et al. [21], Shobana and Poonkuzhali [26]), the application of decision tree algorithms with advanced and domain-specific feature selection methods remains underexplored.

**Research methodology.** The research methodology to guide this research is presented in Fig. 1. It consists of six steps research methodology, namely dataset collection, dataset preprocessing, applying machine learning, comparing feature selection algorithms, combining both machine learning and feature selection algorithms, and finally comparing the results to produce a new hybrid machine learning model.

The first step is dataset collection. The raw data used is a dataset from the BoT-IoT dataset which includes normal network traffic and several attacks traffics [28].

Its main feature of representing a realistic IoT environment is the main reason to use BoT IoT dataset for the study. It mimics several attacks such as DDoS, DoS, Reconnaissance and Theft attacks [28]. Recent studies also employ the dataset which reflects the usability of the dataset in current research [29–31].

There are 46 features in this dataset with a total data of 19,056. Fig. 2 shows the distribution of total data of 19,056 into five labels, DoS, DDoS, Reconnaissance, Normal dan Theft.

The second step is pre-processing the dataset. The aim of this step is to assess dataset completeness, whether missing data exist or not, the number of features, the number of rows, and many others. This step is important to prepare dataset ready to further steps.

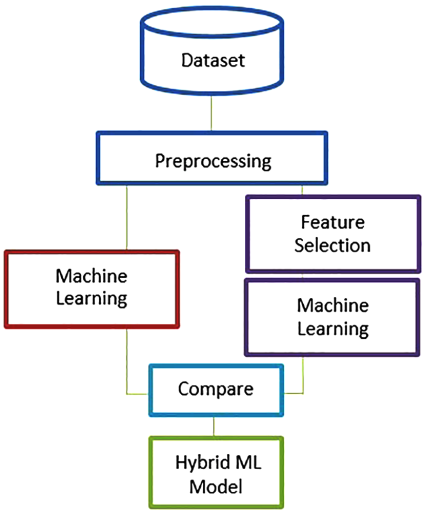


*Fig. 1. Research Methodology*

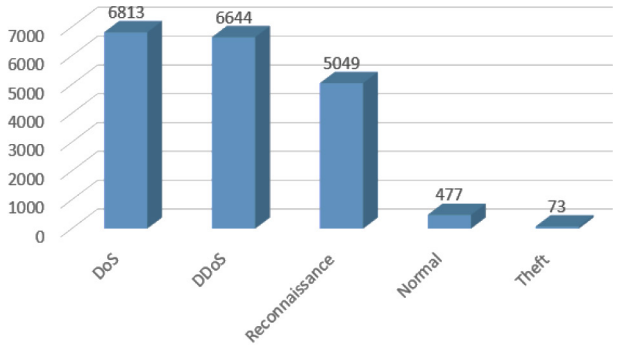


*Fig. 2. Distribution of labels in the dataset*

Then, around 70 % of dataset is selected for testing step with two different approaches in the next step. Firstly, it will be applied directly to Decision Tree algorithm. This first approach is conducted to view overfitting issue that commonly occurs when many unimportant features of dataset are used. Secondly, the selected 70 % of the dataset will undergo Feature Selection mechanisms and then to Decision Tree algorithm. In this second approach there are two Feature Selection mechanisms which will be applied (Forward Selection and Backward Elimination) in order to mitigate overfitting in machine learning models.

Finally, the results from Decision Tree, Forward Selection with Decision Tree and Backward Elimination with Decision Tree are compared in terms of several aspects such as accuracy and processing time to decide the best performing model.

**Findings.** The machine learning analysis is performed over Rapidminer software (Fig. 3). There are three models applied in the study, namely Decision Tree (Model1), Forward Selection Decision Tree (Model2) and Backward Elimination Decision Tree (Model3).

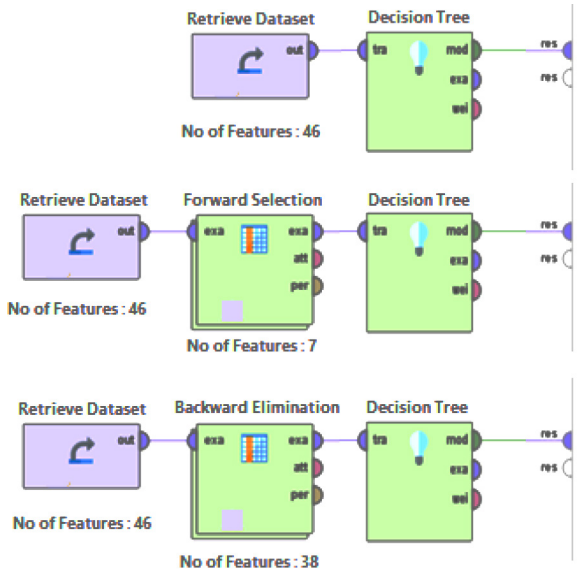


*Fig. 3. Rapidminer implementation of the models*

As can be seen in Fig. 3, the beginning step is retrieving dataset into each of the three models. The retrieved dataset is actually the dataset that has passed preprocessing stage without any missing data and other errors. In the first model, the dataset (with 46 features) directly supplied to Decision Tree, while in the second and third model the dataset should undergo feature selection process to reduce the number of features. Model 2 uses Forward Selection algorithm while model 3 employs Backward Elimination algorithm before being supplied to Decision Tree algorithm. It is clearly seen that Forward Selection algorithm reduces the number of features from 46 to 7, while Backward Elimination reduces the number of features from 46 to 38.

Then, in the next step approximately 70 % of dataset is selected for training with the three machine learning models established before. For the first model, the training dataset is directly proceeded to Decision Tree algorithm, while for the second and third models the selected 70 % of the dataset will undergo Feature Selection mechanisms before applied to Decision Tree algorithm.

In this second approach there are two Feature Selection mechanisms which will be applied (Forward Selection and Backward Elimination). The results of the three models are then analyzed and the results are presented in the form of confusion matrix (Figs. 4, 5 and 6).

Fig. 4 shows the confusion matrix for model 1 (Decision Tree) obtained from Rapidminer software. The accuracy of model 1 is calculated as follows.

**accuracy: 97.56% +/- 0.76% (micro average: 97.56%)**

| | true DDoS | true Reconnaissance | true DoS | true Normal | true Theft | class precision |
|---|---|---|---|---|---|---|
| pred. DDoS | 6644 | 0 | 0 | 0 | 36 | 99.46% |
| pred. Reconnaissance | 0 | 5049 | 0 | 429 | 0 | 92.17% |
| pred. DoS | 0 | 0 | 6813 | 0 | 0 | 100.00% |
| pred. Normal | 0 | 0 | 0 | 48 | 0 | 100.00% |
| pred. Theft | 0 | 0 | 0 | 0 | 37 | 100.00% |
| class recall | 100.00% | 100.00% | 100.00% | 10.06% | 50.68% | |

*Fig. 4. Confusion matrix for ML model 1 (Decision Tree)*

accuracy: 99.82% +/- 0.11% (micro average: 99.82%)

| | true DDoS | true Reconnaissance | true DoS | true Normal | true Theft | class precision |
|---|---|---|---|---|---|---|
| pred. DDoS | 6636 | 0 | 3 | 0 | 0 | 99.95% |
| pred. Reconnaissance | 0 | 5042 | 1 | 2 | 1 | 99.92% |
| pred. DoS | 8 | 1 | 6807 | 2 | 1 | 99.82% |
| pred. Normal | 0 | 4 | 1 | 468 | 2 | 98.53% |
| pred. Theft | 0 | 2 | 1 | 5 | 69 | 89.61% |
| class recall | 99.88% | 99.86% | 99.91% | 98.11% | 94.52% | |

*Fig. 5. Confusion matrix for ML model 2 (Forward Selection and Decision Tree)*

accuracy: 98.95% +/- 0.33% (micro average: 98.95%)

| | true DDoS | true Reconnaissance | true DoS | true Normal | true Theft | class precision |
|---|---|---|---|---|---|---|
| pred. DDoS | 6491 | 0 | 36 | 0 | 0 | 99.45% |
| pred. Reconnaissance | 1 | 5049 | 6 | 0 | 3 | 99.80% |
| pred. DoS | 152 | 0 | 6769 | 0 | 0 | 97.80% |
| pred. Normal | 0 | 0 | 0 | 477 | 0 | 100.00% |
| pred. Theft | 0 | 0 | 2 | 0 | 70 | 97.22% |
| class recall | 97.70% | 100.00% | 99.35% | 100.00% | 95.89% | |

*Fig. 6. Confusion matrix for ML model 3 (Backward Elimination and Decision Tree)*

**DT** (1 results Process results)
Completed Oct 26 2020 8.33.14 AM (execution time 4

**forward selection DT** (1 results Process results)
Completed Oct 26 2020 8.14.59 AM (execution time 1

**backward elimination DT** (1 results Process results)
Completed Oct 26 2020 8.23.04 AM (execution time 1

*Fig. 7. Processing time for all models*

$$Accuracy\,M1 = \frac{6{,}644 + 5{,}049 + 6{,}813 + 48 + 37}{19{,}056} = 97.56\,\%.$$

The model achieves a high overall accuracy of 97.56 %, indicating its effectiveness in predicting these behaviors. In this study while accuracy is the main evaluation for consideration, we also describe other noteworthy results. Class-specific metrics reveal that the model performs exceptionally well in identifying DDoS, Reconnaissance, and DoS attacks, achieving 100 % recall for these categories.

Furthermore, the model exhibits high precision, especially for DDoS (99.46 %) and DoS (100 %), ensuring that most predictions for these classes are correct.

In the case of Reconnaissance, the model maintains a 100 % recall but has a slightly lower precision of 92.17 %, suggesting a minor occurrence of false positives. This indicates the model which occasionally misclassifies other behaviors as Reconnaissance, though it remains highly reliable in identifying actual instances of this behavior. For the Normal and Theft categories, while precision remains high (100 %), recall drops to 10.06 % for Normal traffic and 50.68 % for Theft, showing the model's difficulty in detecting these classes, especially for normal network activity.

Overall, it is clear that the first model demonstrates strong capabilities in detecting network attacks, particularly DDoS, Reconnaissance, and DoS, with near-perfect precision and recall in most cases. However, the lower recall for Normal and Theft categories suggests room for improvement, particularly in fine-tuning the model to enhance its detection rate for non-attack behaviors while maintaining high precision across all classes.

Fig. 5 shows the confusion matrix for model 2 (Forward Selection and Decision Tree) using from Rapidminer software. The accuracy of model 2 is calculated as follows.

$$Accuracy\,M2 = \frac{6{,}636 + 5{,}042 + 6{,}807 + 468 + 69}{19{,}056} = 99.82\,\%.$$

The confusion matrix presented demonstrates the model's refined performance in classifying network behaviors, achieving an impressive overall accuracy of 99.82 %. The classification results for DDoS, Reconnaissance, and DoS attacks are particularly noteworthy, as the model achieves near-perfect precision and recall for these categories.

Specifically, the model's precision for DDoS reaches 99.95 % with a recall of 99.88 %, and similarly, Reconnaissance is detected with a 99.92 % precision and a 99.86 % recall. These results signify the model's capability to reliably identify these attack types with virtually no false positives or missed attacks, highlighting its robustness in security-focused applications.

Furthermore, the model has made substantial progress in detecting Normal traffic, which was previously a challenge. With a 98.53 % precision and 98.11 % recall for the Normal category, the model now effectively differentiates between benign network behaviors and malicious activities. This improvement indicates a more balanced performance, where both attack and non-attack behaviors are recognized with high accuracy. The false positive and false negative rates for normal traffic have been significantly reduced compared to earlier performance, contributing to a more comprehensive detection capability.

The Theft category, while still lagging behind other classes, shows reasonably good performance with 89.61 % precision and 94.52 % recall. Although this class has a slightly higher rate of false positives and false negatives, the model is still effective in identifying the majority of theft cases. Given the high complexity of detecting subtle theft behaviors in network traffic, these results are promising. However, further optimization of features or techniques specific to theft detection could help enhance the performance in this category to match the success seen in identifying DDoS, Reconnaissance, and DoS attacks.

Fig. 6 shows the confusion matrix for model 3 (Backward Elimination and Decision Tree) obtained from Rapidminer software. The accuracy of model 3 is calculated as follows.

$$Accuracy\,M3 = \frac{6{,}491 + 5{,}049 + 6{,}769 + 477 + 70}{19{,}056} = 98.95\,\%.$$

The confusion matrix presented shows a strong overall accuracy of 98.95 %, reflecting the model's solid performance across different categories. Accuracy measures the proportion of total predictions (both true positives and true negatives) that were correct. The model performs exceptionally well in detecting key network behaviors such as Distributed Denial of Service (DDoS), Reconnaissance, Denial of Service (DoS), Normal, and Theft.

In terms of class-specific accuracy, the model excels at detecting Reconnaissance with a perfect 100 % recall and 99.80 % precision, indicating that it correctly identifies all instances of Reconnaissance without incorrectly labeling other categories as Reconnaissance. DDoS detection is also highly accurate, with 97.70 % recall and 99.45 % precision, signifying that almost all true DDoS instances are caught by the model, and only a minimal number of non-DDoS events are misclassified as DDoS.

The model shows similar high accuracy for DoS attacks, achieving 99.35 % recall and 97.80 % precision. While it correctly identifies nearly all true DoS instances, a small number of DDoS cases (152) are misclassified as DoS. For Normal traffic, the accuracy is perfect, with both 100 % precision and 100 % recall, which indicates the model's ability to fully differentiate between normal and attack behaviors.

Lastly, Theft detection, with 95.89 % recall and 97.22 % precision, shows slightly lower performance compared to other classes but remains strong, accurately identifying most theft cases while maintaining a low rate of false positives. Overall, the model's accuracy across the various categories demonstrates its effectiveness in both attack detection and classification of benign traffic.

In addition to accuracy, the study also concerns the calculation of the processing time required by each model to finish the process of detecting botnet attack within IoT networks and then comparing them.

The result processing time for all models is shown in Fig. 7. It is clearly seen that model 1 requires 4 seconds of execution time, while both remaining models require only 1 second to finish their processes. This means the application of feature selection technique (both forward selection and backward elimination) significantly reduces the execution time in comparison to Decision Tree only.

The Table concludes the main findings. It is clearly seen that a new hybrid model, a combination of Forward Selection and Decision Tree which we called *FoSDeT Model,* outperforms other models by accounting for 99.82 % accuracy and requires only 1 second for processing time.

In addition, our *FoSDeT Model* also shows better result in comparison to previous studies. For example, a study by Chiba, et al. who applied Decision Tree and other machine learning algorithms for network based IDS has 96.66 % accuracy for Decision Tree [32]. Then a new Decision Tree model proposed in [33] to develop fuzzy signature-based intrusion detection systems achieved 96.70 % accuracy. In addition, another Decision Tree implementation to establish intelligent network intrusion detection by [34] showed accuracy level of 99.42 %. More recently, the research employing Decision Tree to provide intelligent security for smart home has accounted for 99.28 % accuracy [35].

*Table*

Final results

| Comparison | Accuracy, % | Execution Time, s |
|---|---|---|
| Model 1 Decision Tree | 97.56 | 4 |
| Model 2 (Forward Selection & Decision Tree) | 99.82 | 1 |
| Model 3 (Backward Elimination & Decision Tree) | 98.95 | 1 |

Finally, our *FoSDeT Model* is proven as a novel hybrid machine learning model for accurately and timely detecting any cyber attacks caused by botnet on Internet of Things networks.

**Conclusions.** Botnet attacks have been considered a serious problem in Internet of Things networks that often hamper their benefits. To address the challenges posed by botnet attacks in Internet of Things (IoT) networks, we have proposed a novel hybrid machine learning model, termed the *FoSDeT Model***.** This model combines the Forward Selection algorithm with the Decision Tree algorithm, leveraging the strengths of Forward Selection for dimensionality reduction and the interpretability and efficiency of Decision Trees. Our results demonstrate that the *FoSDeT Model* significantly outperforms other models considered in this study as well as several models presented in previous studies, achieving a remarkable accuracy rate of 99.82 % and a rapid detection time of just 1 second. These novel findings underscore the potential of the *FoSDeT Model* as a robust and efficient solution for real time intrusion detection in resource constrained IoT networks.

**Acknowledgement.** *The authors express their gratitude to the CAICE Center for Applied Informatics and Computer Engineering (Politeknik Negeri Ujung Pandang) and The United Kingdom Simulation Society, Nottingham Trent University for technical support and also to P3M Politeknik Negeri Ujung Pandang to finalize the study.*

**References.**

**1.** Ozmen, M. O., Song, R., Farrukh, H., & Celik, Z. B. (2023, January). Evasion attacks and defenses on smart home physical event verification. *Network and Distributed System Security Symposium (NDSS). Internet Society*. https://doi.org/10.48550/arXiv.2401.08141.

**2.** Sadeghi-Niaraki, A. (2023). Internet of Thing (IoT) review of review: Bibliometric overview since its foundation. *Future Generation Computer Systems, 143*, 361-377. https://doi.org/10.1016/j.future.2023.01.016.

**3.** Almazrouei, O. S. M. B. H., Magalingam, P., Hasan, M. K., & Shanmugam, M. (2023). A review on attack graph analysis for iot vulnerability assessment: challenges, open issues, and future directions. *IEEE Access, 11*, 44350-44376. https://doi.org/10.1109/ACCESS.2023.3272053.

**4.** Meidan, Y., Bohadana, M., Shabtai, A., Guarnizo, J. D., Ochoa, M., Tippenhauer, N. O., & Elovici, Y. (2017, April). ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis. *Proceedings of the symposium on applied computing,* (pp. 506-509). https://doi.org/10.1145/3019612.301987.

**5.** Zhao, H., Shu, H., & Xing, Y. (2021, January). A review on IoT botnet. *The 2nd International Conference on Computing and Data Science,* (pp. 1-7). https://doi.org/10.1145/3448734.34509.

**6.** Razdan, S., Gupta, H., & Seth, A. (2021, April). Performance analysis of network intrusion detection systems using j48 and naive bayes algorithms. *2021 6th International Conference for Convergence in Technology (I2CT),* (pp. 1-7). IEEE. https://doi.org/10.1109/I2CT51068.2021.9417971.

**7.** Kotak, J., & Elovici, Y. (2023). IoT device identification based on network communication analysis using deep learning. *Journal of Ambient Intelligence and Humanized Computing, 14*(7), 9113-9129. https://doi.org/10.1007/s12652-022-04415-6.

**8.** Syamsuddin, I., Nur, R., Olivya, M., Irmawati, & Saharuna, Z. (2020). Evaluation of a Novel Intelligent Firewall Simulator for Dynamic Cyber Attack Lab. *Artificial Intelligence and Bioinspired Computational Methods: Proceedings of the 9th Computer Science On-line Conference 2020, 29*, (pp. 257-267). Springer International Publishing.

**9.** Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends, 1*(1), 56-70. https://doi.org/10.38094/jastt1224.

**10.** Beigi, E. B., Jazi, H. H., Stakhanova, N., & Ghorbani, A. A. (2014, October). Towards effective feature selection in machine learning-based botnet detection approaches. *2014 IEEE Conference on Communications and Network Security,* (pp. 247-255). IEEE. https://doi.org/10.1109/CNS.2014.6997492.

**11.** Singh, K., Guntuku, S. C., Thakur, A., & Hota, C. (2014). Big data analytics framework for peer-to-peer botnet detection using random forests. *Information Sciences, 278*, 488-497. https://doi.org/10.1016/j.ins.2014.03.066.

**12.** Alejandre, F.V., Cortés, N.C., & Anaya, E.A. (2017, February). Feature selection to detect botnets using machine learning algorithms. *2017 international conference on electronics, communications and computers (CONIELECOMP),* (pp. 1-7). IEEE. https://doi.org/10.1109/CONIELECOMP.2017.7891834
**13.** Miller, S., & Busby-Earle, C. (2016, December). The role of machine learning in botnet detection. *2016 11th international conference for internet technology and secured transactions (ICITST),* (pp. 359-364). IEEE. https://doi.org/10.1109/ICITST.2016.7856730.
**14.** Pektaş, A., & Acarman, T. (2017, July). Effective feature selection for botnet detection based on network flow analysis. *International Conference Automatics and Informatics,* (pp. 1-4).
**15.** Gadelrab, M.S., ElSheikh, M., Ghoneim, M.A., & Rashwan, M. (2018). BotCap: Machine learning approach for botnet detection based on statistical features. *International Journal of Communication Networks and Information Security, 10*(3), 563.
**16.** Hoang, X.D., & Nguyen, Q.C. (2018). Botnet detection based on machine learning techniques using DNS query data. *Future Internet, 10*(5), 43. https://doi.org/10.3390/fi10050043.
**17.** Mathur, L., Raheja, M., & Ahlawat, P. (2018). Botnet detection via mining of network traffic flow. *Procedia computer science, 132*, 1668-1677. https://doi.org/10.1016/j.procs.2018.05.137.
**18.** Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity, 2*(1), 1-22. https://doi.org/10.1186/s42400-019-0038-7.
**19.** Nõmm, S., & Bahşi, H. (2018, December). Unsupervised anomaly based botnet detection in IoT networks. *2018 17th IEEE international conference on machine learning and applications (ICMLA),* (pp. 1048-1053). IEEE. https://doi.org/10.1109/ICMLA.2018.00171.
**20.** Shafiq, M., Tian, Z., Bashir, A.K., Du, X., & Guizani, M. (2020). IoT malicious traffic identification using wrapper-based feature selection mechanisms. *Computers & Security, 94*, 101863. https://doi.org/10.1016/j.cose.2020.101863.
**21.** Baig, Z.A., Sanguanpong, S., Firdous, S.N., Nguyen, T.G., & So-In, C. (2020). Averaged dependence estimators for DoS attack detection in IoT networks. *Future Generation Computer Systems, 102*, 198-209. https://doi.org/10.1016/j.future.2019.08.007.
**22.** Bovenzi, G., Aceto, G., Ciuonzo, D., Persico, V., & Pescapé, A. (2020, December). A hierarchical hybrid intrusion detection approach in IoT scenarios. *GLOBECOM 2020-2020 IEEE global communications conference,* (pp. 1-7). IEEE. https://doi.org/10.1109/GLOBECOM42002.2020.9348167.
**23.** Shaukat, K., Luo, S., Chen, S., & Liu, D. (2020, October). Cyber threat detection using machine learning techniques: A performance evaluation perspective. *2020 international conference on cyber warfare and security (ICCWS),* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCWS48432.2020.9292388.
**24.** Soe, Y.N., Feng, Y., Santosa, P.I., Hartanto, R., & Sakurai, K. (2020). Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features. *Electronics, 9*(1), 144. https://doi.org/10.3390/electronics9010144.
**25.** Ullah, I., & Mahmoud, Q. H. (2020). A two-level flow-based anomalous activity detection system for IoT networks. *Electronics, 9*(3), 530. https://doi.org/10.3390/electronics9030530.
**26.** Shobana, M., & Poonkuzhali, S. (2020, December). A novel approach for detecting iot botnet using balanced network traffic attributes. *International Conference on Service-Oriented Computing,* (pp. 534-548). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-76352-7_48.
**27.** Syamsuddin, I., & Barukab, O.M. (2022). SUKRY: suricata IDS with enhanced kNN algorithm on raspberry Pi for classifying IoT botnet attacks. *Electronics, 11*(5), 737. https://doi.org/10.3390/electronics11050737.
**28.** Qian, G., Hu, L., Zhang, W., & He, W. (2023). A new intrusion detection model for industrial control system based on hierarchical interval-based BRB. *Intelligent Systems with Applications, 18*, 200239.
**29.** AlHaddad, U., Basuhail, A., Khemakhem, M., Eassa, F.E., & Jambi, K. (2023). Ensemble model based on hybrid deep learning for intrusion detection in smart grid networks. *Sensors, 23*(17), 7464. https://doi.org/10.3390/s23177464.
**30.** Karmous, N., Aoueileyine, M.O.E., Abdelkader, M., & Youssef, N. (2023, March). Enhanced Machine Learning-Based SDN Controller Framework for Securing IoT Networks. *International Conference on Advanced Information Networking and Applications,* (pp. 60-69). Cham: Springer International Publishing.
**31.** Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems, 100*, 779-796. https://doi.org/10.1016/j.future.2019.05.041.
**32.** Chiba, Z., Abghour, N., Moussaid, K., El omri, A., & Rida, M. (2019). Intelligent approach to build a deep neural network based IDS for cloud environment using combination of machine learning algorithms. *Computers & Security, 86*, 291-317.
**33.** Masdari, M., & Khezri, H. (2020). A survey and taxonomy of the fuzzy signature-based intrusion detection systems. *Applied Soft Computing, 92*, 106301.
**34.** Guezzaz, A., Benkirane, S., Azrour, M., & Khurram, S. (2021). A reliable network intrusion detection approach using decision tree with enhanced data quality. *Security and Communication Networks, 2021*(1), 1230593.
**35.** Batool, S., Abid, M.K., Salahuddin, M.A., Aziz, Y., Naeem, A., & Aslam, N. (2024). Integrating IoT and Machine Learning to Provide Intelligent Security in Smart Homes. *Journal of Computing & Biomedical Informatics, 7*(01), 224-238.

## FoSDet: нова гібридна модель машинного навчання для точного та швидкого виявлення ботнету інтернету речей

*І. Сиамсуддін*[*1], *Д. Аль-Дабасс*[2]

1 – Державна політехніка Уджунг Панданг, м. Макассар, Індонезія
2 – Університет Ноттінгем Трент, м. Ноттінгем, Велика Британія
[*] Автор-кореспондент e-mail: irfans@poliupg.ac.id

**Мета.** Це дослідження спрямоване на впровадження нової гібридної моделі машинного навчання для підвищення точності та швидкості виявлення ботнет-атак у мережах Інтернету речей. Нова модель є результатом інтеграції алгоритму дерева прийняття рішень та алгоритмів вибору ознак для створення нового гібридного машинного навчання з метою підвищення ефективності виявлення ботнетів Інтернету речей.

**Методика.** У дослідженні застосована методологія дослідження на основі шести кроків. Вона складається зі збору масивів даних, попередньої обробки масивів даних, застосування машинного навчання, порівняння алгоритмів виділення ознак, поєднання машинного навчання та алгоритмів виділення ознак і, нарешті, порівняння результатів.

**Результати.** Нова гібридна модель машинного навчання (ML) під назвою FoSDeT була отримана в результаті поєднання алгоритму дерева прийняття рішень та алгоритму відбору ознак під назвою Forward Selection (Прямий відбір), що демонструє значне покращення виявлення ботнетів Інтернету речей у порівнянні зі стандартною моделлю дерева прийняття рішень.

**Наукова новизна.** Робота пропонує простий, але потужний гібридний підхід, що інтегрує алгоритм дерева прийняття рішень із двома попередньо визначеними алгоритмами відбору ознак, а саме: прямим відбором і зворотним виключенням. Нова гібридна модель під назвою FoSDeT демонструє значне підвищення ефективності виявлення ботнетів Інтернету речей.

**Практична значимість.** Гібридна модель, отримана в результаті даного дослідження, може бути використана фахівцями з IT-безпеки при розробці реальних систем виявлення вторгнень для захисту мереж Інтернету речей від ботнет-атак.

**Ключові слова:** *Інтернет речей, ботнет, кібератака, машинне навчання, точність виявлення, швидкість виявлення*

*The manuscript was submitted 06.09.24.*