Comparison of Machine Learning Algorithms on Noisy Data

Dijana Oreski, Dunja Visnjic, Nikola Kadoic
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varazdin
{dijana.oreski, dunja.visnjic, nikola.kadoic}@foi.unizg.hr

Abstract*. Having noisy data in datasets is not a rare situation. Noisiness in data can influence the results and accuracy of machine learning algorithms applications. This paper focuses on the question of which machine learning algorithm will perform the best when identifying noisy data. Answering this question brings us one step closer to a meta-learning recommendation system, the main goal of project SIMON. To answer this question, experiments are conducted on the publicly available datasets Bot-IoT, which consists of real and simulated IoT network traffic. Several approaches to implementing behavioral IoT botnet attack detection have been explored in the literature, including machine learning. In this paper, we are exploring four different types of machine learning algorithms: (i) machine learning algorithms based on error, (ii) machine learning algorithms based on information, (iii) machine learning algorithms based on similarity, and (iv) machine learning algorithms based on probability. The results show that the highest accuracy is achieved by a machine learning algorithm based on error – an artificial neural network that achieved the highest accuracy. However, even though the decision tree achieved slightly lower results, there is no statistically significant difference between the artificial neural network and the decision tree.*

Keywords. Machine learning algorithms, noisy data, meta-learning, network forensics, IoT devices

1 Introduction

This paper is prepared under the scope of project SIMON: Intelligent system for automatic selection of machine learning algorithms in social sciences. The main goal of the project SIMON is to develop an intelligent system for the automatic selection of machine learning algorithms in the social sciences that perform better on a given data set, taking into account the specific characteristics of the data. The research on the project involves a comparative analysis of a large number of machine learning algorithms on a large number of datasets.

In the last few years, there is a significant increase in the usage of meta-learning to enable the selection of the best machine learning algorithm for a given data set. Meta-features are used to explain the properties of the datasets and the performance of machine learning algorithms. One of the meta-features is noise. Noise is a measurement disparity between a setting that is empirical and what a dataset claims about it. Empirical data may be noisy for a variety of reasons, such as a lack of accuracy in the dataset's construction or the addition by mistake of extra attributes to some items, or the omission of certain objects while characterizing the scope of an attribute. Machine learning algorithms' ability to predict outcomes can suffer with the presence of noise in data sets. Bot-IoT data is characterized by noise.

In recent years, the wide adoption of the modern Internet of Things (IoT) paradigm has led to the development of fog computing, which improves the collection and processing of enormous amounts of data, when cloud computing features, such as networking, data storage, administration, and analytics, are placed very close to the edge of networks. IoT development has also led to an increase in security challenges. Many IoT devices have various security flaws in their implementation and design, making them a prime target for botnet attacks. One of the best options is intrusion detection systems, especially those created with artificial intelligence.

The billions of physical items that IoT consists of, can connect with each other, including only a little human involvement. IoT has developed into one of the most widespread technologies and a fascinating area of study in the business and research sector. IoT is becoming increasingly popular and in demand. Numerous organizations are providing funding in this area for their purposes and as a service to other organizations. Although the IoT is expanding rapidly due to technological advancements, the proliferation of IoT devices, and the activation of services, there are significant security threats and financial harm caused by the actions of numerous botnets. A botnet is a group of software robots, or bots, that operate automatically

and autonomously. Botnets function on networks of zombie computers under the remote control of attackers. IoT has been impacted and infected by intelligent botnet activity, including distributed denial of service (DDoS) attacks, spamming, and phishing. There hasn't been a network forensics technique that can perfectly classify, detect, and trace the activities of sophisticated botnets to date, even though botnets with such attacks have posed a severe security concern to the Internet infrastructure for years. Consequently, it's critical to correctly identify these botnets' actions and there is a strong need to create new approaches for detecting attacks, predicting attacks, and preventing attacks.

Machine learning algorithms can be applied to protect data from cyber security threats. Machine learning algorithms are used in a variety of ways to stop and spot network outbreaks and security holes. In recent years, numerous studies have trained and validated predictive models developed employing machine learning algorithms to define these botnet attacks. Artificial neural networks are among the machine learning algorithms which were mostly used so far across various research papers.

Prior research focused primarily on achieving the highest level of accuracy in separating legitimate from malicious IoT communications, with little attention paid to identifying the specific sort of attack that was being performed, or investigating data characteristics, neither comparing various machine learning approaches nor investigating their applicability for IoT dataset and specifies of such data.

Bot-IoT data is characterized by noise. Recently, a few research papers proposed pre-processing techniques which support data noise cleaning on IoT data. Jane and Arockiam (Jane & Arockiam, 2021) proposed a technique called Detection and Removal of Noise (DaRoN) that removes “the null values, error values, repeated values, incomplete values, and irrelevant values”. By doing that, the authors removed noisy IoT data. Bobulski and Kubanek (Bobulski & Kubanek, 2022) discussed the need for automatization of data cleaning and suggested a data cleaning method.

Although such initiatives provided a promising avenue to deal with noise in data, cleaning approaches imply deleting data. Part of the data disappears during the cleaning phase, which results in the loss of information from the initial dataset. This is especially important when dealing with small datasets, so the decision maker/data analyst may decide not to clean the data so that potentially important information would not be lost.

In this paper we did not perform data cleaning, instead, we are employing four machine learning-based predictive models on raw data to preserve information and discover which machine learning approaches work best on such data.

An aim is to identify IoT botnet assaults that not only aid in separating legitimate traffic from malicious traffic but also identify the specific IoT botnet attack type. To accomplish this, four different types of machine learning algorithms were applied to publicly available IoT datasets. Such a dataset is shown to be noisy, with numerous erroneous attribute values and missing values. From a data mining point of view, noisy data requires careful examination and preparation before the development of predictive models. By exploring and analyzing that dataset, we are striving to answer the question: which of the four different approaches to machine learning gives the most accurate predictive models on noisy IoT data?

The rest of the paper is organized as follows. In Section 2, a related literature review is provided. Section 3 explains the data and research methodology and gives research results focusing on predictive model accuracy. Finally, Section 4 concludes the article and provides guidelines for future research activities.

2 Literature review

Finding the intelligent intrusion detection system in IoT-based environments for many sorts of applications has been the subject of a vast body of work and study so far in the literature. Hereinafter, are listed some of them.

Bagui, Stevens and Bagui (Bagui et al., 2021) present a thorough examination of the benchmark dataset NSL-KDD to create a powerful network-based intrusion detection system. The uniqueness of this study lies in the discovery of the bare minimum set of information necessary for the automated classification of each assault as well as each attack type in the NSL-KDD dataset. There hasn't yet been any analysis done on the specific attack's level. Following the usage of Information Gain for feature selection are machine learning techniques like J48 Decision Tree and Naive Bayes. With each method, high classification accuracy is attained.

Hyun (Hyun, 2021) conducted a comparative study on how well some machine learning algorithms perform at spotting botnet activities. k-nearest neighbours (k-NN) is the most effective and efficient machine learning method for DDoS, DoS, and reconnaissance attack detection, according to experimental findings utilizing the Bot-IoT dataset.

Malathi and Padmaja (Malathi & Padmaja, 2023) performed research to provide security using a variety of machine learning methods, which are primarily intended to detect an attack on an interconnected (IoT) network right away. Different recognition algorithms are estimated using specific metadata, or Bot-IoT. Several different machine learning algorithm types were handled during this execution step, and the majority of them achieved outstanding results. Metadata from the Bot-IoT was used to collect novel factors, but implementation and the new features created produced more trustworthy results.

Wiyono and Cahyani (Wiyono & Cahyani, 2020) were motivated to develop a novel classification

algorithm for network forensics that could monitor suspected botnet activity in the compromised network and is based on the identification of network traffic. Based on the performance study and experimental findings, the authors concluded that decision tree C4.5 algorithm and network flow identification combined with feature selection and proper classification approaches are sufficient to identify and classify attacks and help track botnet activity in the IoT environment.

Motylinski et.al. (Motylinski et al., 2022) presented a method for classifying the different attack types that were included in the IoT-Bot dataset and pre-processing phase. The authors compared the results obtained with the GPU-accelerated versions of the cuML library's Random Forest, k-NN, Support Vector Machine (SVM), and Logistic Regression classifiers, as well as the pre-processing steps taken to prepare the data for training. The best-trained models achieved 0.99 scores for accuracy, precision, recall, and f1-score, by using their approach. Additionally, the training and estimation times were greatly shortened by using feature selection and training models on GPU.

Alothman, Alkasassbeh, and Al-Haj (Alothman et al., 2020) tested various classifiers, and the results from the best three: J48, Random Forest (RF), and Multilayer Perceptron (MLP) networks were explained. The outcomes demonstrated the superiority of the RF and J48 classifiers over MLP networks and other cutting-edge technologies. The best binary classifier revealed in this study had an accuracy of 0.99, while the best classifications of main attacks and subcategories had accuracy values of 0.96 and 0.93, respectively. In this study, authors evaluated results also in terms of False Negative (FN) rates. This time around, J48 and RF classifiers performed better than the MLP network classifier and were able to classify subcategories with a maximum micro FN rate of 0.076.

A study by Bagui et.al (Bagui et al., 2019) classifies cyberattacks in the UNSW-NB15 dataset using a hybrid feature selection process and classification techniques. In order to identify the best subset of features, k-means clustering, and correlation-based feature selection were combined. After feature selection, two classification methods - one probabilistic, Naive Bayes (NB), and the other based on decision trees (J48), were used in the modeling phase. According to their findings, the NB model in combination with a hybrid feature selection approach was able to increase the classification accuracy of the majority of attacks, particularly the unusual attacks. With this feature selection and NB model combination, the false alarm rates were lower for the majority of attacks, especially the unusual attacks. Although the J48 decision tree model's classification rate for all attack families was already quite high, with or without feature selection, it did not perform any better.

Naaz (Naaz, 2021) used the IoT dataset to test the effectiveness of the machine learning algorithms random forest classifier, support vector machine, and logistic regression for the detection of phishing attacks. The results were then compared to earlier studies that had used the same dataset. Based on the accuracy, error rate, precision, and recall, the outputs of these algorithms have then been compared.

Yudhana, Riadi, and Ridho (Yudhana et al., 2018) performed the classification of DDoS attacks by utilizing naive Bayes and neural networks to analyze network traffic. Based on the results, it was discovered that naive Bayes had a 99.9% accuracy rate while artificial neural networks had a 95.23% accuracy rate. The conclusions of the experiments demonstrate that the naive Bayes approach outperforms the neural network. The authors conclude that their analysis along with experimental results can be utilized as proof in the trial process.

Alrashdi et.al. (Alrashdi et al., 2019) propose an Anomaly Detection-IoT (AD-IoT) system, which is an intelligent anomaly detection based on a Random Forest machine learning algorithm, to handle IoT cybersecurity concerns in a smart city. At dispersed fog nodes, the suggested approach may successfully detect hacked IoT devices. The authors used a current dataset to demonstrate the model's accuracy and evaluate it. Their research demonstrates that the AD-IoT is capable of achieving the maximum classification accuracy of 99.34% with the lowest false positive rate.

Almiani et.al. (Almiani et al., 2020) provided a fully automated, artificially created intrusion detection system against cyberattacks. The suggested model makes use of multi-layered recurrent neural networks that are intended to be used for security in fog computing, which is implemented very close to end users and IoT devices. Using a balanced version of the hard dataset, NSL-KDD, the authors demonstrated their suggested model. The experimental findings and computer simulations indicated the stability and robustness of the suggested model in terms of a range of performance measures.

Lutta, Sedky, and Hassan (Lutta et al., 2021) recently carried out a Systematic Literature Review (SLR) of the most recent IoT forensics research developments. One of their guidelines for future research indicates the need for machine learning algorithms implemented in this domain.

Numerous fields deals with the problem of data with noise and IoT forensics is no exception. There are two main types of noise in data: class noise and attribute noise. Attribute noise is considered to be less harmful to modeling than class noise. Robust machine learning algorithms have proven to be a good approach to dealing with imperfect data. According to the literature review, those algorithms are k-NN (Liu & Zhang, 2012; Moosavi et al., 2010), artificial neural networks (Folleco et al., 2009; Miranda et al., 2009), classification, and regression trees (Folleco et al., 2009; Khoshgoftaar et al., 2011; Miranda et al., 2009).

3 Empirical analysis

The used dataset is described in the first section of this chapter, and the methodologies for data analysis are explained in the second section. The third section provides insights into research results.

3.1 Data description

Data used in this research was created by Koroniotis, Moustafa, Sitnikova, and Turnbull (Koroniotis et al., 2019). The authors created a new Bot-IoT dataset using realistic IoT networks. In (Koroniotis et al., 2019), a full description of simulating IoT sensors is provided along with data understanding. There are 3 668 522 cases in the dataset. It should be mentioned that it is the reduced dataset since the original one contains 72 000 000 instances and its analysis was extremely time and computationally demanding. That is the reason why just 5% of the original dataset has been extracted. In our study, we also used such a reduced dataset. 46 features are present in the dataset for each instance. Some of these features are characterized as noisy. Thus, hereinafter we are dealing with attribute noise.

Since Koroniotis et al. (Koroniotis et al., 2019) performed feature selection and extracted only 10 features from the original dataset, 10 features are also chosen in this research. The ten best features which were determined by Koroniotis et al. (Koroniotis et al., 2019) are as follows: rate, drate, srate, state number, max, mean, min, stddev, gs number, seq. The ten best features from our research are in Table 1, along with noise level measured as amount of irrelevant information.

Table 1. Feature description

Feature	Feature Explanation	Noise level
drate	Destination-to-source packets per second	0.65
Flags number	Numerical representation of feature flags	0.54
max	Maximum duration of aggregated records	0.32
mean	Average duration of aggregated records	0.29
min	Minimum duration of aggregated records	0.34
N IN Conn P DstIP	Number of inbound connections per destination IP.	0.41
N IN Conn P SrcIP	Number of inbound connections per source IP.	0.11
seq	seq Argus sequence number	0.71
State number	Numerical representation of feature state	0.26
stddev	The standard deviation of aggregated records	0.74
attack	Class label: 0 for Normal traffic, 1 for Attack Traffic	0.44

For feature selection in this study, we used a contrast set mining based on the STUCCO algorithm: SfFS (STUCCO for Feature Selection). In our earlier research papers, SfFS application in feature selection produced the best results (Oreski & Androcec, 2018; Oreški & Andročec, 2020). The concept of SfFS is initially introduced in (Oreški & Kliček, 2015). The decision on the implementation of SfFS for feature selection is due to the noisy character of the used data. Missing values along with the outliers bring noise to the data. SfFS has shown to be a good approach for dealing with noisy data (Oreški & Kliček, 2015).

3.2 Methods overview

Koroniotis et al. (Koroniotis et al., 2019) evaluated the performance of network forensic methods by applying three machine learning algorithms. The models that were trained were: Support Vector Machine (SVM), Recurrent Neural Network (RNN), and Long-Short Term Memory Recurrent Neural Network (LSTM-RNN). All those algorithms present similar approaches to the development of predictive models. In this paper, we are expanding the number of different approaches by investigating the effectiveness of four different approaches to learning. Thus, we are comparing algorithms belonging to four different groups of machine learning algorithms: (i) machine learning algorithms based on error, (ii) machine learning algorithms based on information, (iii) machine learning algorithms based on similarity, (iv) machine learning algorithms based on probability. An artificial neural network is used as a machine learning algorithm based on error, the decision tree is used as a machine learning algorithm based on information, the k-NN is used as a machine learning algorithm based on similarity, and Naïve Bayes classifier is used as machine learning algorithm based on probability.

The artificial neural network algorithm develops a predictive model by correcting the weights of links between neurons to reduce the error of the model. The decision tree algorithm seeks the most informative features to develop a predictive model based on such features. k-NN seeks similar instances and classifies new instances in the same class as the nearest feature. The Naïve Bayes classifier is based on the Bayesian theorem of conditional probability.

After the four algorithms were applied to the dataset, several two-matched sampling t-tests were used to compare the algorithms to answer the following research questions:

1. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and decision tree (DT)?

2. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and k-NN?
3. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and Naïve Bayes (NB)?
4. Is there a statistically significant difference in accuracy between decision tree (DT) and k-NN?
5. Is there a statistically significant difference in accuracy between decision tree (DT) and Naïve Bayes (NB)?
6. Is there a statistically significant difference in accuracy between k-NN and Naïve Bayes (NB)?

4 Research results

In data analysis, we initially carried out feature selection and identified relevant features for separating the attack and normal classes (ten features enlisted in Table 1). The proposed methodology's classification accuracy is then evaluated.

Predictive model accuracy can be tested and assessed in a variety of ways. The *k*-fold cross-validation is employed here. The data set is split into *k* subsets using *k*-fold cross-validation. The test set is always one of the *k* subsets, whereas the training set is always the other *k-1* subsets. 10 folds are used in this research.

Table 2 presents the results of four machine algorithms in terms of accuracy.

Table 2. Models accuracy

Algorithm	Accuracy
Artificial neural networks	99,83
Decision tree	99,24
k-NN	98,65
Naïve Bayes classifier	96,33

We may infer several conclusions by using the performance measure from Table 3. Question to be asked, is it possible to generalize the outcomes or are they the result of chance? The goal of statistical significance testing is to determine how well evaluation measures reflect classifier behavior. Two matched sampling t-tests were used to compare the algorithms we tested on one domain. The significance of the mean difference is investigated at the significance level of 0.05. To test whether we can reject the null hypothesis, the assumption in Table 3 is that there is no difference between the mean values of algorithms performances.

The presumptions of the t-test were met.

Table 3. T-test

Hypothesis	Model	T-test
H0: ANN = DT	ANN	0.07
	DT	
H0: ANN = k-NN	ANN	0.04
	k-NN	
H0: ANN = NB	ANN	0.02
	NB	
H0: DT = k-NN	DT	0.06
	k-NN	
H0: DT = NB	DT	0.02
	NB	
H0: k-NN =NB	k-NN	0.02
	NB	

As seen in Table 3. there are statistically significant differences in the performances of artificial neural networks and k-NN and NB. However, differences in performances between ANN and DT are not statistically significant. DT performs statistically significantly better than NB, but not when compared with k-NN. In the end, k-NN performs better than NB and the difference in performances between these machine learning algorithms is statistically significant.

5 Conclusion

In this paper, we have proposed a machine learning-based intrusion detection predictive model for IoT network security. The proposed model adopts four different approaches to machine learning-based development of predictive models. The results of the performance evaluation reveal the effectiveness of the artificial neural networks approach compared to the decision tree, k-NN, and Naïve Bayes classifier.

This paper gives three scientific contributions: i) in the field of machine learning, by investigating how different machine learning approaches handle noisy IoT data, (i) in network forensics, by comparing different machine learning approaches and demonstrating which one achieves the best predictive model in this domain, and (iii) in the field of IoT field, because the models are more accurate than the previous models applied to the same dataset.

There are several limitations of the research presented here. First, only one dataset is used in algorithms comparison. In future research, we will upgrade several datasets. Second, the dataset was highly unbalanced, which could lead to bias in the learner's favor of the attack class. This prejudice arises from the fact that the attack class is much overrepresented in comparison to the normal class. We

will employ several strategies to address the class imbalance in future research. Some of them include comparative analysis of different class imbalanced strategies combined with machine learning algorithms on several datasets to determine if there is a connection between certain machine learning algorithms' performance in terms of accuracy and the level of imbalance and noisiness in data. Third, we will perform data cleaning in the pre-processing phase and compare the results of our models gained on raw data with the results of cleaned data.

Acknowledgments

This paper is supported by Croatian science foundation under the project SIMON: Intelligent system for automatic selection of machine learning algorithms in social sciences, UIP-2020-02-6312.

References

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, *101*, 102031. https://doi.org/10.1016/j.simpat.2019.102031

Alothman, Z., Alkasassbeh, M., & Al-Haj Baddar, S. (2020). An efficient approach to detect IoT botnet attacks using machine learning. *Journal of High Speed Networks*, *26*(3), 241–254. https://doi.org/10.3233/JHS-200641

Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. (2019). AD-IoT: Anomaly detection of IoT cyberattacks in smart city using machine learning. *2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019*, *January*, 305–310. https://doi.org/10.1109/CCWC.2019.8666450

Bagui, S., Kalaimannan, E., Bagui, S., Nandi, D., & Pinto, A. (2019). Using machine learning techniques to identify rare cyber-attacks on the UNSW-NB15 dataset. *Security and Privacy*, *2*(6). https://doi.org/10.1002/spy2.91

Bagui, S., Stevens, J., & Bagui, S. (2021). Analyzing and Classifying Network Attacks Using Machine Learning on the NSL-KDD Dataset. *The Journal of Computing and Technology*, *2*(1), 24–36.

Bobulski, J., & Kubanek, M. (2022). A method of cleaning data from IoT devices in Big data systems. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, *February 2023*, 6596–6598. https://doi.org/10.1109/BigData55660.2022.10020651

Folleco, A. A., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2009). Identifying learners robust to low quality data. *Informatica (Ljubljana)*, *33*(3), 245–259.

Hyun, M.-J. (2021). Hyun, M. (2021). A comparative study of the performance of machine learning algorithms to detect malicious traffic in IoT networks. *Journal of Digital Convergence*, *19*(9), 463–468. https://doi.org/10.14400/JDC.2021.19.9.463

Jane, V. A., & Arockiam, L. (2021). Daron: A technique for detection and removal of noise in IoT data by using central tendency. *Annals of the Romanian Society for Cell Biology*, *25*(2), 3197–3203.

Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *41*(3), 552–568. https://doi.org/10.1109/TSMCA.2010.2084081

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*(November), 779–796. https://doi.org/10.1016/j.future.2019.05.041

Liu, H., & Zhang, S. (2012). Noisy data elimination using mutual k-nearest neighbor for classification mining. *Journal of Systems and Software*, *85*(5), 1067–1074. https://doi.org/10.1016/j.jss.2011.12.019

Lutta, P., Sedky, M., Hassan, M., Jayawickrama, U., & Bakhtiari Bastaki, B. (2021). The complexity of internet of things forensics: A state-of-the-art review. *Forensic Science International: Digital Investigation*, *38*, 301210. https://doi.org/10.1016/j.fsidi.2021.301210

Malathi, C., & Padmaja, I. N. (2023). Identification of cyber attacks using machine learning in smart IoT networks. *Materials Today: Proceedings*, *80*(xxxx), 2518–2523. https://doi.org/10.1016/j.matpr.2021.06.400

Miranda, A. L. B., Garcia, L. P. F., de Carvalho, A. C. P. L. F., & Lorena, A. C. (2009). Use of Classification Algorithms in Noise Detection and Elimination. *4th International Conference, HAIS*, 417–424.

Moosavi, M. R., Fazaeli Javan, M., Zolghadri Jahromi, M., & Sadreddini, M. H. (2010). An adaptive nearest neighbor classifier for noisy environments. *Proceedings - 2010 18th Iranian Conference on Electrical Engineering, ICEE 2010*, 576–580. https://doi.org/10.1109/IRANIANCEE.2010.5507

005

Motylinski, M., MacDermott, Á., Iqbal, F., & Shah, B. (2022). A GPU-based machine learning approach for detection of botnet attacks. *Computers and Security*, *123*(September), 102918. https://doi.org/10.1016/j.cose.2022.102918

Naaz, S. (2021). Detection of phishing in internet of things using machine learning approach. *International Journal of Digital Crime and Forensics*, *13*(2), 1–15. https://doi.org/10.4018/IJDCF.2021030101

Oreski, D., & Androcec, D. (2018). Hybrid Data Mining Approaches for Intrusion Detection in the Internet of Things. *Proceedings of International Conference on Smart Systems and Technologies 2018, SST 2018*, 221–226. https://doi.org/10.1109/SST.2018.8564573

Oreški, D., & Andročec, D. (2020). Genetic algorithm and artificial neural network for network forensic analytics. *43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1200–1205.

Oreški, D., & Kliček, B. (2015). A novel feature selection techniques based on contrast set mining. *14th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'15)*.

Wiyono, R. T., & Cahyani, N. D. W. (2020). Performance Analysis of Decision Tree C4.5 as a Classification Technique to Conduct Network Forensics for Botnet Activities in Internet of Things. *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*, 1–5. https://doi.org/10.1109/ICoDSA50139.2020.9212932

Yudhana, A., Riadi, I., & Ridho, F. (2018). DDoS classification using neural network and naïve bayes methods for network forensics. *International Journal of Advanced Computer Science and Applications*, *9*(11), 177–183. https://doi.org/10.14569/ijacsa.2018.091125