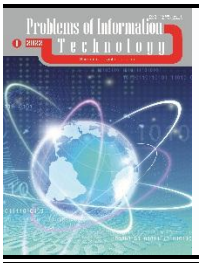

Available online at www.jpit.az

13(1)
2022

Experimental Study of Machine Learning Methods in Anomaly Detection

Makrufa Sh. Hajirahimova[a], Leyla R. Yusifova[b]

[a,b,] Institute of Information Technology, Azerbaijan National Academy of Sciences, B. Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

[a]hmakrufa@gmail.com; [b]yusifova863@gmail.com

0000-0003-0786-5974[a]

0000-0002-9720-8638[b]

ARTICLE INFO

http://doi.org/10.25045/jpit.v13.i1.02

Article history:
Received 9 September 2021
Received in revised form 9 November 2021
Accepted 5 January 2022

Keywords:
Big data
Anomaly
DoS attacks
IDS
Machine learning
Ensemble classification

ABSTRACT

Recently, the widespread usage of computer networks has led to the increase of network threats and attacks. Existing security systems and devices are insufficient in the detection of intruders' attacks on network infrastructure, and they considered to be outdated for storing and analyzing large network traffic data in terms of size, speed, and diversity. Detection of anomalies in network traffic data is one of the most important issues in providing network security. In the paper, we investigate the possibility of using machine learning algorithms in the detection of anomalies – DoS attacks in computer network traffic data on the WEKA software platform. Ensemble model consisting of several unsupervised classification algorithms has been proposed to increase the efficiency of classification algorithms. The effectiveness of the proposed model was studied using the NSL-KDD database. The proposed approach showed a higher accuracy in the detection of anomalies compared to the results shown by the classification algorithms separately.

1. Introduction

The growth of Internet technologies have intensified the cyber-attacks. The intensity of cyber-attacks has neutralized traditional security techniques like signature-based against new types of attacks (Garofalo, 2017; Hajirahimova, 2014). Thus, in the Big Data Era, critical sectors such as government, energy, health, banking and telecommunications, education, transport, research centers, etc. have been destroyed as the result of various types of cyber-attacks (e.g. spam, botnets, Denial of Service (DoS), Denial of Distributed Service (DDoS), phishing, malware, viruses, etc.) to the network infrastructure (Heydari et al., 2015; Almedia, 2017; Wang et al., 2018). These organizations spend a lot of money to protect their infrastructure using various monitoring techniques. However, as the attackers use advanced devices to interfere the infrastructure, existing security and log file analysis methods are considered outdated (Ariyaluran et al., 2019). Urgent processing of the collected data is important to detect potential threats in the network. Instant monitoring of the network infrastructure and detecting abnormal behavior and threats become impossible due to the inability of available traditional monitoring devices for big data processing (Raguseo, 2018).

The anomaly is defined as a given pattern not coincided with the expected behavior. The anomaly detection was first proposed by D.E. Denning in 1987 (Denning, 1987). The main concept of this method is to determine the behavior of a network/system where a predefined behavior is compared to normal behavior. The result will be either to accept it or to launch an alarm management system for further investigation. The goal of the anomaly detection is to detect outlier data (patterns,

templates) that differ from or deviate from the main part of the data (Abdulhammed, 2019). Thus, the anomaly detection refers to the identification of elements or events in the database that cannot be identified by a specialist and does not coincide with the expected pattern.

The big volume creates serious difficulties in detecting anomalies. Because, as the number of variables or attributes increases, the amount of data also increases. According to Cisco, annual global IP traffic reached 3.3 zettabytes in 2021 (Global - VNI Complete Forecast Highlights, 2021). This creates numerous challenges in terms of data security, data transmission reliability, and detection of traffic anomalies and attacks. Traditional methods/algorithms for detecting anomalies in big data couldn't provide sufficient accuracy (Srikanth, et al., 2020; Rehman, 2016). In recent years, the anomaly detection has become a major research topic in the field of machine learning and has been the focus of numerous papers (Dua, Du, 2011; Buczak & Guven, 2016). Computer network anomalies are understood as unusual and significant changes in network traffic (Garofalo, 2017). Human mistakes made during data entry, errors in measuring devices, errors in data processing (data manipulation), mechanical errors in the system behavior, and other anomalies are the common causes.

The main goal of the anomaly detection approach is to establish a statistical model to describe normal traffic. Any deviation from this model can be considered an anomalous event and an attack. Anomalies are understood as the examples distinguishing from most of the data presented at the abstract level. Three types of anomalies are distinguished in the literature: *point anomalies*, *contextual anomalies*, *collective anomalies*. This type of anomaly has been considerably analyzed in various studies so far (Hodge, Austin, 2004; Chandola, Banerjee, Kumar, 2009; Nassif et al., 2021; Gogoi et al., 2011).

The anomaly detection studied for many years and becoming the subject of scientific research has been applied in various fields so far. The detection of data fraud of credit card transactions (Juan et al., 2018; Husejinović, 2020; Phua et al., 2010; Dash, & Ng, 2010; Chaudhary, Yadav., & Mallick, 2012), anomalous traffic in computer networks (Aliguliyev, Hajirahimova, 2019; Shon, Moon, 2007; Alguliyev, Aliguliyev, Imamverdiyev, Sukhostat, 2017; Zhang et al., 2008), suspicious cyber activity (Ariyaluran et al., 2019), detection of anomalies in medical data (eg, anomalous MRI description, etc.) (He, Wang, Graco, Hawkins, 1997; Saneja, Rani, 2017; Antal, Hajdu, 2014; Schlegl, Seeböck et al., 2017; Varian, 2020) of anomalous indications of sensor devices (e.g. abnormal readings of the apparatus sensor may indicate a malfunction in some components of the spacecraft) (Fujimaki et al., 2005) and etc. are major issues in Big Data analytics (Chandola, Banerjee, Kumar, 2005; Wang, Jones, 2017). From this point of view, the detection of non-standard data (or previously unobserved data) in big data, especially its prevention is of great importance. A network anomaly is potentially harmful traffic that affects network security. These are the major challenges for both government and corporate entities, as they can lead to financial losses, network disruption, and even threaten national security (Ariyaluran, 2019).

The main purpose of this study is to detect anomalies in network traffic, that is DoS attacks. DoS attacks are real threats to the network and cyber infrastructure (Denning, 1987; Garofalo, 2017). DoS attacks can paralyze or fail a service using network and service servers, connecting networks, and network devices (routers, etc.) resulting in significant losses. DoS attacks use the vulnerabilities in the communication protocols to hinder or completely fail the target host response to the legitimate user. This malicious activity is implemented by sending numerous requests to the "victim" server.

This paper tests several machine learning algorithms over the NSL-KDD database (NSL-KDD) in the detection of anomalies in big network data, that is DoS attacks, and proposes a classification ensemble model to increase the efficiency of classification.

The following structure of the paper is organized as follows. Section 2 summarizes the related work. Section 3 describes the methodology of the proposed approach. Section 4 experimentally testa the proposed approach and discusses the experiment results. Section 5 summarizes the results of the study.

2. Related work

This section focuses on the current state of the study in the field of anomaly detection. We should note that both review (Hodge, Austin, 2004; Agrawal, Agraual, 2015; Chandola, Banerjee, Kumar, 2005; Srikanth, Philip, Jiong, et al., 2020; Gupta, Gao, Aggarwal, Han, 2014; Nassif et al., 2021; Patcha, Park, 2007), and numerous innovative studies (Ariyaluran, 2019; Aliguliyev, Hajirahimova, 2019); Sukhostat et al, 2018; Akoglu et al., 2015; Wei et al., 2019; Aggarwal, 2005) have been conducted in the field of anomaly detection so far.

Intrusion Detection Systems (IDS) have been very corroborative for network administrators (Akbar, 2010) over the years. Two approaches to the detection of intrusion are distinguished: signature-based misuse detection and anomaly detection (Mukherjee, Heberline, Levitt, 1994). The main idea of misuse detection is to present the attacks in the form of templates or signatures.

Obviously, most of the IDSs are based on a set of rules specified by security experts (Tsai et al., 2009). The major shortcomings of these approaches are the difficulty of detecting unknown new attacks. Moreover, due to the large volume of network traffic, the rules coding is slowed down, a lot of time is spent, and dependence on expert knowledge is observed, etc. Anomaly detection methods produce high levels of false alarms (Zhang et al., 2008). Intellectual analysis methods are used to overcome such limitations of IDS and more accurately detect new intrusions, normal and anomalous network traffic (Lee, Stolfo, Mok, 2000; Agarwal, Mittal, 2012).

Parametric, nonparametric statistical methods based on distribution, proximity measurement are considered to be the primary methods for the anomaly detection (Chandola, Banerjee, Kumar, 2005). Anomaly detection systems based on statistical methods consist of two stages. The system first observes and collects one or more statistical features of network traffic, then compares the current situation with the stored situation using a stochastic method to detect behavioral changes. The key issue is to detect various types of malware, such as DoS, phishing, and spam, frequently through bots placed on target hosts. To avoid this problem, an anomaly detection method is proposed using the Ensemble Empirical Mode Decomposition algorithm (Marnerides et al., 2015).

Recently, machine learning (ML) and deep learning methods have been widely used in the identification of anomalies (Goldtein, Abdulhammed, Buczak, 2016). ML was defined in 1959 by Arthur Samuel as "a learning field providing computers to be taught without explicit programming" (Samuel, 1959). Thus, ML detects latent correlation patterns through iterative learning based on selected sample data (or past experience) instead of explicit programming. Note that there are supervised, unsupervised, and semi- supervised types of ML methods (Chandola, Banerjee, Kumar, 2005; Schlegl, et al., 2017).

Zhang and others apply a rule-based machine learning algorithm to solve IDS problems, especially unsupervised random forests to detect new intrusions (Zhang et al., 2008).

Camacho and others propose an approach considering four characteristics of big data (volume, diversity, reliability, and speed) to detect anomalies in the network of judicial systems, identify and interpret abnormal behaviors (Camacho et al., 2014). Principal Component Analysis (PCA) is used to eliminate both uncertainty (low accuracy) and high volume problem. The authors used a nuclear calculation preventing the dimensional problem of volume (number of observations) in PCA databases and allowing parallelism. Hierarchical models are also proposed when volume is excessive. Finally, the Exponentially Weighted Moving Average (EWMA) approach is used to provide high speed of data flows analysis.

Hybrid methods have also been proposed for the detection of anomalies so far (Agarwal, Mittal, 2012; Kim et al., 2014; Shon, Moon, 2007). B.Agarwal and his colleague propose a hybrid method consisting of a combination of network features entropy and Support Vector Machine (SVM) algorithms (Agarwal, Mittal, 2012). As a result, in the network anomalies' detection, they approved the superiority of the entropy-based detection method compared to the SVM-based detection system.

A new detection algorithm (MSD - k-means)

combining the statistical method MSD (Mean and Standard Deviation) and the k-means learning method is proposed to improve detection accuracy in big data by minimizing the impact of noisy data (Wei et al., 2019). MSD-k-means comprises two stages: 1) the application of MSD algorithm to eliminate noisy data in the data; 2) the application of k-means algorithm to obtain locally optimal clusters. MSD-k-means performs higher accuracy than other methods.

Approaches based on the k-means clustering algorithm are been proposed detecting network anomalies (Munz, 2007), (Kumari et al., 2016). Munz applies the k-means algorithm to groups the training data containing NetFlow records into normal and abnormal traffic groups. Kumari and others develops a methodology to prevent cyber-attacks based on k-means clustering using the Apache Spark analytics.

Obviously, computer systems have limitations in terms of storage, processing, and analysis of big data due to their volume, velocity, and variety. In order to overcome this restrictions, an optimization approach based on weighted clustering is proposed to detect anomalies (Alguliyev, Aliguliyev, Imamverdiyev, Sukhostat, 2018; Alguliyev, Aliguliyev, Imamverdiyev, Sukhostat, 2017). The weight of each point is determined by its position relative to the center of the entire data set. The weighted clustering algorithm detects anomalies more accurately than the k-means algorithm proposed in experiments using seven large-scale databases. Due to the increasing computational complexity and diversity of data, the selection of devices for all types of anomalies is difficult. In the second study, an improved optimization approach for clustering is proposed (Alguliyev, Aliguliyev, Imamverdiyev, Sukhostat, 2017). Experimental results on three databases (Australian credit card applications, heart disease database, and NSL-KDD database) showed that the proposed algorithm detects anomalies more accurately than the k-means algorithm.

The financial sector actively uses big data analytics, including fraud detection (Chaudhary, 2012; Phua et al., 2010) and transaction processing (Dash & Ng, 2010). The growing problem of card payments in misuse cases is in the focus of bank payments and payment service providers. Thus, there often occurs credit card fraud which eventually results in huge financial losses. Criminals use the technologies, such as Trojans or phishing, to steal other people's credit card information (Juan et al., 2018; Husejinović, 2020). Juan and others uses a random forest algorithm to train the behavioral characteristics of normal and abnormal operations (Juan et al., 2018). In his study, Husejinović uses Naive Bayes, C4.5 decision tree, and Bagging ensemble machine learning algorithms to predict the results of fraud operations. C4.5 decision tree algorithm demonstrates more accurate forecast (92.74%) than others in forecasting fraud operations.

Another study uses a deep learning method based on the restricted Boltzmann machine (RBM) to detect DoS attacks (Imamverdiyev, Abdullayeva, 2018). To improve the DoS attack detection accuracy, seven layers are added between the visible and hidden layers of the RBM. Precise results in DoS attack detection are obtained by optimizing the hyperparameters of the proposed deep RBM model.

3. Proposed approach

Problem statement. Assume that $D = \{x_1, x_2, \ldots, x_n\}$ database representing network traffic is provided and $M = \{m_1, m_2, \ldots, m_k\}$ number of classification algorithms are selected. The detection of malicious traffic (DoS attacks) with high accuracy on the network is required.

Research methodology. The classification ensemble model providing the detection of anomalies in computer network traffic, that is DoS attacks, is proposed to increase the efficiency of the work of IDS. Base classifiers such as NaiveBayes (NB), Decision tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron, K-Nearest Neighbor (KNN) are selected. Below a brief explanation of these classifiers is provided:

In **Naive Bayes** model, the probability of which class an object refers to is calculated according to the Bayes theorem (Yang et al., 2016). In the essence of Bayes' theorem stands the conditional probability. Conditional probability is understood as the probability of an event occurring, given that another event has already occurred. The Naive Bayes classifier considers the n characters as conditionally

independent of each other. First, the conditional probability of each given features is calculated, then the Bayes theorem is applied to predict the class label of the sample. Equation 1 is the mathematical expression of Bayes' theorem.

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} \quad (1)$$

Where, (A) is the occurance probability of the event A. P(A|B) –occurance probability of A when B occurs. P(B|A) –occurance probability of B when A occurs. P(B) –occurrence probability of the event B.

Decision tree (DT) is a tree structural classifier proposed in the 80s of the last century, consisting of roots, branches and leaves. The root is where the tree begins. This represents all databases and then divides into two or more branches. The branches represent the symbols in the database, and the leaves indicate the class symbols. The decision tree is divided into branches based on questions/answers. The decision tree is a non-parametric learning method for classification and regression analysis. The decision tree consists of two stages - learning and forecasting. The model is taught using the training data provided in the learning process and is used to forecast the result of the test data shown in the prognosis stage (Dewan et al., 2010). Note that one of the essential issues in the implementation of the decision tree is the selection of an important attribute for the root node or sub-nodes. Two common attribute selection methods in problem solving exist, namely Information Gain and Gini Index.

Random Forests (RF) was proposed by L.Brayman in the early 2000s based on the decision trees ensemble (Breyman - 2001). The reason for being random is that not all atributes are involved in the construction of an arbitrary tree, but random ones are involved. That is, the RF creates several decision trees organized from a random subset of data. The aim is to increase the accuracy of the classification. As seen, the algorithm is performed in two stages: selection of symbols and classification. The object belongs to the class in which most of the trees vote. That is, the final decision is made on the basis of decisions tree aggregation, and determined by a majority vote. Beside DoS attacks, RF is also used to detect attacks such as Probe, U2R & R2L, and botnets with high accuracy (Farnaaz & Jabbar, 2016). We should note that when working with big data, RF requires a large amount of computing.

K-Nearest Neighbor (KNN) is a machine learning algorithm used to determine a new object in the system. For each new input object, some neighboring objects are identified belonging to one of the classes. The KNN algorithm is a non-parametric method commonly used for classification and regression problems (Zhang, Zhou, 2015; Wauters et al., 2017). In this algorithm, selection of the correct K parameter and the solution of the proximity metrics issue is important. Various distance functions (Euclidean distance, Minkowski, cosine distance) are used in the literature to measure the distance between objects geometrically. When the most commonly used Euclidean distance function (formula 2) is used, the sum of the distances from a new object to k number of objects belonging to this class is calculated. The new object refers to the class with the smallest distance.

$$dist(x, y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (2)$$

Where x and y denote input vectors with m number of traits. We should note that different traits should be normalized before applying the distance function, as they are generally measured in different sizes.

Support Vector Machine (SVM) is one of the most reliable forecasting methods based on the theory of statistical learning, proposed by V. Vapnik and A. Chervonenkis. Binary line classifier is widely used in classification and regression problems. Assume that, the training set consists of (x_i, y_i) elements, where x is a trait vector, and y is the corresponding class sign: $y \in \{+1, -1\}$. We should find such a hypermode which separates $y_i = 1$ $y_i = -1$ points and passes at the maximum distance from the nearest points of the training set. $w \cdot x - b = 0$ describes a hypermode that groups the traits space into classes. Here, w is a

normal vector of hypermode. If dot product of w vector with x_i is greater than the permitted value of b, $w \cdot x_i > b \Rightarrow y_i = 1$, then the point refers to the first category, if its smaller, $w \cdot x_i < b \Rightarrow y_i = -1$, then it refers to second one (Carlos et al., 2012; Erfani et al., 2016).

Multilayer Perceptron (MLP). MLP refers to the class of multilayer (more than one) perceptron artificial neural networks. Obviously, artificial neural networks are based on the simulation of biological neural networks. Neurons are a key computing component in these networks. The MLP consists of at least three layers: input layer, hidden layer, and output layer. Except the input layer, each node is a neuron that uses a nonlinear activation function. The input layer includes input trait vectors corresponding to one element of the trait vector of each neuron.

Ensemble learning algorithms. Though it is difficult to follow the history of ensemble models, it has been studied by researchers from various fields since the 1990s. Ensemble learning algorithms are meta algorithms that integrate multiple machine learning algorithms into a predictive model to reduce variance or improve predictions (Zhou, 2015). In the context of training theory, a meta-algorithm aims to create a new "strong algorithm" by combining the results of individual "weak algorithms" and giving weight to each. Today, there exist meta-algorithm samples such as "multiplicative weights", "weighted majority", "boosting", "bagging", "ensemble averaging", "voting", etc. The ensemble is more accurate than a separate training and is more practical in hiding the weaknesses of individual models. Figure 1 illustrates the general architecture of the ensemble algorithm.

As seen from the diagram, the ensemble model uses several machine learning algorithms together to provide more accurate predictions for the database. Separate algorithms are taught on the database, and each algorithm gives a separate forecast. The predictions of the algorithms are combined in an ensemble model and the final prediction is determined based on the sounds.

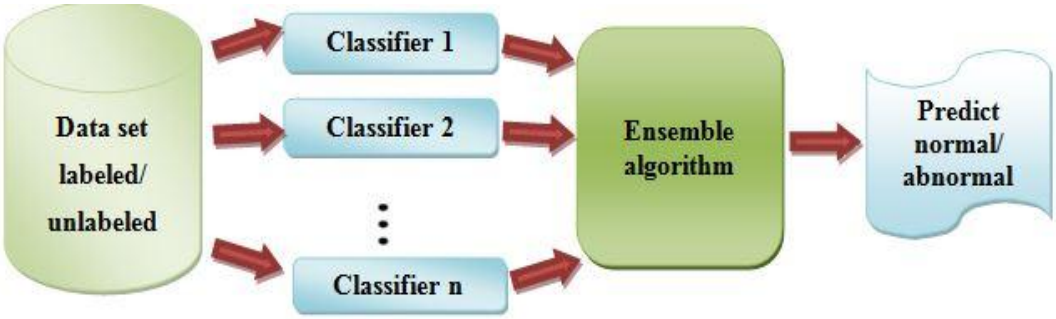

Figure 1. General architectural scheme of the ensemble algorithm

Experiments

This section discusses experiments and obtained results. The experiments are performed on a computer with Windows 8.1 (64bit) operating system, Intel (R) Core (TM) i7-4510 processor, 8GB of RAM. In this work, an experimental study of the proposed classification ensemble algorithm for the detection of DoS attacks on network traffic data is performed in the WEKA environment and over the NSL-KDD database.

NSL-KDD database. NSL-KDD database *train.arff* and *test.arff* files are used to detect anomalies in network traffic. NSL-KDD stores 125,973 recording samples in the training database and 22,544 in the test database. Each record contains 42 traits. The last trait is labeled as "anomalous" or "normal" for each record (NSL-KDD data set). In general, we can get a list of NSL-KDD traits and comprehensive information about them from (NSL-KDD data set; Dhanabal, Shantharajah, 2015; Akbar, 2010).

Evaluating the detection performance of classifiers is an essential issue in machine

learning. In the evaluation of detection performance, the metrics such as precision, recall, false positive rate (FPR), true positive rate (TP), f-measure, and accuracy are used.

Confusion Matrix (table 1) is one of the easiest and simplest approaches used to evaluate the accuracy and precision of a model (Fawcett, 2006; Holz, 2008; Gu, et al., 2006; Aliguliyev, Hajirahimova, 2019).

Table 1. Confusion matrix

		Actual	
		Positive	Negative
Predicted	Positive	True Positive TP	False Positive FP
	Negative	False Negative FN	True Negative TN

- true positive rate (TPR)

$$TPR = \frac{TP}{TP + FN} \quad (3),$$

- true negative rate (TNR) –

$$TNR = \frac{TN}{TN + FP} \quad (4),$$

- false positive rate (FPR)

$$FPR = \frac{FP}{(FP + TN)} \quad (5),$$

- false negative rate (FNR)

$$FNR = \frac{FN}{(FN + TP)} \quad (6).$$

The confusion matrix and the detection metrics of the classification algorithms based on it are calculated using formulas (7-10) (Huang, Charles, 2005):

$$precision = \frac{TP}{(TP + FP)} \quad (7),$$

$$recall = \frac{TP}{(TP + FP)} \quad (8),$$

$$F - measure = \frac{2*(precision*recall)}{(precision+recall)} \quad (9),$$

$$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (10).$$

The Naive Bayes, Decision tree, Random forests, SVM, Multilayer perceptron and KNN classifiers used to create the proposed machine learning ensemble model for detecting DoS attacks in network traffic data are tested in the WEKA software environment. Table 2 illustarted a comparison of test results.

Table 2. Comparison of the classifiers results

Methods	Accuracy	TP rate	FPRate	Prec-sion	Recall	F-meas.	ROC Area
DT	0,815	0,696	0,027	0,971	0,696	0,811	0,84
NB	0,761	0,633	0,069	0,924	0,633	0,751	0,917
KNN	0,793	0,666	0,038	0,959	0,666	0,786	0,814
RF	0,804	0,677	0,027	0,971	0,677	0,798	0,959
SVM	0,759	0,635	0,076	0,917	0,635	0,75	0,779
MLP	0,736	0,592	0,071	0,917	0,592	0,719	0,814
Ensemble (Vote)	**0,977**	**0,953**	**0,001**	**0,999**	**0,952**	**0,975**	**0,981**

As seen from Table 2, the proposed ensemble algorithm performs best results in all metrics in detecting DoS attacks. Thus, the highest result on the accuracy metric is 97.7% and the lowest on the false precision (FP) is 0.1%. The lowest result in the accuracy metric is perfomed by the MLP algorithm.

Figure 2 visually presents the results of the classifiers applied to the test data and the proposed algorithm for the various metrics. As seen from the diagram, the proposed ensemble algorithm performs highest result in TP metrics (0.953), and the lowest results are performed by MLP, NB and SVM algorithms. DT and RF algorithms perform best result (2.7%) following the proposed algorithm FP metrics. The worst result (7.6%) was performed by the SVM algorithm.

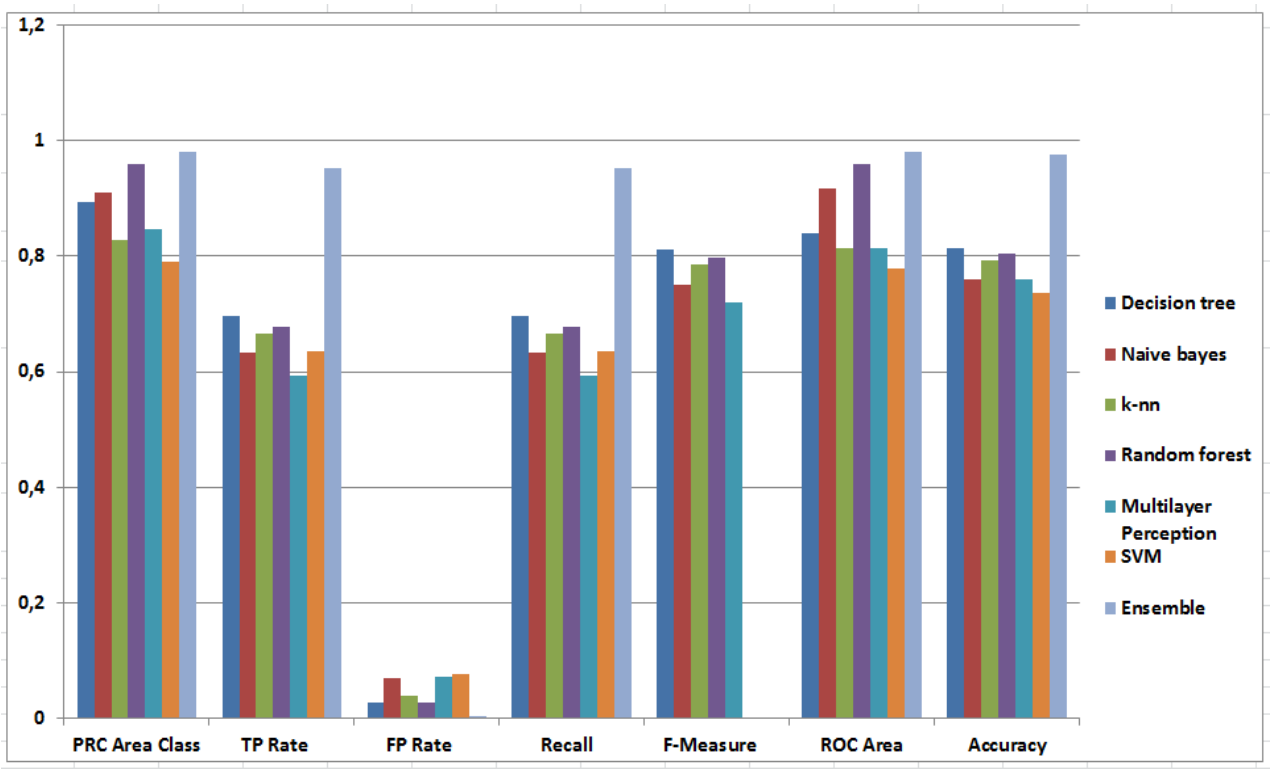

Figure 2. Comparison of the classifiers results on test data

The most important evaluation metrics, such as the AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve is used to test or visualize the effectiveness of a two- or more-class classifier model (Huang, Charles, 2005; Ling et al., 2005; Fawcett, 2006). That is, the ROC curve is a visualization method of the classifiers results. ROC diagrams describe the compromise between the degree of correct determination of classifiers and the degree of false alarm (Figure 3). The diagram demonstrates true positive cases on the ordinate axis and false positive cases on the abscissa axis.

We observe from the description of the ROC curve in Figure 3 that this curve is very close to 1, considering the ensemble algorithm with the highest value (0.981) for all methods applied in the database. The SVM performs the lowest result (0.779) on the ROC Area metric.

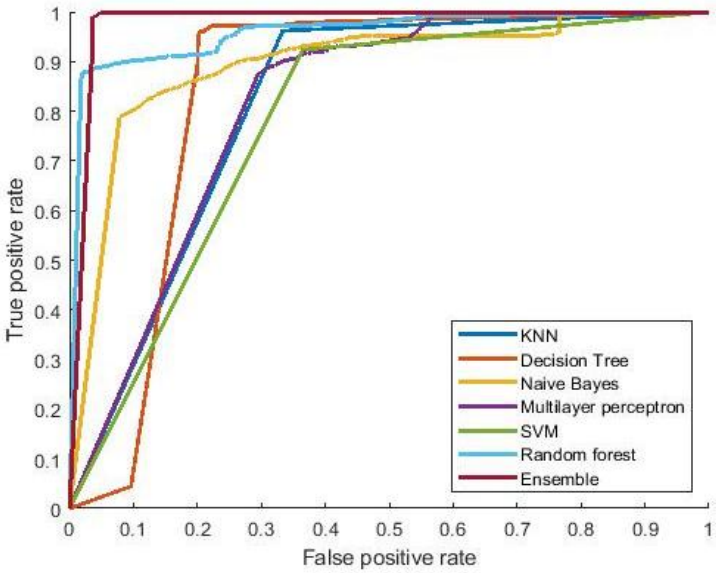

Figure 3. ROC curve created on the basis of test data

Conclusion

Wide usage of network services and applications in state and private organizations requires adequate security measures against network and computer intrusions. The study developed an ensemble model consisting of basic training algorithms such as Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, Multilayer Perceptron for the detection and prevention of network intrusions known as DoS attacks on computer networks. It aimed to create a strong classifier with a lower probability of learning error than a weak classifier with a relatively high probability of learning error taken separately in the detection of anomalous traffic. The ensemble model is noticeable with higher accuracy. The results of our tests showed that the ensemble-based approach performed higher accuracy than other basic training classifiers in the detection of network traffic anomalies, i.e., DoS attacks. The obtained results showed that 99.9% precision, 95.2% recall and 97.7% accuracy and 0.981 AUC can be accepted as a reliable approach to detecting DoS attacks.

Though the proposed approach provided sufficient results, but we can achieve higher efficiency by applying deep learning and optimization strategies, etc. Moreover, the real time use of real data is significant. This is the future research area of our investigation.

Reference

Abdulhammed R., Faezipour M., Abuzneid A., and AbuMallouh A. (2019). Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network trafc. IEEE Sensors Lett., Jan. 2019 3(1), pp. 1-4. https://doi.org/10.1109/LSENS.2018.2879990

Agarwal B., Mittal N. (2012). Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques. Procedia Technology, 6. pp. 996-1003. http://dx.doi.org/10.1016/j.protcy.2012.10.121

Aggarwal CC, Philip SY. (2005). An effective and efficient algorithm for high-dimensional outlier detection. VLDB J. 14(2), pp. 211–221. https://doi.org/10.1007/s00778-004-0125-5

Agrawal S, Agrawal J. (2015). Survey on anomaly detection using data mining techniques. Procedia Computer Science, 60, pp. 708–713. https://doi.org/10.1016/j.procs.2015.08.220

Akbar S., Nageswara R. K., Chandulal J. A. (2010). Intrusion detection system methodologies based on data analysis. International Journal of Computer Applications. 5(2), pp. 10–20. http://dx.doi.org/10.5120/892-1266

Akoglu L., Tong H., Koutra D. (2015). Graph based anomaly detection and description: a survey. Data Mining Knowl Discov. 29(3), pp. 626–88. https://doi.org/10.1007/s10618-014-0365-y

Alguliyev R., Aliguliyev R., İmamverdiyev Y. N., Sukhostat L. (2018). Weighted Clustering for Anomaly Detection in Big Data. Statistics, Optimization & Information Computing, 6(2), pp. 178–188. https://doi.org/10.19139/soic.v6i2.404

Alguliyev R. M., Aliguliyev R. M., Imamverdiyev Y. N. and Sukhostat L. V. (2017). An anomaly detection based on optimization. International Journal of Intelligent Systems and Applications, 9(12), pp. 87-96. DOI: 10.5815/ijisa.2017.12.08

Aliguliyev R. M., Hajirahimova M. Sh. (2019). Classification Ensemble Based Anomaly Detection in Network Traffic. Review of Computer Engineering Research, vol. 6(1), pp. 12-23. DOI:10.18488/journal.76.2019.61.12.23

Almeida V. A., Doneda D., & de Souza Abreu J. (2017). Cyberwarfare and Digital Governance. IEEE Internet Computing. 21(2), pp. 68-71. https://doi.org/10.1109/MIC.2017.23

Antal B. and Hajdu A. (2014). An ensemble-based system for automatic screening of diabetic retinopathy. Knowl.-Based Syst., vol. 60, pp. 20–27. https://doi.org/10.1016/j.knosys.2013.12.023

Ariyaluran R. A. H., et al. (2019). Real-time big data processing for anomaly detection: A Survey. International Journal of Information Management, vol.45, pp. 289-307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006

Bellman R. (2013). Dynamic programming. Chelmsford: Courier Corporation.

Breiman, L. (2001). Random forests. Machine Learning. 45(1), pp. 5–32. https://doi.org/10.1023/A:1010933404324

Buczak A. L., & Guven E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), pp. 1153-1176. https://doi.org/10.1109/comst.2015.2494502

Camacho J., Macia-Fernandez G., Diaz-Verdejo J., Garcia-Teodoro P. (2014). Tackling the big data 4 vs for anomaly detection. In: Computer communications workshops (INFOCOM WKSHPS), 2014 IEEE conference on. IEEE. pp. 500–505. https://doi.org/10.1177/1550147720921309

Carlos A. Catania, Facundo Bromberg, Carlos Garcна Garino (2012). An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection Expert Systems with Applications. 39(2), pp. 1822–1829. https://doi.org/10.1016/j.eswa.2011.08.068

Chandola V., Banerjee A., Kumar V. (2009). Anomaly detection: a survey. ACM Computing Surveys, 41(3), pp. 71-97. https://doi.org/10.1145/1541880.1541882

Chaudhary K., Yadav J., & Mallick B. (2012). A review of fraud detection techniques: Credit card. International Journal of Computers and Applications, 45(1), pp. 39–44. DOI: 10.5120/6748-8991

Dash M., & Ng W. (2010). Outlier detection in transactional data. Intelligent Data Analysis, 14(3), pp. 283–298. DOI: 10.3233/ida-2010-0422

Denning D. E. (1987). An Intrusion-Detection Model. IEEE transactions on software engineering, 13(2), pp. 222 – 232. https://doi.org/10.1109/TSE.1987.232894

Dewan Md. F., Nouria Harbi, and Mohammad Zahidur Rahman (2010). Combining naive bayes and decision tree for adaptive intrusion detection. International Journal of Network Security & Its Applications (IJNSA), 2(2), pp. 1-12. DOI: 10.5121/ijnsa.2010.2202

Dhanabal, Shantharajah S. P. (2015). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. International Journal of Advanced Research in Computer and Communication Engineering 9(4) (2015) pp. 446–452. http://dx.doi.org/10.4236/jcc.2016.44008

Dua S., Du X. (2011). Data mining and machine learning in cybersecurity. Boca Raton, FL, CRC Press, 256 p. https://doi.org/10.1201/b10867

Erfani S. M., Rajasegarar S., Karunasekera S., Leckie C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. Pattern Recogn. 58 pp. 121–34. https://doi.org/10.1016/j.patcog.2016.03.028

Farnaaz, N., & Jabbar, M. (2016). Random Forest Modeling for Network Intrusion Detection System. Procedia Computer Science, 89, pp. 213-217. https://doi.org/10.1016/j.procs.2016.06.047

Fawcett T. (2006). An Introduction to ROC Analysis. Pattern Recognition Letters, 27 (8), pp. 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fujimaki R., Yairi T., Machida K. (2005). An approach to spacecraft anomaly detection problem using kernel feature space. In Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM Press, New York, NY, USA, pp. 401–410. https://doi.org/10.1145/1081870.1081917

Garofalo M. (2017). Big data analytics for Flow-based anomaly detection in high-speed networks. PhD Thesis. http://dx.doi.org/10.6093/UNINA/FEDOA/11617

Global - VNI Complete Forecast Highlights https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2021_Forecast_Highlights.pdf.

Gogoi P., Bhattacharyya D. K., Borah B., and Kalita J. K. (2011). A survey of outlier detection methods in network anomaly identification. The Computer Journal, 54(4), pp. 570-588. https://doi.org/10.1093/comjnl/bxr026

Goldstein M, Uchida S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS ONE. 11(4). https://doi.org/10.1371/journal.pone.0152173

Gu G., Fogla P., Dagon D., Lee W., Skori B. (2006). Measuring intrusion detection capability: An information-theoretic approach. Proceedings of the ACM Symposium on Information, Computer and Communications Security, pp. 90–101. https://doi.org/10.1145/1128817.1128834

Gupta M., Gao J., Aggarwal C.C., Han J. (2014). Outlier detection for temporal data: a survey. IEEE Trans Knowl Data Eng. 26(9), pp. 2250–67. https://doi.org/10.1109/TKDE.2013.184

Makrufa S. Hajirahimova (2016). Big data technologies and information security challenges. Problems Information Technologies, №1, pp. 41–46. http://dx.doi.org/10.25045/jpit.v07.i1.06

He H., Wang J., Graco W., and Hawkins S. (1997). Application of neural networks to detection of medical fraud. Expert Systems with Applications 13(4), pp. 329–336. https://doi.org/10.1016/S0957-4174(97)00045-6

Heydari A. et al. (2015). Detection of review spam: a survey. Expert Syst Appl, 42(7) pp. 3634–42. https://doi.org/10.1016/j.eswa.2014.12.029

Hodge V., Austin J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), pp. 85–126. https://doi.org/10.1007/s10462-004-4304-y

Holz T. (2008). Security measurements and metrics for networks. Lecture Notes in Computer Science, vol. 4909, pp. 157–165. http://dx.doi.org/10.4236/ijcns.2013.61004

Husejinović A. (2020). Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers. Periodicals of Engineering and Natural Sciences 8(1), pp. 1-5. http://pen.ius.edu.ba

Xuan S., Liu G., Li Z., Zheng L., Wang S., & Jiang C. (2018). Random forest for credit card fraud detection. IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). https://doi.org/10.1109/ICNSC.2018.8361343

Imamverdiyev Y., Abdullayeva F. (2018). Deep Learning Method for Denial of Service Attack Detection Based on Restricted Boltzmann Machine Big Data, 6(2), pp. 159-169. https://doi.org/10.1089/big.2018.0023

Jin Huang and Charles, Ling X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms IEEE transactions on knowledge and data engineering, 17(3), pp. 299-310. https://doi.org/10.1109/TKDE.2005.50

KDD data set, 1999. http://kdd.ics.uci.edu/databases/kddcup99

Kim G., Lee S., and Kim S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Syst. Appl., 41(4), pp. 1690-1700. https://doi.org/10.1016/j.eswa.2013.08.066

Kumari R., Sheetanshu, Singh M. K., Jha R. & Singh N. K. (2016). Anomaly detection in network traffic using K-mean clustering. 3rd International Conference on Recent Advances in Information Technology (RAIT). https://doi.org/10.1109/RAIT.2016.7507933

Lee W., Stolfo S. J., Mok K. W. (2000). Adaptive intrusion detection: A data mining approach. Artificial Intelligence Review, 14(6), pp. 533-567. https://doi.org/10.1023/A:1006624031083

Marnerides A. K., Spachos P., Chatzimisios P., and Mauthe A. U. (2015). Malware detection in the cloud under Ensemble Empirical Mode Decomposition. In 2015 Int. Conf. Comput. Netw. Commun. IEEE., pp. 82–88. https://doi.org/10.4018/IJESMA.2018070104

McHugh J. (2000). Testing Intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory. ACM Transactions on Information and System Security, 3(4), pp. 262–294. https://doi.org/10.1145/382912.382923

Mukherjee B., Heberline L. T., & Levitt K. (1994). Network instruction detection. IEEE Network, 8, pp. 26–41. https://doi.org/10.1007/978-0-387-33112-6_8

Münz G., Li S., Carle G. (2007). Traffic anomaly detection using k-means clustering. In: GI/ITG Workshop MMBnet. pp. 13-14. DOI:10.1.1.323.6870

Nassif A. B. et al. (2021). Machine Learning for Anomaly Detection: A Systematic Review. IEEE Access, vol.9, pp. 78658- 78700. https://doi.org/10.1109/access.2021.3083060

NSL-KDD data set for network-based intrusion detection systems [Electronic resource]. 2017. Access mode: http://nsl.cs.unb.ca/NSL-KDD/

Patcha A., Park J.M., (2007). An overview of anomaly detection techniques: existing solutions and latest technological trends. Comput Netw. 51(12), pp.3448–3470. http://dx.doi.org/10.1016%2Fj.comnet.2007.02.001

Phua C., Lee V., Smith-Miles K., & Gayler R. (2010). A comprehensive survey of data miningbased fraud detection, Research Computing Research Repository https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1016%2Fj.chb.2012.01.002&v=67e7929e

Raguseo E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. International Journal of Information Management, 38(1), pp. 187-195. https://doi.org/10.1016/j.ijinfomgt.2017.07.008

Rehman M. H., Liew C. S., Abbas A., Jayaraman P. P., Wah T. Y., & Khan S. U. (2016). Big data reduction methods: a survey. Data Science and Engineering, 1(4), pp. 265-284. https://doi.org/10.1007/s41019-016-0022-0

Samuel A. L. (1959). Some studies in Machine Learning using the game of checkers. IBM Journal of research and development, 3(3), pp. 210–229. https://doi.org/10.1147/rd.33.0210

Saneja B., Rani R. (2017). An efficient approach for outlier detection in big sensor data of health care. International Journal of Communication Systems, 30(17), pp. 1-10. https://doi.org/10.1002/dac.3352

Schlegl T., Seeböck P., Waldstein S. M., Schmidt-Erfurth U., and Langs G. (2017). Unsupervised Anomaly Detection With Generative Adversarial Networks to Guide Marker Discovery. International Conference on Information Processing in Medical Imaging, pp. 146-157. https://doi.org/10.1007/978-3-319-59050-9_12

Shon T., Moon J. (2007). A hybrid machine learning approach to network anomaly detection. Information Sciences, 177(18), pp. 3799-3821. https://doi.org/10.1016/j.ins.2007.03.025

Srikanth T., Philip B., Jiong J. et al. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. Journal of Big Data, 7(42), pp. 1-30. https://doi.org/10.1186/s40537-020-00320-x

Tsai C. F., Hsu Y. F., Lin C. Y., Lin W. Y. (2009). Intrusion detection by machine learning: A review. Expert Syst. Appl., 36(10), pp. 11994-12000. https://doi.org/10.1016/j.eswa.2009.05.029

Varian I. (2020). IMRT (Intensity Modulated Radiation Therapy). 26 June 2020. https://patient.varian.com/en/treatments/radiation-therapy/treatment-techniques

Wang C., Zhao Z., Gong L., Zhu L., Liu Z., & Cheng X. (2018). A Distributed Anomaly Detection System for In-Vehicle Network Using HTM. IEEE ACCESS, 6, pp. 9091-9098. https://doi.org/10.3390/s20143934

Wang L. & Jones R. (2017). Big data analytics for network intrusion detection: A survey. International Journal of Networks and Communications, 7(1), pp. 24-31 doi: 10.5923/j.ijnc.20170701.03

Wauters, M., & Vanhoucke, M. (2017). A Nearest Neighbour extension to project duration forecasting with Artificial Intelligence. European Journal of Operational Research, 259(3), pp. 1097-1111. https://doi.org/10.1016/j.ejor.2016.11.018

Wei Y. et al. (2019). MSD-Kmeans: A Novel Algorithm for Efficient Detection of Global and Local Outliers, pp. 1-12. https://doi.org/10.1145/3459930.3469523

Yang T. et al. (2016). Improve the Prediction Accuracy of Naive Bayes Classifier with Association Rule Mining. IEEE 2nd International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance, and Smart Computing, IEEE International Conference on Intelligent Data and Security, pp. 129-133. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2016.38

Zhang J., Zulkernine M., and Haque A. (2008). Random-Forests-Based Network Intrusion Detection Systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 38(5), pp. 649–659. https://doi.org/10.1109/TSMCC.2008.923876

Zhang M.L., Zhou Z.H., (2005). A k-nearest neighbor based algorithm for multi-label classification / Proc. of the International Conference on Granular Computing, pp. 718–721. https://doi.org/10.1109/GRC.2005.1547385

Zhou Z., (2012). Hua Ensemble Methods. Foundations and Algorithms. CRC Press, p.234. https://doi.org/10.1201/b12207