APPLICATION OF MACHINE LEARNING FOR INTRUSION DETECTION SYSTEM

Nitin Pise

Dr. Vishwanath Karad MIT World Peace University, Pune, 411038, India
nitin.pise@mitwpu.edu.in

Abstract

Due to Covid-19 pandemic, the most of the organizations have permitted their employees to work from home. Also, it is every essential to have security at the highest level so that information will flow in the safe and trusted environment between the different organizations. There is always threat of misuses and different intrusions for communication of the data securely over the internet. As more and more people are using online transactions for the different purposes, it is found that the cyber attackers have become more active. Three in four organizations have faced the different cyber-attacks in the year 2020. So, the detection of intrusion is very important. The paper introduces the intrusion detection system and describes its classification. It discusses the different contributions to the literature in literature review section. The paper discusses the application of the different feature selection techniques for reducing the number of features, use of the different classification algorithms for the intrusion detection and it shows how machine learning is used effectively. KDD99 benchmark dataset was used to implement and measure the performance of the system and good results are obtained and the performance of the different classifier algorithms was compared. Tree based classifiers such as J48 and ensemble techniques such as random forest give the best performance on KDD99 dataset.

Keywords: machine learning, intrusion detection system, feature selection, classifiers, attacks

1. Introduction

As per [1], "Computer security is a science as well as an art". It is an art because before using every system, it has to be tested for security. It is a science as it is based on mathematical constructions, proofs and analysis. The field of security is very challenging because technology is changing every day. Thus, we need to secure computer system, hardware devices like switches, routers and computer networks from the hackers by developing new algorithms and protecting the systems from them.

Now everybody uses the word "security" in daily life. Secure means the system is safe. It means that the system is secure. This indicates that the object, system or software is free from attack or hazard. The system needs to be protected from the malicious users or attackers who can harm, may be unintentionally or intentionally. In the network security, the security is to protect the network and only the users with some authorization can access the network. The security is required at the different layers as shown below for making the organization secure.

- Physical security layer: This layer takes care of the physical objects. It includes the access mechanism which allows authorized person to access the different physical devices such as hard disk, pen drive, compact disk or digital video disk (DVD) or the computer system.
- Private security: Using this layer the individual or a group is protected.
- Project security: Here the security is provided to the entire project consisting of design, code, etc.

Everywhere internet is used widely with the different devices and smart phones. In this Covid19 Pandemic, the users are making more digital transactions for their payments and other operations. A number of software are available on internet which are used by the attackers for

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

attacking computer systems, mobile phones and other devices, etc. For this, it is not necessary to have much domain knowledge in the field. Therefore, information security (IS) is a hot research area. IS helps in protecting people from the various attacks and involves creating awareness among the users.

Today Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) are buzzwords. Everybody is talking the use of AI & ML for different applications. AI is a major research area in computer science dealing with making machines intelligent. ML is subarea of AI, which is used for prediction, forecasting based on past data or history. DL is again subarea of ML where artificial neural networks (ANN) with more than one hidden layer are used for processing involved to complete or solve any problem. Neural networks (NN) can be used for intrusion detection. The more details about how intrusions are detected and these systems are covered in the next section.

Intrusion detection system (IDS)

Now a days due to increased use of Internet and company networks, the network traffic is increasing day by day. Access to company networks should be given only to the authorized users, so it is necessary to detect unauthorized entities or intruders, who are trying to access company networks, this is called intrusion detection. There are a lot of methods or techniques for intrusion detection. Ideally, IDS should detect all attacks which are attempted or intrusions and continue work by stopping the attacks. If the company network is using firewall, IDS has to modify the firewall rules so as to stop the attacks. Thus, IDS protects the system, analyses and predicts the behaviour of users. Then based on the behaviour, it classifies as an attack or intrusion into the system.

Recently, wireless ad hoc sensor network is widely used in civil as well as military jobs. Sensor network is deployed mostly in open as well as unprotected environment, so security is a very important challenge.

2. Related Work

Network traffic analysis (NTA) is used for evaluating the performance and security of network operations and management. This section discusses different machine learning approaches for network traffic analysis. Network traffic is increasing day by day, also there is a lot of development in Artificial Intelligence (AI) which requires new ways for detecting intrusions, classify internet traffic then analyse the behaviour of malware. A survey of techniques used in the network traffic analysis is presented in this section.

Network traffic analysis involves the process of interception, recording and analysing network traffic communication patterns for detecting and responding to security threats [5]. [5] describes a review of network traffic analysis and how the prediction techniques are used. [15] gives a brief overview and comparison among some existing ML approaches used in traffic analysis. In today's world, e-commerce, banking and business related highly confidential and valuable data is communicated over the network. So, it is necessary to analyse the network traffic to provide proper information security. Network traffic analysis and prediction is a practical approach, in which prevention of security breaches is required so the network is monitored continuously. It is used to classify whether network packets are normal or malicious. These techniques try to divide and allocate the network resources as per the forecasted traffic.

The predictability of network traffic is important in dynamic bandwidth allocation. network security, planning of network, predictive congestion control and so on. Two types of predictions namely long and short period predictions can be done. A detailed forecasting of traffic models for evaluating future capacity requirements is obtained in long term prediction. It allows us to do minute to minute planning and take better decisions. Short period prediction which varies from few milli seconds to few minutes is used for dynamic resource allocation. It improves quality of service (QoS) mechanisms, packet routing, congestion control, and optimal management of the available resources. Several different techniques are used for network traffic analysis which includes time series models, modern data mining techniques, soft computing approaches and neural networks. A short description about network traffic analysis is

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

given and then several available network analysis methods are reviewed in the next section. Section three reviews various network traffic prediction methods. In the last section, conclusion is given.

A generic framework for NTA involves pre-processing, actual analysis and observations to find out patterns from the network data. Fig.1 shows these main phases of NTA. The detailed description of these three phases is provided in the subsequent subsections.

Evaluation Metrics or performance measures

Different machine learning techniques along with the different pre-processing techniques used are described in [5]. Many different performance measures are used for analysis of the results. Accuracy, the detection rate, false positive rate, accuracy and testing time are used for measuring the performance of ML algorithms on the different datasets. Here the metrics based on confusion matrix can be used. Thus, the analysis of network traffic and prediction is an important research field in information security domain.

Digital transformation has become the priority agenda for most governments and this vision cannot be achieved without a fully-developed cybersecurity infrastructure in place. The Cyber Security adopts the Machine Learning which removes the limitations of algorithms which are based on traditional rules. The authors in [10] describe how several Machine Learning algorithms can overcome the popular issues faced in cyber security.

[18] describes the different categories of cybersecurity threats and the different challenges posed by them. How Bigdata tools are used for solving these challenges, how they can detect and prevent attacks in real-time is also explained. How Apache Spark, Services Fortscale, IBM Security QRadar are used as Bigdata tools is explained. QRadar helps for revealing hidden relations resided in large amounts of security data. It uses analytics for reducing billions of security events to a set of prioritized incidents. Fortscale provides a Big data solution against APT attacks while the victim organization is not aware of the attack.

Available Benchmark Datasets

Testing ad evaluating using standard benchmark datasets is very important in network traffic analysis. A lot of standard datasets are available in the literature. E.g., the KDD cup dataset is used for intrusion detection which consists of 4,900,000 training instances and 41 features. The test dataset consists of 24 training and testing attacks. It contains 38 types of attack which are classified into four types: Probing, Denial of Service (DoS), Remote to User (R2L), and User to Root (U2R).

Other datasets are NSL-KDD dataset, CAIDA dataset, Waikto dataset, Berkeley Lab dataset, ACM SIGCOMM'01 dataset, ISCX data set, etc.

Pre-processing

The real-world data to be used for network traffic analysis is incomplete, inconsistent. It may contain missing, duplicate values or noise. So pre-processing is very important step which should be carried out before actual network traffic analysis. This is because insights or output will depend on the input to the system. If the input is noisy, incomplete or inconsistent, it will not give good results or insights from the data.

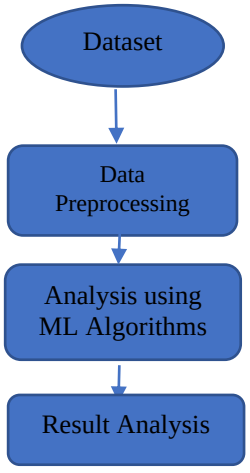

Figure 1. General framework of Network Traffic Analysis

[11] describes the different machine learning algorithms for Intrusion Detection. The authors describe the use of bigdata processing tool called as Apache Spark [13] and the four different algorithms. They have algorithms such as Support Vector Machines (SVM), Naïve Bayes and tree-

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

based machine learning algorithms decision tree and random forest are tested using Apache for detecting intrusion activities in the network traffic. A recent public dataset for intrusion detection, UNSW-NB15 with all 42 features is used for experimental work and the authors conclude that random forest gives the best performance when the different performance measures are compared such as accuracy and prediction time. The UNSW-NB15 data set was created at the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) using the AXIA Perfect Storm tool to create a hybrid of modern normal and abnormal network traffic [12].

[16] describes key literature surveys on machine learning and deep learning methods for network analysis of intrusion detection. The authors provide a brief tutorial description of each ML/ DL method in the survey report. The paper discusses the use of machine learning methods such as support vector machine, k-nearest neighbor, decision tree, deep belief network (DBN). DBN is a probabilistic generative model which consists of multiple layers of stochastic and hidden variables. [10] also shows how the deep learning techniques such as convolutional neural networks (CNN), recurrent neural networks (RNN) and LSTM RNN are used for classifying network traffic which provide cyber security.

[17] shows visualization of the knowledge map of Artificial Intelligence applications in cyber security. Emerging trends of AI applications in cyber security have been identified by the authors in the paper. According to a 2014 Consumer News and Business Channel (CNBC) report [17], "the global economy suffers a loss of US $400 billion every year due to cyber-crimes. The human factor is the weakest link and is the main reason for cyber security failure." To address this weakness, automated systems, such as artificial intelligence applications are used in cyber security. A visual analysis of the hotspots and emerging trends regarding the use of AI in cyber security applications through scientometric techniques is also conducted by the authors.

[19] discusses intrusion detection systems that are adapted to allow routers, network defense and systems to detect and report malicious network traffic signal. The paper also presents an innovative approach known as SCDNN, which combines spectral clustering (SC) and DNN algorithm carried out on six KDD-CUP99 and NSL-KDD datasets and a sensor network dataset to test performance of the model. [20] comprehensively review and compare the key previous deep learning-focused cybersecurity surveys. Through an extensive review, the survey provides a novel fine-grained taxonomy that categorizes the current state-of-the-art deep learning-based IDSs with respect to different facets, including input data, detection, deployment, and evaluation strategies. Adaptation a SAE and denoising auto-encoders (DAE) for the deep neural network is suggested so as to overcome the limitations of the traditional auto-encoders. Finally, the authors identify open research challenges, and recommend future research directions for deep learning-based IDSs.

Deep learning is now buzzword in machine learning. Deep learning is composed of multiple layers of hidden layers for learning. The layers are connected through neurons, which act as the processing unit in the learning process. Like neural network, deep learning networks must learn first before they are used for the application in intrusion detection. Deep learning can cope with very large data and has been widely used in various fields such as face recognition, speech processing, etc. Therefore, researchers have used deep learning methods for intrusion detection.

[21] presents the global architecture of IDS as well as a few commercially available tools. Also, research directions to improve IDS's performances and particularly the application of Neural Networks to Intrusion Detection are recommended in the paper. Most IDS commercial tools refer to the misuse detection model, and vendors must update signatures of intrusions continuously.

A guidance document on Intrusion Detection Systems is available from National Institute of Standards and Technology (NIST) organization [2]. Figure 2 shows intrusion detection system which consists of the different components such as firewall, router which are used to connect internal computer network with the internet.

Intrusion Detection Systems can be classified into three categories:

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

· host-based IDS, evaluate information found on a single or multiple host systems, including contents of operating systems, system and application files.
·network-based IDS, evaluate information captured from network communications, analyze the stream of packets which are traveling across the network. Packets are captured through a set of sensors.
· vulnerability-assessment IDS, detect vulnerabilities on internal networks and firewalls.

IDS based on anomaly detection model can detect symptoms of attacks without specifying model of attacks, but their drawback is that they are very sensitive to false alarms. This should be avoided. Intrusion detection system sits between a set of networked users and firewall and router which connects the users with the outside world through Internet. IDS detect anomaly activities and prevents from attacks from unauthorized or unauthenticated persons. Another variation of IDS is shown in figure 3 where trained model of bigdata dataset and feature extraction are used to provide anomaly-based IDS model.

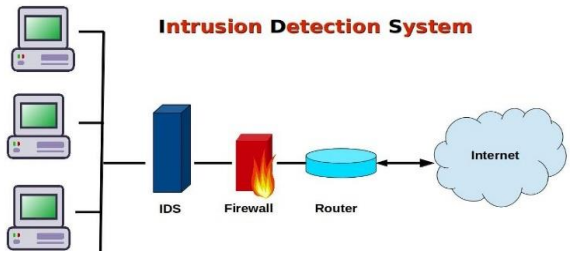

Figure 2. One type of Intrusion Detection System, Image credits to www.peerlyst.com

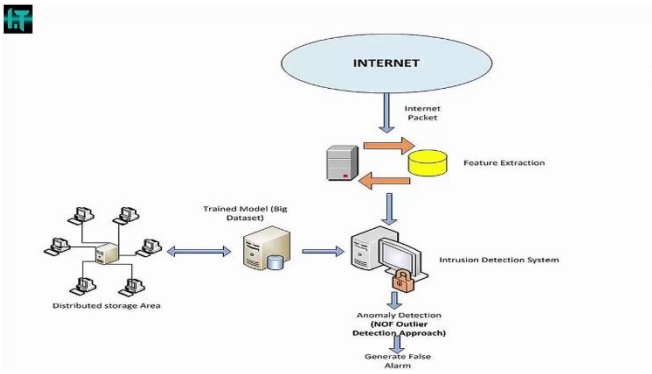

Figure 3. Second type of Intrusion Detection System, Image credits to www.youtube.com

IDS can be classified as shown in figure 4 based on analysis method used and the source of data. Based on analysis method, we classify into misuse IDS and anomaly IDS. Host based IDS and network-based IDS is another categorization of IDS based on source of data.

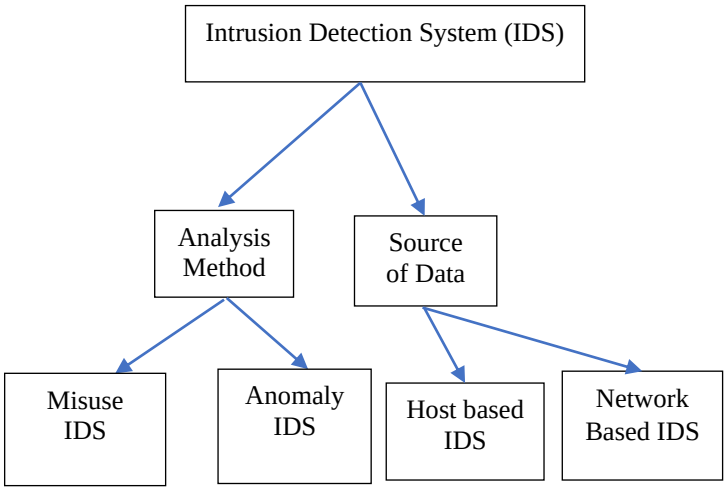

Figure 4. Classification of Intrusion Detection Systems

Two primary models are used for analyzing events so as to detect attacks:
· misuse detection model: IDS detect intrusions by looking for activity corresponding to known signatures of intrusions or vulnerabilities.
· anomaly detection model: IDS detect intrusions by searching for abnormal network traffic.

Commercially available tools

A Jackson [22] of Los Alamos National Laboratory has written a complete survey of IDS products. Characteristics for each of the seventeen products are studied according to nine major features which are as follows: suitability, flexibility, protection, interoperability, comprehensiveness, event management, active response, support.

Commercially available Intrusion Detection tools today are surveyed in [23]. Examples of IDS tools are classified according to the three models: host-based, network-based and vulnerability-assessment tools are also presented. [24] proposes a hybrid intrusion detection system using Naïve Bayes and decision tree classifier.

3. Proposed Work

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

The objective of the proposed work is to improve accuracy and other performance measures such as precision, recall for detecting intrusions. The proposed work consists of the steps such as dataset selection, feature selection and application of machine learning algorithms or classifiers and then comparison of results for the different classifiers.

KDD-99 dataset is used which is having total 42 attributes or features including class. As KDD-99 dataset is widely used dataset in the literature of intrusion detection systems, the proposed system also uses the same benchmark dataset. The results are shown as follows after applying Best First search method. Greedy stepwise and attribute ranking methods are also used for feature selection. As all forty-one features are not important for selecting the class and if all features are used, it takes a lot of time for processing. So, feature selection is an important step before applying the different machine learning algorithms for classification.

Four different classifiers such as ZeroR, J48, Naïve Bayes (NB), and Random Forest are used in the proposed work. Naïve Bayes is a statistical classifier, it performs probabilistic prediction, i.e., predicts class membership probabilities. It is based on Bayes' Theorem. A simple Bayesian classifier, naïve Bayesian classifier, has comparable performance with decision tree and selected neural network classifiers. It is easy for implementation and gives good results in the most of the cases. So Naïve Bayes is used in the proposed wok. J48 [26], and random forest [27] are tree-based classifiers. Random forest is an ensemble-based classifier. Each classifier in the ensemble is a decision tree classifier which is generated using a random selection of attributes at each node to determine the split. Each tree votes and the most popular class is returned during classification. Random forest classifier is more robust to errors and outliers. It gives more accuracy as compared to a single classifier. So random forest is used in the experimental work. The decision trees produced by J48 is utilized for classification. At every node of the tree, J48 selects the attribute of the data which most effectively splits its arrangement of tests into subsets improved in one class or the other. The splitting criterion used in J48 is the standardized information gain. J48 helps to make accurate predictions as well as it explains the patterns in it. It deals with the problem of missing values, numeric attributes, etc. So, it is considered for experimentation in the proposed work. ZeroR is the simplest, rule-based classifier and predicts the majority class. ZeroR is a baseline classifier, hence it can be used for comparing performance with the other classifiers. Four different classifiers used in the experimental work are from three different category of classifiers such as rule-based, tree based and statistical or probabilistic classifiers. The support vector machine takes a lot of time for training, so it is not used in the experimental work. The study can be extended to include other classifiers. 10-fold cross validation mode was used for evaluation of the results. Table 1 shows 30 features selected from 42 features of KDD99 dataset.

Table 1. 30 out of 42 features in KDD99 Dataset

Feature no.	Attribute	Feature no.	Attribute
1	duration	16	num_root
2	protocol_type	17	num_file_creations
3	service	18	num_shells
4	**flag**	19	num_access_files
5	**src_bytes**	20	num_outbound_cmds
6	**dst_bytes**	21	is_host_login
7	land	22	is_guest_login

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

8	wrong_fragment	23	count
9	urgent	24	srv_count
10	hot	25	serror_rate
11	num_failed_logins	**26**	**srv_serror_rate**
12	**logged_in**	27	rerror_rate
13	num_compromised	28	srv_rerror_rate
14	root_shell	29	same_srv_rate
15	su_attempted	**30**	**diff_srv_rate**

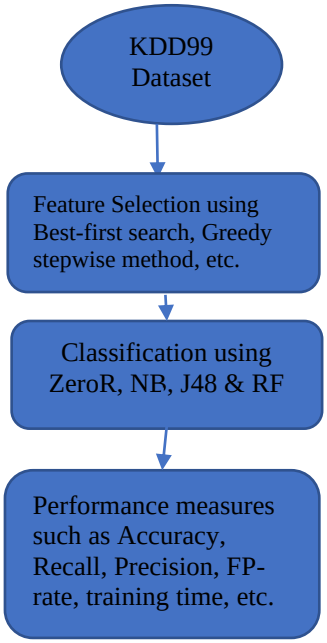

Figure 5. The Flow of Proposed Work

The tasks such as dataset collection, preprocessing, application of machine learning algorithms and result evaluation described in detail as above are shown in figure 5.

Performance measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) are used for comparing the performance of four different classifiers. Also, the classification measures such as training time, accuracy, precision, recall and F-measures are considered for the performance comparison. False Positive Rate (FP-rate) is also considered as it involves intrusion detection system and it should be lower.

IV. Results and Discussion

The results after applying the Best first method is shown as follows:

Evaluation mode: evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 398
Merit of best subset found: 0.568
Selected attributes: 4,5,6,12,26,30: 6
flag
src_bytes
dst_bytes
logged_in
srv_serror_rate
diff_srv_rate

The best-first search and greedy stepwise methods give the same result. i. e. the six features are selected by both the methods which are shown in bold in Table 1. Whereas the third method, i. e. the ranker method using information gain select features as follows:

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Selected attributes in descending order of weightage:
5,3,6,4,30,29,33,34,35,38,12,39,25,23,26,37,32,36,31,24,41,2,27,40,28,1,10,8,13,16,19,22,17,15,14,18,11,7,21,20,9: 41

Table 2 and figure 6 shows the results for different classifiers before application of feature selection process.

Table 2. Results for different classifiers

Parameter	ZeroR	J48	Random Forest
Accuracy (%)	53.45	99.91	99.92
MAE	0.4976	0.0017	0.0028
RMSE	0.4988	0.029	0.0285
Training time(s)	0.02	11.37	31.64
Precision	0.535	0.999	0.999
Recall	1.000	1.000	0.999
F-Measure	0.697	0.999	0.999
FP Rate	0.535	0.001	0.001

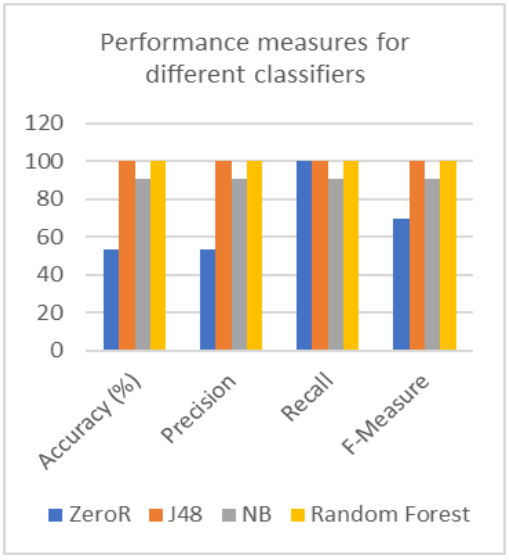

Figure 6. Performance measures for different classifiers before application of feature selection

Table 3 and figure 7 shows the results for different classifiers after application of feature selection process.

Table 3. Results for different classifiers after feature selection process (with six attributes)

Parameter	ZeroR	J48	NB	Random Forest
Accuracy (%)	53.45	99.38	81.89	99.42
MAE	0.4976	0.0106	0.1809	0.01
RMSE	0.4988	0.074	0.4232	0.0712
Training time(s)	0.01	1.37	0.08	15.57
Precision	0.535	0.994	0.851	0.994
Recall	1.000	0.994	0.819	0.994
F-Measure	0.697	0.994	0.812	0.994
FP Rate	0.535	0.007	0.205	0.006

From table 2 and 3, it is clear that as redundant features are removed, the time required to build the model, i. e. training time reduces drastically, but there is not much improvement in accuracy, precision, recall, etc. So still some work is required regarding efficient feature selection, e. g. genetic algorithm may be used. The experimental results show that Random forest and J48 which are tree- based classifiers work better as compared to ZeroR and Naïve Bayes classifiers.

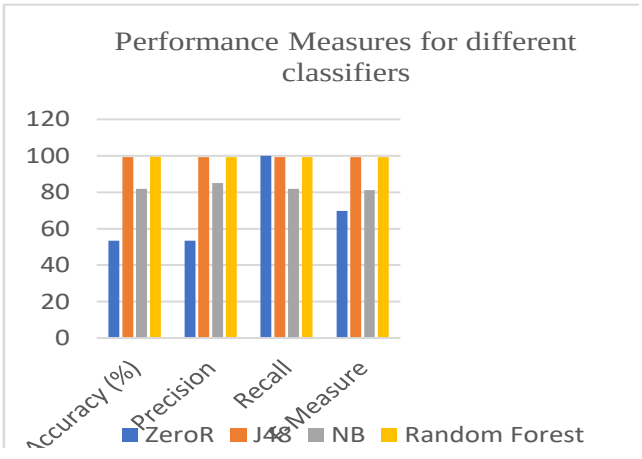

Figure 7. Performance measures for different classifiers after performing feature selection

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Performance measures such as percentage accuracy, precision, recall and F-measure are shown graphically in figures 6 and 7 respectively before applying feature selection and after applying feature selection. Tree based classifiers such as J48 and ensemble techniques such as random forest give the best performance on KDD99 dataset.

5. Conclusion and Future Work

Feature selection is very important step before applying machine learning algorithms in intrusion detection system. It eliminates the redundant attributes so the less time is required for processing the dataset. Random forest and J48 perform the best as compared to the other classifiers such as Naïve Bayes and ZeroR. They also keep the false positive rate to 0.001. The intrusion detection system should keep false positives at acceptable level for the different types of network attacks. Thus, the proposed work gives good results as discussed in the paper on KDD99 dataset. Still there is a need of more experimentation as well as more and more classifiers should be included in the experimental work to enhance the intrusion detection system. Also, there is need to use the better methods of feature selection or optimization methods such as genetic algorithms or particle swarm optimization (PSO). Also, the work is to be completed for testing on latest available datasets for intrusion detection. The author wants to make the intrusion detection system adaptive in future.

Acknowledgments

The authors would like to thank the anonymous reviewers who have given a lot of valuable suggestions, so the manuscript is improved significantly.

References

[1] V. K. Pachghare, “Cryptography and Information Security”, PHI learning, (2015).
[2] Bace R & Mell P, “NIST Special publication on Intrusion Detection Systems”, http://csrc.nist.gov/publications/drafts/idsdraft.pdf.
[3] Scarphone K, Mell P, “Guide to Intrusion Detection and Prevention Systems (IDPS)(Draft)”, thttps://csrc.nist.gov/ctsrc/media/publications/sp/800-94/rev-1/draft/documents/draft_sp800-94-rev1.pdf.
[4] Exploiting machine learning in cybersecurity, https://techcrun"ch.com/2016/07/01/exploiting-machine-learning-in-cybersecurity/.
[5] ManishJoshi, Theyazn Hassn Hadi, “A Review of Network Traffic Analysis and Prediction Techniques”, https://arxiv.org/abs/1507.05722.
[6] Naved Naeem Abbas, Tanveer Ahmed, “Investigating the applications of artificial intelligence in cyber security”, Scientometrics, https://oi.org/10.1007/s11192-019-03222-9.
[7] Mallika, Vikas Deep, Purushottam Sharma, “Analysis and Impact of Cyber Security Threats in India using Mazarbot Case Study”, 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, pp.493-503.
[8] Yao-Wen Huang, D. T. Lee, “Web Application Security—Past, Present, and Future *, Book on Computer Security in the 21st Century”, January 2005, DOI: 10.1007/0-387-24006-3_12.
[9] “Solving the Top 10 Application Security Threats”, https://www.mrc-productivity.com/research/solving-application-ecurity.pdf.
[10] Sumit Soni, Bharat Bhushan, “Use of Machine Learning algorithms for designing efficient cyber security solutions”, 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies [ICICICT], 2019, pp. 1496-1501.
[11] Mustapha Belouch, Salah El Hadaj, Mohamed Idhammad, “Performance evaluation of intrusion detection based on machine learning using Apache Spark”, Procedia Computer Science, Vol. 127 (2018), pp. 1–6.
[12] B. A. Tama, K. H. Rhee, “A combination of PSO-based feature selection and tree-based classifiers ensemble for intrusion detection systems”, in Advances in Computer Science and Ubiquitous Computing, (2015), pp 489–495.
[13] H. Karau, A. Konwinski, P. Wendell & M. Zaharia. “Learning spark: lightning-fast big data analysis", O'Reilly Media, Inc. (2015).
[14] ShaluMall, Sushil Kumar Saroj, A New Security Framework for Cloud Data, Procedia Computer Science 143 (2018) pp.765–775.
[15] Nour Alqudaha, Qussai Yaseen, “Machine Learning for Traffic Analysis: A Review”, Procedia Computer Science 170 (2020), pp. 911–916.
[16] Yang Xin, Lingshuang Kong, “Machine Learning and Deep Learning Methods for Cyber

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

security", IEEE Access, VOL. 6 (2018), pp. 35365-81.
[17] Naveed Naeem Abbas, Tanveer Ahmed, Syed Habib Ullah Shah, Muhammad Omar & Han Woo Park, "Investigating the applications of artificial intelligence in cyber security", Scientometrics, https://doi.org/10.1007/s11192-019-03222-9.
[18] Prajakta Joglekar, Nitin Pise, "Solving cyber security challenges using big data", International Journal of Computer Applications, Vol. 154 (4), 2016, pp. 9-12.
[19] T. Ma, F. Chang, J. Chong, Y. Yu, X. Chen, "A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks", sensors, Vol. 16 (10), 2016, pp. 1701.
[20] Arwa Aldweesh, Abdelouahid Derhab, Ahmed Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues", Knowledge-Based Systems 189 (2020) pp.105-124.
[21] Jean-Philippe Planquart, "Application of Neural Networks to Intrusion Detection", SANS Institute Information Security Reading Room.
[22] Jackson A, "Intrusion Detection System (IDS) product survey", Los Alamos National Laboratory - New Mexico, http://lib-www.lanl.gov/la-pubs/00416750.pdf.
[23]Infosecurity magazine, July 2001, available on-line at http://www.infosecnews.com.
[24] Dewan Md. Farid, Nouria Harbi, Mohammad Zahidur Rahman, "Combining Naive Bayes and decision tree for adaptive intrusion detection", International Journal of Network Security & Its Applications (IJNSA), Volume 2, Number 2, (2010).
[25] Hettich, S. and Bay, S. D., The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Department of Information and Computer Science, (1999).
[26] Quinlan J, C45 programs for machine learning. Morgan Kaufmann Publishers, San Francisco, (1993).
[27] Leo B, Random forests. Machine Learning, Vol. 45, No. 1, (2001), pp.5–32.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731