Machine learning approach for automated defense against network intrusions

Farhaan Noor Hamdani [Corresp., 1], **Farheen Siddiqui** [Corresp. 1]

[1] Department of Computer Science, Jamia Hamdard University, Delhi, New Delhi, India

Corresponding Authors: Farhaan Noor Hamdani, Farheen Siddiqui
Email address: farhaanoor@outlook.com, fsiddiqui@jamiahamdard.ac.in

With the advent of the internet, there is a major concern regarding the growing number of attacks, where the attacker can target any computing or network resource remotely Also, the exponential shift towards the use of smart-end technology devices, results in various security related concerns, which include detection of anomalous data traffic on the internet. Unravelling legitimate traffic from malignant traffic is a complex task itself. Many attacks affect system resources thereby degenerating their computing performance. In this paper we propose a framework of supervised model implemented using machine learning algorithms which can enhance or aid the existing intrusion detection systems, for detection of variety of attacks. Here KDD (knowledge data and discovery) dataset is used as a benchmark. In accordance with detective abilities, we also analyze their performance, accuracy, alerts-logs and compute their overall detection rate.

These machine learning algorithms are validated and tested in terms of accuracy, precision, true-false positives and negatives. Experimental results show that these methods are effective, generating low false positives and can be operative in building a defense line against network intrusions. Further, we compare these algorithms in terms of various functional parameters

1. INTRODUCTION

The security of the computer systems mainly relies on three core pillars: confidentiality, integrity and availability [1]. Safeguarding these services itself is a challenging task. With the abrupt increase in the number of smart-end devices and technologies, threat of intrusions also arises. Any action which results from penetrating into a system or a virtual environment unlawfully is known as an intrusion. An intrusion detection system is a contrivance system of detecting any malicious attempt to disrupt any communication between two secure parties. So, for refining security structure, it is necessary that a detection system should be able to identify novel outbreaks. Conventionally, intrusion detection system methods plunge into two classes: Anomaly and signature detection. Techniques which examines the known tessellations of incidents are called signature-based methods. While as establishing a normal profile for baseline activities and then comparing it against various patterns, is practiced in anomaly-based techniques [2]. These methods are further classified as host and network based depending upon the installed location. These methodologies are used to counter against many intrusions. Out of these attacks DDOS attacks have an impact on a large scale. The main purpose of denial of service attacks is to affect compound devices at a time degrading their computing performance. The affected systems are called botnets or zombies [3]. These systems from a botnet network and are then controlled remotely by the aggressor [4]. There is a frailty in the network layer, which includes tampering with the services like ICMP (internet control message protocol), UDP (user data- gram protocol) and TCP (transmission control protocol).

DDOS attacks take recompenses from these services in flooding a network with unnecessary traffic [5]. Within the context of the attacks, the main aim of this paper to highlight the security concern regarding various attacks, in safeguarding computing resources along with a need to establish a defense structure to counter these attacks. Here section wise integration of the chapters includes, Section 2 which explains the recent/existing intrusion detection systems in which their pros and cons are discussed. In Section 3 the machine learning algorithms are discussed, their working and properties as well. Further, Section 4 the benchmark dataset KDD is discussed in terms of its usage as a benchmark for IDS, which concludes in section 5, implementing the proposed methodology using machine learning algorithms. The KDD dataset contains many attacks, out of which it has majority instances of ddos attacks. The implementation of a machine learning model using different supervised algorithms, while using KDD dataset as a baseline for training these models is carried out. Further, we compute their intrusion detection rate and compare their accuracies alongside with a calculated graph.

2. INTRUSION DETECTION SYSTEM

An intrusion detection system refers to a software or a system which continuously monitors the network systems for any malignant data traffic or policy desecrations. Any alert generated via these systems is directly reported back to the administrator or collected centrally via SIEM (security information and event management). A SIEM collectively matches the patterns from multiple output sources and utilizes an alarm-filtering methods for differentiating between normal and malignant traffic [6]. Types detection systems include:

2.1 Host-based detection: In this detection system the host machine is monitored for all its activities. Host-based IDS is installed on the local host machine. It contains a host-handler which acts as a sensor. Host-based sensor collects system logs, logging activities and other operating system-based logs [7]. They mainly hinge on audit tracks for their functionalities, that enables them to identify elusive outlines of violations, which otherwise would not be tracked in sophisticated level of perception [8]. Host-based methods generally provide more detailed information about an intrusion event than network -based detections, they can provide exact statistics like commands used by attackers, the aim of an attacker, files modified or accessed etc. [9]. However, some disadvantages of HIDS include inability to detect network traffic, dependability, on audit imprints which consume a lot of system resources and privation of traverse platform functionality [10].

2.2 Network-based detection: Monitors the activities of the whole network environment in which it is instated. It accumulates the information from a network rather than from a single host [11]. NIDS inspects the packets stirring across the network and is used in packet-level examination of computer systems in a network via IP pattern checking, investigating both transport and application procedure level conducts. By doing so it can detect many IP centered DDOS assaults which include flooding SYN attacks, disintegrated malignant packets strikes etc. [12]. NIDS are generally low-cost equipment and have a faster retort than HIDS, since there no requirement of sensors to operate on a local host level [13]. Certain complications regarding NIDS include: their constrained discernibility into the local systems, inability to detect encrypted attack network stream etc. [12]

2.3 Application-based detection: It examines the operative conduct and analyzes the proper working of a protocol. The sensor system or intermediate application is employed amongst various process and cluster of servers which observes and examines the exchange of protocols among participating devices [14]. Results of application-based monitoring is extremely accurate in the identification of any malignant activities in applications it safeguards. However, these IDS may fail to spot assaults which are not explicitly targeted at that application [15].

3. MACHINE LEARNING APPROACH

Machine learning is the systematic study of algorithms along with arithmetical models that enable computing systems to accomplish a particular task, without exercising unequivocal operations. The term “Machine Learning” was devised by Arthur Samuel in the year 1959 [16]. These techniques generally depend on data for learning, training and then performing predictive analysis based on some mathematical model [17]. Machine learning systems are broadly classed into three categories: supervised, unsupervised and reinforced learning. In this paper, we have used supervised learning models for resolving the issue regarding detection of attacks.

3.1 Supervised Learning: is a machine learning method in which a model learns to map an input to a particular output based on some examples on which it trained earlier [18]. The data on which the machine is trained is usually a labelled data and these models depend on human interactions for learning, training and testing purpose. Here user acts as a teacher, which teaches the machine based on some mathematical models. This approach is mainly used in solving of classification problems. Some of the models implemented in this paper are:

a) Decision Trees: are well-organized, nonparametric procedures which can be applied to various cataloging and reversion tasks. The structure of decision tree consists of a hierarchical anatomy of nodes, with one node considered as the root node on top, following different leaf nodes at the bottom. Thus, forming a tree like arrangement [19]. It consists of graphs, edges and vertices [20]. These include features like pruning, Gini-index and information gain for accuracy and best split parameters applied on the resultant tree.

b) Artificial Neural Networks: are models which are identical to the biological neurons in configuration and working. Neural networks learn from historical events and evolves periodically [21]. Usually neural networks consist of different layers that are constructed via many interconnected nodules in the presence of an activation function for manipulating the output range. The two main operations involve in these networks are feedforward and back propagation. Their execution time is fast and provide excellent accuracy in predicting events.

c) Recurrent Artificial Neural Networks: belong to the class of artificial neural networks, having same working and features as well. They have an internal memory state which handles the series of inputs, this makes them recall the previous output of the hidden layers and process them in the next state of input from the layers as well. Suitable for speech and handwriting detection [22].

d) Support Vector Machine: are the models which build on arithmetic learning with the task of establishing the position of separating margin known as hyperplane. Hyperplane results in the proper separation of classes in various dimensions [23]. The tuning parameters include kernel trick features, gamma and C. These techniques take a lot of time in compiling and require high end system resources depending upon the size of the dataset. Eg: svm linear, polynomial, Gaussian etc.

e) Naive Bayes: It assumes that the features of a particular data are independent conditionally and then calculates the total-class conditional probability [24]. Naive Bayes, in classification problems yields good results the condition being there exists simple relationships among entities. The Bayes classifier working drives on a robust individuality conjecture, which means different attributes have no effect on the probability of each other [25].

4. KDD-99 Dataset

The KDD-99 dataset is a compiled data regarding computer intrusion detection developed in the Lincoln research labs, Massachusetts. KDD stands for “Knowledge Discovery and Data”. The data are collected in a controlled military environment. It contains various distinct features which were obtained and categorized as ‘normal’ and ‘attackers’ type. KDD is considered as a standard benchmark for testing, IDS, favored and recognized by Defense Advanced Research Projects Agency (DARPA). Among these 41 attributes they mainly cluster into three groups: simple characteristics of separate connection, context attributes within a connection and transport descriptions. KDD-99 data set encompasses nearly 4,900,000 occurrences of captured network data traffic. Also, it consists

normal and 22 diverse types of attacks. These instances are classified further into 41 attributes with a categorized labeling as ‘attack =1’ or ‘normal=0’ [26].

Attack Categories in KDD-Dataset:

i) DOS attacks: Here the main aim of the attacker is to disrupt the services of the legitimate users, rendering them service less. The attacker continuously monitors the target network systems for any vulnerability. The infected machines here are called botnets and are controlled by the attacker to carry out further attacks on a large scale.

ii) User to Root attack: are different attacks as compared to DOS attacks. The attacker accesses the account of the legit user which is perhaps attained by using some hacking tools, exploiting the vulnerability, brute-force, dictionary attacks or social engineering skills etc. After gaining root access of the system it uses it as a launchpad for further attacks.

iii) Remote to User attack: Here the attacker sends malignant packets to the target system over the internet to gain access to the victim machine.

iv) Probing: belongs to a category where the attacker continuously monitors the network to gain some info and vulnerabilities related to the targeted system**.** It mainly involves passive scanning. This info is then utilized by the attacker to launch a planned attack [26].

5. PROPOSED DETECTION SYSTEM

To extend the efficacy of existing detection techniques against network intrusions, we examine how different machine learning algorithms can be implemented to counter the attacks in an automated manner. For this, we propose and design an extensible framework which can be used to test these algorithms.

The proposed model can enhance and facilitate the existing detection systems like firewalls etc. The detection apparatus is observed as, the inspection of the incoming traffic from the internet through the desired implemented firewall. The filtered traffic is passed through the extractor, which extracts the features of the incoming data as per KDD-99 dataset format. Further, applied with normalization of conversion of string values to numerical for ease of classification. The categorized data obtained through preprocessing is then passed through machine learning algorithms for advance detection of any malignant traffic content. Also, it can also exhibit offline inspection of the captured incoming data. In our research we have used six of the machine learning algorithms, which are trained and tested on KDD dataset. The following experiment is implemented via python programming language in Anaconda software distribution platform, carried on 4th Gen windows10 system with 8gb of RAM.

6. NORMALIZATION

In this method, the string values of KDD-99 dataset both training and tested sets are converted into numerical values for ease in numerical computing of the model. Various string values include:

Attacks= *teardrop, pod, land, smurf, neptune, Back, rootkit, perl, loadmodule, Buffer_overflow, guess_passwd, Ftp_write, phf, multihop, spy, imap, warezmaster, warezclient, portsweep, Nmap, ipsweep, Satan.*

Protocol type = *tcp, udp, http, icmp, ftp, smtp*

Flags=*rsto, s0, sh, s1, s0, rej, sf, rstr, s3, oth, s2.*

Services=*smtp,ntp_u,shell,aol,imap4,urh_i,netbios_ssn,tftp_u,mtp,uup,echo,tim_i,ssh,iso_tsap,time,netbios_ns,systat,hostnames,login,efs,supdup,http_8001,kshell,vmnet,http_2784,Z39_50,courier,ctf,finger,nntp,ftp_data,red_i,ldap,http,ftp,pm_dump,http_443,exec,klogin,auth,netbios_dgm,other,link,X11,discard,private,remote_job,IRC, daytime,nnsp,pop_3,pop_2,gopher,sunrpc,name,rje,domain,uucp_path,domain_u,csnet_ns,whois,eco_i,bgp,sql_net,printer,telnet,ecr_i,urp_i,netstat,harvest.*

These string attributes are mapped and converted into numerical values 0 or 1. Example tcp is converted to 0, http is converted to 0, SF is converted to 0, icmp is converted to 1. Conversion is done based on probability of occurrence of the string value in one type of instance. If the probability of occurrence of the tcp is more in the normal packet, then it is normalized as 0 otherwise 1. Similarly, other string values are mapped to these normalized values.

7. TRAINING PHASE AND TESTING PHASE

a) Artificial Neural Network: trained on 1048575 instances of captured data, each instance contains 41 attributes and is further classified as 'attack or 'normal'. The dataset here is a labelled dataset. The machine learns to distinguish between an attack and normal traffic. Training results are excellent with 94.73% accuracy and error 5.26%. Testing results with 97.92% accuracy performed on 199999 instances. The ANN structure contains 41 attributes as input with one hidden layer containing 17 nodes along with the applied sigmoid function to manipulate the output results. Feedforward and back-propagation methods are used for calculating the error.

b) Decision Tree: trained on 1048575 instances as well. Here the original dataset is then split into 80% training data and 20% tested data. 209715 instances are tested and 838860 for training the model. Both Gini -index and information gain features are used to match the accuracy of detecting intrusions. Accuracy results, 99.98%. *

c) SVM Linear and Gaussian: both are trained on 206997 instances of the captured KDD dataset type. The dataset is split into two i.e. 80% for training and 20% for testing. Results obtained: 98.5% for linear and 99.86% for gaussian. Kernel trick is used for classification, separation of the attributes through a hyperplane. Hyperplane divides the attributes along the kernel selected i.e. linear line separating the features in svm linear and non-linear line separating the features in a dimensional space in case of gaussian svm.

d) Recurrent Artificial Neural Network: trained on 1048576 instances of KDD dataset. Again, the dataset is split into two, 80% training and 20% tested. The dataset is a labelled and categorized as '0' for normal and '1' for attack.

e) Naive Bayes: trained on 1000000 instances of labelled KDD data. It calculates the likelihood probability of each attributes and then predicts the outcome results. The data here is splitted into two 799999 trained and 200000 tested instances. Results 95.32% detection rate. *

*K-fold cross validation technique is used in both decision trees and naive bayes model. The main purpose is to ensure the validation, as to how well our model is trained on a given dataset and tested on unobserved data.

182

183 **8. COMPARISON AND RESULTS**

184

185

186

OUTPUT PREDICTED RESULTS:

The detection results of machine learning algorithms obtained are quite satisfactory, as the average detection rate is greater than 85%. Predicted class have classification values '0' and '1'. Highest accuracy is obtained with decision trees and lowest with RNN. Since RNN stores previous output, it requires extra memory optimization and decision trees operate well on a hierarchical topology, which gives them options of considering only relevant branch rather than whole architecture. However, these algorithms have different execution time, apart from SVM all other algorithms took few minutes to execute. Since SVM has to classify all the dataset at once that too with separable distance in dimensionality scenario to classify the attributes. SVM algorithms produce good results but overall computation costs more time and resources. RNN and ANN, though both utilize neural network model, the overall accuracy of ANN is proved to be superior. Thus, it concludes that depending upon the computing resources or the size/input of the dataset the following algorithms can be implemented selectively.

9. CONCLUSION

The perceptive intrusion detection methodology used in this paper based on machine learning algorithms proved significantly effective in separation/detection of attacks. we have used algorithms of our own choice and also built a novel framework that is easily extensible with other machine learning algorithms as well. The accuracy obtained is effective in identifying legitimate traffic from non-legitimate one in case of test dataset. As compared with the traditional firewalls, they tend to produce less false alarms and can save from unnecessary alert logs, which in turn helps in utilizing productive resources to another important tasks. Thus, machine learning algorithms that are tested and validated can be incorporated within the existing intrusion detection systems for their enhancement. Since it is easy to implement, deploy and is cost efficient as compared to other systems. But this method requires continuous training of the models based on different parameters or datasets. However, the extraction of various features from the live traffic and categorizing them into a dataset format is a strenuous task itself. The growing number of attacks still pose a threat and challenge in the modern world and in-turn place burden on strategizing an effective method to counter these attacks.

REFERENCES

[1] M. Bishop, Introduction to Computer Security. Professional, 2004 Addison-Wesley.

[2] S. Kumar, A. Yadav. Increasing Performance of Intrusion Detection System Using Neural Network. 2014 IEEE International Conference on Advanced Communication Control and Technologies (ICACCCT), pp. 19 35-1939.

[3] M. Reed Denial of Service attacks and mitigation techniques: Real time implementation with detailed analysis. [Online] SANS Institute InfoSec Reading Room 2011. Available from: ⟨http://www.sans.org/reading-room/whitepapers/detection⟩.

[4] Troj/Flood-IM.Backdoor DDoS Trojan. Detected by Sophas. Available from: ⟨https://secure2.sophos.com⟩.

[5] B. B. Gupta, M. Misra, R. C. Joshi, ―FVBA: A Combined Statistical Approach for Low Rate Degrading and High Bandwidth Disruptive DDoS Attacks Detection in ISP Domain, in the proceedings of 16th IEEE International Conference on Networks (ICON-2008), DOI: 10.1109/ICON.2008.4772654.

[6] Martellini, Maurizio; Malizia, Andrea (2017-10-30). Cyber and Chemical, Biological, Radiological, Nuclear, Explosives Challenges: Threats and Counter Efforts. Springer. ISBN 9783319621081.

[7] Bace, Rebecca- "An Introduction to Intrusion Detection &Assessment"- Infidel, Inc. for ICSA, Inc.

[8] Rebecca Gurley Bace "Intrusion Detection"- Macmillan Technical publishing- 2000.

[9] Intrusion detection system buyer's guide.

[10] Bace, Rebecca: An Introduction to Intrusion Detection & Assessment. Infidel Inc.

[11] Global Information Assurance Certification Copyright SANS institute.

[12] Modi, C., et al., A survey of intrusion techniques in cloud. Journal of Network and Computer-Applications,2013. p.42-57.

[13] Wu, M. Protocol-based classification for intrusion detection. in WSEAS International Conference. Proceedings Mathematics and Computers in Science and Engineering. 2008. World Scientific and Engineering Academy and Society.

[14] Karthikeyan. K.R and A. Indra- "Intrusion Detection Tools and Techniques a Survey".

[15] Brien M. Posey Choosing-an-intrusion-detection-system- Network-host-or application-based-IDS. searchenterprisedesktop.techtarget.com/tip/

[16] Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3 (3): 210-229. doi:10.1147/rd.33.0210

[17] Bishop, C. M. *(2006),* Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

[18] Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.

[19] E. Alpaydin, Introduction to Machine Learning (MIT Press, Cambridge, 2010). ISBN:026201243X, 9780262012430.

254 [20] S. Safavian, D. Landgrebe, A survey of decision tree classifier methodology. IEEE Trans.Syst. Man Cybern.
255 21(3), 660–674. ISSN: 0018–9472.

256 [21] D. Prashanth Kumar et al, (IJCSIT) International Journal of Computer Science and Information Technologies,
257 Vol. 5 (6), 2014, 7041-7044.

258 [22] Sak, Hasim; Senior, Andrew; Beaufays, Francoise (2014)."Long Short-Term_Memory_recurrent neural
259 network architectures for large scale acoustic modelling".

260 [23] V. Vapnik and C. Cortes, "Support Vector Network," Machine Learning, vol. 20, pp. 273-297.

261 [24] Dewan Md. Farid, M. Z. (2011). Adaptive Intrusion Detection based on Boosting and. International Journal
262 of Computer Applications.

263 [25] Wafa'S.Al-Sharafat and Reyadh "Development of genetic-based machine learning for network intrusion
264 detection" world academy of science, engineering and technology.

265 [26] kdd.ics.uci.edu/databases/kddcup99/task.

266

Table 1(on next page)

Types of attack categories in the KDD dataset [26].

Attacks	**Categories**
DOS attacks	Teardrop, pod, land, Smurf, Neptune.
Probing	portsweep, Nmap, ipsweep, Satan.
Remote2 Local attacks	guess_passwd, Ftp_write, phf, multihop, spy, imap, warezmaster, warezclient.
User 2 Root attacks	Rootkit, Perl, load module, Buffer overflow.

Table 2(on next page)

Various string conversions.

Name	***Value***	***Name***	***Value***	***Name***	***Value***
SF	*0*	*icmp*	*1*	*tcp*	*0*
ecr_i	*1*	*Private*	*1*	*netstat*	*1*
S0	*1*	*udp*	*0*	*daytime*	*1*
Finger	*1*	*RSTR*	*1*	*name*	*1*
Smtp	*0*	*ftp_data*	*0*	*whois*	*1*
domain_u	*0*	*pop_3*	*0*	*REJ*	*1*

Table 3(on next page)

Overall comparison of Algorithm

Model	*Training set*	*Tested set*	*Training result*	*Testing result*	*Execution time*	*False Positive rate*
ANN	*1048575*	*199999*	*94.73%*	*97.92%*	*1 minute*	*615*
DT	*838860*	*209715*	*NA*	*99.98%*	*1 minute*	*29*
SVM-L	*165596*	*41400*	*NA*	*98.50%*	*11 hours*	*289*
SVM-G	*165596*	*41400*	*NA*	*99.86%*	*9 hours*	*61*
RNN	*999999*	*206996*	*89.75%*	*86.55%*	*1 Minute*	*671*
NB	*799999*	*200000*	*NA*	*95.32%*	*1 minute*	*3132*

Figure 1(on next page)

Detection Model

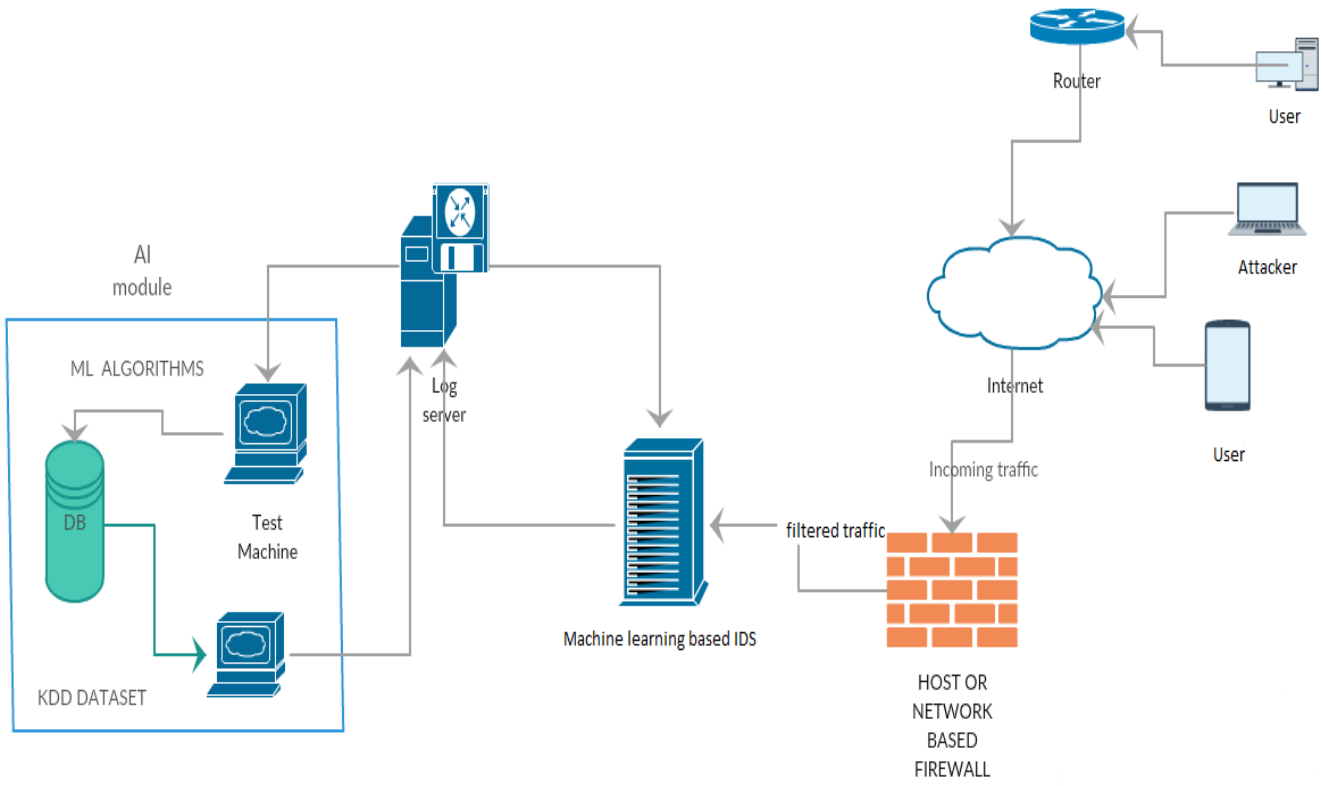

Router
User
Attacker
Internet
User
Incoming traffic
filtered traffic
HOST OR
NETWORK
BASED
FIREWALL
Machine learning based IDS
Log
server
AI
module
ML ALGORITHMS
DB
Test
Machine
KDD DATASET

Figure 2(on next page)

Detection rate in different algorithms

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27777v1 | CC BY 4.0 Open Access | rec: 3 Jun 2019, publ: 3 Jun 2019

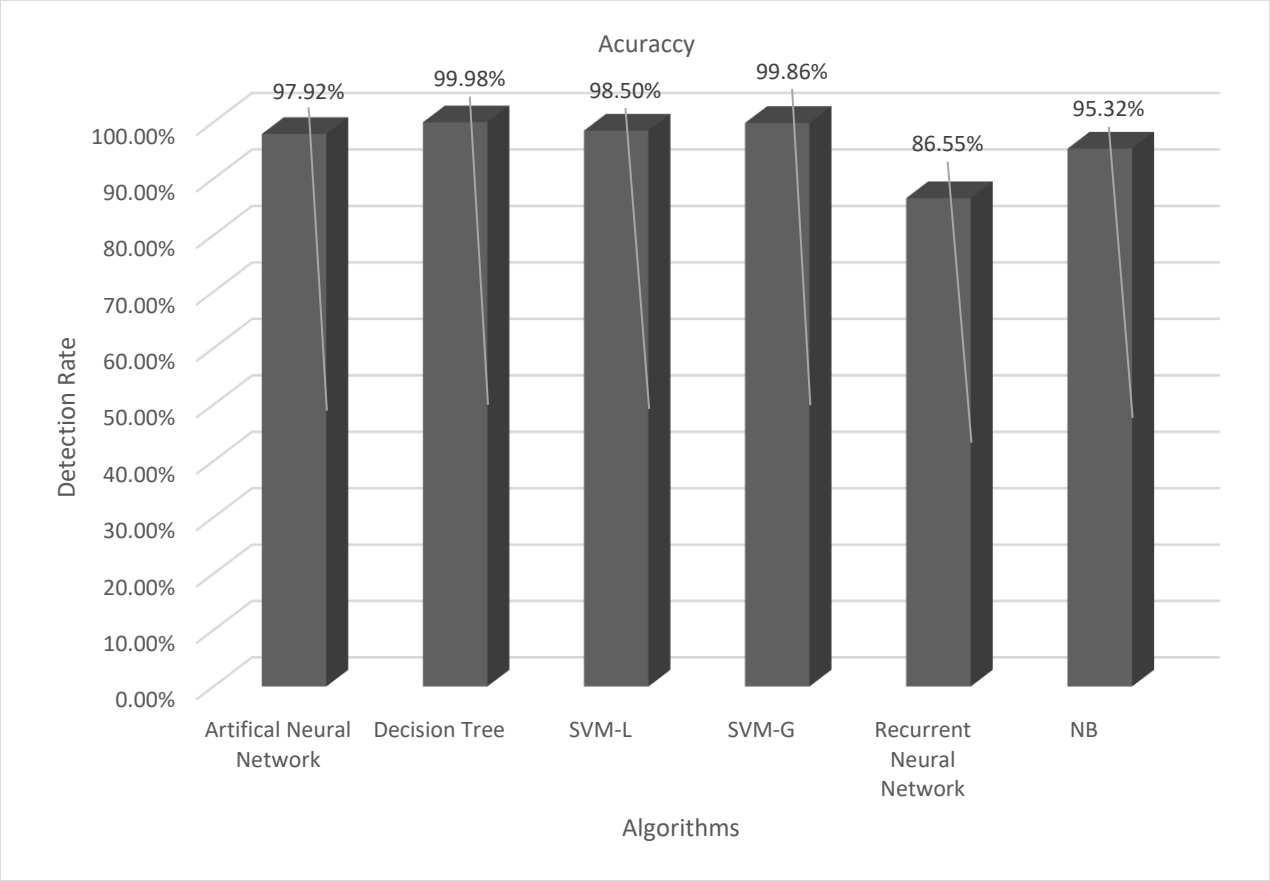
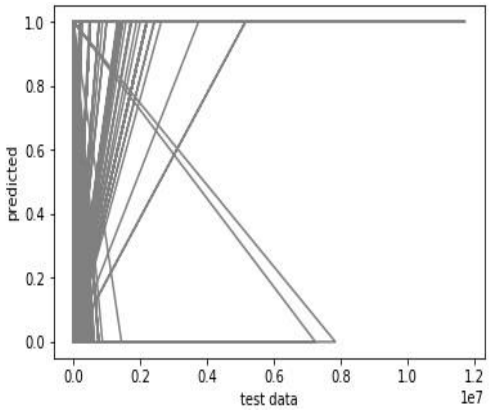
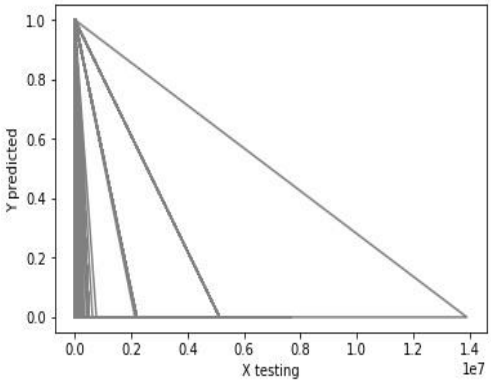


Figure 3(on next page)

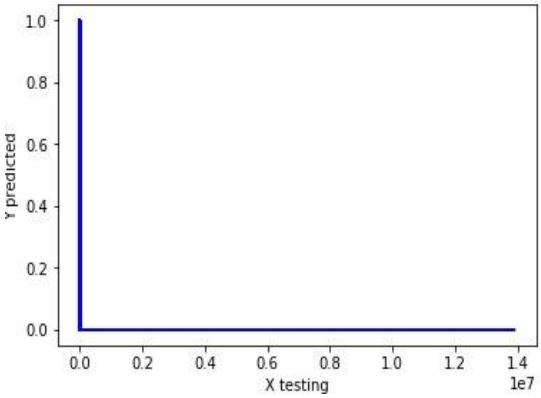
Output Graphs



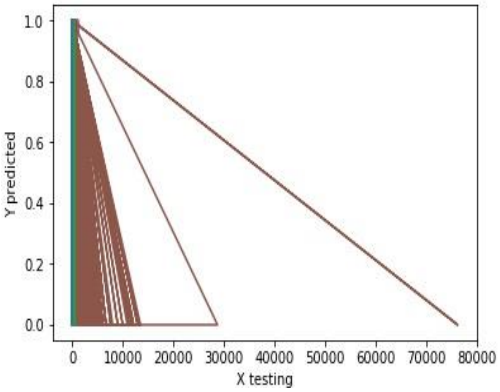
Artificial Neural Network



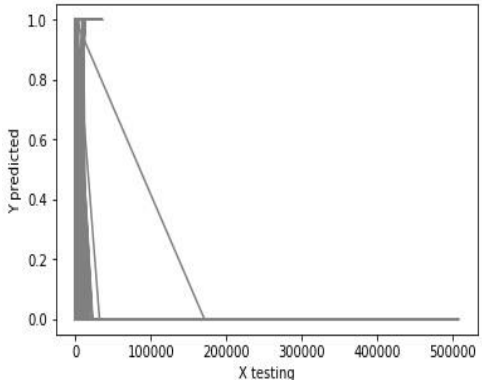
SVM_G



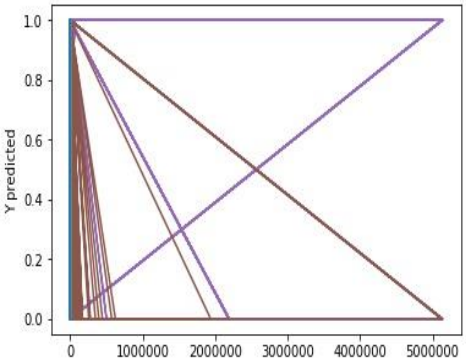
Naïve Bayes



Recurrent Neural Network



Decision Tree



SVM-L