Network Intrusion Detection - Comparison across classifiers

Gaurav Yadav

Student,
Department of Computer Science & Engineering,
Institute of Engineering & Technology, Alwar, India.

Abstract— With increasing use of networks, intrusions across the Internet have become a major threat in our world. This research study has been influenced by the different intrusion threats on internet and the ways to detect them. In this research we have studied and analyzed the famous network traffic data -NSL KDD dataset and its various features. The model aims to classify the network data into attack and non-attack and the attack class is further classified into DoS, Probe, R2L and U2R attack types. The work includes the classification of network traffic data by applying various classifier and analyze their accuracy. The model is being trained with different dataset sizes and have been evaluated on test data set. We have observed that the more we train the model, the more accurate classification results have been produced.

Index Terms— **IDS, TCP/IP, KNN, security.**

I. Introduction

The demand of an intrusion detection system in various applications has increased in the recent years since huge amount of data is available to be stored and processed every day. The networking systems are generating huge amount of data by monitoring the surroundings of applications in which they are deployed. Any kinds of suspecting behaviors are detected by the devices. Any kinds of vulnerabilities in any computer network can be found by an intruder that aims to harm the users using that device. For preventing the entry of intrusions, the best solution is to protect the system or its resources.

II. Network Intrusion and Machine Learning

Security is the main problem in the internet today however there are new technologies emerging every day in the same way new types of attacks and tools are being updated and upgraded by the intrudes. Internet brings a vast amount of user to the website. This cause the internet companies to have big data of users accessing the site and sharing the information, but the security of the user's data is still a big question when comes to internet security. New security methods are proposed to secure the data of the user and the organization [1].

The major attacks can be classified as under,

1) "External Penetrations: this is an unauthorized user of the system
2) Internal penetrations: as an authorized system user who uses the system in an unauthorized manner
3) Finally, Misfeasors: as an authorized user who tries to exploit their accessing privileges".

To prevent attackers from stealing data and users from falling prey to security attacks, most of the computer infrastructure make use of Firewalls, "Virtual Private Networks" (VPN), Antivirus and communication over secure layer [2]. Below are some use cases wherein machine learning can be used:

- Machine learning has found huge application towards self-driving cars, trains, planes etc. The main intent is to avoid accidents via human error.
- Machine learning is also used heavily on E-Commerce and news domains, wherein based on user history, certain recommendations or relevant details can be shown.
- Detecting fraud during a payment transaction in ecommerce or any other sites or in the payment processing sites.

The attention in machine learning is due to factors like Bayesian analysis and data mining are the more popular one. Mechanisms like increasing volumes computational processes which is cheaper and powerful, varieties of available data which has less data storage.

IDS is the practically applicable solutions against unsafe attacks in cyberspace. Furthermore, attackers always keep updating and changing their techniques and tools [3]. The implementing an accepted Intrusion detection system is also a very challenging task. Several experiments have been conducted and evaluated to assess various ML classifiers based on KDD dataset [4]. The main goal of all these researches has been to increase performance and accuracy of intrusion detection system.

The machine learning has workflow that need to be taken to have a good prediction model.

- Gathering data
- Data pre-processing
- Researching for a suitable model
- Training and testing the model

Machine Learning and artificial have come a long way and great advancements have been made in the field. The machine learning models can learn the patterns via train data and then use those learnings to identify anomalies in the test data. Machine learns from preceding computations to deliver repetitive and reliable results. It is a new technology which has gained evolution due to computing power and datasets.

III. Literature Survey

Gong Shang-fu, et.al (2012) proposed a method in which R-SVM was applied into IDS. The various selection solutions and rough feature extractions were compared in this research [5]. The results showed that the R-SVM method outperformed other methods.

However, this was not the best solution through which the irregular behaviors could be discriminated. The methods through which the TPR could be increased while managing the minimum training time and forecast time were still needed to be designed.

V. K. Pachghare et.al (2009) presented a study in which a neural networks-based algorithm was to be developed which provided high suitability for IDSs. It was commonly named as SOM (Self Organizing Maps) [6]. A promising method that was implemented in different classification problems was called the neural network. Based on the assumption that every user was unique and left a footprint on the computer system, it was implemented through the neural network component. The security officer could be alerted about the possible security breach if the footprints of a user did not match their reference footprint designed on the basis of normal system activities. The advantages and disadvantages of using SOM were discussed towards the end of this research. Further, the importance of SOM is generating an efficient IDS was also explained.

Zhang Ran, (2012) proposed an integrated intrusion detection model in this paper. The multi-agent approach was used as a base to design this model [8]. Through the coordination of agents in collaborative manner, this model detected the suspicious behaviors using all kinds of agents. Further, the measures through which the danger could be stopped were studied here. Thus, adapting the changes in the environment and attack was possible through this method. To construct dynamically adaptive IDS by defining and analyzing the formal intrusion detection model, a theoretical foundation is presented in this paper.

Dheeraj Pal, et.al (2014) proposed a study which aimed to design an IDS by making improvements in the genetic algorithm [7]. Based on the information gain, the attribute subset reduction was applied here. This method also aimed to considerably reduce the training time and complexity. Further, in comparison to the hard-computing approach applied in existing genetic algorithm, the rule was more efficient by embedding a soft computing approach in rule generation. The attack could be detected with higher efficiency using the generated rule. The KDD'99 data set was used to perform verifications of this model. It was seen that higher detection rates and low false positive rates were achieved as per the empirical results.

IV. EXPERIMENT RESULTS AND CONCLUSIONS

Using NSL KDD Data set as base, we compare the accuracy and "cross validation mean score" against various ML classifiers. We train the model using training data set and then check the classification of attack and non-attack traffic using test data set.

It can be observed that almost all the classifiers have an accuracy of less than 99% which can be improved further.

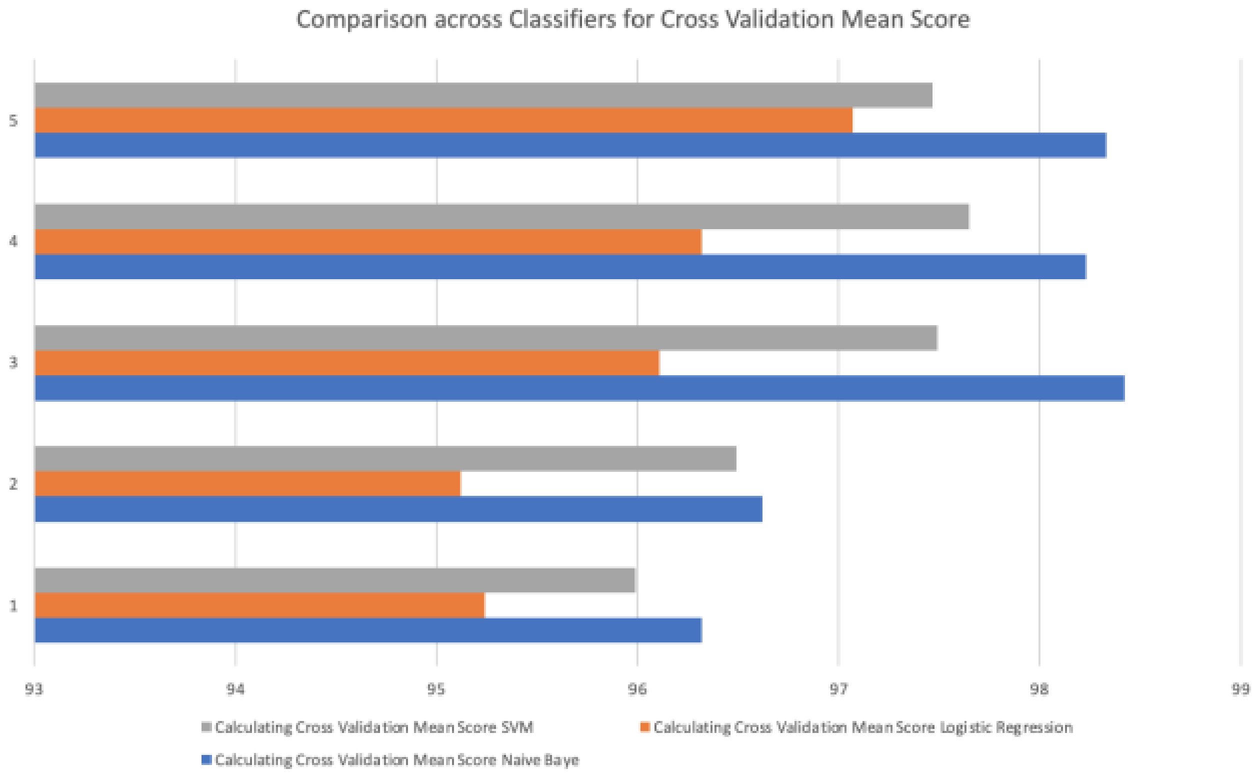

Fig. 1: Comparing "Cross Validation Mean score" across Classifiers.

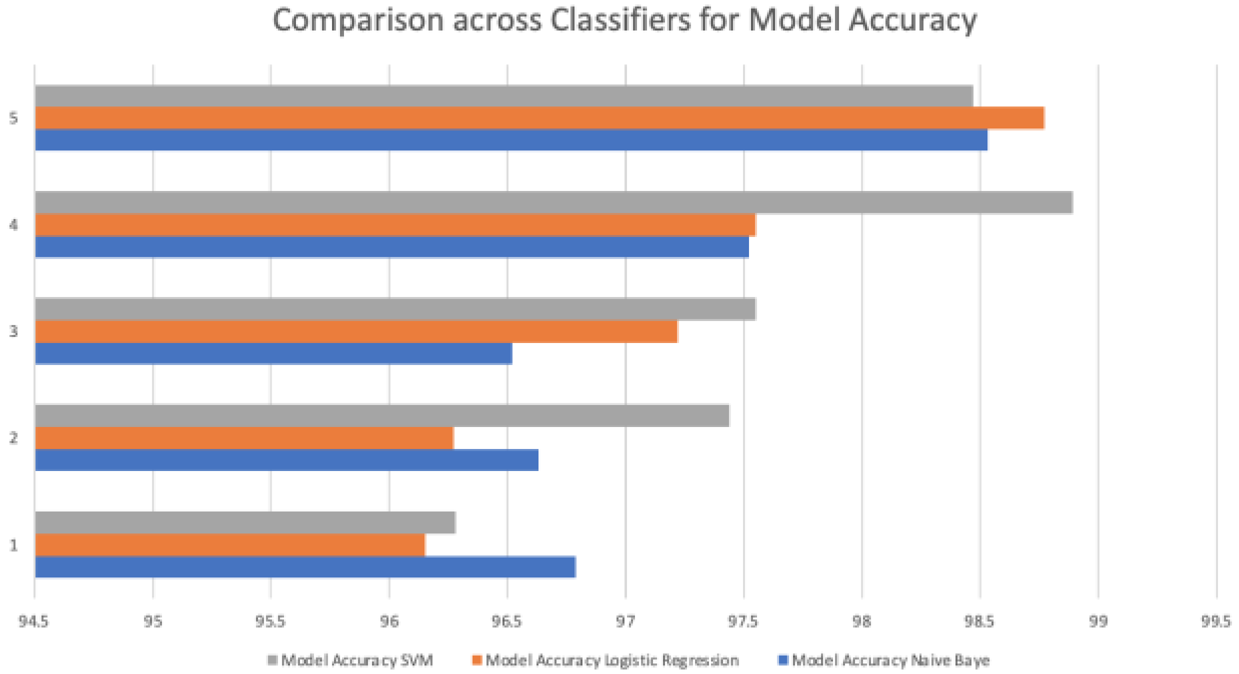

Fig. 2: Comparing "Model Accuracy" across Classifiers.

REFERENCES

[1] Yu-Xin Meng, The practice on using machine learning for network anomaly intrusion detection,2011, International Conference on Machine Learning and Cybernetics, Volume: 2, Pages: 576 – 581

[2] Yi Yi Aung, Myat Myat Min, A collaborative intrusion detection based on K-means and projective adaptive resonance theory, 2017, 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Pages: 1575 – 1579

[3] V. K. Pachghare ; Parag Kulkarni ; Deven M. Nikam, Intrusion Detection System using Self Organizing Maps, 2009, International Conference on Intelligent Agent & Multi-Agent Systems, Pages: 1 – 5

[4] V. Jaiganesh, P. Sumathi, S. Mangayarkarasi, An analysis of intrusion detection system using back propagation neural network, 2013, International Conference on Information Communication and Embedded Systems (ICICES), Pages: 232 – 236

[5] Gong Shang-fu, Zhao Chun-lan, Intrusion detection system based on classification, 2012, IEEE International Conference on Intelligent Control, Automatic Detection and High-End Equipment, Pages: 78 – 83

[6] V. K. Pachghare ; Parag Kulkarni ; Deven M. Nikam, Intrusion Detection System using Self Organizing Maps, 2009, International Conference on Intelligent Agent & Multi-Agent Systems, Pages: 1 – 5

[7] Dheeraj Pal, Amrita Parashar, Improved Genetic Algorithm for Intrusion Detection System,2014, International Conference on Computational Intelligence and Communication Networks, Pages: 835 – 839

[8] Zhang Ran, A Model of Collaborative Intrusion Detection System Based on Multi-agents, 2012 International Conference on Computer Science and Service System, Pages: 789 – 792