International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Intrusion Detection in Network with the help of Supervised Machine Learning Technique alongside Feature Selection

Sanchita Hegde[1], Shalini K V[2], Sheethal S[3], Shree Raksha R[4], Sanjaykumar J H[5]

[1,2,3,4]UG student, [5]Assistant Professor, Dept. of Information Science and Engineering, Sapthagiri College of Engineering, Bengaluru, India

---

ABSTRACT- A new supervised machine learning system is developed to classify network traffic data whether it is malicious or benign. We developed a supervised machine learning model that can classify unseen network traffic based on what is learnt from the seen traffic. We used both SVM and ANN learning algorithm to find the best classifier with higher accuracy and success rate.

Keywords- *Supervised machine learning, Support vector machine (SVM), Artificial neural network (ANN)*

1. INTRODUCTION

With the wide spreading usages of internet and increases in access to online contents, cybercrime is additionally happening at an increasing rate. Intrusion detection is that the opening move to forestall security attack. Hence the protection solutions like Firewall, Intrusion Detection System (IDS), Unified Threat Modeling (UTM) and Intrusion Prevention System (IPS) are becoming much attention in studies. IDS detect attacks from a range of systems and network sources by collecting information and so analyze the data for possible security breaches. The network based IDS analyzes the information packets that travel over a network and this analysis are disbursed in two ways. Till today anomaly based detection is way behind than the detection that works supported signature and hence anomaly based detection still remains a serious area for research. The challenges with anomaly based intrusion detection are that it must cater to novel attack that there's no prior knowledge to spot the anomaly. Hence the system somehow must have the intelligence to segregate which traffic is harmless and which one is malicious or anomalous and for that machine learning techniques are being explored by the researchers over the previous couple of years. IDS however isn't a solution to all or any security related problems. for instance, IDS cannot compensate weak identification and authentication mechanisms or if there's a weakness within the network protocols.

2. LITERATURE SURVEY

Intrusion detection system plays a major role in network security. Intrusion detection model are often a predictive model that has ability of predicting the network having normal data traffic. Machine Learning algorithms are accustomed build accurate models for clustering, classification and prediction. during this paper classification and predictive models for intrusion detection are built by using machine learning technique by the usage of algorithms like Logistic Regression, Gaussian Naive Byes, Support Vector Machine and Random Forest. These algorithms are tested with NSL-KDD data set. The results shows, Random Forest Classifier performs the other methods in identifying the information traffic and detecting as normal or an attack.

Advantages: Prevent attacks and maintains the privacy of users as IPS records the network activity only when it finds an activity that matches the list of known malicious activities.

Disadvantage: disadvantage is that the intrusion software can create an outsized number of false alarms.[9]

The four key steps of a Feature selection process are feature subset generation, subset evaluation, stopping criterion and result validation. The feature subset generation is a heuristic search process which results in the selection of a candidate subset for evaluation. It uses searching strategies like data into both binary and multi classes and maximize it accuracy.

Advantage: Very efficient and fast to compute.

Disadvantage: A feature that is not useful by itself can be very useful when combined with others.[20]

Since the primary introduction of anomaly-based intrusion detection to the research community in 1987, the sphere has grown tremendously. A range of methods and techniques introducing new capabilities in detecting novel attacks were developed. Most of those techniques report a high detection rate of 98% at the low warning rate of 1%. While a range of anomaly-detection techniques are

proposed, adequate comparison of those methods' strengths and limitations which will cause potential commercial application is difficult. Since the validity of experimental research in academic engineering, in general, is questionable, it's plausible to assume that research in anomaly detection shares the above problem. The concerns about the validity of those methods may partially explain why anomaly-based intrusion-detection methods don't seem to be adopted by industry. To analyze this issue, we review the present state of the experimental practice within the area of anomaly-based intrusion detection and survey 276 studies during this area published during the amount of 2000-2008. We summarize our observations and identify the common pitfalls among surveyed works.

Advantages: it's possible that several of the identified pitfalls were avoided within the conducted research, but not reported.

Disadvantages: we are able to lose its value behind an ambiguous, unclear and unsound presentation. [4]

We can observe the clear picture of macro-level opportunity indicators affecting cyber-theft victimization. Supported the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to work out if variation in system impacted cyber-victimization across states. the present study found that structural conditions like unemployment and non-urban population are related to where users access the net. Also, this study found that the proportion of users who access the net only reception was positively related to state-level counts of cyber-theft victimization. The theoretical implications of those findings are discussed.

Advantages: We examine effects of macro-social opportunity factors on state-level cyber-theft victimization. Supported theoretical arguments and prior research findings associated with COT and cybercrime victimization, we hypothesized that online routine activities associated with where users access the net would affect cyber-theft victimization.

Disadvantages: Sample size within the current study could weaken statistical power, which is that the probability of rejecting a false null hypothesis, and result in the insignificance of those two online routine activities. [1]

3. PROPOSED METHODOLOGY

A. Data Collection

We are collecting the KDD cup training and test network intrusion data from UCI-Machine learning repository. Which it's almost 42 attributes

B. Data Normalization

The information normalization is that the process of cleaning data during this module we are removing the repeated data and removing the empty rows.

C. Feature selection

In the first part, we extracted most relevant features using different feature selection (FS) methods. within the wrapper method we used SVM classification algorithm with cross-validation to avoid over fitting and under fitting problem. within the filter method a ranker algorithm is employed to seek out the most effective result suitable for our proposed classifier. The training data we used from NSL-KDD dataset contains 25,191 labeled instances.

D. Classification of intruders:

With the features found in feature selection part, total four models are in-built Weka software suite using the training dataset. Classification using supervised machine learning first requires training the model using training dataset. We used 20% of NSL-KDD dataset as training data that have 25,191 labeled data instances. To training the model we used SVM, Decision tree, Random Forest and logistic regression learning algorithm for every form of feature selection method. Supported the training models will generate different results.

E. Performance Evaluation:

While comparing the performance of the proposed model with the others works, we picked works having hypothesis of comparable aspects associated with learning algorithm and benchmarking datasets. But there are other aspects like attribute reduction, number of instances, the quantity layers and learning rates used. The detection success rate of the proposed model is additionally compared with other existing models.

4. RESULT

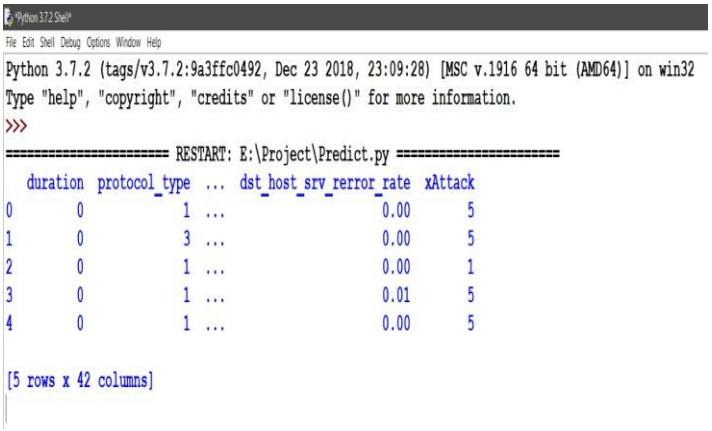

Figure 1. Detection of attacks

Figure 1 shows how attacks are detected.

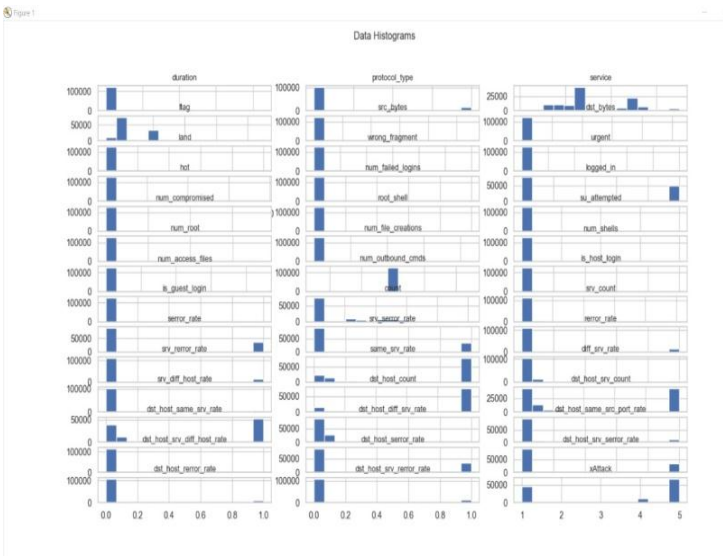

Figure 2. Data histograms

Figure 2 shows the data histograms for different data sets.

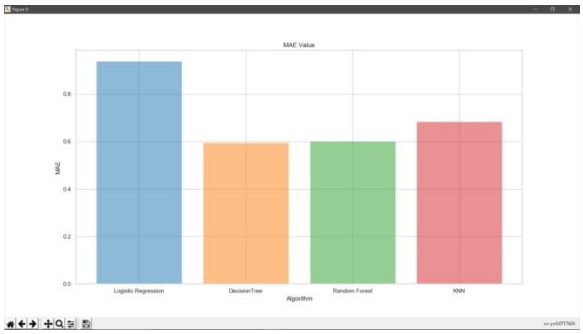

Figure 3. Mean absolute error value

Figure 3 shows the mean absolute error value for different algorithms.

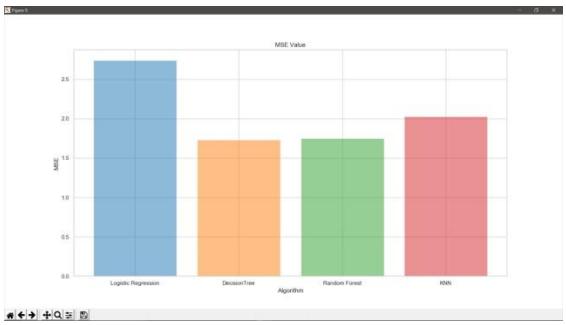

Figure 4. Mean squared error value

Figure 4 shows the mean squared error value for different algorithms.

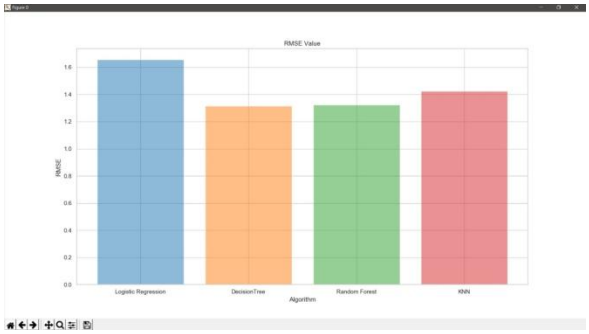

Figure 5. Root mean square error value.

Figure 5 shows the root mean square error value for different algorithms.

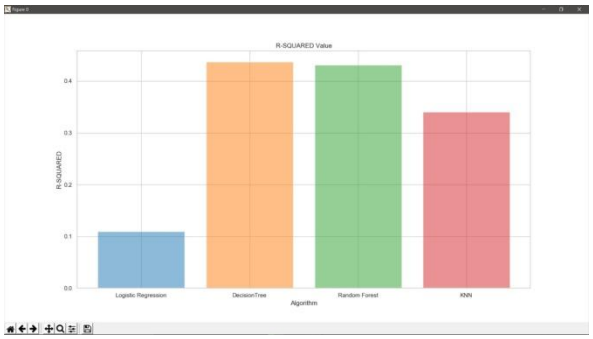

Figure 6. R squared value

Figure 6 shows R squared value for different algorithms.

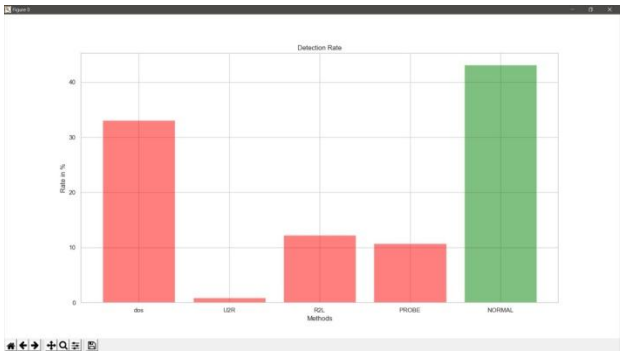

Figure 7. Detection rate

Figure 7 shows the detection rate for different algorithms.

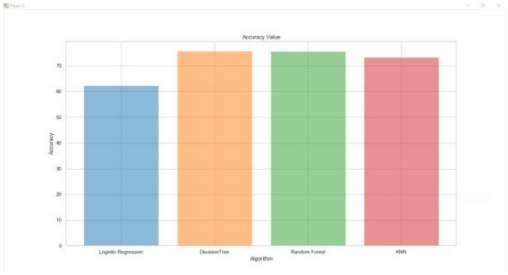

Figure 8. accuracy value

Figure 8 shows the accuracy rate for different algorithms.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.

[10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in Communication Networks

and Services Research, 2007. CNSR'07. Fifth Annual Conference on, 2007, pp. 350–358.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[13] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and Information Systems Conference (MilCIS), 2015, 2015, pp. 1–6.

[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.

[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.

[16] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016, pp. 21–26.

[17] M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial naive bayes for network intrusion detection," in Information Assurance and Security (IAS), 2010 Sixth International Conference on, 2010, pp. 5–10.

[18] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on, 2015, pp. 92–96.

[19] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," Journal of Engineering Science and Technology, vol. 8, no. 1, pp. 107–119, 2013.

[20] Saraya.k, Prabhu R, Dr Ramesh Kumar M, Preethi P "Network Based Intrusion Dtection System using Filter Based Feature Selection Algorithm" Journal of engineering and technology (IRJET).