Scientific Journal of Impact Factor (SJIF): 5.71

E-ISSN (O): 2348-4470
P-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 8,Issue 12,December 2021

Increasing the Performance of MACHINE LEARNING-Based MODELs on an Imbalanced and up-to-date dataset.

Prof. Pritam Ahire[1], Ankit Rathod[2], Rahul Sanap[3], Omkar Thorat[4]

Department of Computer Engineering, D Y Patil Institute of Engineering and Technology, Pune, India

Abstract —*In growing times, the use of internet is spreading at a lightning speed and which as a result N/Wed computer has been increasing in our daily lives. This expanding chain of N/Wed computer weakens the servers which enable hackers to intrude on computer by using various means which may be know as well as unknown and makes them even harder to detect. So as a protection to the computers the Intrusion Detection System (MODEL) is introduced which is trained with some MACHINE LEARNING techniques by making use of previous available data. . The used datasets were collected during a limited period in some specific N/W and generally don't contain up-to-date data. In this paper, we propose six machine-learning-based MODELs by using Random Forest, Gradient Boosting, Adaboost, Decision Tree, and Linear Discriminant Analysis algo. To implement a more realistic MODEL, an up-to-date security dataset, CSE-CIC-MODEL2018, is used instead of older and mostly worked datasets. Therefore, to increase the efficiency of the system depending on attack types and to decrease missed intrusions and false alarms, the imbalance ratio is reduced by using a synthetic data generation model called Synthetic Minority Oversampling Technique. Experimental results demonstrated that the proposed approach considerably increases the detection rate for rarely encountered intrusions*

Keywords: MODEL, INTRISION DETECTION, SMOTE, MACHINE LEARNING, CSE-CIC- MODEL2018, IMBALANCED DATASET.

I. INTRODUCTION

Due to technological developments, most of the real-world transactions have been made available in the cyber world. Thus, many operations, such as banking, shopping, online examinations, electronic commerce, and communication are used extensively within this new environment. With the widespread use of smartphones, people can connect to this global N/W and perform transactions at any time and from anywhere. Although this digitalization facilitates the daily work of human beings, due to the weakness of the servers and the newly emerged intrusion techniques, N/Ws are often attacked by the intruders who take advantage of the anonymous nature of the Internet not only to steal some information or money but also to slow down the operation of N/W services. Security administrators traditionally prefer password protection mechanisms, encryption techniques, and access controls in addition to firewalls as a means of protecting the N/W. However, these techniques are not sufficient for protecting the system

II. MOTIVATION

N/W security plays an essential role in secure communication and avomodel financial loss and crippled services due to N/W intrusions. Intruders generally exploit the flaws of popular software to mount a variety of attacks against N/W computer systems. The damage caused in the N/W attacks may vary from a little disruption in service to on developing financial loss. Recently, intrusion detection systems (MODELs) comprising MACHINE LEARNING techniques have emerged for handling unauthorized usage and access to N/W resources. With the passage of time, a wide variety of MACHINE LEARNING techniques have been designed and integrated with MODELs. Still, most of the MODELs reported poor intrusion detection results using false positive rate and detection rate. For solving these issues, we have proposed this system to increase the performance of intrusion detection through various MACHINE LEARNING algo. to give higher accuracy in intrusion detection than already proposed systems.

III. PROBLEM STATEMENT

Our aim is to develop a system that Increases the Performance of MACHINE LEARNING-Based MODELs on a Imbalanced and Up-to-Date Dataset.

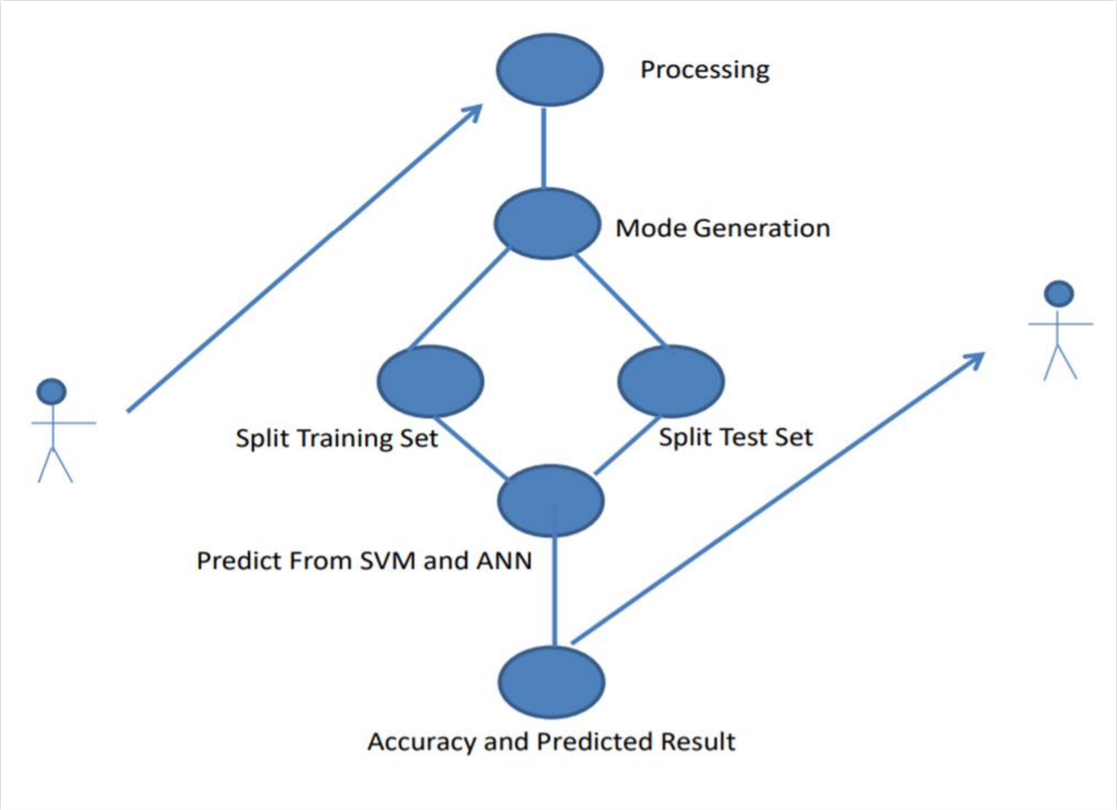
IV. LITERATURE SURVEY

Sr. No.	Name of Paper	Authors	Publication Name	Published On	Elements
1.	Increasing the Performance of MACHINE LEARNING-Based MODELs on an Imbalanced and up-to-date dataset.	GOZDE KARATAS, ONDER DEMIR, OZGUR KORAY SAHINGOZ	IEEE	2020	Compared with other algo. of the same kind, the effect of the algo. is obviously improved, and it has a great practical value.
2.	detailed investigation and analysis of using MACHINE LEARNING techniques for intrusion detection.	P. Mishra. V.Varadharajan, U.Tupakula E. S. Pilli.	IEEE	2019	Even if an optimal feature set is sufficient for analyzing the behavior of an attack, it is not good for analyzing the behavior of other attacks. Hence, there is a need to define the optimal feature subset and a suitable technique for each type of attack.
3.	Using MACHINE LEARNING to detect DoS attacks in wireless sensor N/Ws.	A. I. Al-issa, M. Al-Akhras, M. S. Alsahli, and M. Alawairdhi	IEEE	2019	The desicion trees technique achieved better(higher) true positive rate and better(lower) false positive rate than support vector machine.
4.	An adaptive ensemble MACHINE LEARNING model for intrusion detection.	X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu.	IEEE	2019	Ensemble MACHINE LEARNING has a. good generalization effect, which is Worthy of continuous promotion and optimization in the field of N/W security research and application

V. PROPOSED SYSTEM

Proposed systems were implemented in Keras/Tensorflow using the Python programming language, and Scikit learn libraries. To measure the performance metrics, experiments are executed on a workstation that has the properties shown in Table 9. Proposed systems were executed on the Multicore structure of the NVIDIA Ge Forcer GTX 1080 Ti Graphic card, whose specifications are detailed in Table 10. To calculate the performance measure of the proposed systems; Accuracy, Precision, Recall, F1-Score and Error Rate values are used.

VI. USE CASE DIAGRAM



VII. CONCLUSION

However, these minority classes are generally positive classes. Therefore, the imbalance ratio should be decreased to increase the efficiency of the system and to decrease its average accuracy. In this paper, six different MACHINE LEARNING models (Decision Tree, Random Forest, K Nearest Neighbor, Adaboost, Gradient Boosting, and Linear Discriminant Analysis) were implemented using a recent dataset (CSE-CIC-MODEL2018). O decrease the imbalance-ratio, a data sampling model was used by increasing the data size of the minority groups. The experimental results showed that the implemented models have a very good accuracy level when compared with recent literature. The use of a sampled dataset caused the average accuracy of the models to increase between 4.01% and 30.59%.

VIII. ACKNOWLEDGMENT

This paper is supported by many people, some of whom have a direct role and some of them have an indirect way by publishing their research online which helped to understand this concept easily. We express our deepest gratitude, sincere thanks and deep feeling of appreciation to our Project Guide Professor Pritam R. Ahire, his presence at any time throughout the Semester, important guidance, opinion, comment, critics, encouragement, and support greatly improved this project work. We thank the college administration for providing the necessary infrastructure and technical support. Finally, we extend our heartfelt thanks to our friends and family members.

REFERENCES

[1]. J. M. Johnson and T. M. Khoshgoftaar, ``Survey on deep learning with class imbalance," J. Big Data, vol. 6, no. 1, p. 27, 2019.
[2]. X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ``An adaptive ensemble MACHINE LEARNING model for intrusion detection," IEEE Access, vol. 7, pp. 82512_82521, 2019.
[3].R. Abdulhammed, M. Faezipour, Abumallouh, ``Deep and MACHINE LEARNING approaches for anomaly-based intrusion detection of imbalanced N/W traffic," IEEE Sens. Lett., vol. 3, no. 1, pp. 1_4, Jan. 2019.

[4].Mohammed Yasin Jisan, and M. M. Rahman, ``N/W intrusion detection using supervised MACHINE LEARNING technique with feature selection," in Proc. Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST), Jan. 2019, pp. 643_646.

[5]. A. I. Al-issa, M. Al-Akhras, M. S. Alsahli, and M. Alawairdhi, ``Using MACHINE LEARNING to detect DoS attacks in wireless sensor N/Ws," in Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT), Apr. 2019, pp. 107_112.

[6]. E. Kurniawan, F. Nhita, A. Aditsania, and D. Saepudin, ``C5.0 algo. and synthetic minority oversampling technique for rainfall forecasting in Bandung regency," in Proc. 7th Int. Conf. Inf. Commun. Technol. (ICoICT), Jul. 2019, pp. 1_5.