RESEARCH ARTICLE OPEN ACCESS

Comparative Evaluation of IDS using Machine Learning Algorithms

Om Ambavkar[1], Amit Chaurasiya[2], Roshan Chauhan[3], Prathmesh Bharti[4], Mahalaxmi Palinje[5]

Electronics & Telecommunication Department, ACE, Mumbai University Malad(W), Mumbai, India

Abstract:

In This paper we are going to train and test our dataset with different machine learning algorithm with different attacks, and also we are going to compare it's results with various parameters. This paper presents an all-inclusive survey on various examination article that utilize single, hybrid and ensemble classification algorithms. The outcomes measurements, weaknesses and dataset involves by the concentrated on articles in the improvement of IDS were look at.

Keywords: Machine learnings, Intrusion Detection Sytem, Network, and Algorithms.

I. INTRODUCTION: -

Intrusion Detection is the issue of distinguishing unapproved use, abuse, and maltreatment of PC frameworks by both system insiders and outer intruders. Most of the existing commercial IDS products are signature-based but not adaptive or self-learning. Many techniques were underway to detect the anomalies but had less success. For detecting illicit or abnormal behaviour, IDS is used. It is a security technology that is designed to automatically detect unauthorized access to computer systems, networks, or devices. It analyzes data from various sources, such as network traffic, system logs, and application data, to identify potential security threats and alert security personnel. The goal of an IDS is to prevent malicious activity, such as data theft, malware infections, or unauthorized access, by detecting anomalies in system behavior and generating alerts or taking other protective measures.

II. LITERATURE SURVEY

Comparative evaluation of Intrusion Detection Systems (IDS) using different machine learning algorithms is a research area that compares the performance of various machine learning algorithms in detecting cyber-attacks. This involves evaluating the accuracy, false positive rate, false negative rate, and other performance metrics of different machine learning algorithms in detecting various types of cyber-attacks. Some popular machine learning algorithms used for intrusion detection are:

1) Decision Trees (DT)
2) K-Nearesrt Neighbours (KNN)
3) Random Forest (RF)

The evaluation is performed by training the algorithms on a labelled dataset, which contains both normal and attack data, and then testing their ability to correctly identify new unseen instances as either normal or attack. The results of such evaluations are useful for security practitioners, researchers, and practitioners to make informed decisions about the best machine learning algorithms for intrusion detection. It is important to note that the performance of a machine learning algorithm for intrusion detection is highly dependent on the quality and size of the training dataset, the feature representation of the data, and the overall system architecture.

An intrusion detection system (IDS) is a security software that monitors a network or systems for malicious activities or policy violations and produces alerts when such incidents are detected. A literature survey of IDS involves reviewing existing research studies and articles to gain a comprehensive understanding of the state of the field. Some of the key topics covered in a literature survey on IDS include, Types of IDS: Network-based IDS, Host-based IDS, Hybrid IDS, Anomaly-based IDS, Signature-based IDS, etc.

IDS evaluation metrics: Accuracy, False The paper proposed Intrusion detection method using K-Nearest Neighbor, Decision tree, Random forest. The performance measure of Three different machine learning algorithms in detecting systems, The Three types of attack such as DoS, R2L &Probe. As a result, shows that the K- Nearest neighbor performs best in prediction accuracy compared to other algorithms.

III. RESULTS

To compare this model, you can look at the following metrics:

Accuracy: This measures how many instances are correctly classified out of the total number of instances. A higher accuracy is generally desirable.

Precision: This measures the fraction of positive instances that are correctly classified as positive. A high recall means that most positives instances are correctly classified.

Recall: This measures the fraction of positive instances that are correctly classified as positive out of all positive instances. A high recall means that most positive instances are correctly classified.

F1 Score: This is the harmonic mean of precision and recall. It is a measure of a balance between precision and recall.

Support: Support is the number of instances that each algorithm was tested on.

Denial of Service

Trained Model Parameters of kdd dataset for DOS Attack

Algorithm Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest	99.99%	100%	100%	100%	134686
Decision Tree	99.99%	100%	100%	100%	134686
KNeighbors Classifier	99.65%	100%	100%	100%	134686

Figure 3.1

Test Model Parameters of kdd dataset for DOS Attack

Algorithm Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest	85.13%	88%	83.5%	85%	17169
Decision Tree	81.65%	82%	82%	82%	17169
KNeighbors Classifier	86.66%	87%	87%	86%	17169

Figure 3.2

Remote to Local

Trained Model Parameters of kdd dataset for R2L Attack

Algorithm Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest	99.98%	100%	100%	100%	134686
Decision Tree	99.97%	100%	100%	100%	134686
KNeighbors Classifier	99.75%	100%	100%	100%	134686

Figure 3.3

Test Model Parameters of kdd dataset for R2L Attack

Algorithm	Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest		77.90%	61%	78%	68%	12456
Decision Tree		78.20%	75%	78%	75%	12456
KNeighbors Classifier		78.69%	77%	79%	71%	12456

Figure 3.4

Probe

Trained Model Parameters of kdd dataset for probe Attack

Algorithm	Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest		99.99%	100%	100%	100%	134686
Decision Tree		99.99%	100%	100%	100%	134686
KNeighbors Classifier		99.82%	100%	100%	100%	134686

Figure 3.5

Test Model Parameters of kdd dataset for probe Attack

Algorithm	Parameter →	Accuracy	precision	Recall	F1 Score	Support
Random Forest		82.05%	83%	88%	89%	12132
Decision Tree		80.17%	87%	80%	82%	12132
KNeighbors Classifier		85.52%	85%	86%	85%	12132

Figure 3.6

IV. CONCLUSION

Based on these metrics, K-Nearest Neighbors seems to perform the best overall, with the highest accuracy, precision, recall, and F1 score both in the trained and test sets. Decision Tree has lower accuracy and F1 score compared to Random Forest, but its precision and recall are almost equal. Random Forest Classifier has the second highest accuracy in the test set but its precision, recall and F1 score are lower than K-Nearest Neighbors and Decicion Tree Classifier has the lowest accuracy in the test set but its precision, recall and F1 score.

V. REFERENCES

[1] M. Alkasassbeh and M. Almseidin, “Machine Learning Methods for Network Intrusion Detection.” World Academy of Science, Engineering and Technology International Journal

of Computer and Information Engineering Vol:12, No:8, 2018.

[2] Usman Shuiabu, Musu, Sudesha Chakraboty," A Review on Intrusion Detection System Using Machine Learning Techniques".2021 International conference on computing, Communication & Intelligent System vol:12, No:9, 2018

[3] Prachiti Parkar's "A Survey On Cyber Security IDS Using ML Methods". 2021 5th International Confereence on Intelligent company and control system. CFP21K74-ART; ISBN: 978-0-7381-1327-2

[4] I. Abrar, Z. Ayub, M. Faheem, Alwi Bamhdi , "A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset," no. 4, 2020,doj:10.1109/ICOSEC49089.2020.9215232

[5]Hamza Nachan , Pratik Kumhar , Simran Birla , Dristi Poddar, Sambhaji Sarode, " Intrusion Detection System: A Survey,". Volume 10 Issue05(May2021),DOI:10.17577/IJERTV10I S050479

[6] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using Bagging with Partial Decision Tree base classifier," Procedia Comput. Sci., vol. 49, no. 1, pp. 92–98, 2015, doi: 10.1016/j.procs.2015.04.231.

[7]W. C. Lin, S. W. Ke, and C. F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," Knowledge-Based Syst., vol. 78, no.1,pp.13–21,2015,doi: 10.1016/j.knosys.2015.01.009.

[8]A. S. Amira, S. E. O. Hanafi, and A. E. Hassanien,"Comparison of classification techniques applied for network intrusion detection and classification," J. Appl. Log., vol. 24,pp.109–118,2017,doi: 10.1016/j.jal.2016.11.018.

[9]Zena Khalid Ibrahim, Mohammed Younis Thanon, "Performance Comparison of Intrusion Detection System Using Three Different Machine Learning Algorithms,". ISBN: 978-1- 7281-8501-9, Doj: 10.1109\ICIT50816.2021