ISSN: 2278-4632
Vol-13, Issue-04, No.06, April : 2023

CYBER-ATTACK DETECTION IN NETWORK USING MACHINE LEARNING TECHNIQUES

Ch. Sowjanya[1] Assistant professor, Department of Computer Science and Engineering, Amrita Sai institute of science and technology, Paritala, Andhra Pradesh India.
, **N. Anusha**[2], Assistant professor, Department of Computer Science and Engineering, Amrita Sai institute of science and technology, Paritala, Andhra Pradesh India.
D. Srikanth[3]**,** Assistant professor, Department of Computer Science and Engineering, Amrita Sai institute of science and technology, Paritala, Andhra Pradesh India.
M. Aadil[4], Assistant professor, Department of Computer Science and Engineering, Amrita Sai institute of science and technology, Paritala, Andhra Pradesh India.
Sk. shehnaz[5] Assistant professor, Department of Computer Science and Engineering, Amrita Sai institute of science and technology, Paritala, Andhra Pradesh India.

Abstract
Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of new innovations give incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults. In this paper, this problem is solved by proposing a intrusion detection system through ML models with majority voting technique using Random Forest, Decision Tree, Logistic Regression, Support Vector Machine to detect the attack in network using specific parameters with high accuracy and efficiency.
Keywords: Decision tree, Random Forest, Support Vector Machine (SVM), Logistic Regression**.**

1.Introduction

Today, political and commercial entities are increasingly engaging in cyber-attacks to damage, disrupt, or censor information content in computer networks. In designing network protocols, there is a need to ensure reliability against intrusions of powerful attackers that can even control a fraction of parties in the network. Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly attacks like denial of service. most systems built based on such techniques suffer from high false positive and false negative detection. Firewall only controls every access from network to network, which means prevent access between networks. But it does not detect in case of an internal attack. In recent days, cyber-security and protection against numerous cyber-attacks are becoming a burning question. The main reason behind that is the tremendous growth of computer networks and the vast number of relevant applications used by individuals or groups for either personal or commercial use, especially after the acceptance of the Internet of Things (IoT). So, it is obvious to develop accurate defense techniques such as machine learning-based intrusion detection system (IDS) for the system's security in general, an intrusion detection system (IDS) is a system or software that detects infectious activities and violations of policy in a network or system. An IDS identifies the inconsistencies and abnormal behavior on a network during the functioning of daily activities in a network or system used to detect risks or attacks related to network security, like denial-of- service (Dos). An intrusion detection system also helps to locate, decide, and control unauthorized system behavior such as unauthorized access, or modification and destruction.

2.Literature Review

The following research papers were referred by us before doing our project. While referring each of these papers we have come across various different findings discussed below.

DOI: 10.36893.JK.2023.V13I04N16.0023-0026

ISSN: 2278-4632
Vol-13, Issue-04, No.06, April : 2023

Title: Port Scanning techniques and the defense against them.
Author: R. Christopher
Title: Practical automated detection of stealthy port scans.
Author: S. Staniford, J.A. Hoagland, J.M. McAlerney.
Title: Combined analysis of support vector machine and principle component analysis for ids.
Author: M.C. Raja, M.M.A. Rabbani
Method Used: support vector machine.
Title: Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model.
Author: S. Aljawarneh, M. Aldwairi, M.B. Yassein.
Method Used: knn, support vector machine, decision tree.

3 Problem Definition

The continuous attacks in network, protection of significant data, security of put away information stages, accessibility of information and so forth. upon these issues, digital fear is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues in people and establishments. To overcome this problem by using machine learning algorithms with intrusion detection system has been created to maintain a distance from digital attacks.

4. Methodology

While implementing the project, the following steps were implemented in order to achieve the results.

Data Cleaning and Preprocessing

One of the first steps is to make sure that the dataset we are using is accurate. The dataset should not have any missing values and if the dataset does have missing values, they should be replaced by the appropriate values. The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed.

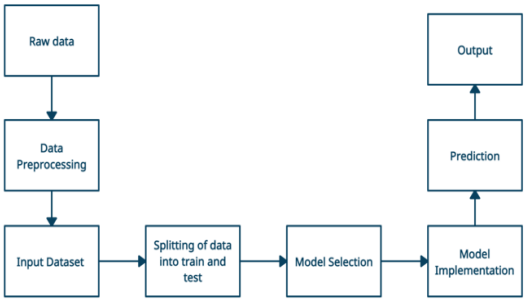
Feature Selection

It is important that we select only those features that will be necessary to determine the type attack class. For this, we have created a correlation matrix that shows the linear relationship of a feature with every other features. If features are highly correlated then that feature should be dropped.

Model Building

The next step is building the machine learning model. While building the machine learning model, first we need to split our dataset into 2 parts i.e.: training data and test data. We have split the data in the ratio of 80-20. Taking the training data, we apply our machine learning algorithms on the features of the dataset. We have used 4 machine learning algorithms on our training dataset and the algorithms that gives us the highest accuracy will be selected for recommendation.

Architecture diagram of our machine learning model



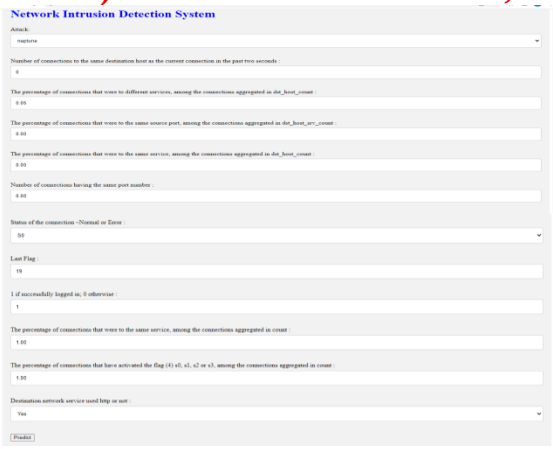
Building a UI

In the next step, we have built a UI for a user to input his data. so that once he enters the information in the intrusion detection system such as last flag, logged_in, connection status, count, serror_rate etc., the model will process the data and will gives the results of which type of the attack class occurred in the network.

Below is a screenshot of the UI is the intrusion detection system.

DOI: 10.36893.JK.2023.V13I04N16.0023-0026

ISSN: 2278-4632
Vol-13, Issue-04, No.06, April : 2023



After putting the values by the user, we get the results of the attack class in the way mentioned below

5. Dataset

The dataset for this topic was taken from Kaggle.

6. Machine Learning Algorithms Used

Random Forest

Random Forest is a supervised ensemble machine learning algorithm used in both classification as well as regression problems. It contains various decision trees and an average of it is taken so as to give the output

Decision Tree

Decision Tree is one of the most popular machine learning algorithms used mostly in classification problems but can also be used for regression type of problems. The working of it is based on a simple technique, wherein a yes/no question is asked and according to the answer the tree is split in smaller nodes.

Support Vector Machine:

put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

Logistic Regression

It is one of the simplest algorithms in machine learning. It is used for solving classification problems. It uses a sigmoid function to mathematically calculate the probability of an observation and accordingly, the observation is then put into its respective class.

7. Results

The algorithms we have used are Decision Tree, Random Forest, Logistic Regression, Support vector machine. After implementing the algorithms on our dataset, we can see that Random Forest and Decision Tree gives us the highest accuracy out of all the algorithms.

8. CONCLUSION

Nowdays, everyone are using internet so security is very important. If they know any attack is happened or not, like if attack is happened further actions in order to provide a security will be taken.

We are using machine learning algorithms like random forest, decision tree, logistic regression, support vector machine to train the machine regarding attacks. So that we are giving information regarding attack. By using dataset we are going to predict whether cyber attack is done or not. Using the higher

DOI: 10.36893.JK.2023.V13I04N16.0023-0026

accuracy rate algorithm that helps to predict best results for the early and efficient detection of the cyber attack in the network

REFERENCES

K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.
R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.
M. Baykara, R. Das¸, and I. Karado ˘gan, "Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.
S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.
S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.
K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.
N. Moustafa and J. Slay, "The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems," in Building Analysis Datasets and Gathering
Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.
L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.
S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.
M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.
S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.
I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.
D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.
N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.
P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.
C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.
R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.

 DOI: 10.36893.JK.2023.V13I04N16.0023-0026