International Journal for Modern Trends in Science and Technology, 8(06): 199-202, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0806032
Available online at: http://www.ijmtst.com/vol8issue06.html

Detection of Cyber Attack in Network using Machine Learning Techniques

A. Pavan Kumar[1], P. Divya[2], J. Bhumika[2], Y.V.V.Thanusha[2], N.Kesava[2]

[1]Assistant Professor, Department of Computer Science & Engineering, Sri Vasavi Institute of Engineering & Technology, Pedana, A.P, India
[2]Department of Computer Science & Engineering, Sri Vasavi Institute of Engineering & Technology, Pedana, A.P, India

To Cite this Article
A. Pavan Kumar, P. Divya, J. Bhumika, Y.V.V.Thanusha and N.Kesava. Detection of Cyber Attack in Network using Machine Learning Techniques. International Journal for Modern Trends in Science and Technology 2022, 8(06), pp. 199-202. https://doi.org/10.46501/IJMTST0806032

Article Info
Received: 09 May 2022; Accepted: 02 June 2022; Published: 07 June 2022.

ABSTRACT

The expanded use of cloud administrations, developing number of web applications clients, changes in network framework that interfaces gadgets running versatile working frameworks and continually advancing organization innovation cause novel difficulties for digital security. Traditional techniques for interruption discovery and profound parcel assessment, while still generally utilized and suggested, are not, at this point adequate to satisfy the needs of developing security dangers. As processing power increments and cost drops, Machine Learning is viewed as an elective technique or an extra component to shield against malwares, botnets, and different assaults. The significant commitment of the paper is the suggestion of Machine Learning way to deal with model typical conduct of utilization and to identify digital attacks

Keywords: Cyber Attack, attack detection, security

1. INTRODUCTION

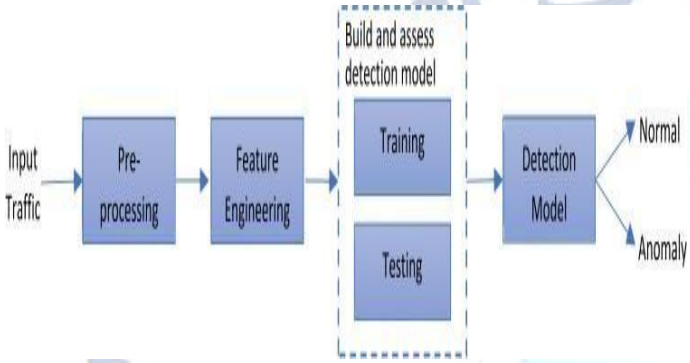
Computers and IP networks try, through illegal access or unlawful use of an asset, to expose, alter, disable, damage, corrupt, or acquire information. an attack is committed. An attack is an aggressive action against computer systems, infrastructures, computer networks or individual computing devices. An intruder is an individual or process who, without authorization and with a possible harmful purpose, seeks access to data, functions or other limited systems areas. Cyber threats may be part of information warfare and cyber warfare depending on circumstances. Sovereign nations, persons, groups, societies, or enterprises might use a cyber-warfare and it could come from an anonymous source. Sometimes a cyber weapon is a product that helps a cyberwarfare .By accessing into a vulnerable system, a cyber threat can steal, change or destruct a specific target. Cyber invasions can vary from the installation of malware on a computer to the destruction of whole nations' infrastructure. Legal scholars try to restrict the usage of the word to physical damage events and differentiate it from more common privacy violations and more comprehensive hacking. Increasingly complex and deadly are cyber intrusions.

Types Of Cyber attacks:
Injection attacks
DNS Spoofing
Phishing
Denial of Service
Man in the middle attacks

PROPOSED SYSTEM :

The proposed system primarily follows a set of steps. We can achieve our goal if we follow those steps. The algorithm's key steps are listed below. 1) Every dataset should be normalised. 2) Subdivide the original dataset into testing and training datasets. 3) Create IDS models using Logistic Regression, Decision Tree, Random Forest, and MLP. 4) Evaluate the performance of each model.

ARCHITECTURE:



DataSet:

For our application in order to run effectively we've chosen NSL-KDD dataset which is having 42 features initially and maintain 4 attack category's.

NSL-KDD is a useful benchmark data set for comparing different intrusion detection methods.

Attribute Number	Features	Description
1	duration	Length of the time duration of the connection
2	protocol_type	Protocol used
3	service	Service used by destination network
4	flag	Status of the connection (Error or Normal)
5	src_bytes	Number of data bytes transferred from source to destination
6	dst_bytes	Number of data bytes transferred from destination to source
7	land	If source and destination port no. and IP addresses are same then it will set as 1 otherwise 0
8	wrong_fragment	Total number of wrong fragments in a connection
9	urgent	Number of urgent packets (these packets with urgent bit activated)
10	hot	Number of 'hot' indicators means entering in a system directory
11	num_failed_logins	Number of failed login attempts
12	logged_in	Shows login status (1- successful login, 0- otherwise)
13	num_compromised	Number of compromised conditions
14	root_shell	Shows root shell status (1-if root shell obtained otherwise 0)
15	su_attempted	Set as 1 if 'su_root' command used otherwise set as 0
16	num_root	Number of operations performed as root
17	num_file_creations	Number of file creation operations
18	num_shells	Number of shell prompts in a connection
19	num_access_files	Number of operations on access control files
20	num_outbound_cmds	Number of outbound commands in a ftp session
21	is_host_login	If login as root or admin then this set as 1 otherwise 0
22	is_guest_login	Set as 1 if login as guest otherwise 0
23	count	Number of connections to the same destination host
24	srv_count	Number of connection to the same service (port number)
25	serror_rate	Percentage of connections that have activated flag (#4) s0,s1,s2 or s3, among the connections aggregated in count (#23)
26	srv_serror_rate	Percentage of connection that have activated flag (#4) s0,s1,s2 or s3, among the connections aggregated in srv_count (#24)
27	rerror_rate	Percentage of connections that have activated flag (#4) REJ, among the connections aggregated in count (#23)
28	srv_rerror_rate	Percentage of connections that have activated flag (#4) REJ, among the connections aggregated in srv_count (#24)
29	same_srv_rate	Percentage of connections that were to the same services, among the connections aggregated in count (#23)
30	diff_srv_rate	Percentage of connections that were to the different services, among the connections aggregated in count (#23)
31	srv_diff_host_rate	Percentage of connections that were to different destination machines among the connections aggregated in srv_count (#24)
32	dst_host_count	Number of connections having the same destination host IP address
33	dst_host_srv_count	Number of connections having same port number
34	dst_host_same_srv_rate	Percentage of connections that were to the same service among the connections aggregated in dst_host_count (#32)
35	dst_host_diff_srv_rate	Percentage of connections that were to different service among the connections aggregated in dst_host_count (#32)
36	dst_host_same_src_port_rate	Percentage of connections that were to the same source port among the connections aggregated in dst_host_srv_count (#33)
37	dst_host_srv_diff_host_rate	Percentage of connections that were to the different destination machines among the connections aggregated in

Implementation:

Logistic Regression

Logistic regression is one of the most common algorithms of machine learning under the Supervised Learning Technology. It is used with a group of individual variables to predict the category dependent variable. The output of a categorically dependent variable is forecast by logical regression. The result must thus be either categorical or discrete. It can be either YES or No, 0 or 1 true or False, but it offers the probabilistic values between 0 and 1 instead of giving the precise value of 0 and 1

Random Forest

Random Forest is a prominent method for Machine learning through the controlled learning process i.e.., supervised learning. It may be utilized in ML for issues like classification and regression. It is built on the notion of ensemble learning, a process through which numerous classifications are combined in order to resolve a complicated problem and to enhance model performance

Decision Tree

Decision Tree is a Supervised learning method that might be used to solve regression and classification tasks, and it is also most commonly further used solve nonlinear equations. A Decision tree will have two set of nodes: Decision nodes are often used to make major decisions and also have numerous

branches, while other Leaf nodes seem to be the outcome of all those decision making but have no additional branch offices. The decisions or tests are based on the features of the data set.

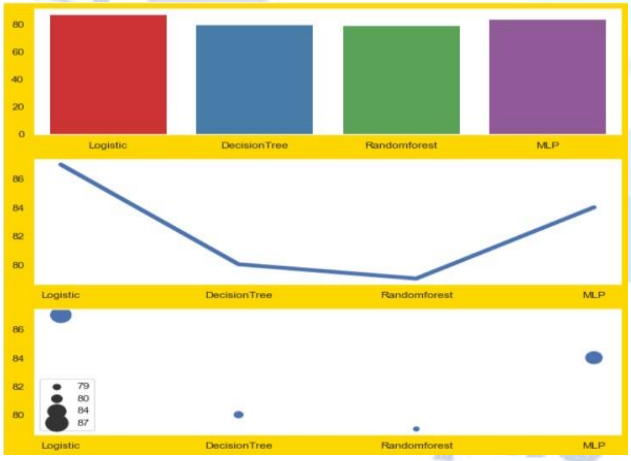
MLP classifier

A multilayer perceptron (MLP) is a type of machine learning that uses feed forward (ANN). The term MLP is used vaguely; it can apply to any feed forward ANN, or it can refer to channels made up of various layer upon layer of perceptrons (with threshold activation).Multilayerperceptrons were often meant to refer to as "vanilla" neural network models – particularly when only one convolution layer is present.

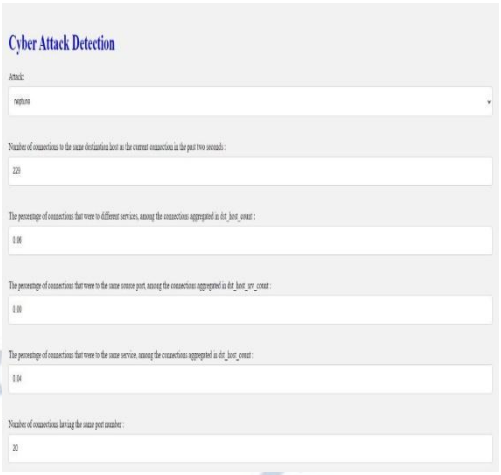
Data Visualization

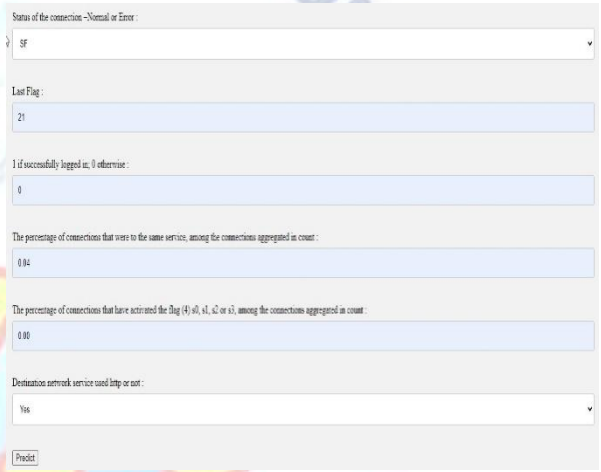
Data Visualization is the discipline of attempting to comprehend data by presenting it in a visual context in order to uncover patterns, trends, and connections that might otherwise go undetected.Python has a number of excellent graphing packages that are packed with useful functionality. Python offers a great library for creating interactive, real-time, or highly customizable plots.

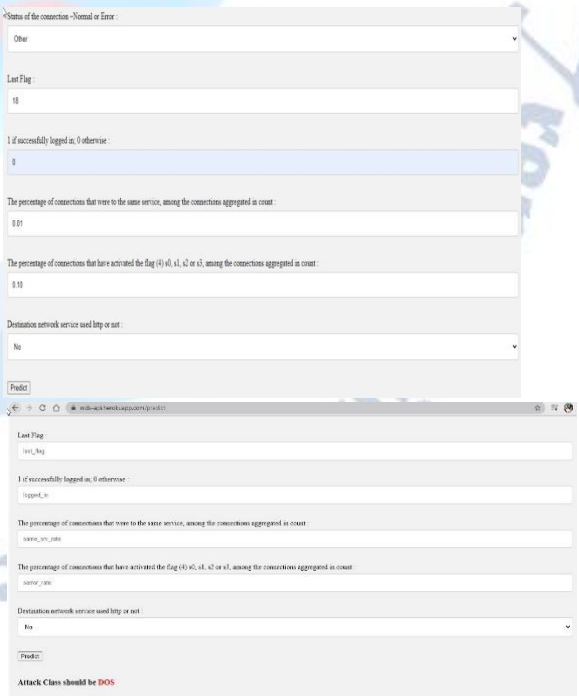
The below figure shows the visualization of ML Algorithms and their accuracies in the form of bar-plot , line-plot and scatterplot.

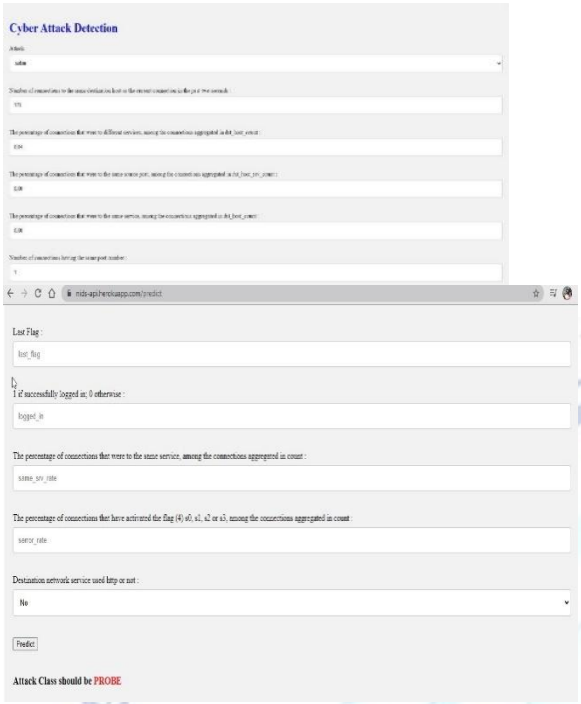


Below figures shows the running and execution of project









CONCLUSION:

In conclusion, our objective is to create and comparison models capable of detecting attacks in a practical NSL-KDD network traffic. A thorough examination of the data and a lot of network security paper studies resulted in the extraction of the essential characteristics. Displaying a Logistical Regression, Random Forestry, Decision Tree and MLP classification methods were then evaluated. In all other cases, Logistic Regression was used to detect attacks, yielding a detection accuracy of greater than 85%.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] V. Ford and A. Siraj, "Applications of machine learning in cyber security," in Proceedings of the 27th International Conference on Computer Applications in Industry and Engineering, 2014.

[2] H. Jiang, J. Nagra, and P. Ahammad, "Sok: Applying machine learning in security-a survey," arXiv preprint arXiv:1611.03186, 2016.

[3] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson, "Shallow and deep networks intrusion detection system: A taxonomy and survey," arXiv preprint arXiv:1701.02145, 2017.

[4] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," in 2018 10th International Conference on Cyber Conflict (CyCon), 2018: IEEE, pp. 371-390.

[5] S. Sheikhi, M. Kheirabadi, and A. Bazzazi, "An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network," International Journal of Engineering, vol. 33, no. 2, pp. 221-228, 2020.