WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

INTRUSION DETECTION USING MACHINE LEARNING ALGORITHMS

A.srinivasan[1] , Ms. G.Umamaheswari[2]

II MCA[1] , Department of Computer Science and Applications,
Assistant Professor[2] , Department of Computer Science and Applications,
Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur, Tamil Nadu, India.
srinivasanmca2910@gmail.com

ABSTRACT:

A software programme called an intrusion detection system (IDS) is designed to watch over network or system activity and detect any harmful activity. The massive expansion and use of the internet creates questions about how to securely store and transmit digital information. Hackers now employ a variety of techniques to obtain vital information. New things like viruses and worms are imported as the internet enters society. In order to weaken the system, malicious individuals employ various methods like password cracking and the detection of unencrypted information. Therefore, security is required for users to protect their system from hackers. One of the common types of defence is the firewall technique. the private network from the public network, and it serves this purpose. IDS are utilised by insurance companies, medical applications, credit card fraud, and network-related operations. These attacks can be found using a variety of intrusion detection techniques, methods, and algorithms. The primary goal of this project is to present a comparative analysis of several machine learning and deep learning algorithms for intrusion detection. In real-time network datasets like Intrusion Detection System (IDS) datasets and UNSW datasets, a variety of machine learning algorithms, including Back Propagation, Feed Forward, Recurrent Neural Network, and Multilayer Perceptron (MLP), have been utilised to build IDs. . MLP is a popular neural network classifier based on the number of classes (output) and hidden layers. MLP employs weights for each node at the neural network; the most useful attributes will receive large weights, whereas variables that have little impact on predicting class will receive smaller weights. The proposed system may be implemented in a Python tool for performance analysis and its error rate and accuracy numbers can be evaluated.

Keywords: Intrusion, Network, Data, Detection

1.Introduction:

An intrusion detection system (IDS) is a tool or software programme that keeps an eye out for hostile activities or rule violations on a network or in a system. Any unlawful behaviour or violation is often recorded either centrally using a security information and event management (SIEM) system or notified to an administrator. In order to separate harmful activity from false alerts, a SIEM system integrates outputs from many sources and employs alarm filtering mechanisms. IDS types come in a variety of sizes, from small networks to many machines. Network intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS) are the two most used categories. An example of a HIDS is a system that keeps track of crucial operating system files, whereas an example of an NIDS is a system that examines incoming network traffic. IDS can also be categorised according to detection methods. The most well-known variations are anomaly-based detection (which frequently uses machine learning) and signature-based detection (which recognises deviations from a model of "good" traffic, such as malware). An intrusion detection system (IDS) is a tool or software programme that keeps an eye out for hostile activities or rule violations on a network or in a system. Any unlawful behaviour or violation is often recorded either centrally using a security information and event management (SIEM) system or notified to an administrator. In order to separate harmful activity from false alerts, a SIEM system integrates outputs from many sources and employs alarm filtering mechanisms. IDS types come in a variety of sizes, from small networks to many machines.An example of a HIDS is a system that keeps track of crucial operating system files, whereas an example of an NIDS is a system that examines incoming network traffic. IDS can also be categorised according to detection methods. The most well-known variations are anomaly-based detection (which frequently uses machine learning) and signature-based detection (which recognises deviations from a model of "good" traffic, such as malware).

2.PURPOSE OF THE PROJECT

The IDS can be identified based on the location of the detection and the method or methodology used to make the detection. IDS are divided into two specific niches: host intrusion detection systems (HIDS) and network intrusion detection systems (NIDS). The first system stated assists in the analysis of incoming networking traffic, whereas HIDS operation depends on operating system activity. Clustering and classification were the primary data mining on IDS topics that were initially covered. Since the initial data set for the clustering problem lacked a label, the object produced by the clustering algorithm

was given the same class as records with comparable data. Depending on the qualities and characteristics of the already existing data, the behaviour of the packet was classified as either normal or abnormal. This method of classification uses data that has already been grouped. This suggests that the information is labelled. A data mining technique called classification is employed to examine a data set. . Clustering and classification were the primary data mining on IDS topics that were initially covered. Since the initial data set for the clustering problem lacked a label, the object produced by the clustering algorithm was given the same class as records with comparable data. Depending on the qualities and characteristics of the already existing data, the behaviour of the packet was classified as either normal or abnormal. This method of classification uses data that has already been grouped. This suggests that the information is labelled. A data mining technique called classification is employed to examine a data set.

3.SYSTEM ANALYSIS

Software Engineering

Software engineering is a subfield of engineering that involves creating software products according to established scientific concepts, practises, and guidelines. A reliable and effective software product is the result of software engineering.

4.EXISTING SYSTEM

The IDS can be distinguished based on the location of the detection and the method or methodology used to make the detection. IDS are divided into two specific niches: host intrusion detection systems (HIDS) and network intrusion detection systems (NIDS). The first system stated assists in the analysis of incoming networking traffic, whereas HIDS operation depends on operating system activity. Clustering and classification were the primary data mining on IDS topics that were initially covered. Since the initial data set for the clustering problem lacked a label, the object produced by the clustering algorithm was given the same class as records with comparable data. Depending on the qualities and characteristics of the already existing data, the behaviour of the packet was classified as either normal or abnormal. This method of classification uses data that has already been grouped. This suggests that the information is labelled. A data mining technique called classification is employed to examine a data set. The classification of the data is crucial in today's environment of continuously flowing data. To categorise the data, a variety of methods are utilised, including decision trees, rule-based induction, Bayesian networks, evolutionary algorithms, etc. To detect the intrusion from network datasets, machine learning techniques including Random Forest, Naives Bayes, and Support Vector Machine algorithms are implemented in the current framework. The current framework can offer significant false alarm rates and poor accuracy.

5.SOLUTION OF THESE PROBLEMS IN PROPOSED SYSTEM

Artificial neural network learning. We can process a big number of objects in the application domain using deep learning methodology in order to train. Millions of data points are subjected to a process. Deep learning uses the data to discover features. If there is a lot of data available, the system's performance may suffer. Deep learning is a learning technique that is highly suited for improving accuracy in terms of performance. Three main categories of learning exist: supervised, semi-supervised, and unsupervised. In this case, the deep learning approach is taken into consideration when performing the intrusion detection. The term "intrusion" refers to a breach in a network's or computer system's security. Another is the process of detecting intrusions, which is called intrusion detection. Anomaly detection and abuse detection are the two categories under which intrusion detection techniques are categorised. Security has emerged as a critical concern for computer systems due to the recent decade's rapid proliferation of computer networks. In recent years, various machine learning-based approaches for the creation of intrusion detection systems have been developed. A neural network approach to intrusion detection is presented in this study. Based on an off-line analysis strategy, an intrusion detection system called a Multi-Layer Perceptron (MLP) is deployed. This research tries to tackle a multi-class problem in which the type of assault is additionally recognised by the neural network, whereas most prior studies have concentrated on classifying records in one of the two main classes—normal and attack. Layered feed forward networks, such as MLP, are frequently trained using static back propagation (BP). Numerous applications requiring static pattern classification have included such networks. One input layer, one or more hidden layers, and one output layer make up the flexible MLP model.

6.SYSTEM ARCHITECTURE

A system architecture, sometimes known as a systems architecture, is a conceptual model that describes a system's behaviour, structure, and other aspects. A formal description and representation of a system that is set up to facilitate analysis of its structures and behaviours is called an architecture description. System architecture might include system elements, those elements' outwardly perceptible characteristics, and the connections (like behaviours) between them. It can offer a blueprint from which systems and products that will cooperate to implement the whole system can be developed. The architecture description languages (ADLs) collectively refer to efforts to formalise languages that describe system architecture.

Various organizations define systems architecture in different ways, including:

A physical element allocation that gives the design solution for a consumer product or life-cycle process that aims to satisfy the functional architecture's and the requirements baseline's needs.

The most significant, widespread, top-level strategic innovations, decisions, and related justifications for the overall structure (i.e., the fundamental components and their relationships), together with related traits and behaviours, are comprised of architecture.
If it is documented, it could contain details like a thorough list of the hardware, software, and networking capabilities now in use; an outline of long-term objectives and priorities for future purchases; and a strategy for updating and/or replacing out-of-date gear and software.

The synthesis of product design architectures and life-cycle procedures.

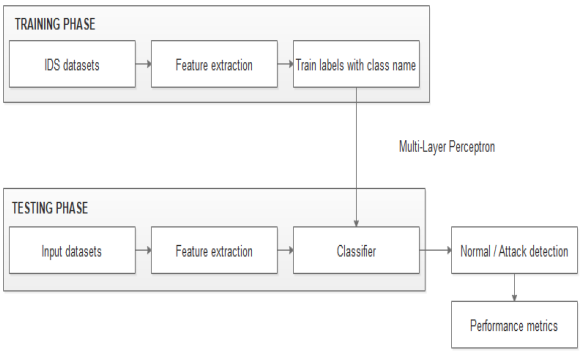


Fig.1 system architecture

7.SOFTWARE REQUIREMENT SPECIFICATION

A. DATASETS ACQUISITION

In our studies, we make use of the KDD Cup dataset, which is designed to benchmark intrusion detection problems. The dataset is a collection of simulated raw TCP dump data collected over a LAN for nine weeks. From seven weeks of network traffic, the training data was processed to produce around 5 million connection records, and from two weeks of testing data, approximately 2 million connection records were produced. Upload the UNSW datasets as well. The IXIA Perfect Storm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) created the raw network packets of the UNSW-NB 15 dataset in order to produce a hybrid of genuine current normal activities and synthetic contemporary attack behaviours. We can publish the network datasets in this module to the csv file

B. PREPROCESSING

The [data mining] process includes a crucial phase called data pre-processing. The adage "garbage in, garbage out" is especially relevant to initiatives involving data mining and machine learning. The methods used to collect data are frequently not tightly regulated, which leads to out-of-range numbers, impossible data combinations, missing values, etc. Therefore, before performing an analysis, it is crucial to consider the representation and quality of the data. Knowledge discovery during the training phase is more challenging if there is a lot of redundant, irrelevant information available, or noisy data. The tasks involved in data preparation and filtering can take a long time to process. Remove any unnecessary or missing values from datasets that have been submitted in this module.

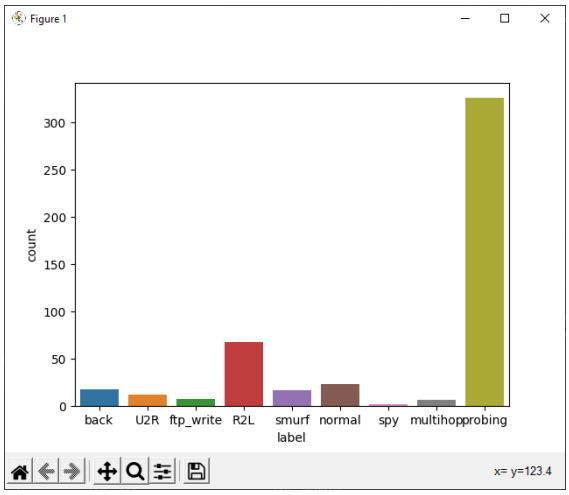
C. FEATURES EXTRACTION

A way of creating combinations of the variables to get past these issues while still accurately describing the data is known as feature extraction. Many machine learning experts think that the secret to efficient model creation is well optimised feature extraction. The process of selecting a portion of the first characteristics is known as feature selection. In order to do the intended task using this reduced representation rather of the whole starting data, it is expected that the selected features will contain the pertinent information from the input data. From pre-processed datasets, we can choose from a wide range of attributes in this module.

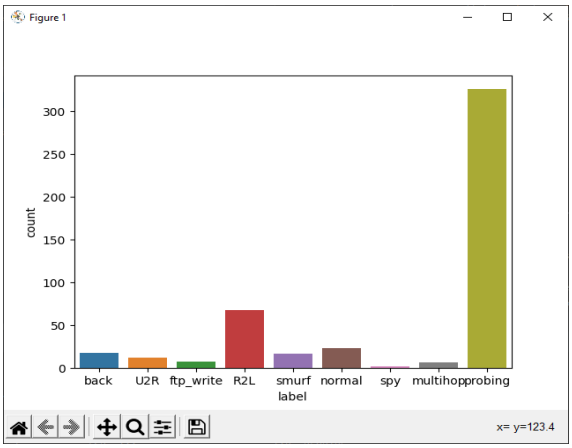
D. CLASSIFICATION

More and more organisations are becoming vulnerable to a larger range of attacks as computer network activities and sensitive information on network systems multiply. Protecting network systems from infiltration, interruption, and other anomalous behaviours by unauthorised attackers becomes of utmost importance. Use machine learning and deep learning techniques in this module to find the intrusion. A type of feed-forward artificial neural network is called a multilayer perceptron (MLP). At least three layers of nodes make up an MLP. Each node, with the exception of the input nodes, is a neuron that employs a nonlinear activation function. Back propagation is a learning method that is used by MLP during training. . MLP differs from a linear perceptron due to its numerous layers and non-linear activation. It can discriminate between data that cannot be separated linearly. Especially when they feature a single hidden layer, multilayer perceptron neural networks are commonly referred to as "vanilla" neural networks. A perceptron is a linear classifier, meaning it uses a straight line to divide two groups in order to categorise input. Select the classify option in the Python tool and then select the functions choices to perform a multilayer perceptron based on the class attribute.

TRAIN THE DATASETS



TRAINED SETS:



8.CONCLUSION & FUTURE ENHANCEMENT

Given how frequently apps and their behaviour change, intrusion detection is crucial to network security. In recent years, there has been a lot of study done on network intrusion detection, and many different methods, including machine learning and deep learning methods, have been proposed. The requirement for precise network flow classification grew as a result. For the precise categorization of intrusion detection, we have here suggested a deep learning model using a multi-layer perceptron with feature selection. In this project, we've shown how to build a thin neural network that can identify network intrusions in real time. We have also shed more light on the various classification systems' processes as a result of this procedure. We talked about potential data processing methods., which are transferable to other supervised machine learning techniques. We also described a quick technique for locating crucial neural network properties based on link weights. further contrasted the BPN, FNN, and RNN algorithms with the deep learning algorithm (MLP). Comparisons are made using accuracy and error measurements. Compared to the current machine learning algorithms, MLP can give lower error metrics and the best accuracy.

To increase intrusion detection accuracy and lower the false alarm rate in the future, we may also apply sophisticated neural network algorithms as convolutional neural networks using MATLAB Toolbox.

9.REFERENCES:

1. Van Rossum, Guido, and Fred L. Drake. The python language reference manual. Network Theory Ltd., 2011.
2. Van Rossum, Guido, and Fred L. Drake. The python language reference manual. Network Theory Ltd., 2011.
3. Dierbach, Charles. Introduction to Computer Science using Python: A Computational Problem-Solving Focus. Wiley Publishing, 2012.
4. James, Mike. Programmer's Python: Everything is an Object Something Completely Different. I/O Press, 2018.
5. Reges, Stuart, Marty Stepp, and Allison Obourn. Building Python Programs. Pearson, 2018.