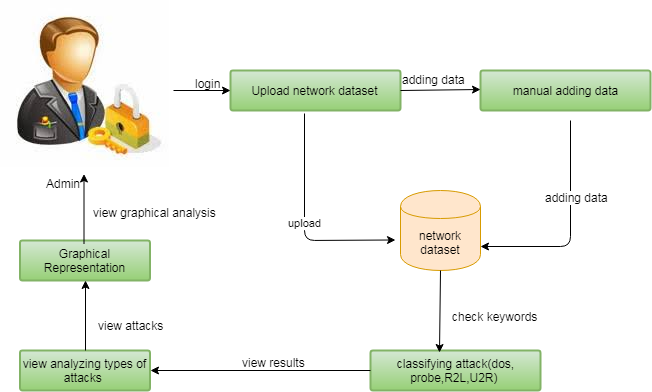
**Machine Learning and Deep Learning Methods for Cybersecurity**

**ABSTRACT:**

**T**he development of the Internet, cyber-attacks are changing rapidly and the cyber security situation is not optimistic. This survey report describes key literature surveys on machine learning (ML) and deep learning (DL) methods for network analysis of intrusion detection and provides a brief tutorial description of each ML / DL method. Papers representing each method were indexed, read, and summarized based on their temporal or thermal correlations. Because data are so important in ML / DL methods, we describe some of the commonly used network datasets used in ML / DL, discuss the challenges of using ML / DL for cyber security and provide suggestions for research directions.

**ARCHITECTURE:**

****

**EXISTING SYSTEM:**

Cybersecurity is a set of technologies and processes designed to protect computers, networks, programs and data from attacks and unauthorized access, alteration, or destruction. A network security system consists of a network security system and a computer security system. Each of these systems includes firewalls, antivirus software, and intrusion detection systems (IDS). IDSs help discover, determine and identify unauthorized system behavior such as use, copying, modification and destruction. The purpose of this paper is for those who want to study network intrusion detection in ML/DL. Thus, great emphasis is placed on a thorough description of the ML/DL methods, and references to seminal works for each ML and DL method are provided. Examples are provided concerning how the techniques were used in cyber security. Similarly to ML methods, DL methods also have supervised learning and unsupervised learning. Learning models built under different learning frameworks are quite different. The benefit of DL is the use of unsupervised or semi-supervised feature learning and hierarchical feature extraction to efficiently replace features manually. With the increasingly in-depth integration of the Internet and social life, the Internet is changing how people learn and work, but it also exposes us to increasingly serious security threats. How to identify various network attacks, particularly not previously seen attacks, is a key issue to be solved urgently.

**PROPOSED SYSTEM:**

The concept of DL was proposed by Hinton et al. based on the deep belief network (DBN), in which an unsupervised greedy layer-by-layer training algorithm is proposed that provides hope for solving the optimization problem of deep structure. Then the deep structure of a multi-layer automatic encoder is proposed. In addition, the convolution neural network proposed by Lecun et al. is the first real multi-layer structure learning algorithm that uses a space relative relationship to reduce the number of parameters to improve the training performance. Feature processing is the process of putting domain knowledge into a feature extractor to reduce the complexity of the data and generate patterns that make learning algorithms work better. Feature processing is time-consuming and requires specialized knowledge. In ML, most of the characteristics of an application must be determined by an expert and then encoded as a data type. Features can be pixel values, shapes, textures, locations, and orientations. The performance of most ML algorithms depends upon the accuracy of the features extracted. Dataset Intrusion Detection Dataset was applied to the 3rd International Knowledge Discovery and Data Mining Tools Contest. This model identifies features between intrusive and normal connections for building network intrusion detectors. In the NSL-KDD dataset, each instance has the characteristics of a type of network data. It contains 22 different attack types grouped into 4 major attack types. Dos back, Neptune, smurf, teardrop, land, pod Probe Satan, portsweep, ipsweep, nmap. R2L Warezmaster, warezclient, ftpwrite, guesspassword, imap, multihop, phf, spy U2R Rootkit, butteroverflow, loadmodule, perl.

**ALGORITHM:**

**SUPPORT VECTOR MACHINE**

Support Vector Machine (SVM) is one of the most robust and accurate methods in all machine-learning algorithms. It primarily includes Support Vector Classification (SVC) and Support Vector Regression (SVR). The SVC is based on the concept of decision boundaries. A decision boundary separates a set of instances having different class values between two groups. The SVC supports both binary and multi-class classifications. The support vector is the closest point to the separation hyperplane, which determines the optimal separation hyperplane. In the classification process, the mapping input vectors located on the separation hyperplane side of the feature space fall into one class, and the positions fall into the other class on the other side of the plane. In the case of data points that are not linearly separable, the SVM uses appropriate kernel functions to map them into higher dimensional spaces so that they become separable in those spaces

**K-NEARESTNEIGHBOR**

The kNN classifier is based on a distance function that measures the difference or similarity between two instances. The standard Euclidean distance d(x, y) between two instances x and y is defined as : n 2 k k k=1 d(x,y)= (x -y ) where, xk is the kth featured element of instance x, yk is the kth featured element of the instance y and n is the total number of features in the dataset. Assume that the design set for kNN classifier is U. The total number of samples in the design set is S. Let C = {C1 ,C2 ,…CL} are the L distinct class labels that are available in S. Let x be an input vector for which the class label must be predicted. Let yk denote the kth vector in the design set S. The kNN algorithm is to find the k closest vectors in design set S to input vector x. Then the input vector x is classified to class Cj if the majority of the k closest vectors have their class as Cj

**MODULES:**

1. **UPLOAD DATASET**

Users search the any link notably, not all network traffic data generated by malicious apps correspond to malicious traffic. Many malware take the form of repackaged benign apps; thus, malware can also contain the basic functions of a benign app. Subsequently, the network traffic they generate can be characterized by mixed benign and malicious network traffic. This dataset is upload.

1. **MANUAL ADDING DATA**

User handling for some various times of smart phones ,desktops laptops and tablets .If any kind of devices attacks for some unauthorized malware softwares.In this malware on threats for user personal dates includes for personal contact, bank account numbers and any kind of personal documents are hacking in possible. So add the network data in manualy.

1. **CLASSIFYING ATTACKS**

Here, we compare the classification performance of SVM with other popular machine learning algorithms. We have selected several popular classification algorithms. For all algorithms, we attempt to use multiple sets of parameters to maximize the performance of each algorithm. Using SVM algorithms classification for malware bag-of-words weightage.

1. **GRAPHICAL REPRESENTATION**

The main part of the project is to analysis the attack types in the network dataset. The user data analysis of the data can be done by charts format. This is the place where admin have ability to come for particular solution about proposed system. The pictorial representations of collected data are shown in the form of graphs. The different graphs give the best analysis of the system.

**REQUIREMENT ANALYSIS**

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

**REQUIREMENT SPECIFICATION**

**Functional Requirements**

* Graphical User interface with the User.

**Software Requirements**

For developing the application the following are the Software Requirements:

1. Python
2. Django
3. MySql
4. MySqlclient
5. WampServer 2.4

**Operating Systems supported**

1. Windows 7
2. Windows XP
3. Windows 8

**Technologies and Languages used to Develop**

1. Python

**Debugger and Emulator**

* Any Browser (Particularly Chrome)

**Hardware Requirements**

For developing the application the following are the Hardware Requirements:

* Processor: Pentium IV or higher
* RAM: 256 MB
* Space on Hard Disk: minimum 512MB

**CONCLUSION:**

This paper presents a literature review of ML and DL methods for network security. The paper, which has mostly focused on the last three years, introduces the latest applications of ML and DL in the field of intrusion detection. Unfortunately, the most effective method of intrusion detection has not yet been established. Each approach to implementing an intrusion detection system has its own advantages and disadvantages, a point apparent from the discussion of comparisons among the various methods. Thus, it is difficult to choose a particular method to implement an intrusion detection system over the others. Datasets for network intrusion detection are very important for training and testing systems. The ML and DL methods do not work without representative data, and obtaining such a dataset is difficult and time-consuming. However, there are many problems with the existing public

dataset, such as uneven data, outdated content and the like. These problems have largely limited the development of research in this area. Network information update very fast, which brings to the DL and ML model training and use with difficulty, model needs to be retrained long-term and quickly. So incremental learning and lifelong learning will be the focus in the study of this field in the future.