ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org
Vol 14, Issuse4,Dec2023

LONG SHORT-TERM MEMORY (LSTM) DEEP LEARNING METHOD FOR INTRUSION DETECTION IN NETWORK SECURITY

ADILAXMI 1, K. SNEHA2, N. SUCHITHA3, P. ANUSHIKA4, M. SREEJA5

ABSTRACT:

Nowadays, large numbers of people were affected by data infringes and cyber-attacks due to dependency on internet. India is lager country for any resource use or consumer. Over the past ten years, the average cost of a data breach has increased by 12%. Hacking in India is take share of 2.3% of global criminal activity. To prevent such malicious activity, the network requires a system that detects anomaly and inform to the admin or service operator for taking an action according to the alert. System used for intrusion detection (IDS) is software that helps to identify and observes a network or systems for malicious, anomaly or policy violation. Deep learning algorithm techniques is an advanced method for detect intrusion in network. In this paper, intrusion detection model is train and test by NSL-KDD dataset which is enhanced version of KDD99 dataset. Proposed method operations are done by Long Short-Term Memory (LSTM) and detect attack. So admin can take action according to alert for prevent such activity. This method is used for binary and multiclass classification of data for binary classification it gives 99.2% accuracy and for multiclass classification it gives 96.9% accuracy.

Keywords: Intrusion detection; Deep Learning Method; LSTM algorithm; Network Security, NSL- KDD dataset.

INTRODUCTION

Security of data is very important aspect of internet in recent years. For illegal right to use or information from network, intruder made an intrusion in system. An intrusion is nothing but attack, hacking, packet sniffing or stilling of data. Attacks are an aim to tear down system privacy or networks in way to extort money, other malicious intentions or acquiring essential records. Intrusion modify program, data or logic in computer by the use of malicious code resulting in difficulty a few consequences that can give and take the institutes private data to formulate it accessible for cybercriminal. Many different attacks come under Cyber attacks which consist of hacking of data, Denial of services, Malware, Phishing and theft. The percentage of cyber attackers or illegal activities increases in the world and defender of cyber-security are experience a lot of threats from these cyber attacker. It could probably leave in to enormous and a major impact on human lives for that to take security measures are important. And these measures can be done by intrusion detection system (IDS). Intrusion Detection can be done by collecting of data packets, analyzing it and detecting any unwanted, suspicious or malicious things in traffic to inform administrator.

1ASSISTANTPROFESSOR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD.

2,3,4&5 UG SCHOLAR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

This device is prepared for securing our data from any attack or unwanted use. The digital world is especially vulnerable to security threats. Hackers are gradually more hacked websites for various reasons. That creates many security threats that had made numerous companies re-evaluate their security measures. Hackers ascertain the loopholes in the website to break the system and achieve their offender ideas. Intrusion detection system also knows as security information and event management system. Two methods are generally used ie. Signature based and anomaly based. And some types are based on network intrusion, host based, perimeter, VM based intrusion detection system. The system detect network ie. A data packet travels from one to another destination.

II. RELATED WORK This section includes recent development in intrusion detection system. Many methods of machine learning are used for detection of anomalies. As development in technology the techniques also upgrade to accomplish requirements. And now those techniques used machine learning that converted in deep learning methods for more precision and accuracy. Anish Halimaa and Dr. K.Sundarakantham [1] used SVM and Naïve bayes method for intrusion detection. This paper shows comparative analysis of SVM and Naïve bayes with the help of NLS-KDD dataset. Both methods used for solving the classification problem. Accuracy and misclassification rate get calculated and on that basis SVM works better than Naïve bayes. Normalization and feature reduction are also applied to makes comparative analysis. Performance of model is depends on accuracy, main motive is to reduce false alarm rate (FAR) and to increase detection rate by increasing accuracy. SVM algorithm is used for image processing and pattern reorganization application on the other hand Naïve bayes is used for statistical classification which is based on Bayes' theorem. This system includes preprocessing of raw data, classification according to set of rules and result evolution on the base of method used. Mohammed Ishaque et al [2] used deep learning which is an region of Machine Learning research. Deep learning approach is used to selecting a subset of relevant feature from unknown information. These types of property are useful in analysing highly complex information to detect anomaly from data present on internet or web system. Deep learning method is used for feature extraction to reduce dimensionality of the dataset obtained from the web system. This paper used unlabeled training data and web system data are given to preprocessing unit then this data is compared with the help of stacked denoising autoencoder, after this data will be split into attacking and normal traffic. Jin Yang et al [3] proposed the method to address the challenge of unbalanced positive and negative learning samples, we propose using deep convolutional generative adversarial networks (DCGAN), which allows features to be extracted directly from the raw-data, and then generates new training-sets by learning from the raw-data. This paper applies long short-term memory (LSTM) to automatically learn the features of network intrusion behaviours. To remove such dependency and enable intrusion detection in real time, we propose a simple recurrent unit based (SRU)-based model. The proposed model in this paper was verified by wide-ranging experiments on the standard datasets for intrusion detection which is KDD'99 and NSL-KDD that effectively recognizes normal and malicious network activities. It achieves 99.73% accuracy for the KDD'99 dataset and 99.62% on the NSL-KDD dataset. Gozde Karatas et al [4] this paper aimed to survey deep learning based intrusion detection system approach by making a comparative work of the literature contains three main components: data collection, feature selection/ conversion and decision engine. To extend the pliability of the system, rather than signature-based

detection, it's required to implement the system as anomaly detection with a learning system. Therefore, during this paper, it's aimed to supply a brief survey of deep learning-based intrusion detection systems with the overview of varied aspects of intrusion detection and deep learning algorithms. Additionally, this work lists and provides details about some publicly available datasets with their characteristics and shortcomings. Dimitar Nikolov et al [5] this paper presents the effects of problem based learning project on a high-school student in Technology school. The intrusion detection system is predicated on a recurrent neural network classifier namely long-short term memory units. The intrusion detection system (IDS) consists of three modules: monitoring, processing and learning module. Learning module creates the LSTM recurrent neural network and finds the necessary structure, weights and biases. For learning, the dataset consisting of system call sequences is used. Brian Lee et al [6] this paper presents a comparative evaluation of deep learning approaches to network intrusion detection. The paper present a comparative evaluation of deep learning approaches to network intrusion detection. Their performance is evaluated using the network intrusion dataset provided by Knowledge Discovery in Databases (KDD). The results of the analysis between the deep learning models suggests that the utilization of deep learning in NIDS would be a appropriate solution to improving detection accuracy on unclean data; however, building an environment that is specifically designed for this purpose would go a long way to further improve and could greatly impact the decision on which model would work best in a given environment.

III. LSTM: LONG SHORT-TERM ALGORITHM

Long short-term memory (LSTM) is special sort or superior version of a man-made recurrent neural network (RNN) architecture utilized within in the sector of deep learning. LSTM has feedback connections and design to avoid long term dependencies. It can't only process only data points, but also whole sequences of knowledge. For instance, LSTM is applicable to tasks like unsegmented data, connected recognition pattern, speech recognition and anomaly detection in network traffic or IDS's (intrusion detection systems). A common LSTM unit consists four main parts: 1) cell, 2) input gate, 3) output gate and 4) forget gate. The cell memories values over random time intervals and therefore the three gates standardize the flow of data or information into and out of the cell. LSTM networks are complementary to classifying, processing and making predictions based on time series data, since there are often lags of unknown duration between important events in a time series. LSTMs were developed to affect the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is a plus of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications

LITERATURE SURVEY

Title: The Internet of Things - How the Next Evolution of the Internet is Changing Everything.

Year: 2011

Author: D. Evans

Methodology

This research describes the methodology and the development process of creating an IoT platform. This paper also presents the architecture and implementation for the IoT platform. The goal of this research is to develop an analytics engine which can gather sensor data from different devices and provide the ability to gain meaningful information from IoT data and act on it using machine learning algorithms.

Advantage

The proposed system is introducing the use of a messaging system to improve the overall system performance as well as provide easy scalability.

Disadvantage

Low cost devices are easily able to connect wirelessly to the Internet, from handhelds to coffee machines, also known as Internet of Things (IoT).

Title: The Internet of Things: A survey
Year: 2010.

Author: L. Atzori, A. Iera, and G. Morabito

Methodology

The object unique addressing and the representation and storing of the exchanged information become the mostchallenging issues, bringing directly to a third, ‘‘Semantic oriented’’, perspective of IoT.

Advantage

People are informed of the scope and the way in which their movements are tracked by the system (taking peo-ple informed about possible leaks of their privacy is essential and required by most legislations).

Disadvantage

The user can set the preferences of the proxy. When sensor networks and RFID systems are in-cluded in the network, then the proxy operates between them and the services.

Title: Addressing the Class Imbalance Problem in Medical Datasets

Year: 2012

Author: M. Mostafizur Rahman and D. N. Davis

Methodology

A balanced dataset is very important for creating a good training set. They aim to optimize the overall accuracy without considering the relative distribution of each class. Typically real world data are usually imbalanced and it is one of the main causes for the decrease of generalization in machine learning algorithms.

Advantage

The aim was to reduce the ratio gap between the majority classes with the minority class. The proposed method is found to be useful for such datasets where the class labels are not certain and can also help to overcome the class imbalance problem of clinical datasets and also for other data domains.

Disadvantage

The outcome labels of most of the clinical datasets are not consistent with the underlying data. The conventional over-sampling and under-sampling technique may not always be appropriate for such datasets.

Title: A survey on cloud computing security
Year: 2010.
Author: R. Kanday

Methodology

This survey paper provides a general overview on Cloud Computing. The topics that are discussed include characteristics, deployment and service models as well drawbacks.

Advantage

The major part of countermeasures focuses on Intrusion Detection Systems. Moving towards Mobile Cloud Computing and Internet of Things, this survey paper gives a general explanation on the applications and potential that comes with the integration of Cloud Computing with any device that

has Internet connectivity as well as the challenges that are before it.

Disadvantage

Several security issues and countermeasures are also discussed to show the major issues and obstacles that Cloud Computing faces as it is being implemented further.

Title: Data Mining: Practical Machine Learning Tool and Technique with Java Implementation

Year: 2000.

Author: Ian H. Witten and Eibe Frank

Methodology

The convergence of computing and communication has produced a society that feeds on information. Yet most of the information is in its raw form: data. If data is characterized as recorded facts, then information is the set of patterns, or expectations, that underlie the data. There is a huge amount of information locked up in databases—information that is potentially important but has not yet been discovered or articulated. Our mission is to bring it forth. The weather data (Tables 1.2 and 1.3) presents a set of days together with a decision for each as to whether to play the game or not.

Advantage

In these cases the output took the form of decision trees and classification rules, which are basic knowledge representation styles that many machine learning methods used.

Disadvantage

The weather problem is a tiny dataset that we will use repeatedly to illustrate machine learning methods.

EXISTING SYSTEM

HACKING incidents are increasing day by day as technology rolls out. A large number of hacking incidents are reported by companies each year. The existing system doesn't effectively classify and predict the attack which is presented in the network.

DISADVANTAGES

- Doesn't Efficient for handling large volume of data.
- Theoretical Limits
- Incorrect Classification Results.
- Less Prediction Accuracy.

PROPOSED SYSTEM

The proposed model is introduced to overcome all the disadvantages that arises in the existing system. This system will increase the accuracy of the classification results by classifying the data based on the social network mental disorders and others using C4.5 Decision tree classification algorithm. It enhances the performance of the overall classification results.

ADVANTAGES

- High performance.
- Provide accurate prediction results.
- It avoid sparsity problems.
- Reduces the information Loss and the bias of the inference due to the multiple estimates.

IMPLEMENTATION

MODULES

- Data Selection and Loading
- Data Preprocessing
- Splitting Dataset into Train and Test Data
- Feature Extraction
- Classification
- Prediction
- Result Generation

ISSN2454-9940www.ijsem.org
Vol 14, Issuse4,Dec2023

DATA SELECTION AND LOADING

- The data selection is the process of selecting the data for detecting the attacks.
- In this project, the KDDCUP dataset is used for detecting attacks.
- The dataset which contains the information about the duration, flag, service, src bytes, dest bytes and class labels.

DATA PREPROCESSING

- Data pre-processing is the process of removing the unwanted data from the dataset.
- Missing data removal
- Encoding Categorical data
- Missing data removal: In this process, the null values such as missing values are removed using imputer library.
- Encoding Categorical data: That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables. That an integer and one hot encoding is used to convert categorical data to integer data.

SPLITTING DATASET INTO TRAIN AND TEST DATA

- Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes.
- One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.
- Separating data into training and testing sets is an important part of evaluating data mining models.
- Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

FEATURE EXTRACTION

- Feature scaling. Feature scaling is a method used to standardize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data pre-processing step.
- Feature Scaling or Standardization: It is a step of Data Pre Processing which is applied to independent variables or features of data. It basically helps to normalise the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

CLASSIFICATION

The C4.5 algorithm is used in Data Mining as a Decision Tree Classifier which can be employed to generate a decision, based on a certain sample of data (univariate or multivariate predictors). A decision tree is a tool that is used for classification in machine learning, which uses a tree structure where internal nodes represent tests and leaves represent decisions. C4.5 makes use of information theoretic concepts such as entropy to classify the data. For each dataset there should be two files, one that describes the classes and attributes and one that consists of the actual data. The file for attributes and classes should contain all the classes in first line and after that, line by line the attributes and their possible values if the attribute is discrete. For continuous (numerical) attributes, possible values would be "continuous". Check the iris dataset folder for actual data and more specific syntax.

PREDICTION

- It's a process of predicting the attacks in the network from the dataset.
- This project will effectively predict the data from dataset by enhancing the performance of the overall prediction results.

RESULT GENERATION

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

- True Positive
- True Negative
- False Positive
- False Negative
- Accuracy
- Precision
- Recall
- F1-Score

CONCLUSION

We reviewed several influential algorithms for intrusion detection based on various machine learning techniques. Characteristics of ML techniques makes it possible to design IDS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviours. We divided these algorithms into two types of ML-based schemes: Artificial Intelligence (AI) and Computational Intelligence (CI). Although these two categories of algorithms share many similarities, several features of CI-based techniques, such as adaptation, fault tolerance, high computational speed and error resilience in the face of noisy information, conform the requirement of building efficient intrusion detection systems.

REFERENCES

1. S. Yinbiao and K. Lee, "Internet of Things: Wireless Sensor Networks Executive summary," 2014. [5] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci
2. "A survey on sensor networks," IEEE Commun. Mag., vol. 40, no. 8, pp. 102–105, 2002. [6]
3. X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," IEEE Commun. Surv. Tutorials, vol. 11, no. 2, pp. 52–73, 2009. [7]
4. A.-S. K. Pathan, H.-W. Lee, and C. S. Hong, "Security in wireless sensor networks: issues and challenges," 2006 8th Int. Conf. Adv. Commun. Technol., vol. 2, p. 6 pp.-pp.1048, 2006. [8]
5. P. Yi, Y. Jiang, Y. Zhong, and S. Zhang, "Distributed Intrusion Detection for Mobile Ad Hoc Networks," 2005 Symp. Appl. Internet Work. (SAINT 2005 Work., pp. 94–97, 2005. [9]
6. H. Sedjelmaci and M. Feham, "Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network," Int. J. Netw. Secur. Its Appl. (IJNSA), Vol.3, No.4, July 2011, vol. 3, no. 4, pp. 1–14, 2011. [10]
7. L. Khan, M. Awad, and B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering," VLDB J., vol. 16, no. 4, pp. 507–521, 2007. [11]
8. S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 1348–1353, 2014. [12]
9. O. Can, C. Turguner, and O. K. Sahingoz, "A Neural Network Based Intrusion Detection System For Wireless Sensor Networks," Signal Process. Commun.

Appl. Conf. (SIU), 2015 23th, pp. 2302–2305, 2015. [13]

10. F. Lu and L. Wang, “Intrusion Detection System Based on Integration of Neural Network for Wireless Sensor Network,” J. Softw. Eng. 2014. [14]
11. Y. Y. Li and L. E. Parker, “Intruder detection using a wireless sensor network with an intelligent mobile robot response,” Southeastcon, 2008. IEEE, pp. 37–42, 2008. [15] A. Kulakov and D. Davcev, “Tracking of unusual events in wireless sensor networks based on artificial neural-networks algorithms,” Inf. Technol. Coding Comput. 2005. ITCC 2005. Int. Conf., pp. 534–539, 2005. [16]
12. M. Panda, “Security Threats at Each Layer of Wireless Sensor Networks,” Int. J. Adv. Res. Comput.Sci. Softw. Eng., vol. 3, no. 11, pp. 61–67, 2013. [17]
13. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures,” Proc. First IEEE Int. Work. Sens. Netw. Protoc. Appl. 2003., pp. 113–127, 2003