IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 12, December 2020

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

An Efficient Intrusion Detection System with Deep Learning Model

Madhurika Budaraju[1], Keerthi Manchikanti[2], Rashmi Cigiri[3]

Assistant Professor, Department of Computer Science Engineering, Keshav Memorial Institute Of Technology, Hyderabad, Telangana, India[1]

Assistant Professor, Department of Information Technology, Keshav Memorial Institute Of Technology, Hyderabad, Telangana, India[2]

Assistant Professor, Department of Information Technology, Mallareddy Engineering College for Women, Hyderabad, Telangana, India[3]

ABSTRACT: With Identification of new attacks in one of the enormous difficulties in the Intrusion Detection Framework. The role of any Intrusion Detection System is to monitor system networks for malicious activities or any policy violations. The limitations of traditional IDS are that Consistent programming refreshes are required for signature-based IDS to stay aware of the new dangers or System IDS can just distinguish organize peculiarities which limit the assortment of assaults it can find. So, there is a necessity to build an efficient model for Intrusion Detection System. In this paper, we proposed a deep learning framework for Intrusion Detection System. The data set is taken from Kaggle, which consists of 42 features with 1,35,973 samples. We also applied various shallow learning models to compare the performance of the deep neural network. The deep neural network has given accuracy of 88%. All the implementations are done in python.

KEYWORDS: Intrusion Detection Systems, Kaggle, Deep Learning, Python.

I. INTRODUCTION

Cybersecurity provides computing facilities in a secure and communicative environment [1]. There is various diverse kind of assaults inside the internet nowadays. Any computer user expects common security services like confidentiality (preventing accessing of data from unauthorized access), Availability (Data must be available all the times), Integrity (data cannot be modified or changed), non-repudiation (User does not refuse that he has used that network), access control (controlling accessing of data),(Authentication(Identification of a user). Far-reaching investigates have been executed to conquer these assaults. One normal countermeasure is to utilize the supposed Intrusion Detection System (IDS). An intrusion is simply an action that breaks the security necessities integrity, availability, confidentiality. Applications, for example, firewalls and honeypots are the principal lines of protection against a digital attack. Firewalls go about as a channel for arranging traffic - utilizing a lot of rules, a firewall will keep has from outside the inner system to interface with a protected end framework. Firewalls experience the ill effects of the capacity to look after express; a constant aggressor can without much of a stretch detour a firewall and increase passage into the undertaking system. Honeypots act like a snare for assailants - by promoting the likelihood that it contains delicate data, programmers will attempt to get to the honeypot and are then obstructed from the system. But honeypots are successful only if they can bait the attacker. If the attacker in a situation to identify that the honeypot is trying to identify them, then the attacker just ignores it and continue to attack the network. So, a requirement was created for a computer that can learn the structure of network data and distinguish benign from abnormal traffic. To solve these problems, the Intrusion-detection system comes into the picture. Intrusion Detection System must be capable of inspecting the system network for intrusions and send alarms to the network administrator for corrective measures. The key challenge for any solution framework must distinguish benign traffic and anomaly traffic. There are various Internet Protocols exists and various applications are using these protocols. So, it's difficult to create a single IDS that can detect all the attacks efficiently. For detecting different attacks, IDS can be classified into signature-based, anomaly-based, or hybrid based. In the first model (signature-based), it detects attacks by signatures of attackers. In the second model (namely based), it compares user activities with predefined rules to detect a legitimate user. In the third model(hybrid), it combines two or more models to efficiently detect attacks. Machine Learning is one of the solutions to accomplish this complex task. Arthur Samuel coined the term machine learning in the 1960s. Today machine learning is used in most applications to

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

fulfill their requirements. In machine learning, computers can learn things, and later they worked for humans based on their learning methodology. Deep Learning is a sub-field of machine learning where human brain biological neurons are simulated to solve complex problems. Machine Learning and Deep Learning can be supervised or unsupervised learning. In the supervised learning method, we have both inputs and outputs for all the samples, whereas in unsupervised learning we are having only inputs but there are no class labels. Supervised learning can be further classified as regression and classification. Both regression and classification come under prediction. But in regression, the predicted value is continuous. In classification, the predicted value is the category or class label.

II. LITERATURE SURVEY

Applying machine learning techniques for cybersecurity applications is not a new trend. In the previous decades, various researchers applied different Machine Learning and deep learning techniques to characterize and separate anomaly traffic from benign traffic without earlier information on the assault signature. But machine learning techniques need data to be in the prescribed format. If the input dataset is preprocessed well, then the results are good. otherwise, we may not expect good models. So utmost care is to be taken to the data preprocessing stage before applying machine learning or deep learning models to intrusion detection systems. In the data preprocessing step, all the missing values are removed by replacing them with a mean, median, or constant value. Similarly, categorical attributes are encoded as numbers. Abhishek Verma [2] et.al applied machine learning models for Intrusion Detection System for the Internet Of Things applications. The proposed ensemble learning models for IDS.

Leila Mohammadpour [3] et.al experimented with convolutional neural networks for intrusion detecting systems. They applied the convolution neural network model on the famous dataset "NSL-KDD" and achieved an accuracy of 99%. Mohammad Almseidin [4] et.al applied various machine learning algorithms for intrusion detection systems. They have shown that decision tree classifier performs well with respect to the lowest false negatives and random forest classifier performs well with respect to accuracy. Sinh-Ngoc Nguyen [5] et.al applied convolutional neural networks for the design & implementation of the Intrusion Detection System for DoS attacks. They applied a feed-forward neural network for the KDD99 CUP dataset and achieved good results. Gozde Karatas [7] et.al aimed to survey deep learning techniques for network intrusion detection systems.

They are also given an overview of publicly available datasets for IDS related applications. Quamar Niyaz [8] et.al applied a deep learning-based technique called STL (Self-taught Learning) on the NSL-KDD dataset and achieved good results. T Dhikhi [9] et.al applied a combination of autoencoder and support vector machine on NSL KDD dataset and shown that the recognition rate is enhanced. P. Roshni Mol [10] et.al summarizes various IDS attacks, tools, datasets. They also discussed how machine learning techniques play a major role in IDS systems. Min-Joo Kang [11], Jiayan Zhang [14] et.al proposed an IDS model based on a deep neural network for the security of an in-vehicular network. In [11], they used probability-based feature extraction methods and achieved better results with their model.Vinayakumar R[12] proposed a deep learning model called as 'Scale-Hybrid-IDS-AlertNet' on various benchmark datasets NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, CICIDS 2017 and achieved good accuracy rate with their model. Suad Mohammed Othman [13] et.al built an intrusion detection system based on machine learning models for the bigdata environment. They applied a variant of SVM model called Spark-Chi-SVM on the KDD99 dataset and achieved good results within less training time, which is most important for big data.

III. RESEARCH METHODOLOGY

For building an efficient IDS model, we need a good dataset. Kaggle provides various well-processed datasets for machine learning tasks. We downloaded an IDS dataset from Kaggle [15] for doing our experiments. The dataset contains 42 features with 135093 rows. After downloading the dataset, we checked the consistency of the dataset. We verified any missing values are there in the dataset. If the dataset contains missing values, we need to impute them. But this dataset has no missing values. Next, we checked for categorical features. If the dataset contains categorical features, we need to convert them into numerical elements by applying one hot encoding technique. Our dataset has one categorical attribute named "protocol type". For this feature, there are three options ICMP, UDP, TCP. All these three are categories or labels. We applied label encoding and one-hot encoding and converted these three categories into numerical elements. Now preprocessed dataset is ready for us. Then we applied a feed-forward neural network model on this dataset. We also applied several shallow learning models like decision tree classification, random forest classification, etc. to compare the performance of our deep learning model.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

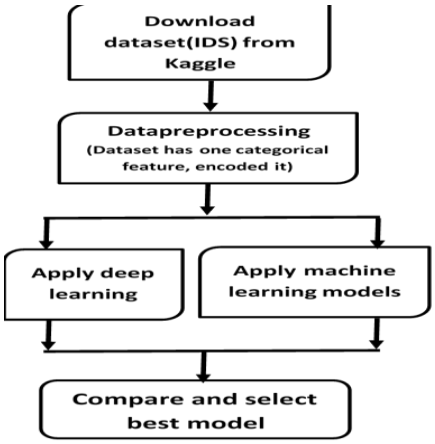

Figure 1: Proposed model

3.1 Machine Learning classification:

In supervised learning, the given dataset is divided into training and testing sets. The model learns important patterns from the training data and then based on the learned model it makes predictions for the testing data. The general splitting criteria for training and the testing split is 70%&30% or 80%&20%. But for our dataset, Kaggle provides two separate training and testing sets. So, there is no need to split the data into training and testing sets. In machine learning tasks, we need to distinguish defendant, independent variables. We divided the two datasets (training and testing sets) into independent and dependent features. The defendant feature is nothing but the class label. The class labels for our dataset are different types of attacks namely normal, dos, probe, r2l, and u2r. All the remaining features are independent features. Based on all independent variables our model needs to find the correct class label. Classification is the task of predicting a label for the class. There are various machine learning classification models available for doing this task. All the Classification algorithms in machine learning come under supervised learning. Here the labels of the class are known to the user. So, after predictions made by the learned model, the user can check the accuracy.

Logistic regression classifier: It is the basic classifier in machine learning. It works in a similar way as regression works. That why its name includes regression. It uses two functions namely logit function and sigmoid function to perform inner calculations.

K-nearest Neighbors: It is a distance-based classifier, in which K number of nearest data points play a major role for deciding class of a new data point.

Support Vector Machine: SVM classifies the data points based on hyperplane. This model gives the user flexibility in deciding relaxation of error allowance.

Decision Tree classifier: It is a tree-based model in which one of the measures (like Gini index) is used to construct a tree. In each step, we use a measure decide the root node and tree is built in a recursive approach.

Naïve Bayes classifier: This classifier is based on probability values. We calculate several probability values to assign a class label.

Random Forest classifier: It is an ensemble approach, in which several decision trees are used for making decisions.

3.2 Artificial Neural Network:

The invention of Artificial Neural Network has inspired from biological neuron in human brain. There are receivers (Dendrites), Transmitters (axons, neurons in this biological structure. Artificial neural network simulates this structure. In ANN, we have input layers (which consists of various features in the dataset), hidden layers, output layer. If we ignore the hidden layer, then this network works like a normal machine learning model. So hidden layers play very important role in ANNs.The hidden layers has the flexibility to consider which features are important for regression or classification tasks. Each input is having weights. So, inputs with weights are feed to network in forward direction and an activation function at each hidden layer processes these combined inputs. At output layer, there is one more activation function, which is used to classify the label. At each round an error is calculated through error function and these errors and back propagated to the network in backward direction to adjust the weights. After adjusting weights once again the information feed forwarded, and error calculated. These forward propagation and backward propagation continued until the network results with less error rate. For doing all these tasks in an effective manner, some algorithms are used like gradient descent, Stochastic gradient descent. The gradient descent algorithm is useful to find

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

the least error value within few calculations only. The researchers also use different algorithms like adam, RMS drop, Momentum, Mini batch gradient descent etc. There is a bias neuron in the neural network model which is added to each and every layer in the network. This neuron stores the value of one. It makes it possible to move the activation function to left or right on the graph. This neuron is not compulsory. In most of the situations, it is required to move the entire activation function to the left or right to form the output values. It can be achieved by the bias. Neural networks can work without bias neurons in reality. The bias values are added and corresponding weights are calculated by the model. This total structure is also called as feed forward neural network model.

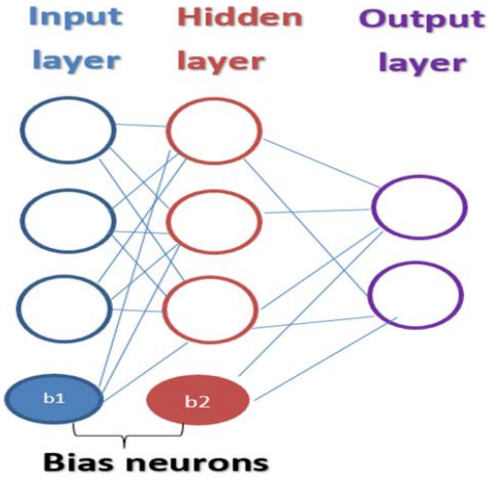

Figure 2: Structure of Artificial Neural Network model

IV. EXPERIMENTAL RESULTS

First, we applied three shallow learning models K-NN, Decision tree classifier, random forest classifier on our dataset. After applying the models, the results are tabulated. The measures used for comparing various classifers are precision, recall and accuracy. If we use only accuracy, sometimes it may lead to wrong interpretations.

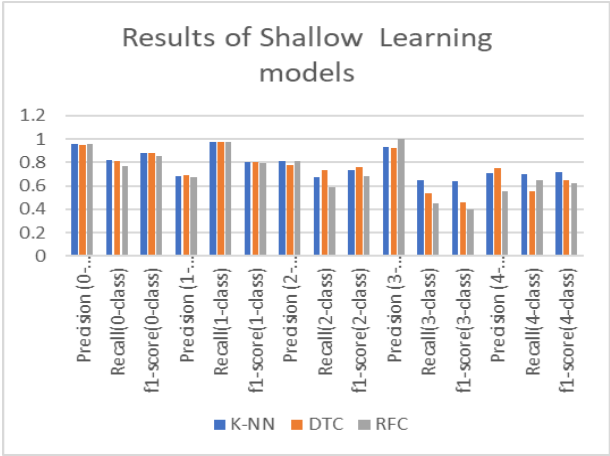

Figure 3: Accuracy of training, testing sets

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

Measure	K-NN	DTC	RFC
Precision (0- class)	0.96	0.95	0.96
Recall(0-class)	0.82	0.81	0.77
f1-score(0-class)	0.88	0.88	0.85
Precision (1- class)	0.68	0.69	0.67
Recall(1-class)	0.97	0.97	0.97
f1-score(1-class)	0.80	0.80	0.79
Precision (2- class)	0.81	0.78	0.81
Recall(2-class)	0.67	0.73	0.59
f1-score(2-class)	0.73	0.76	0.68
Precision (3- class)	0.93	0.92	1.00
Recall(3-class)	0.65	0.54	0.45
f1-score(3-class)	0.64	0.46	0.40
Precision (4- class)	0.71	0.75	0.55
Recall(4-class)	0.70	0.55	0.65
f1-score(4-class)	0.72	0.65	0.62
Accuracy	77.1%	77.3 %	73.6%

Table 1: Results of machine learning models

After applying all the classification models, we get a maximum accuracy of 77.3% with decision tree classification. Next, we applied a feed forward neural network. We used two hidden layers for our model. There is no rule of thumb for selecting number of hidden layers or number of neurons in the hidden layer for the network. We tried with different combinations. Our model has one input layer, two hidden layers and one output layer. The number of neurons in the first hidden layer are 23. The number of neurons in the second hidden layer is 12.

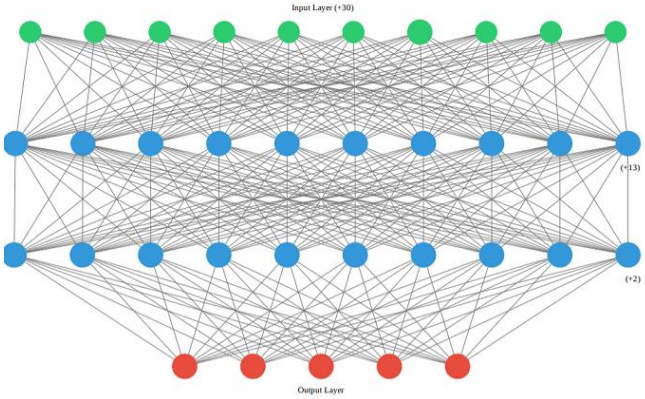

Figure 4: ANN Architecture

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

The output layer has five different outputs. The loss function used in the model is categorical cross-entropy. This loss function can easily distinguish the models. The optimizer used for the model is "adam" optimizer. We also tried SGD (Stochastic Gradient Descent), but we got good results with adam. As this is a multiclass classification problem, we used 'SoftMax' as an output function. The number of epochs is set to 50. We tried with various epochs, but we got the best results with this number.
After applying feed forward neural network, the accuracy of all the models are tabulated.

Model	Accuracy
K-NN	77.1%
DTC	77.3%
RFC	73.6%
Feed Forward NN	86.5%

Table 2: Comparison of results of ML, DL models

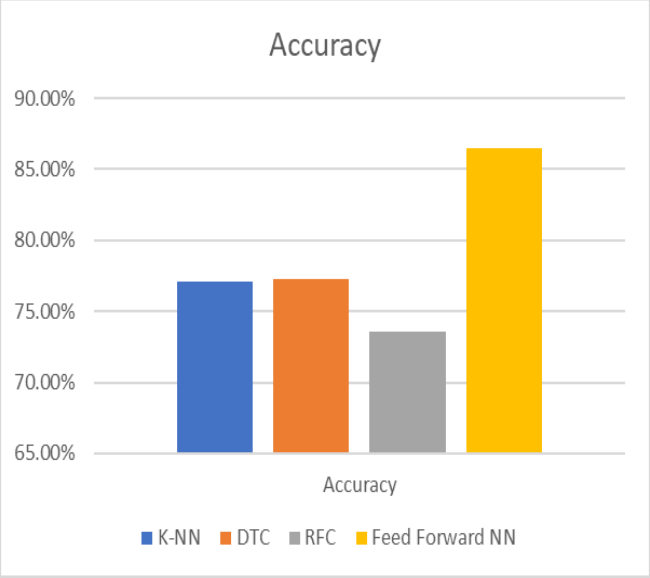

Figure 5: Accuracy comparison ML, DL models

V. CONCLUSION

In this paper, we proposed a feed forward neural network model for Intrusion Detection system. We applied shallow machine learning methods K-NN, Decision Tree classifier, Random Forest for IDS and got maximum accuracy of 77.3%. Later, we applied Neural network model with two hidden layers with different parameter settings. Feed Forward neural network outperforms all the shallow learning models by achieving an accuracy of 86.5%.

REFERENCES

[1] A. Lakshmanarao, M. Shashi, "A Survey on Machine Learning for Cyber Security", International Journal of Scientific & Technology Research Volume 9, Issue 01, January 2020.

[2] Abhishek Verma, Virender Ranga," Machine Learning Based Intrusion Detection Systems for IoT Applications", Wireless Personal Communications, November ,2019 https://doi.org/10.1007/s11277-019-06986-8.

[3] Leila Mohammadpour, Teck Chaw Ling, Chee Sun Liew and Chun Yong Chong "A Convolutional Neural Network for Network Intrusion Detection System", Proceedings of the APAN – Research Workshop 2018.

[4] Mohammad Almseidin, Maen Alzubi, Szilveszter Kovacs∗ and Mouhammd Alkasassbeh, "Evaluation of Machine Learning Algorithms for Intrusion Detection System", arxiv.org, January 2018.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 12, December 2020 ||

[5] Sinh-Ngoc Nguyen, Van-Quyet Nguyen, Jintae Choi, Kyungbaek Kim, "Design and Implementation of Intrusion Detection System using Convolutional Neural Network for DoS Detection", ICMLSC 2018, February 2–4, ACM, https://doi.org/10.1145/3184066.3184089.

[6] Irin Anna Solomon, Aman Jatain and Shalini Bhaskar Bajaj," Intrusion Detection System Using Deep Learning", Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol.8 No.2, 2019, pp. 105-110.

[7] Gozde Karatas, Onder Demir, Ozgur Koray Sahingoz, "Deep Learning in Intrusion Detection Systems," International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism Ankara, Turkey, 3-4 Dec, 2018.

[8] Quamar Niyaz, Weiqing Sun, Ahmad Y Javaid, and Mansoor Alam, "A Deep Learning Approach for Network Intrusion Detection System," BICT 2015, December 03-05, New York City, United States Copyright © 2016 ICST,DOI 10.4108/eai.3-12-2015.2262516.

[9] Dhikhi T, M.S. Saravanan, "An Enhanced Intelligent Intrusion Detection System using Machine Learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.

[10] P. Roshni Mol, Dr. C. Immaculate Mary, "Intrusion Detection System from Machine Learning Perspective," International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181.

[11] Min-Joo Kang, Je-Won Kang, "Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security," PLOS ONE, DOI: 10.1371/journal.pone.0155781 June 7, 2016.

[12] Vinayakumar R, Mamoun Alazab, Soman Kp, Prabaharan Poornachandran, Ameer Al-Nemrat4, And Sitalakshmi Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System", 2169-3536 (c) 2018 IEEE, IEEE Access.

[13] Suad Mohammed Othman, Fadl Mutaher Ba-Alwi1, Nabeel T. Alsohybe, Amal Y. Al-Hashida, "Intrusion detection model using machine learning algorithm on Big Data environment", Journal of Bigdata, Springeropen, https://doi.org/10.1186/s40537-018-0145-4.

[14] Jiayan Zhang, Fei Li, Haoxi Zhang, Ruxiang Li, Yalin Li, "Intrusion detection system using deep learning for in-vehicle-security", Ad-Hoc-Networks,1570-8705,2019,Elsevier, https://doi.org/10.1016/j.adhoc.2019.101974.

[15] https://www.kaggle.com/what0919/intrusion-detection.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details