Intrusion Detection System Using Deep Learning

Presented by

Student name : ABHIJITH S, ABIWAQAS Y

Reg. Number : 181CS102, 181CS105

Department Name : CSE

Guided by

Faculty Name : Mr. SATHISHKUMAR P

Designation : Assistant Professor

Department Name : CSE

Aim and Objectives

- ➔ The primary goal is to build a network intrusion detector, a predictive model capable of distinguishing between adverse connections, called **intrusions or attacks**, and **good** normal connections.
- ➔ It is focused to implement an Intrusion detection and prevention system using Deep Neural Network that can immediately detect the attacks such as DOS, Probe, R2L and U2R.
- ➔ DNN is explored to develop a flexible and effective IDS to detect and classify unforeseen and unpredictable cyberattacks.

Work plan

S.No	Week	Plan
1	Week 1-3	1.) Selection of Domain and Project title. 2.) Study on the Domain and insight view.
2	Week 4-8	3.) Literature Review 4.) Data Set collections using Web Scraping and Open Source Pages 5.) Preprocessing of Datasets.
3	Week 9-12	6.) Data Analysis. 7.) Development and Training Model using Unsupervised Techniques (Classifiers: Decision Tree, Random Classifier)
4	Week 10-15	8.) Development and Training Model using Deep Learning (Bi-LSTM) 9.) Testing and Prediction 10.) Evaluation Metrics

Literature Survey

SI.NO	PAPER TITLE	AUTHOR	YEAR	METHODOLOGY
1.	A Review on Intrusion Detection System using Deep Learning	Ms. Tanushri Jain , Prof. Chetan Gupta	2020	DNN for intruder detection is used to implement the analysis module.
2.	Random-forests-based network intrusion detection systems	J. Zhang, M. Zulkernine, and A. Haque	2019	Equipped Random Forest classifier.
3.	Evaluation of recurrent neural network and its variants for intrusion detection system	R. Vinayakumar, K. Soman, and P. Poornachandran	2020	Simulation on a simple RNN, LSTM and GRU and benchmarked them on the KDD'99 Dataset for multi-category predictions.
4.	Online adaboost based parameterized methods for dynamic distributed network intrusion detection	W. Hu, J. Gao, Y. Wang, O. Wu, and S. Maybank	2018	Employed Adaptive Boosting technique

Proposed Methodology

1. Finalised with NSL-KDD Dataset since up-to-date and contains numerous attacks in multiple protocols.
2. Using Anomaly-based Intrusion Detection approach
3. Data preprocessing to find the percentage of normal connections and attack in the network
4. One-hot encoding and normalizing the dataset
5. Training of ML Model using Logistic Regression, Decision Tree, Random Forest, PLA and Bi-directional LSTM
6. Testing the models and finding the most accurate model - Bi-Directional LSTM
7. Prediction and Evaluation Metrics

Problem Identification in the Project Domain

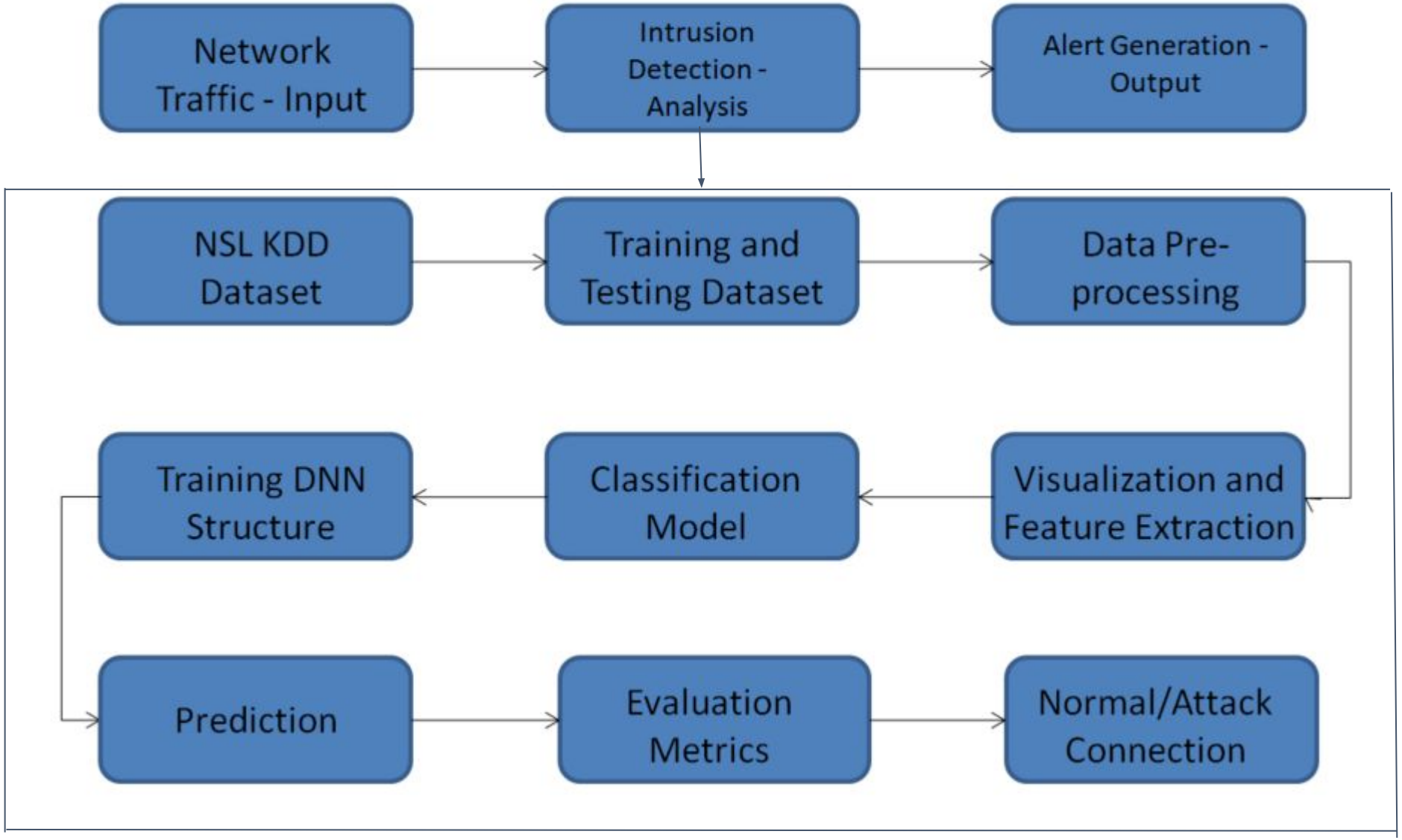
The **continuous change in network behavior and rapid evolution of attacks** makes it necessary to evaluate various datasets which are generated over the years through static and dynamic approaches also to develop an Efficient Intrusion Detection System.

Cyber Threats are becoming increasingly sophisticated and difficult to contain, but the **foremost challenge is to detect new and unknown malwares**, which could be made possible using Deep Learning based IDS.

The need for an up-to-date IDS is crucial, and considering the complexity of malwares in the current stage and existing obsolete system, using Deep Learning Frameworks to build such a system will be immensely beneficial.

An **Anomaly - Based IDS** built using Deep Learning Framework will be efficient in detecting such malwares.

Design

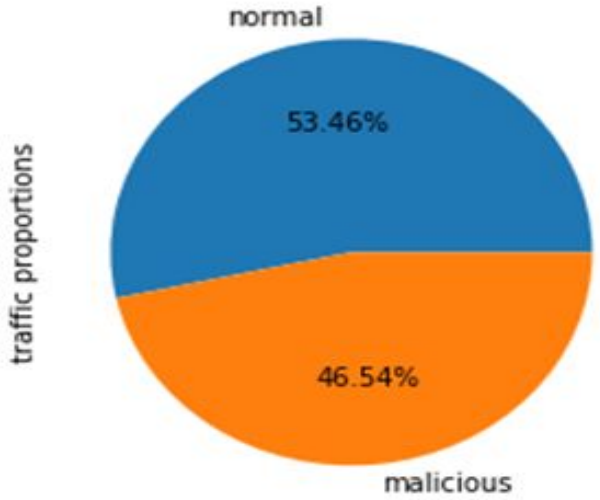

Network Traffic - Input
Intrusion Detection - Analysis
Alert Generation - Output
NSL KDD Dataset
Training and Testing Dataset
Data Pre-processing
Training DNN Structure
Classification Model
Visualization and Feature Extraction
Prediction
Evaluation Metrics
Normal/Attack Connection

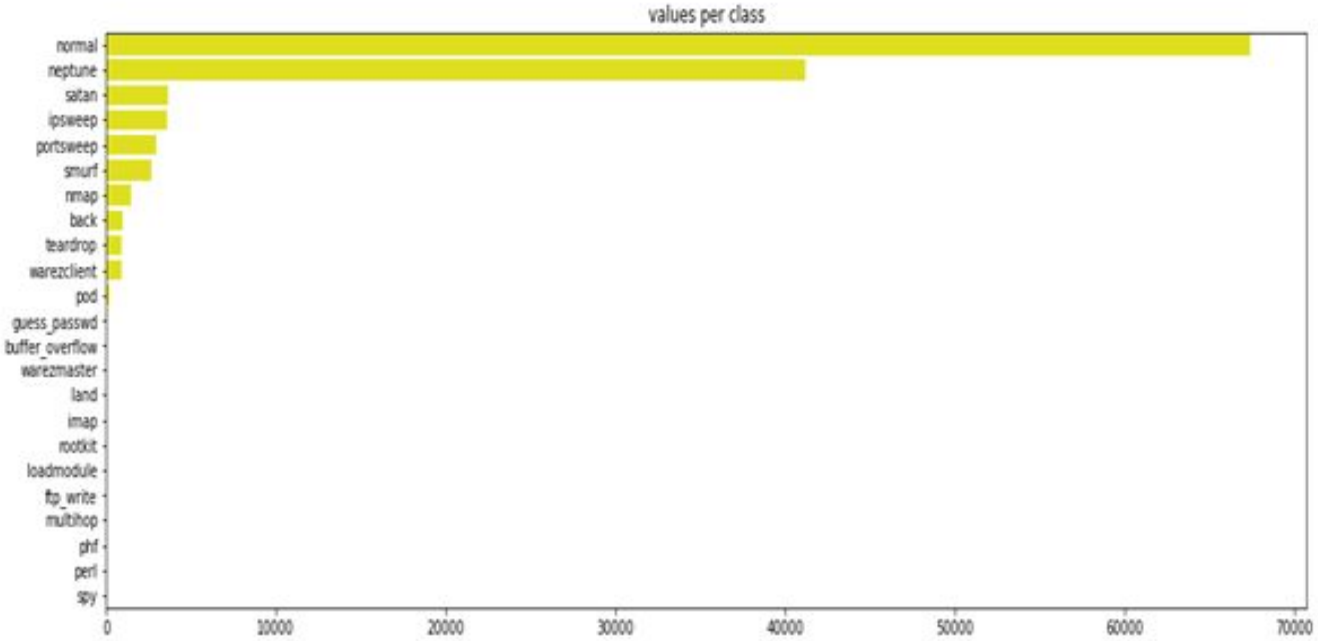
Utilization of Modern Tools

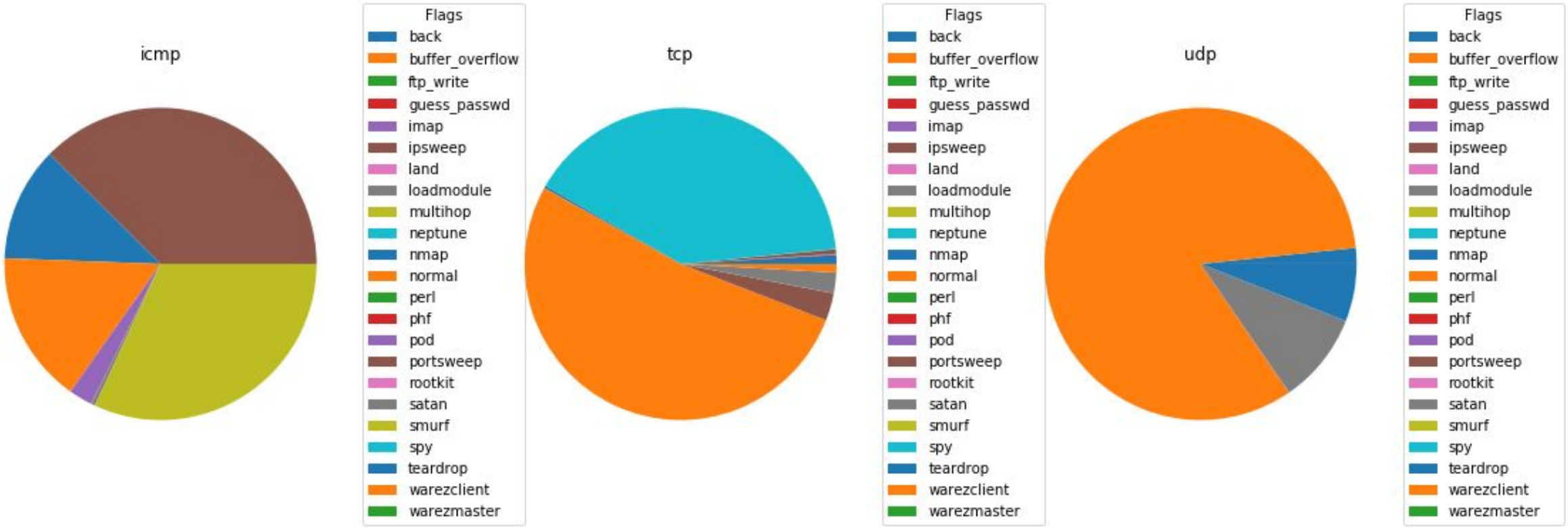
- Utilized Python and Google Colab for the code behind the project
- Machine Learning Frameworks including Tensorflow, Numpy, Pandas, Scikit-Learn for developing the ML Code
- Utilizing compatible hardware - Ryzen 5 & GTX 1650

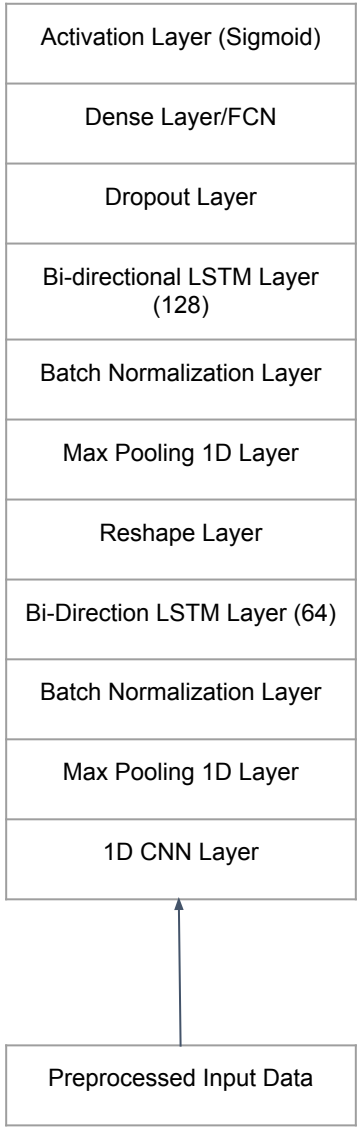
Cost Benefit Analysis

- The project is based on pre-existing open source technologies including Intrusion Detection Systems, Databases containing various known and unknown network traffic attacks, and machine learning algorithms and packages such as Tensorflow and Pytorch.
- The cost of building the model will be relatively low to almost zero since all the technologies involved are open source.


normal
53.46%
46.54%
malicious
traffic proportions


values per class
normal
neptune
satan
ipsweep
portsweep
smurf
nmap
back
teardrop
warezclient
pod
guess_passwd
buffer_overflow
warezmaster
land
imap
rootkit
loadmodule
ftp_write
multihop
phf
perl
spy
0
10000
20000
30000
40000
50000
60000
70000


icmp
tcp
udp
Flags
back
buffer_overflow
ftp_write
guess_passwd
imap
ipsweep
land
loadmodule
multihop
neptune
nmap
normal
perl
phf
pod
portsweep
rootkit
satan
smurf
spy
teardrop
warezclient
warezmaster


Activation Layer (Sigmoid)
Dense Layer/FCN
Dropout Layer
Bi-directional LSTM Layer (128)
Batch Normalization Layer
Max Pooling 1D Layer
Reshape Layer
Bi-Direction LSTM Layer (64)
Batch Normalization Layer
Max Pooling 1D Layer
1D CNN Layer
Preprocessed Input Data

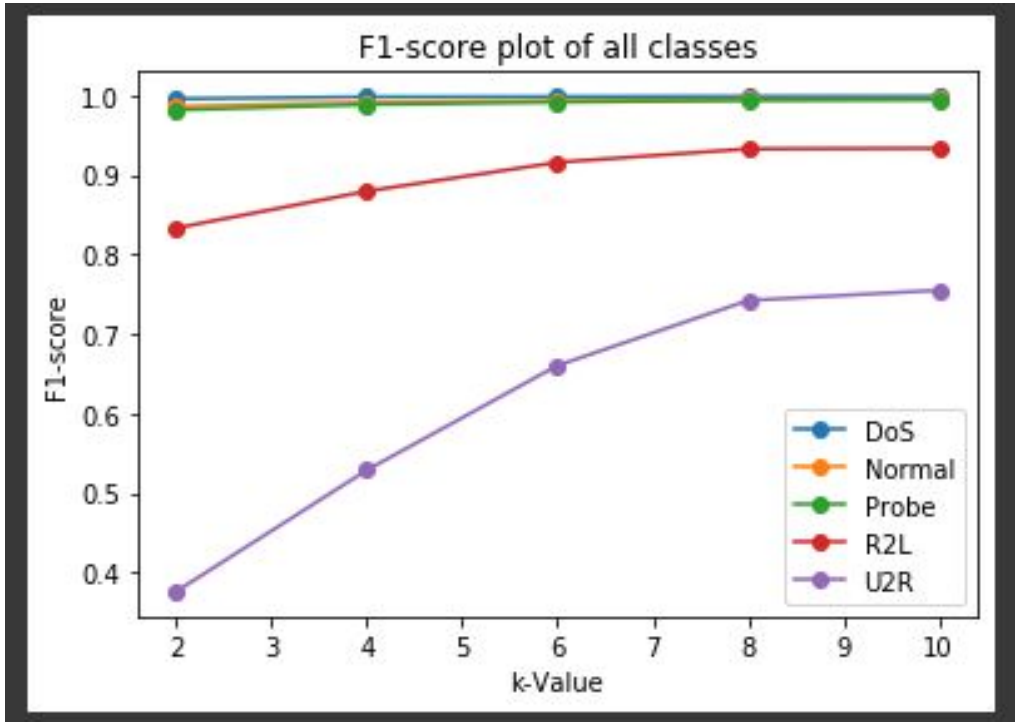
Analysis of Results and Discussion

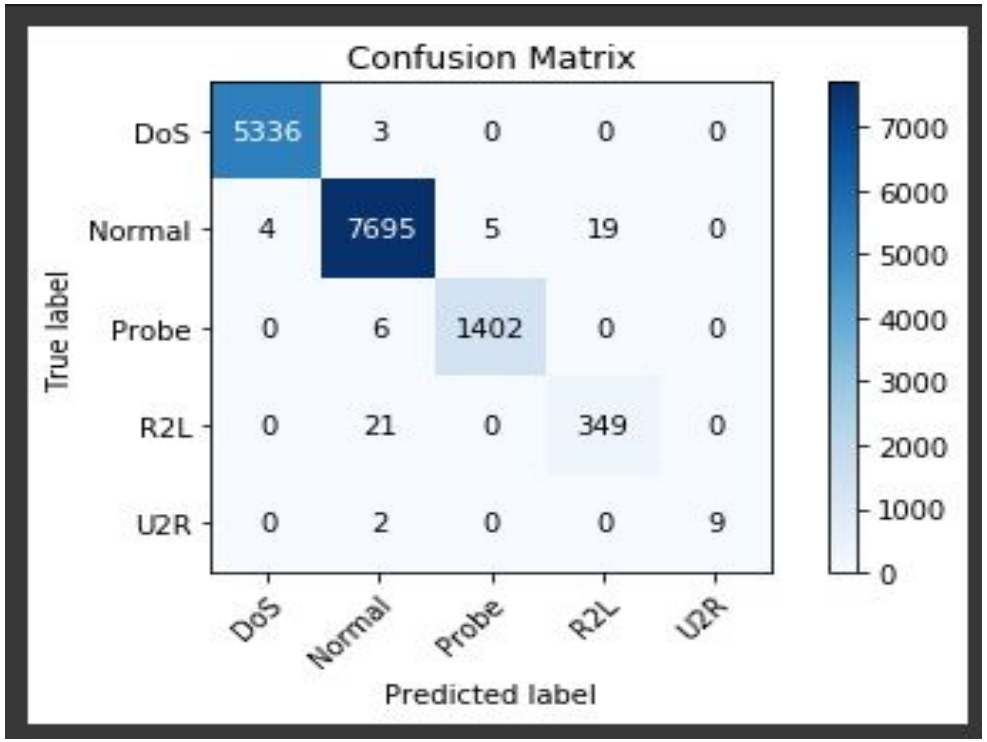
After series of Experiments involving several classifiers and Neural Network, the most accurate of all is found to be Neural Network which is employed with LSTM

K-value	Accuracy%	Detection Rate%	False Positive Rate%
2	0.985106	0.984877	0.378071
4	0.991220	0.990466	0.238356
6	0.993536	0.992910	0.177252
8	0.995529	0.994519	0.137021
10	0.995960	0.994788	0.130288
Average	0.992270	0.991512	0.212198

		Mean Acc	Mean Precision	Mean Recall	Mean F1
0	PLA	0.769784	0.750920	0.577282	0.593128
1	Logistic Regression	0.749956	0.665561	0.490451	0.483570
2	NN	0.809084	0.750800	0.622719	0.655025
3	DTree	0.765215	0.728301	0.544479	0.556670
4	Voting	0.764860	0.851480	0.509835	0.517197
5	Bagging of PLA	0.750754	0.696027	0.507112	0.523922
6	AdaBoost	0.761489	0.782096	0.509929	0.537895
7	Random Forest	0.748714	0.752554	0.479888	0.490832

Analysis of Results and Discussion





Contributions to the Work

ABHIJITH	ABIWAQAS
➔ Research for paper and pre existing models. ➔ Data Collection ➔ IDS and DNN Conceptualization ➔ Data Preprocessing ➔ Researching on the various attacks in the network ➔ Modelling using Unsupervised techniques (Classification Models) ➔ Neural Network Development	➔ Research for paper and pre existing models. ➔ Data Collection ➔ IDS and DNN Conceptualization ➔ Data Preprocessing ➔ Modelling using Unsupervised Techniques (Classification Models) ➔ Neural Network Development ➔ Research on Evaluation metrics

References

1. Chao Liang, Bharanidharan Shanmugam, Sami Azam, Mirjam Jonkman, Friso De Boer, Ganthan Narayansamy, “Intrusion Detection System for Internet of Things based on a Machine Learning approach”, International Conference on Vision Towards Emerging Trends in Communication and Networking, IEEE, 2019.

2. Patel, A., Taghavi, M., Bakhtiyari, K., Júnior, J.C. “An intrusion detection and prevention system in cloud computing: a systematic review”. J. Netw. Comput. Appl. Vol. 36, pp. 25–41, 2018.

3. Y. Farhaoui and A. Asimi, “Performance assessment of tools of the intrusion detection/prevention systems,” International Journal of Computer Science and Information Security, vol. 10, issue 1, pp. 1-7, 2020

4. Ma T et al, “A hybrid methodologies for intrusion detection based deep neural network with support vector machine and clustering technique”, International conference on frontier computing. Springer, 2019.

5. Manjula C. Belavagi and Balachandra Muniyal, Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection, Procedia Computer Science, Elsevier, 2017