Classification and Prediction Technique for DDoS Attacks Using Machine Learning

Dr. NAVANEETHA KRISHNAN M, M.E. Ph.D., Head of the Department,

Department of Computer Science and Engineering

Ms. T. GIFTY, B.E, Student of Computer Science Engineering

Ms. J. SURITHI, B.E, Student of Computer science and Engineering

St. Joseph College of Engineering, Sriperumbudur, Chennai.

ABSTRACT

Distributed network attacks are referred to as Distributed Denial of Service (DDoS)attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such as the framework of the authorized organization's site. In the existing research study.

It is necessary to work with the latest dataset to identify the current state of DDoS attacks. In this presented work, used a machine learning approach to predict DDoS attack types. For this purpose, used Random Forest and XGBoost classification algorithms. To access the research proposed a complete framework for DDoS attack prediction. To meet the proposed objective, we used UNWS-np-15 dataset and Python was used as a simulator.

After applying the machine learning models, we generated a confusion matrix for the identification of the model performance. In the first classification, the results showed that both Precision (PR) and Recall (RE) are 88% for the Random Forest algorithm.

In the second classification, the results showed that both precision(PR) and Recall(RE) are approximately 90% for the XGBoost algorithm.

Key Terms: DDoS – Denial of Service, PR – Precision, RE- Recall, RFL- Random Forest Algorithm, ML – Machine Learning.

INTRODUCTION

Distributed network attacks are referred to, usually, as Distributed Denial of Service (DDoS) attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such as the framework of the authorized organization's website. A DDoS attack sends different requests (with IP spoofing) to the target web assets to exceed the site's ability to handle various requests, at a given time, and make the site unable to operate effectively and efficiently – even for the legitimate users of the network. Typically, the target of various DDoS attacks are web applications and business websites; and the attacker may have different goals

A DDoS attack sends different requests (with IP spoofing) to the target web assets to exceed the site's ability to handle various requests, at a given time, and make the site unable to operate effectively and efficiently_ even for the legitimate users of the network. Typically, the target of various DDoS attacks are web applications and business websites; and the attacker may have different goals. We predict (Bining or DoS hulk or DoSslowloris).

CNN and RNN both are two different algorithms that can be used for different purposes. For example, CNN is used for feature extraction and RNN is used for regression in time series data utilization. Though both CNN and RNN based model producing accurate results, it is very long and time consuming process.

The authors used the CNN and RNN model for intrusion detection. This is a very long and time-consuming process. Therefore, it is very important to perform advanced machine learning techniques to model optimization that train the best model for highly accurate work.

Algorithm:

After preprocessing dataset, that data will be given to the machine learning algorithm. Machine learning algorithm analyzes the data and predict types of DDOSs attack .

Random Forest Classifier

A random forest algorithm is a combination of the decision tree. It is very fast compared to other classification. Now after feature scaling the next step is the machine learning classification model. In our proposed work we used a random forest classification algorithm. The random forest, which is one of the most popular and powerful machine learning classification algorithms, is used for reaching a lot of decisions in the proposed model. In the first classification we observed that both the Random Forest Precision (PR) and Recall (RE) are approximately 89% accurate.

XG Boost

In the era of machine learning and artificial intelligence, the XGBoost algorithm is known as the queen by scientific and academic researchers. Most of the researchers considering as a weapon for big data utilization. This model also working on tree but 100 times faster than other models. The XGBoost learning model have very fast speed, scalability, efficiency and simplicity. This model is more reliable for big data. This model is working on probability. The confusion matrix and outcomes of the classification, are given below, for the XGBoost method, XGBoost Precision (PR) and Recall (RE) are approximately 90% accurate. We noted approximately 90% average Accuracy

Modules:

- Dataset Collection
- Data Pre-process
- Detection

LITERATURE SURVEY

The paper "A machine Learning Approach for predicting DDOS Traffic in Software Defined Networks" by Kshira Sagar Sahoo, Amaan Iqbal, Prasenjit Maiti in 2018 used different machine learning algorithms and found that the KNN model is best for classification as compared to other research work. Nuno Martins et al proposed intrusion detection using machine learning approaches. They used the KDD dataset which is available on the UCI repository. They performed different supervised models to balance un classification algorithm for better performance.

The paper "Detection of DDoS Attacks using Machine Learning Algorithms" by Parvinder Singh Saini, Sunny Behal, Sajal Bhatia in 2020 is a comparative work for network traffic classification. They used machine learning classifiers for intrusion detection. The dataset is taken is CICIDS and KDD from the UCI repository. They found support vector machine SVM one of the best algorithms as compare to others

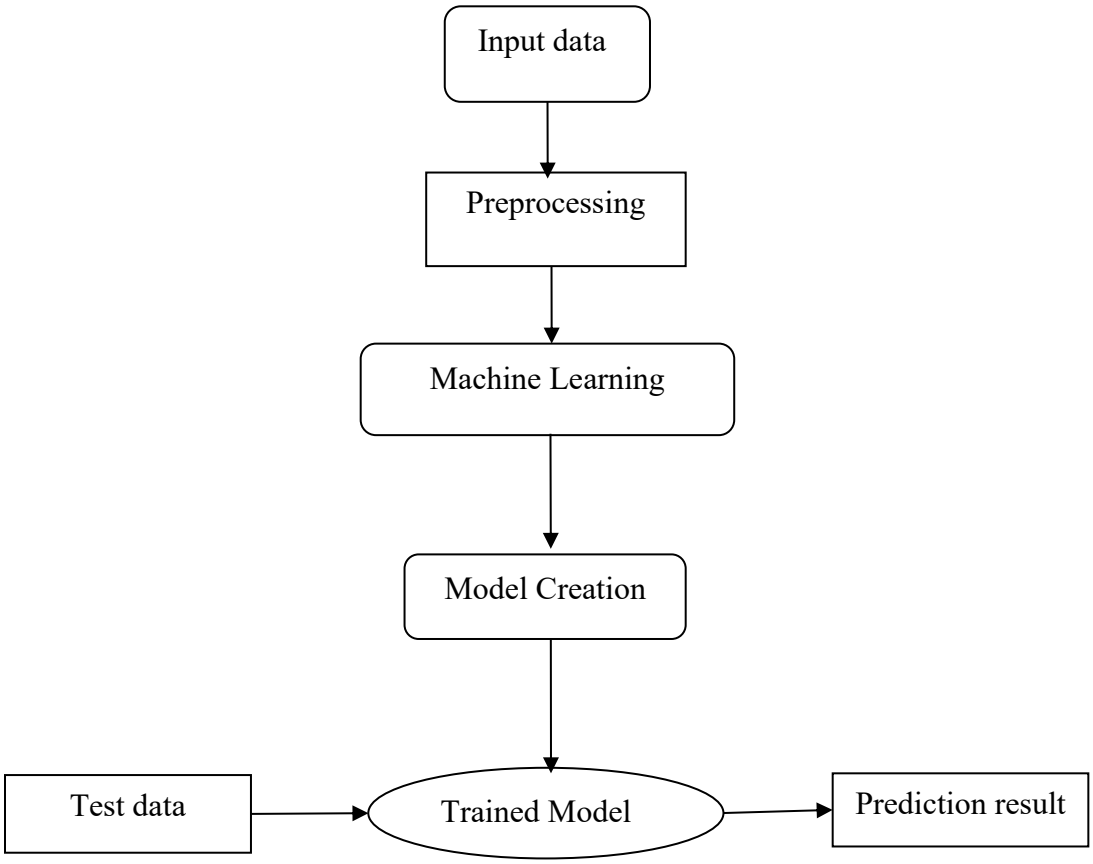
The paper " A Machine Learning-Based Classification and prediction Technique for DDoS Attacks" by Ismail, Hameed Hussain, Ayaz Ali Khan, Ubaid Ullah in 2022 is a systematic review of malware detection using machine learning models. They compared different malware datasets from online resources as well as approaches for the dataset. They found that machine learning supervised models are very effective for malware detection to make a better decision in less time.

The main objectives of this research are twofold: i) Analyze network traffic patterns to accurately identify and classify DDoS attacks, ii) Use machine learning algorithms to predict the likelihood of future DDoS attacks.

SYSTEM DESIGN

CollectedUNSW-nb15 dataset from GitHub1 that contains features' data about the DDoS attacks. This dataset is provided by the Australian Centre for Cyber Security (ACCS). The dataset consists of different features about the DDoS attacks including an ID number, Proto which presents medium of the network, label of the attacks, and attacks' cat which presents the severity of the DDoS attacksThe training dataset is a subset of the overall dataset that is used to train the machine learning model. The model learns from the patterns in the training dataset to make accurate predictions.

Data preprocessing it is very important and time-consuming part of data analysis. here we are going to separate relevant data from irrelevant data and convert it to quality information. For this step we are using statistical techniques to clean data and replace those values which are not important in our experimental analysis.



To design and develop an approach using supervised machine learning classifiers for DDoS attack detection based on different techniques.

We have studied various methodologies which are used for detection of Distributed Denial-of-Service (DDoS) Attacks on Software Defined Network, based on the findings and results we have concluded that the Attribute based Double of Transductive Confidence Machines for Random forest classifier method gives more efficient way to find out anomalous flow in Software Defined Network.

The result of the Machine Learning-Based Classification and Prediction Technique for DDoS Attacks is to easily predict any disturbances or malicious traffic in a network of systems and to identify the particular form of attack and to safeguard the systems connected in the network.

IMPLEMENTATION

Data Preprocessing and Model Creation


```
"cells": [
  {
   "cell_type": "code",
   "execution_count": 2,
   "id": "c4f46635",
   "metadata": {},
   "outputs": [],
   "source": [
    "import numpy as np\n",
    "import pandas as pd\n",
    "import matplotlib.pyplot as plt\n",
    "import seaborn as sns"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 3,
   "id": "3905e537",
   "metadata": {},
   "outputs": [],
   "source": [
    "data=pd.read_csv('./UNSW_NB15_training-set.csv')"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 4,
   "id": "f3de6375",
   "metadata": {},
   "outputs": [
    {
```

```
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "    .dataframe tbody tr th:only-of-type {\n",
    "        vertical-align: middle;\n",
    "    }\n",
    "\n",
    "    .dataframe tbody tr th {\n",
    "        vertical-align: top;\n",
    "    }\n",
    "\n",
    "    .dataframe thead th {\n",
    "        text-align: right;\n",
    "    }\n",
    "</style>\n",
    "<table border=\"1\" class=\"dataframe\">\n",
    "  <thead>\n",
    "    <tr style=\"text-align: right;\">\n",

   ],

{
 "cell_type": "code",
 "execution_count": 5,
 "id": "54899751",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "(82332, 45)"
    ]
   },
   "execution_count": 5,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "data.shape\n"
 ]
},
{
 "cell_type": "code",
 "execution_count": 6,
 "id": "562d27f1",
```

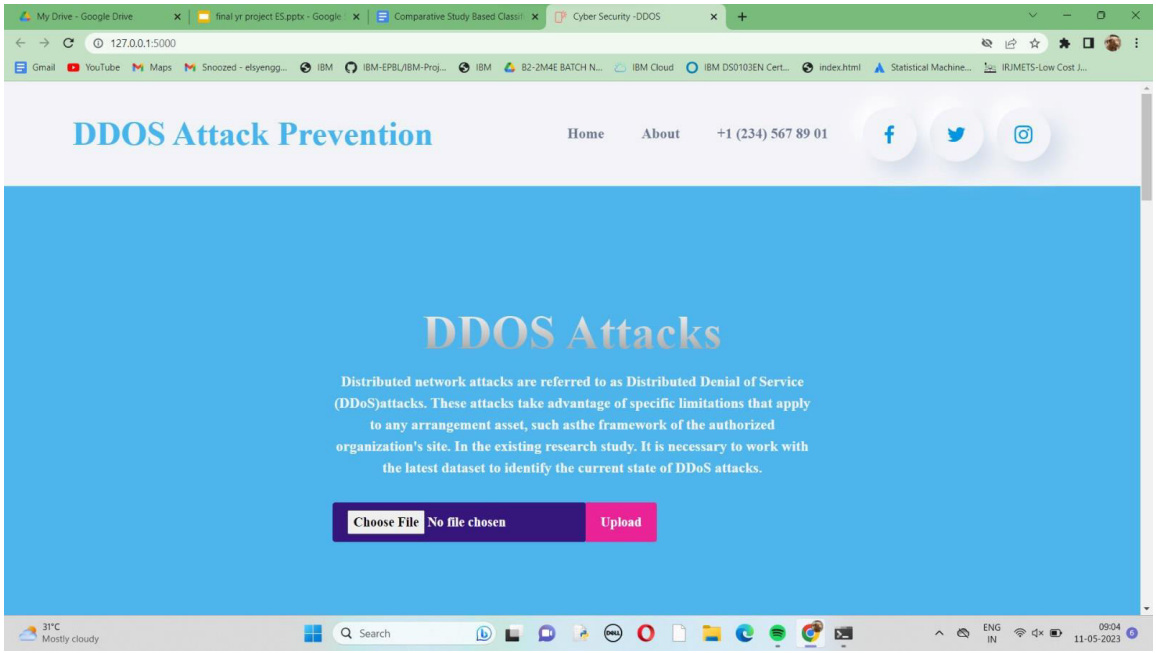
```
 "metadata": {},
 "outputs": [],
 "source": [
  "data=data.drop('label', axis=1 ) "
 ]
},
{
 "cell_type": "code",
 "execution_count": 7,
 "id": "d4211c80",
 "metadata": {},
   },
   "execution_count": 9,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "data['service']"
 ]
},
{
 "cell_type": "code",
 "execution_count": 10,
 "id": "062249c1",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/html": [
     "<div>\n",
     "<style scoped>\n",
     "    .dataframe tbody tr th:only-of-type {\n",
     "        vertical-align: middle;\n",
     "    }\n",
     "\n",
     "    .dataframe tbody tr th {\n",
     "        vertical-align: top;\n",
     "    }\n",
     "\n",
     "    .dataframe thead th {\n",
     "        text-align: right;\n",
     "    }\n",

   },
   "execution_count": 10,
   "metadata": {},
```

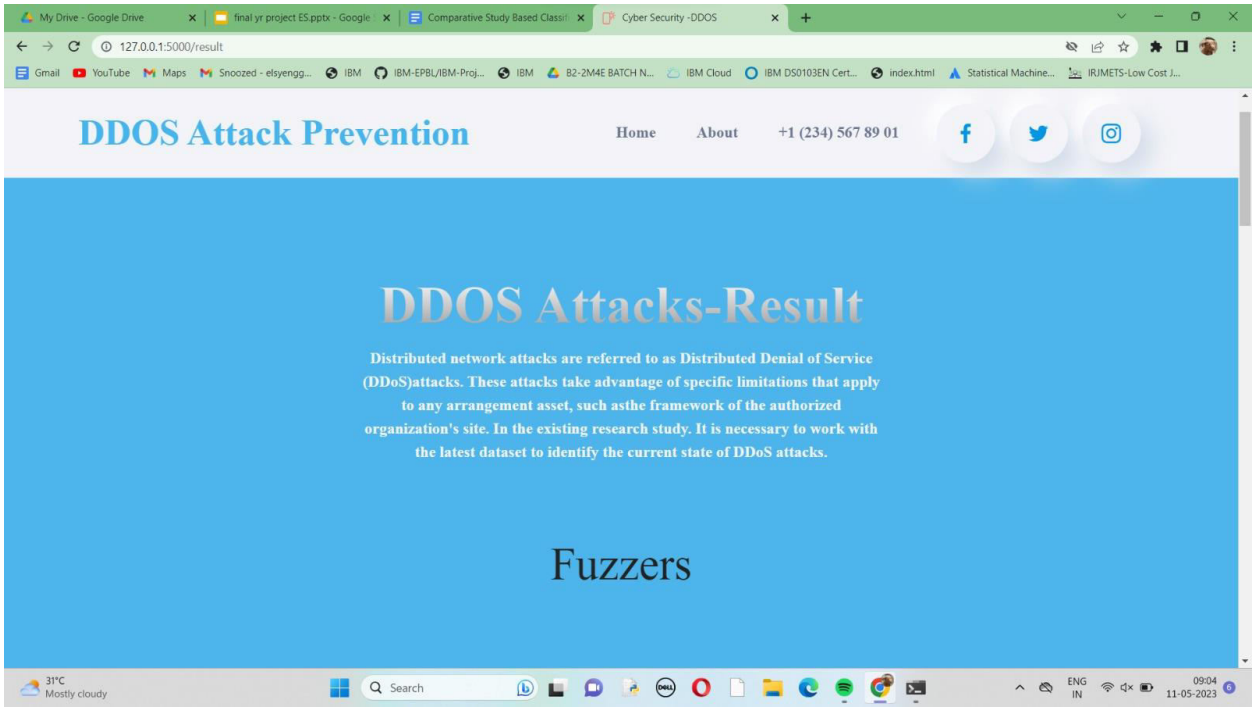
```
   "output_type": "execute_result"
  }
 ],
 "source": [
  "data"
 ]
},
{
 "cell_type": "code",
 "execution_count": 11,
 "id": "bced052e",
 "metadata": {},
 "outputs": [],
 "source": [
  "data_proto=data['proto']\n",
  "data_service=data['service']\n",
  "data_state=data['state']\n",
  "data_proto.value_counts()\n",
  "data_service.value_counts()\n",
  "data['service']=data['service'].fillna(data['service'].mode()[0])"
 ]
},
{
 "cell_type": "code",
 "execution_count": 12,
 "id": "a6079fc8",
 "metadata": {},
 "outputs": [],
 "source": [
  "data_proto=pd.get_dummies(data_proto)\n",
  "data_service=pd.get_dummies(data_service)\n",
  "data_state=pd.get_dummies(data_state)\n"
 ]
},
{
 "cell_type": "code",
 "execution_count": 13,
 "id": "2b605391",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
  "print(data_proto)\n",
  "print(data_service)\n",
  "print(data_state)"
```

```
 ]
},
{
 "cell_type": "code",
 "execution_count": 14,
 "id": "65e695de",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/html": [
     "<div>\n",
     "<style scoped>\n",
     "    .dataframe tbody tr th:only-of-type {\n",
     "        vertical-align: middle;\n",
     "    }\n",
     "\n",
     "    .dataframe tbody tr th {\n",
     "        vertical-align: top;\n",
     "    }\n",
     "\n",
     "    .dataframe thead th {\n",
     "        text-align: right;\n",
     "    }\n",
     "</style>\n",
```


SNAPSHOTS


DDOS Attack Prevention
Home
About
+1 (234) 567 89 01
DDOS Attacks
Distributed network attacks are referred to as Distributed Denial of Service (DDoS)attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such asthe framework of the authorized organization's site. In the existing research study. It is necessary to work with the latest dataset to identify the current state of DDoS attacks.
Choose File
No file chosen
Upload


DDOS Attack Prevention
Home
About
+1 (234) 567 89 01
DDOS Attacks
Distributed network attacks are referred to as Distributed Denial of Service (DDoS)attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such asthe framework of the authorized organization's site. In the existing research study. It is necessary to work with the latest dataset to identify the current state of DDoS attacks.
Choose File
test2.xlsx
Upload



CONCLUSION

In this paper, we proposed a complete systematic approach for detection of the DDOS attack. First, we selected theUNSW-nb15 dataset from the GitHub repository that contains information about the DDoS attacks. This dataset was provided by the Australian Centre for Cyber Security. Then, Python and jyupter notebook are used to work on data wrangling. Secondly, we divided the dataset into two classes i.e. the dependent class and the independent class. Moreover, we normalized the dataset for the algorithm. After data normalization, we applied the proposed, supervised, machine learning approach. The model generated prediction and classification outcomes from the supervised algorithm.

FUTURE ENHANCEMENTS

The paper suggests improving the user-friendliness and speed of the proposed technique to make it more accessible for practical application. The paper suggests exploring the potential of incorporating other features in the dataset to improve the accuracy of the model. The authors suggest testing the proposed technique on larger datasets to evaluate its scalability and generalizability. The paper also recommends investigating the potential of integrating the proposed technique with existing DDoS detection techniques to create a more comprehensive defense

system.

REFERENCES

[1] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ‘‘Adversarial machine learning applied to intrusion and malware scenarios: A systematic review,’’ IEEE Access, vol. 8, pp. 35403–35419, 2020.

[2] G. Karatas, O. Demir, and O. K. Sahingoz, ‘‘Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset,’’ IEEE Access, vol. 8, pp. 32150–32162, 2020.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ‘‘BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset,’’ IEEE Access, vol. 8, pp. 29575–29585, 2020.

[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ‘‘Network intrusion detection based on PSO-xgboost model,’’ IEEE Access, vol. 8, pp. 58392–58401, 2020.

[5] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ‘‘Similarity based feature transformation for network anomaly detection,’’ IEEE Access, vol. 8, pp. 39184–39196, 2020.

[6] A. Agarwal, M. Khari, and R. Singh, ‘‘Detection of DDOS attack using deep learning model in cloud storage application,’’ Wireless Pers. Comun., vol. 2, pp. 1–21, Mar. 2021. [18] Z. Akhtar, ‘‘Malware detection and analysis: Challenges and research opportunities,’’ 2021, arXiv:2101.08429.

[7] D. C. Can, H. Q. Le, and Q. T. Ha, ‘‘Detection of distributed denial of service attacks using automatic feature selection with enhancement for imbalance dataset,’’ in Proc. ACIIDS, 2021, pp. 386–398, doi: 10.1007/978-3-030-73280-6_31.

[8] Q. Tian, J. Li, and H. Liu, ‘‘A method for guaranteeing wireless communication based on a combination of deep and shallow learning,’’ IEEE Access, vol. 7, pp. 38688–38695, 2019. [21] Q. Cheng, C. Wu, H. Zhou, D. Kong, D. Zhang, J. Xing, and W. Ruan, ‘‘Machine learning based malicious payload identification in softwaredefined networking,’’ 2021, arXiv:2101.00847.

[9] S. K. Wanjau, G. M. Wambugu, and G. N. Kamau, ‘‘SSH-brute force attack detection model based on deep learning,’’ Murang’a Univ. Technol., Murang’a, Kenya, Tech. Rep. 4504, 2021. [Online]. Available: http://repository.mut.ac.ke:8080/xmlui/handle/123456789/4504

[10] K. S. Sahoo, B. K. Tripathy, K. Naik, S. Ramasubbareddy, B. Balusamy, M. Khari, and D. Burgos, ‘‘An evolutionary SVM model for DDOS attack detection in software defined networks,’’ IEEE Access, vol. 8, pp. 132502–132513, 2020

AUTHOR 1

Dr.M.Navaneethakrishnan M.E., PhD is a Head of the Department in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. He has completed his Ph.D, in Cyber Security - Computer Science and Engineering in 2017 from Manonmaniam Sundaranar University (MSU) Tirunelveli, Tamilnadu. He has done his M.E, CSE in Anna University Chennai in the year 2008. Dr.M.Navaneethakrishnan has 15 years of teaching experience and has 58 publications in International Journals and Conferences. His research interests include network security, Computer Networks, data science and Machine Learning. He is an active member of ISTE, CSI, IEANG and IEI

AUTHOR 2

Miss.T.GIFTY B.E., Student of Computer Science and Engineering at St.Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I have attended many Workshops and have participated in many Paper Presentations and have won prizes. I am a highly motivated engineering student with great hopes of achieving great things.

AUTHOR 3

Miss.J.SURITHI B.E., Student of Computer Science and Engineering at St.Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I had attended many Workshops and Seminars in the area of Python and Machine Learning. I have also won many Prizes in inter-college Paper Presentation Events. I am an ambitious person with skills required to achieve bigger heights.