Machine Learning Based Prediction of DDOS Attack Types Using Random Forest and XGBOOST Algorithms

Mrs. SINDHU BARATHI, M.E. Assistant professor, Department of Computer Science and Engineering

Mr. S.GOKUL,B.E, Student of Computer Science Engineering

Mr. V.SELVAKUMAR, 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. Distributed Denial of Service is one of the most effective and costly top five cyber-attacks.

Distributed Denial of Service (DDoS) is a type of cyber-attack that prevents legitimate users from accessing network system resources. To minimize major damage, quick and accurate DDoS attack detection techniques are essential. To classify target classes, machine learning classification algorithms are faster and more accurate than traditional classification methods.

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. To access the research proposed a complete framework for DDoS attacks 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 identification of the model performance. In the first classification, the results showed that both Precision (PR) and Recall (RE) are 96% 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: BMI - Body Mass Index, KNN – K Nearest Neighbor, LightGBM – Light Gradient Boosting Machine, ML – Machine Learning, SKLearn – Sci-Kit Learn, SMOTE – Synthetic Minority Oversampling Technique.

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 [1], [2].

Some common types of the DDoS attacks are The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao . We give brief description of each attack in Section I-A. The Internet of Things (IoT) implies the arrangement of interconnected, web-related objects that can collect and interchange information through remote organizations without manual intervention [3].

The ''Things'' can simply be related clinical tools, bio-chip transponders, solar panels, and related vehicles with sensors that can warn the driver of numerous potential problems [4], or any article with sensors that can collect and move information in the organization. Artificial intelligence (AI) is a small tool that transforms information into data. In the past 50 years (approximately), information has had an impact on users privacy and security

LITERATURE SURVEY

In the literature review section we briefly explained all the related model and the closest rival to our proposed study. We studied the latest research papers of the past two years for this research work and also Gozde Karatas et al. [2] proposed a machine learning approach for attacks classification. They 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. [1] 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. In this work, a comparative study was proposed by the use of different classification algorithms and found good results in their work. Laurens D'hooge et al. [6] proposed a systematic review for 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. Xianwei Gao et al. [7] proposed 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. Tongtong Su et al. [3] proposed adaptive learning for intrusion detection. They used the KDD dataset from an online repository. These models are Dtree, R-forest, and KNN classifiers. In this study, the authors found that Dtree and ensemble models are good for classification results. The overall accuracy of the proposed work is 85%.

SYSTEM DESIGN

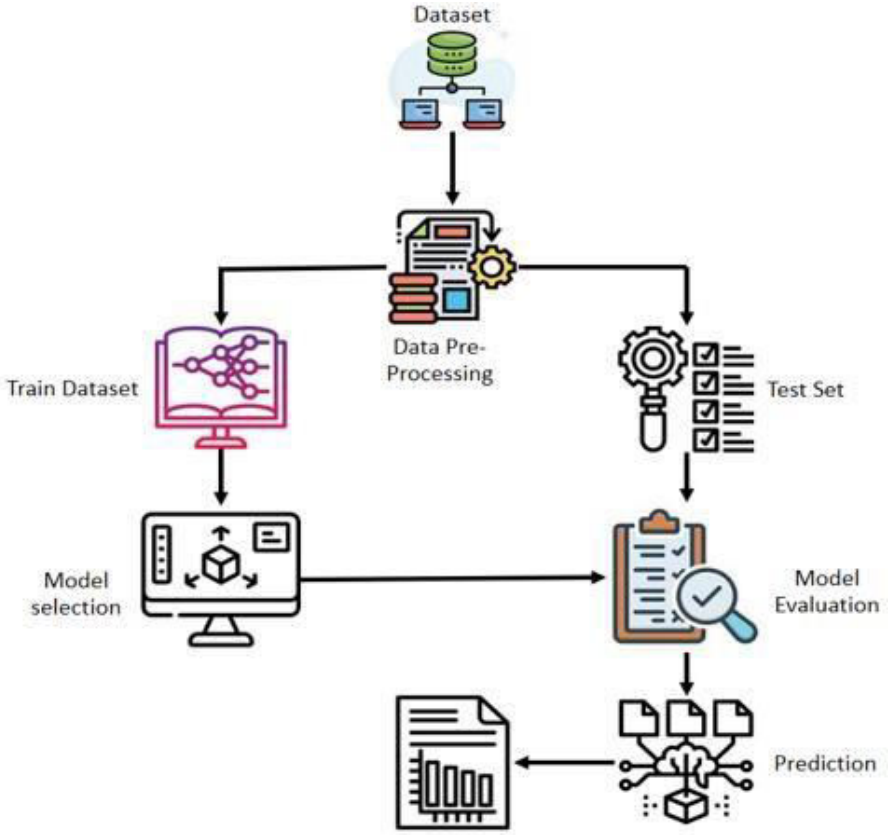
The dataset is the collection of data that is used to train and evaluate the machine learning model. It typically includes features such as age, gender, medical history, lifestyle factors, and other relevant information that may contribute to the risk of stroke.

Data preprocessing involves preparing the dataset for use in the machine learning model. This may include cleaning the data, removing outliers and errors, filling in missing values, and scaling or normalizing the data.

The 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.

The test dataset is a separate subset of the overall dataset that is used to evaluate the performance of the machine learning model. It is used to measure the accuracy of the

model's predictions on new data that it has not seen before.



Model selection involves choosing the appropriate machine learning algorithm for the task. Commonly used algorithms for stroke prediction include logistic regression, decision trees, random forests, and support vector machines.

Model evaluation involves testing the performance of the machine learning model using the test dataset. Common metrics used to evaluate the model include accuracy, precision, recall, and the area under the receiver operating characteristic curve.

Once the machine learning model has been trained and evaluated, it can be used to make predictions on new data. In the case of stroke prediction, the model can be used to predict the risk of stroke for a given patient based on their demographic and medical information.

The result of the machine learning model is the prediction of stroke risk for a given patient. This information can be used by clinicians to make more informed decisions about patient care, such as recommending lifestyle changes or prescribing medication to reduce the risk of stroke.

IMPLEMENTATION:

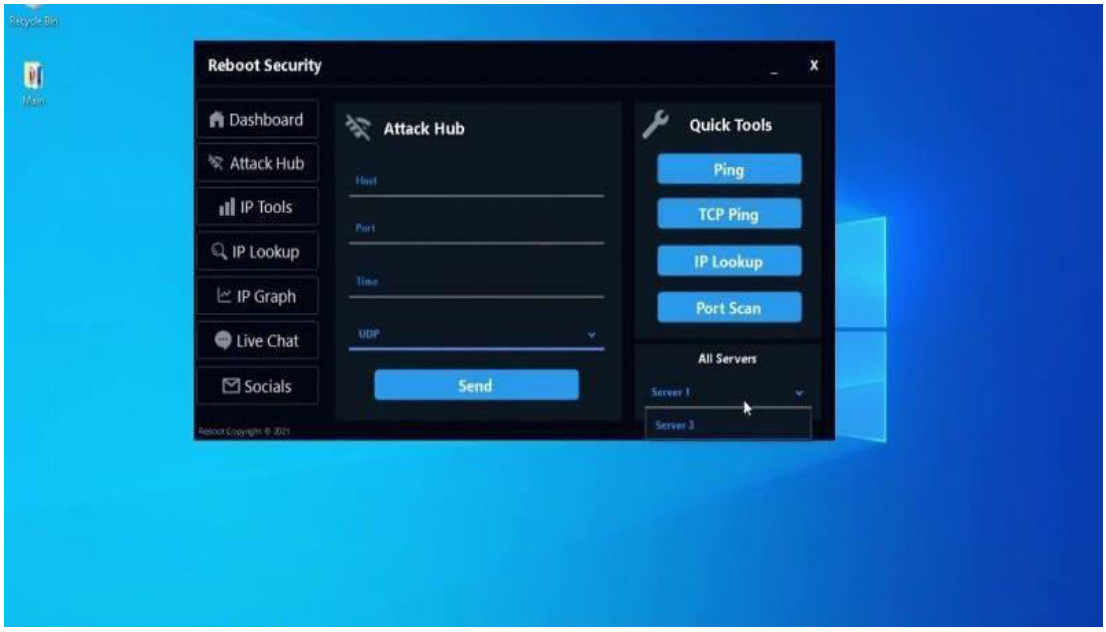
```
def attack():
while True:
    s = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    s.connect((target, port))
    s.sendto(("GET /" + target + " HTTP/1.1\r\n").encode('ascii'),(target,
    port))
    s.sendto(("Host: " + fake_ip + "\r\n\r\n").encode('ascii'), (target, port))
    s.close()
   for i in range(500):
  thread
threading.Thread(target=attack)thread.start(
)
```

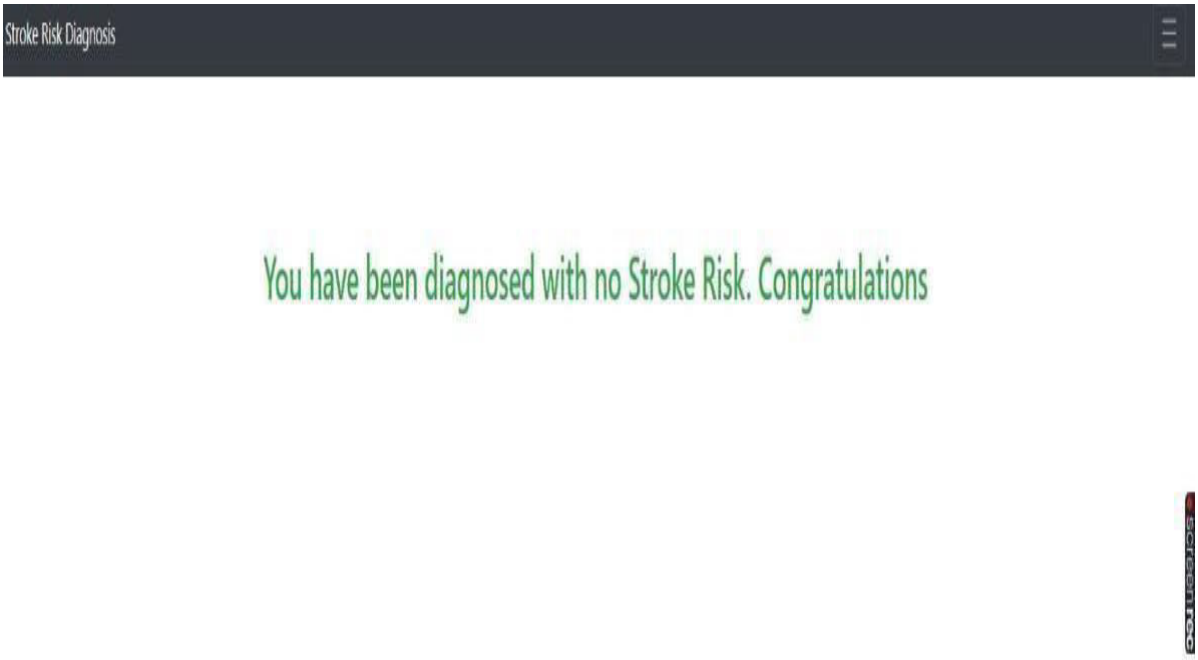
down your attack If you want to see some information, you may print the amounts of requests already sent. Just notice that this will slow.

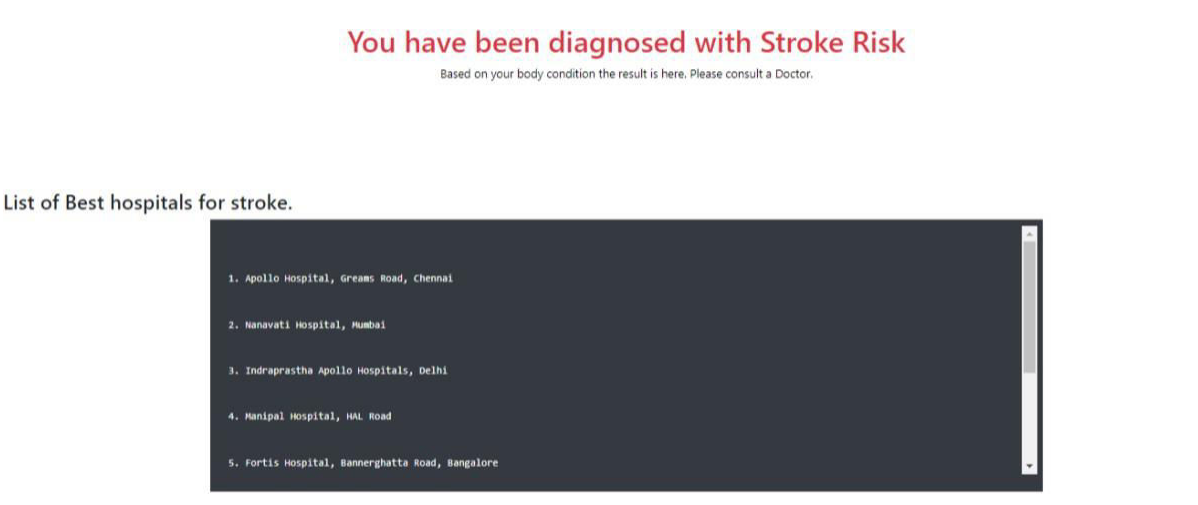
```
attack_num = 0
def attack():
    while True:
        s = socket.socket(socket.AF_INET, socket.SOCK_STREAM) try:
        ddos_host = input("Enter Host URL of the server: ")
        ddos_password= input("Enter the password: ")
        if "http" not in ddos_host: raise Exception("Wrong Host URL.")
        if "/" != ddos_host[-1]:
```

```
ddos_host += "/"

test_data {"password": ddos_password}

agreed requests.post((ddos_host + "get/agreed"), data-
test_data).text

=

if agreed != "False":

raise Exception("Wrong data was given.")

ddos_role tools. question ("Do you want this to be used as a
host?") status = requests.post((ddos_host + "reset"), data-
test_data).text

if status != "200":

print("Something strange happened.")

var.server[1] = ddos_role

var.server[2] ddos_host =

var.server [3] ddos_password =

var.server[0] = True

except Exception as ex:

print("An exception occured.", ex)

print("")

s.connect((target, port))

s.sendto(("GET /" + target + " HTTP/1.1\r\n").encode('ascii'), (target,
port))

s.sendto(("Host: " + fake_ip + "\r\n\r\n").encode('ascii'), (target, port))

global attack_num
attack_num += 1
print(attack_num)
s.close(
```

SNAPSHOTS:


Reboot Security
Dashboard
Attack Hub
IP Tools
IP Lookup
IP Graph
Live Chat
Socials
Attack Hub
Host
Port
Time
UDP
Send
Quick Tools
Ping
TCP Ping
IP Lookup
Port Scan
All Servers
Server 1
Server 3


Stroke Risk Diagnosis
You have been diagnosed with no Stroke Risk. Congratulations


You have been diagnosed with Stroke Risk
Based on your body condition the result is here. Please consult a Doctor.
List of Best hospitals for stroke.
1. Apollo Hospital, Greams Road, Chennai
2. Nanavati Hospital, Mumbai
3. Indraprastha Apollo Hospitals, Delhi
4. Manipal Hospital, HAL Road
5. Fortis Hospital, Bannerghatta Road, Bangalore

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.

FUTURE ENHANCEMENTS

Looking to the future, for functional applications, it is important to provide a more user-friendly, faster alternative to deep learning calculations, and produce better results with a shorter burning time. It is important to work on unsupervised learning toward supervised learning for unlabeled and labeled datasets. Moreover, we will investigate how non-supervised learning algorithms will affect the DDoS attacks detection, in particular, we non-labeled datasets are taken into account.

REFERENCES:

[1] K. Singh, P. Singh, and K. Kumar, “Application layer HTTPGET flood DDoS attacks: research landscape and challenges,” Computers & Security, vol. 65, pp. 344–372, 2017.

[2] M. Masdari and M. Jalali, A Survey and Taxonomy of DoS Attacks in Cloud Computing, 2016, http://onlinelibrary.wiley. com/doi/10.1002/sec.1539/epdf.

[3] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ``An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7,pp. 82512_82521, 2019.

[4] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, ``Toward a lightweight intrusion detection system for the Internet of Things," IEEE Access, vol. 7, pp. 42450_42471, 2019.

[5] X. Larriva-Novo, V. A. Villagrá, M. Vega-Barbas, D. Rivera, and M. S. Rodrigo,

``An IoT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets," Sensors, vol. 21, no. 2, p. 656, Jan. 2021.

[6] 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.

[7] M. Zakarya, ``DDoS veri_cation and attack packet dropping algorithm in cloud computing," World Appl. Sci. J., vol. 23, no. 11, pp. 1418_1424,

2013

AUTHOR 1

Mrs. SINDHU BARATHI M.E., in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. she has done his M.E, CSE in Anna University Chennai in the year 2014 . she is an active member of IEI

AUTHOR 2

Mr.S.GOKUL B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, TamilNadu. I had attended many Workshops, Seminars in Python, Machine Learning. I got placed in Reputed Companies like Click Solutions, Q Spider and some respected companies.

AUTHOR 3

Mr.V.SELVAKUMAR, B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. I had attended many Workshops and Seminars in the area of Python and Machine Learning.