DDOS ATTACK DETECTION USING NETWORK TRAFFIC CLASSIFICATION TECHNIQUES

Project done by:
Hriman Krishna Mahanta
Vikash Bhuyan
Under the supervision of:
Dr. Satyajit Sarmah

ABSTRACT

- The Internet has become an almost indispensable part of the modern world. We are accessing the Internet when using our computers, smartphones and other smart devices.
- Thus, Network Security is increasingly becoming a need of any organisation as the security threats are increasing day by day.
- In this project, we will try to analyze and detect one of the most popular and costly attack namely the Distributed Denial Of Service (DDoS) attack.
- The basic idea of this project is to derive various features from a network traffic and then use those features to classify the network using different machine learning algorithms.

MOTIVATION

- The impacts of a DDoS attack on an organisation or business can be severe. Various DDoS attacks have reportedly caused resources to be offline for 24 hours, multiple days or even a week depending on the severity of the attack.
- During an attack, no employees are able to access network resources and in the case of servers running eCommerce websites, no consumers will be able to purchase products which will lead to huge financial loss for the company.
- Thus, it is very important that a mechanism needs to be created which will help in mitigating the DDoS attack and make the network more secure.
- The aim of this project is to create a model using machine learning which will be able to detect a DDoS attack with a good accuracy.

RELATED WORKS

- A lot of work has been done by various researchers in order to detect DDoS attack using different perspectives. Most of the works on DDoS detection can be broadly categorised into the following approaches:

  (i) Port based detection: This analyzes the port in a server which is congested with lots of requests and detects the anomalies.

  (ii) Payload based detection: This uses pattern recognition of the payload data to detect anomalies in the network traffic.

  (iii) Headers field based detection: This uses the data values present in the headers of the various layers of each packets to detect anomalies in the network traffic.

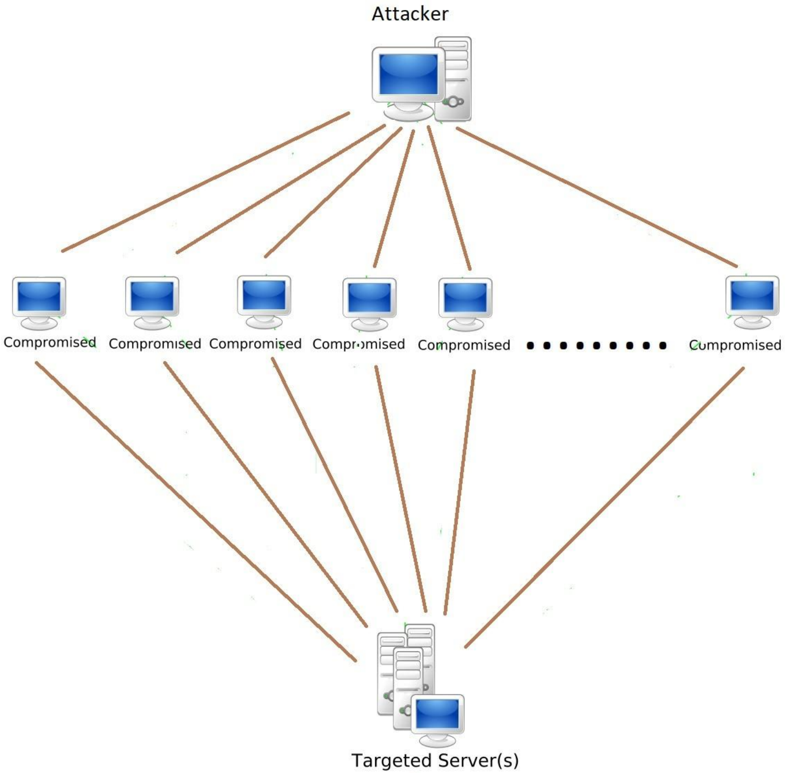
A BRIEF OVERVIEW OF THE PROJECT

- In this project, we have first captured and downloaded pcap files which consists of both normal and attack traffic. Pcap files are network capture files which contains details of the data present in each OSI layer for every packets captured.
- Then, we have extracted various features from the packets by creating a time window. We have used these features later to classify the packets using various machine learning algorithms.
- We have analyzed the packets and derived various interesting features such as the distribution of the protocols on the packets.
- Finally, we have used machine learning algorithm to classify the packets into normal or attack traffic using the features as the parameters for classification.

WHAT IS A DDoS ATTACK ?

- DDoS attack is one of the most prominent types of attacks used today whose aim is to cease all the services offered by a server by flooding the server with a huge amount of request at a very short period of time.
- DDoS attacks mostly happen when an attacker manages to manipulate many distributed network devices by installing malware and then using those devices to carry out a simultaneous flooding of a server.
- The aim of these attacks is to exhaust memory, processor and bandwidth resources of the victim computer.
- As a result of DDoS attack, web servers might become unable to provide HTTP service and email servers might not be able to send or receive emails.

STRUCTURE OF A DDoS ATTACK

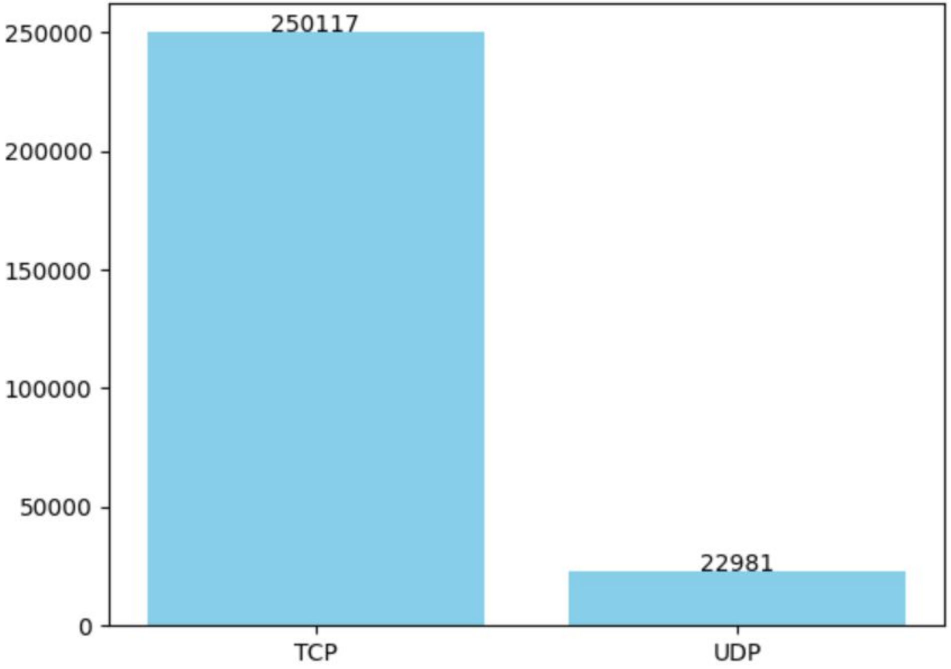


CAPTURING THE PACKETS

- We have used the Wireshark software to capture the packets flowing in a network. Wireshark is a packet capturing tool which captures the network packets flowing into and out from an interface of a computer.
- Wireshark stores the captured packets in a file format “.pcap”. This files contains details of the data values present in each packet on all the layers in the OSI model.
- We have used these data values to analyze the packets and give some interesting statistics about the kind of network traffic that we use when we usually surf through the Internet.

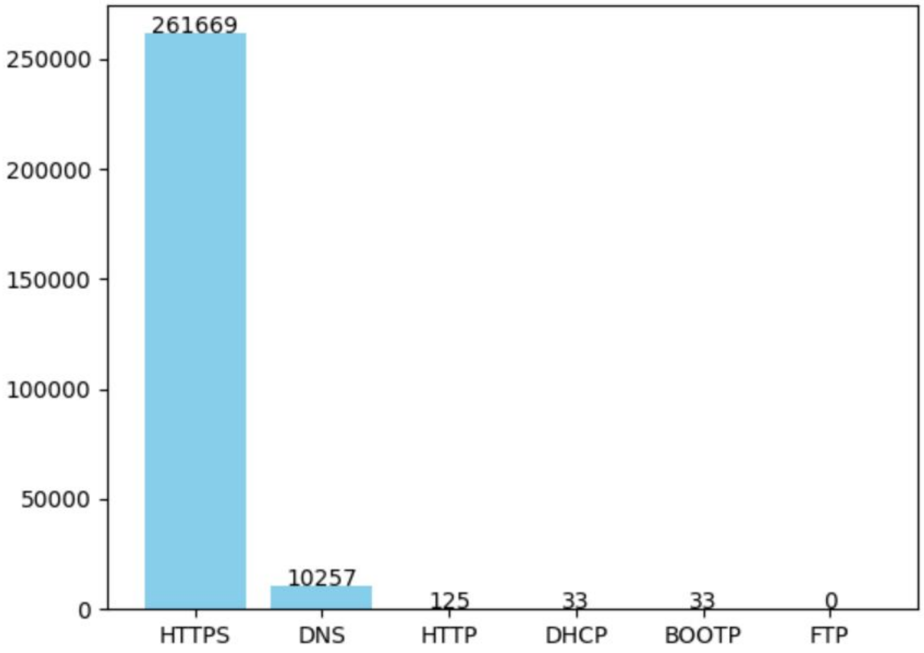
TRANSPORT LAYER PROTOCOLS OF THE PACKETS

- In each packet, the protocol used in the transport layer is present in the "Protocol" field of the IP header.
- We have collected the protocols of each packet on a list in python by extracting the value present in the protocol field.
- Then, we have calculated the number of packets containing TCP and UDP in their transport layer.



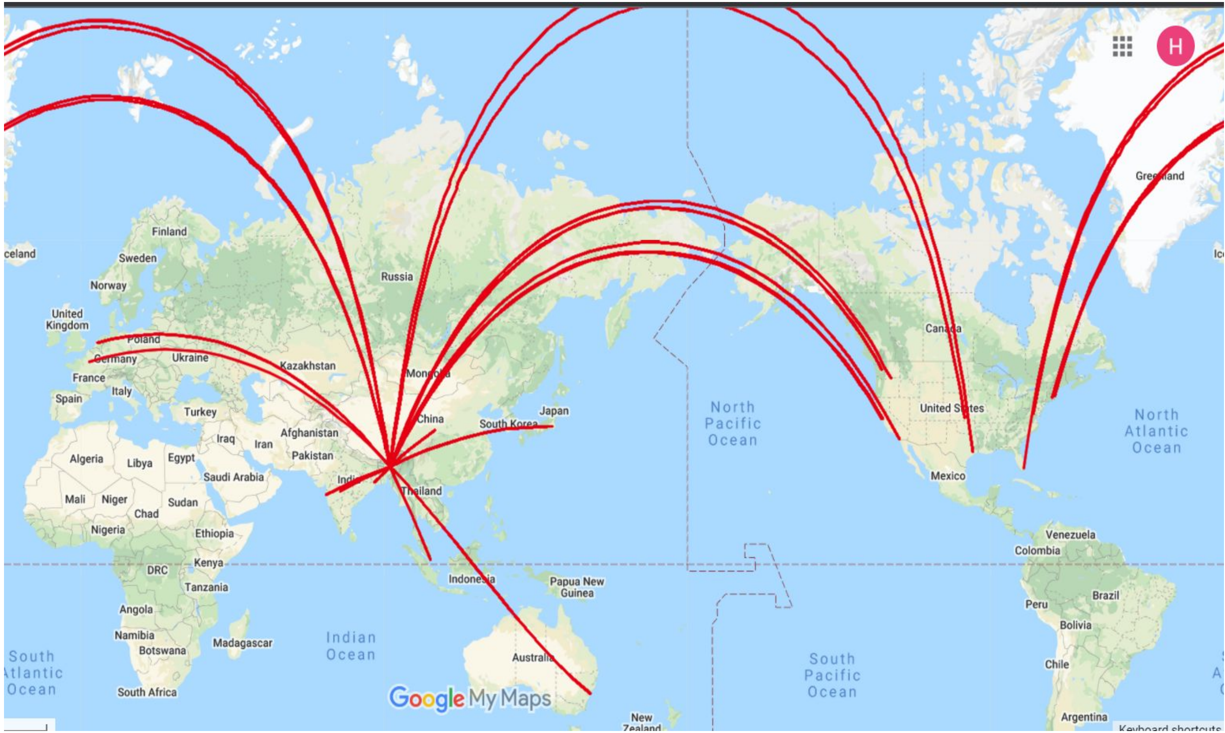
APPLICATION LAYER PROTOCOLS OF THE PACKETS

- The information about the protocol used in application layer can be found from the source port and destination port fields present in TCP and UDP.
- Each application layer protocol has a specific port number assigned to it. The port number is present in the source and destination port fields.
- We have used the port number detail present in each packet to find the corresponding application layer protocols.



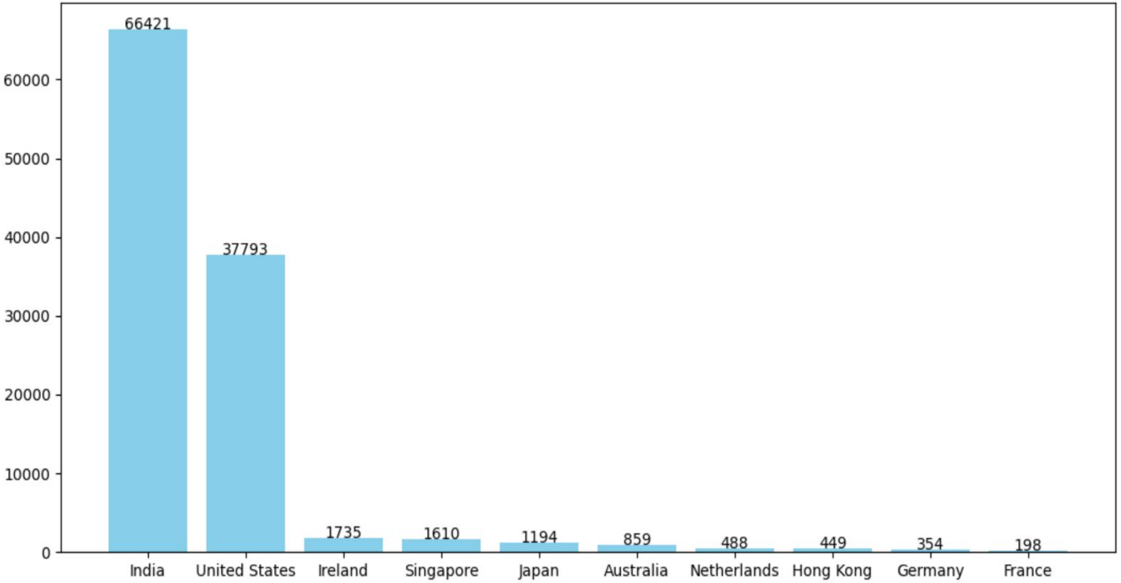
DISTRIBUTION OF NETWORK TRAFFIC FROM DIFFERENT PARTS OF THE WORLD

- To find the corresponding location of different IP address, we have used the “GeoLiteCity” database. This database contains the location of the various computers around the world corresponding to their IP address.
- Using the IP address of each packets that we have captured in the .pcap file, we have plotted a map which displays the route which the packets have taken from the server to our computer.
- We have used the Google My Maps feature in which we have uploaded the latitude and longitude corresponding to each IP address.



AMOUNT OF PACKETS COMING FROM DIFFERENT COUNTRIES

- The GeoLiteCity database also contains the country corresponding to each IP address.
- Using this information, we have tried to count the number of packets coming from different countries.
- The statistics derived here gives a good idea about the amount of network traffic coming from different parts of the world.



CAPTURING THE DATA FOR DDOS ATTACK

- To capture DDoS traffic, we have used a software named Low Orbit Ion Cannon (LOIC). LOIC simulates a DDoS attack by sending a huge number of TCP SYN packets to a particular computer.
- We have then captured the packets in the victim computer and stored them as pcap files.
- In addition to this, we have also downloaded some pcap files online which captured some real world DDoS attacks.
- We have then accumulated all the normal and attack traffic together and labelled them as “normal” and “attack” respectively so that we can use them to train our model for supervised learning.

EXTRACTING THE FEATURES

- After accumulating all the pcap files, we have extracted various features from the network traffic by creating a sliding window size of 500 ms. The features we have extracted are:

(i) tcp_frame_length

(ii) tcp_ip_length

(iii) tcp_length

(iv) udp_frame_length

(v) udp_ip_length

(vi) udp_length

(vii) num_tls

(viii) num_http

(ix) num_dhcp

(x) num_dns

(xi) num_tcp

(xii) num_udp

(xiii) num_igmp

(xiv) num_connection_pairs

(xv) num_ports

(xvi) num_packets

CLASSIFICATION OF NETWORK TRAFFIC

- By using the features, we have tried to classify the packets into belonging to either normal traffic or attack (DDoS) traffic.
- The classification algorithms we have used are:

  (i) Logistic Regression

  (ii) Decision Tree Classifier

  (iii) Random Forest Classifier

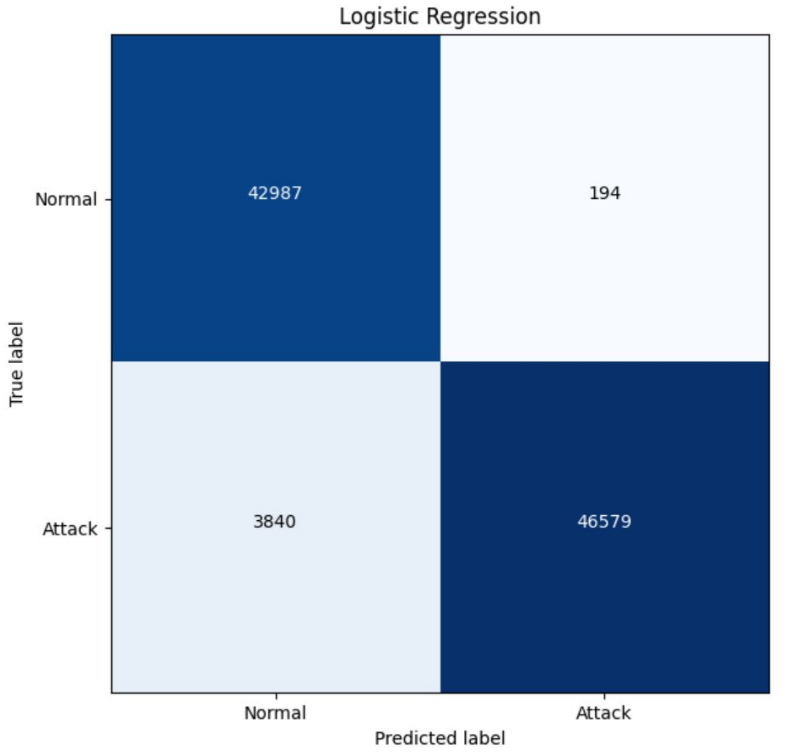
  (iv) Naive Bayes Classifier

  (v) K Nearest Neighbours Classifier

  (vi) Standard Vector Machine
- We have used the “sklearn” package of python to implement these machine learning algorithms.

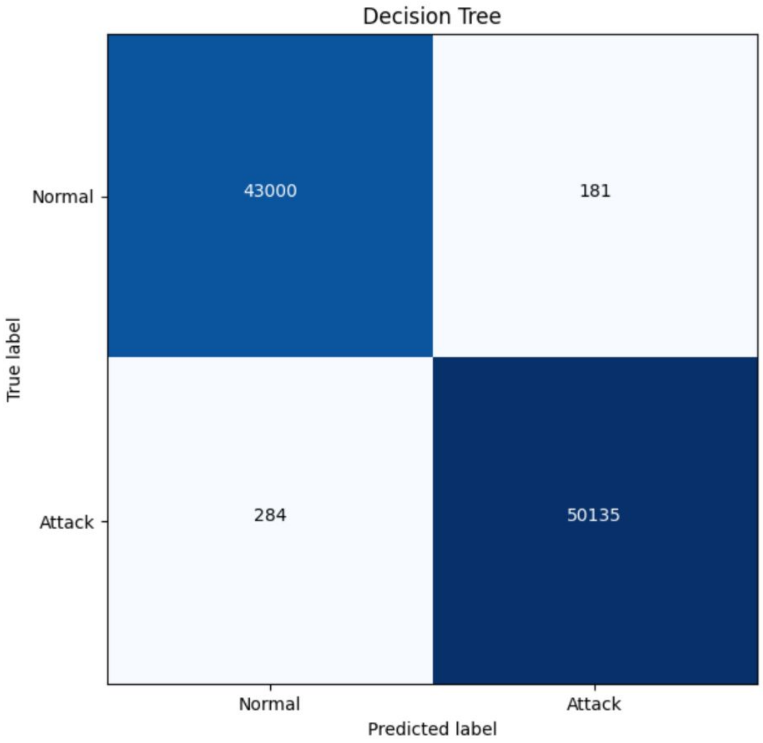
LOGISTIC REGRESSION

- Logistic Regression is used to estimate discrete binary values like 1/0, Yes/No, etc.
- It uses the the sigmoid function in order to classify the data.
- In our result, we have noticed quite a good accuracy but the number of attack packets which are predicted as normal packets are a bit higher.



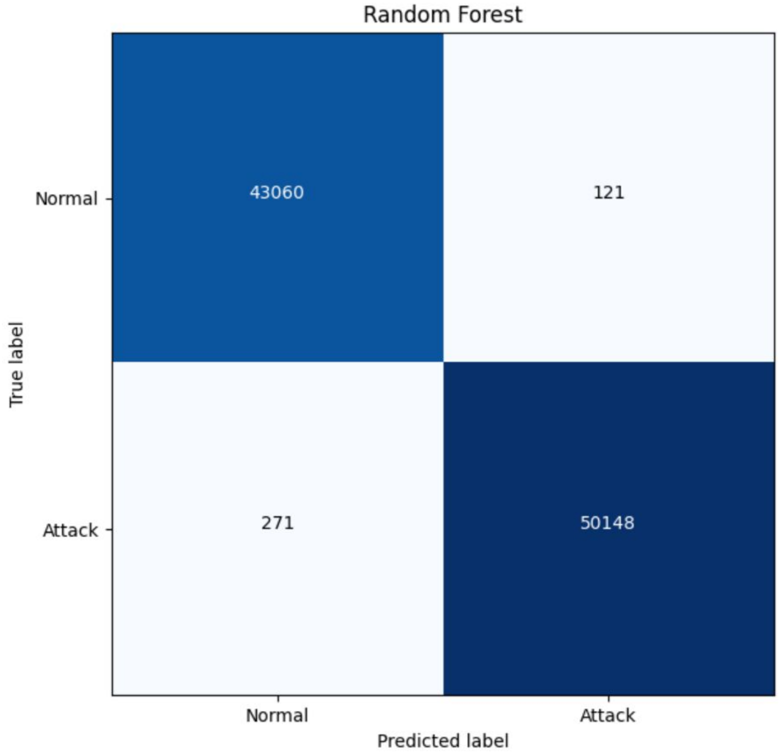
DECISION TREE

- Decision Tree Classifier builds a tree by checking a certain if-else condition at each level of the tree.
- It then makes a decision to either traverse left or right depending if the condition is true or not.
- For our classification, we have seen that decision tree has given a very good accuracy.



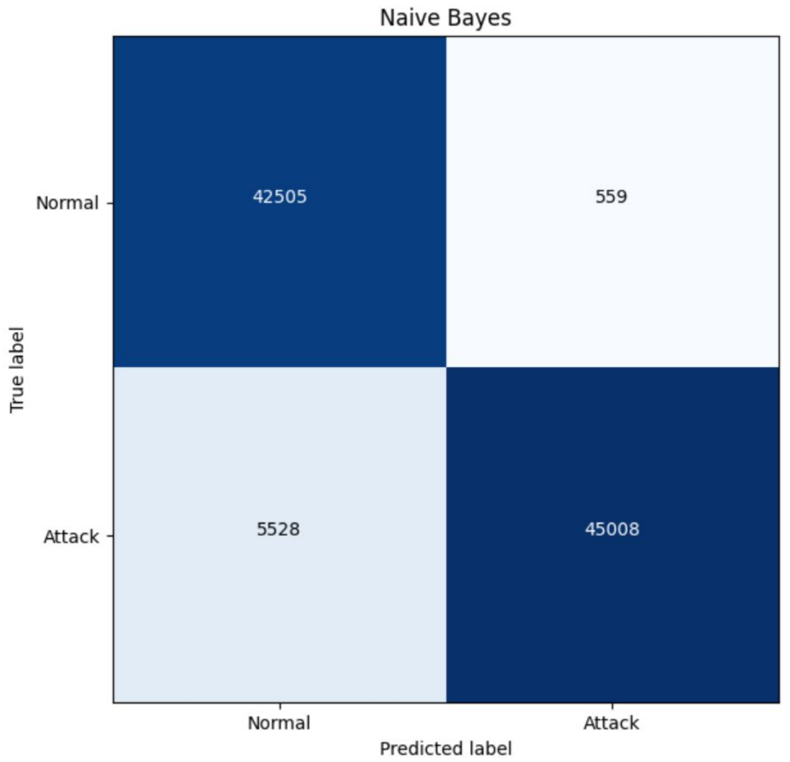
RANDOM FOREST

- A random forest classifier is an ensemble of decision trees. It uses multiple decision trees and chooses the classification given by most decision trees.
- In our case, we have seen that random forest classifier has given an even better accuracy than decision tree classifier.



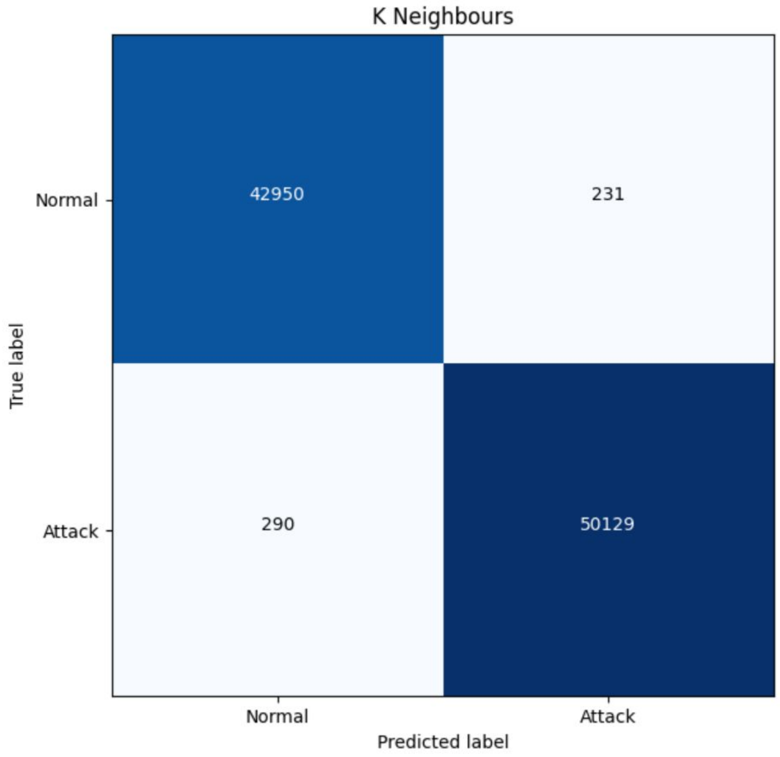
NAIVE BAYES

- Naive Bayes classifier uses the principle of Bayes algorithm to classify the data based on their probabilities.
- We have seen that the results of naive bayes is relatively poor compared to other algorithms. The reason for this might be because it assumes that the features are independent of each other.



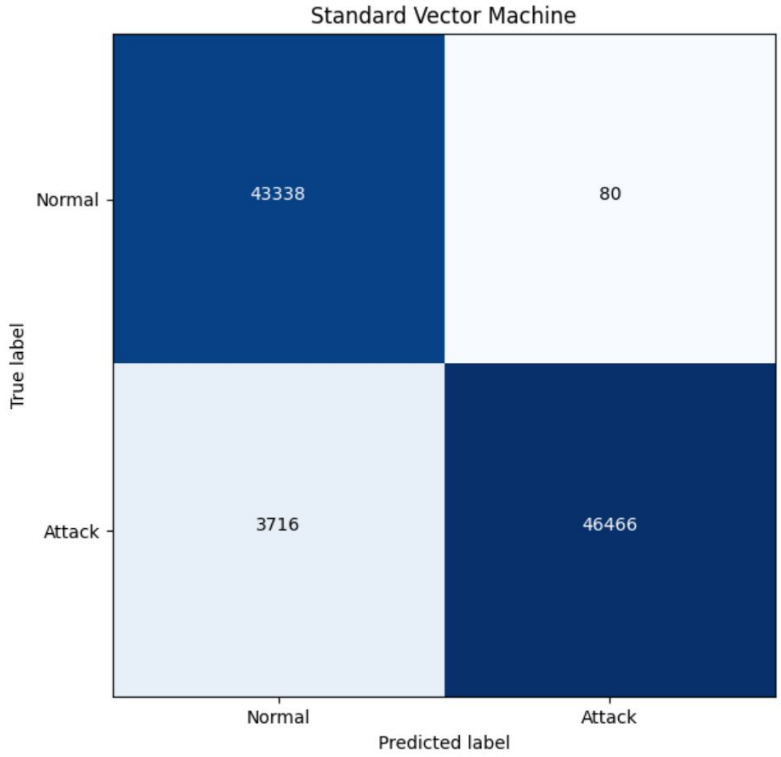
K NEAREST NEIGHBOURS

- K Nearest Neighbours uses the proximity of different points to a data to classify it.
- It looks at k neighbours closest to a data and chooses the most prominent class.
- Our result shows that KNN has also given good accuracy but it took required a much longer running time.

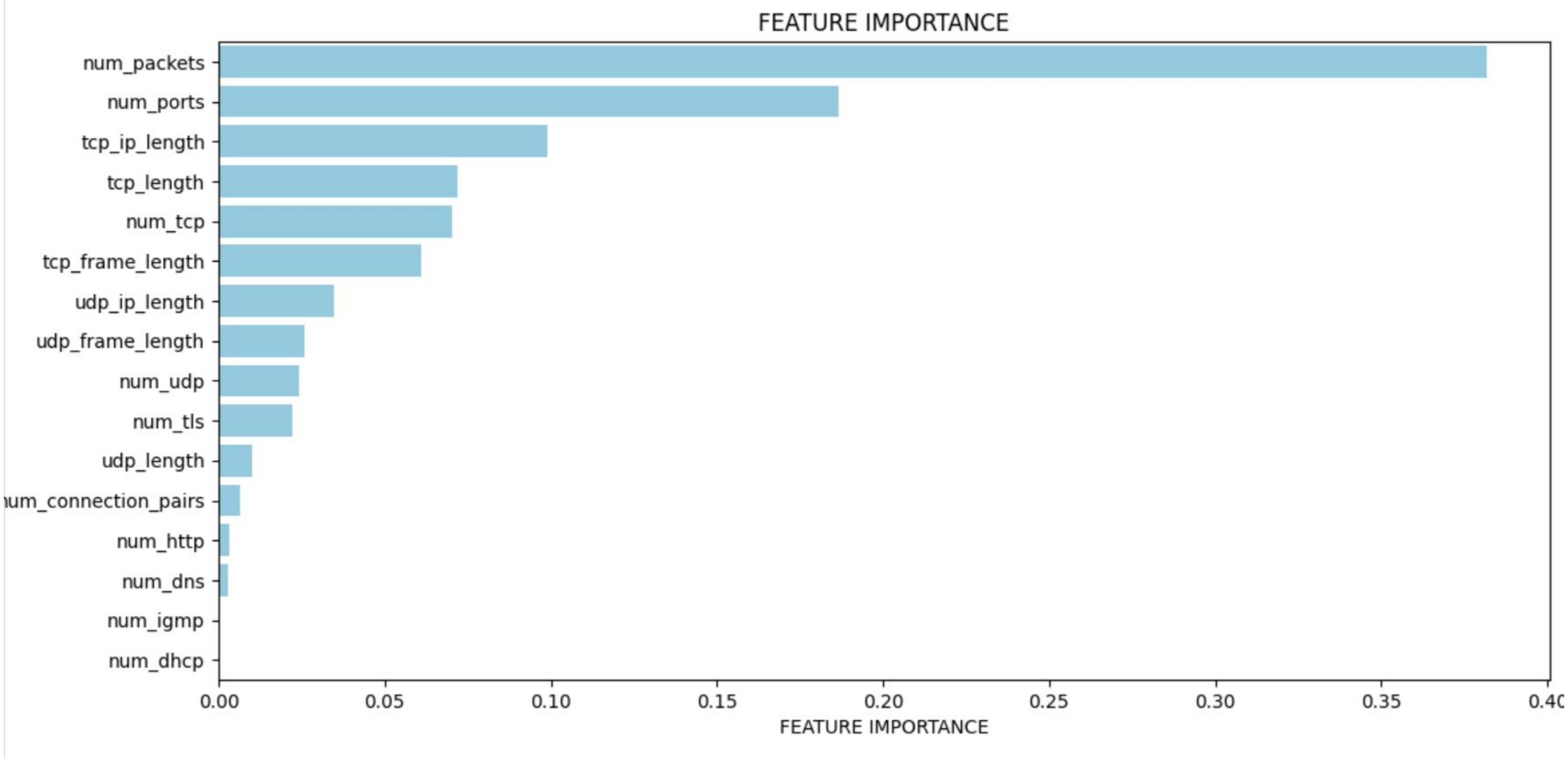


STANDARD VECTOR MACHINE

- Standard Vector Machine plots each data item in an n-dimensional space where n=no of features.
- It then creates a decision boundary and classifies the data according to their side of the boundary.
- In our case, we have seen that SVM has the lowest False Positive compared to other algorithms.



FEATURE IMPORTANCE



CONCLUSION

- In this project, we have captured the packets flowing in a network and derived some useful statistics about the packets.
- We have also tried to simulate a smaller version of a DDoS attack in a lab environment. Then, we extracted some features from the packets by creating a sliding window.
- Finally, we have developed a machine learning model which can be used to detect DDoS traffic coming to a server with a good accuracy.
- From the results of the classification, we have found that “Random Forest” and “K Neighbours” are the best classifiers for this problem.

FUTURE SCOPE

- The classification model we have created using machine learning can be used in some real world server to test its ability to detect DDoS attack. It can also be used to find the correlation of different features in classifying the network traffic in order to further investigate the packets.
- In this project, we have used supervised learning to classify the packets. However, if the information about the kind of traffic are already not available, then unsupervised learning can also be used with the same features to classify the packets.

REFERENCES

[I] Jiahui Chen, Joe Breen, Jeff M. Phillips, Jacobus Van der Merwe, "Practical and configurable network traffic classification using probabilistic machine learning", *Cluster Computing*, 2021.

[II] T. P. Fowdur, B. N. Baulum, Y. Beeharry, "Performance analysis of network traffic capture tools and machine learning algorithms for the classification of applications states and anomalies", International Journal of Information Technology, 2020.

[III] Pranita Mane, Yash Parkar, Jaideep Patel, Viral Sanghavi, Amey Walanje, "Traffic Classification Using Machine Learning", SSRN Electronic Journal , 2019.

[IV] Ons Aouedi, Kandaraj Piamrat, Benoît Parrein, "Performance evaluation of feature selection and tree-based algorithms for traffic classification", Communications Workshops (ICC Workshops) 2021 IEEE International Conference on, pp. 1-6, 2021.

[V] Yoga Durgadevi Goli, R Ambika, "Network Traffic Classification Techniques-A Review", *Computational Techniques Electronics and Mechanical Systems (CTEMS) 2018 International Conference*