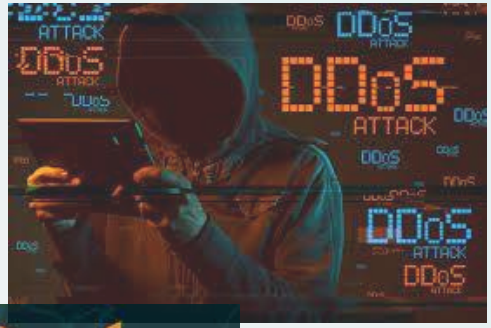
INTROUSION DETECTION



Bootcamp Data Science Project

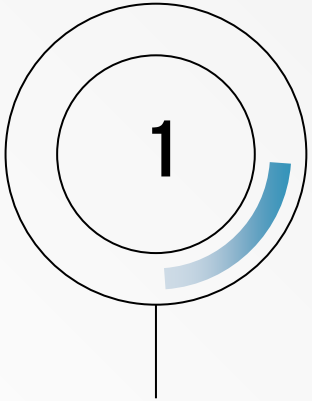
Raghad Abdullah Alrehaili

PRESENTATION CONTENT

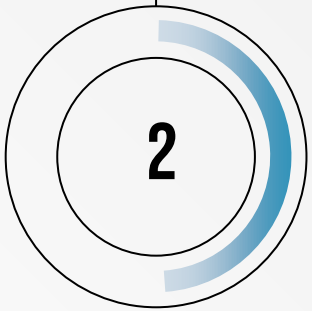
INTRODUCTION

The dataset is a large dataset contains 80 columns , and 1048575 rows, created in 2018 by New Brunswick's University, based on the logs of the University server that aims to analyzing DDoS data. And it is separated into different files dependent on the date. Each file is unbalanced, and I will deal with this problem.

DATA CLEANING

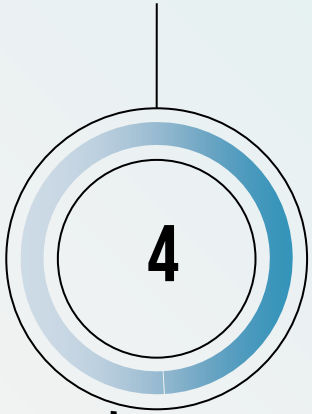


Drop the columns that contain null values



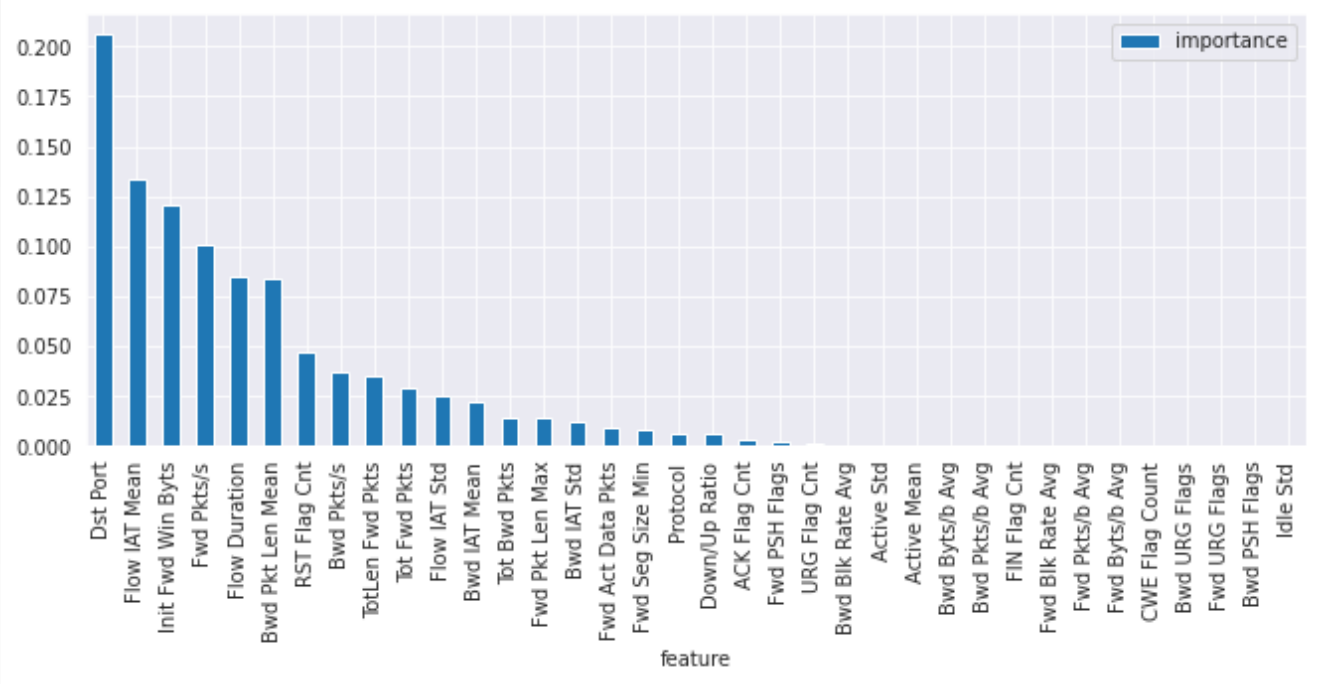
Check for duplicated values and drop them

Detect correlations between features and delete the columns that have correlation higher than 0.69



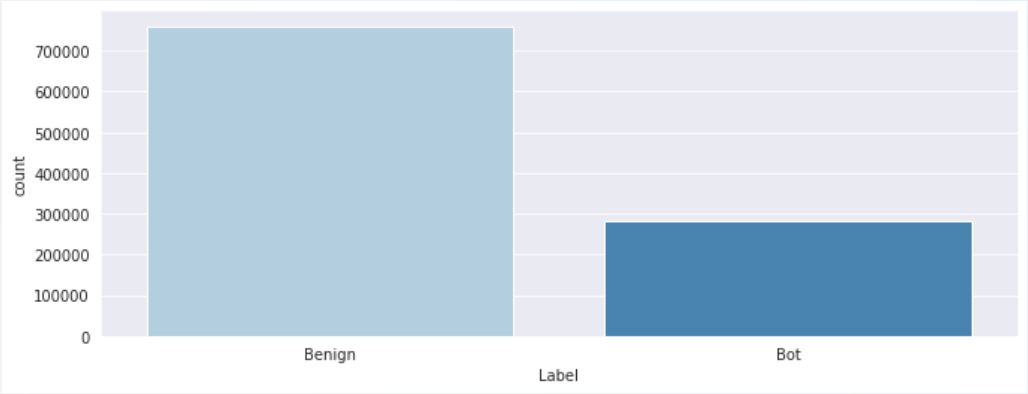
Select only most important features

DATA CLEANING

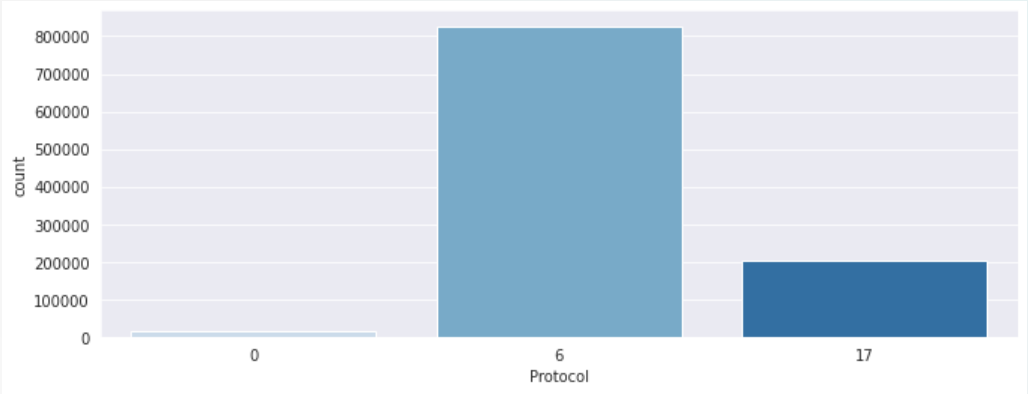


EDA

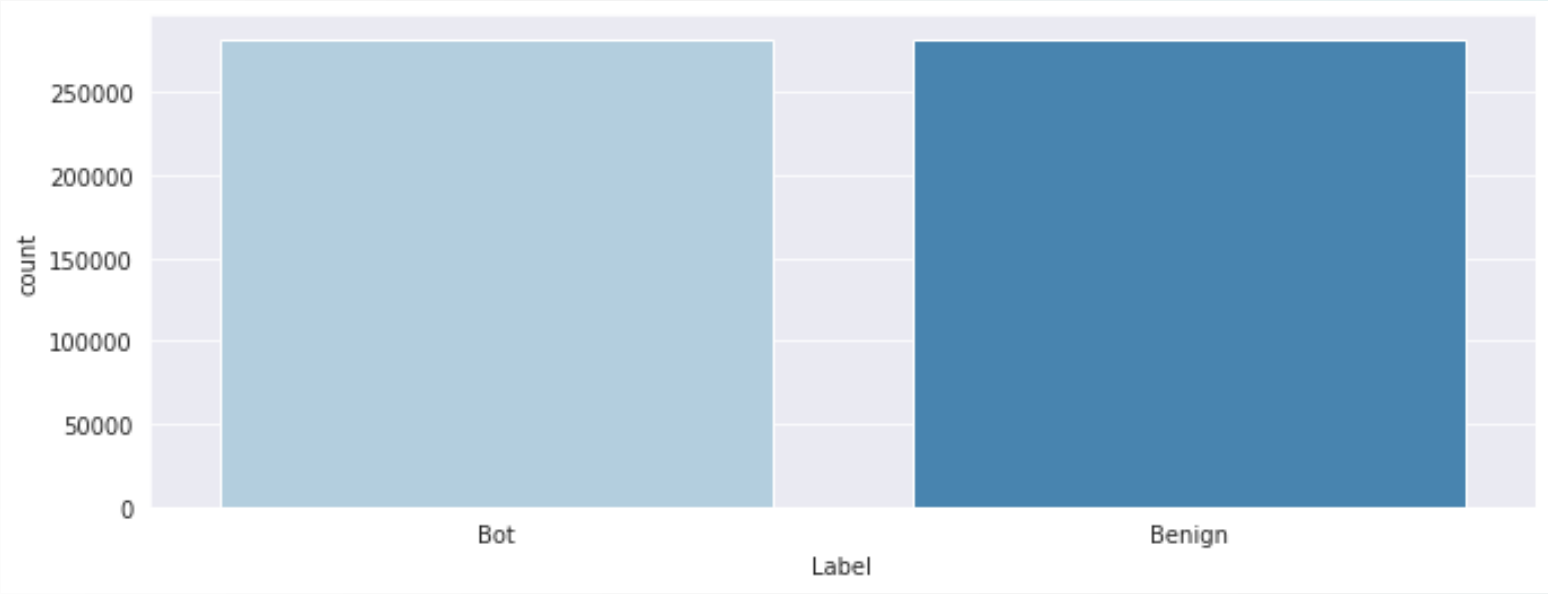
Q1: How many data samples represent a threat?



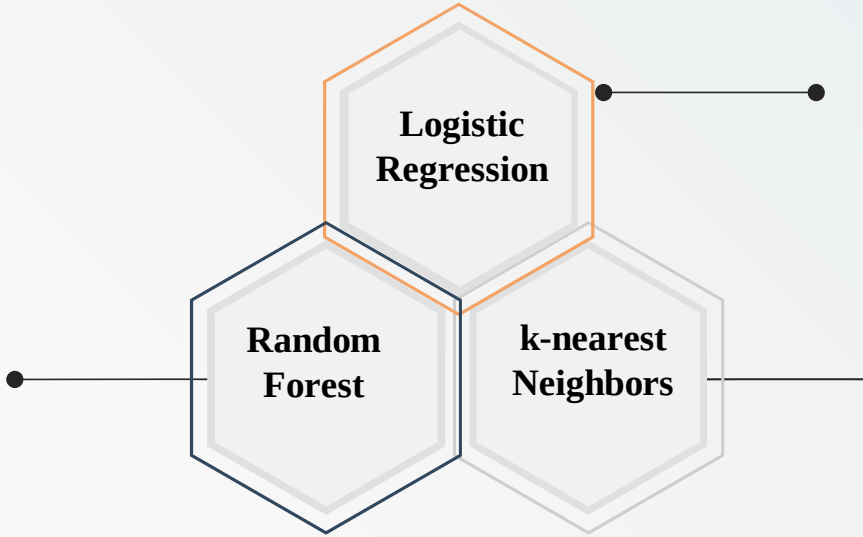
Q2: Which protocol most of malicious packets came from?



PREPROCESSING



MODELS

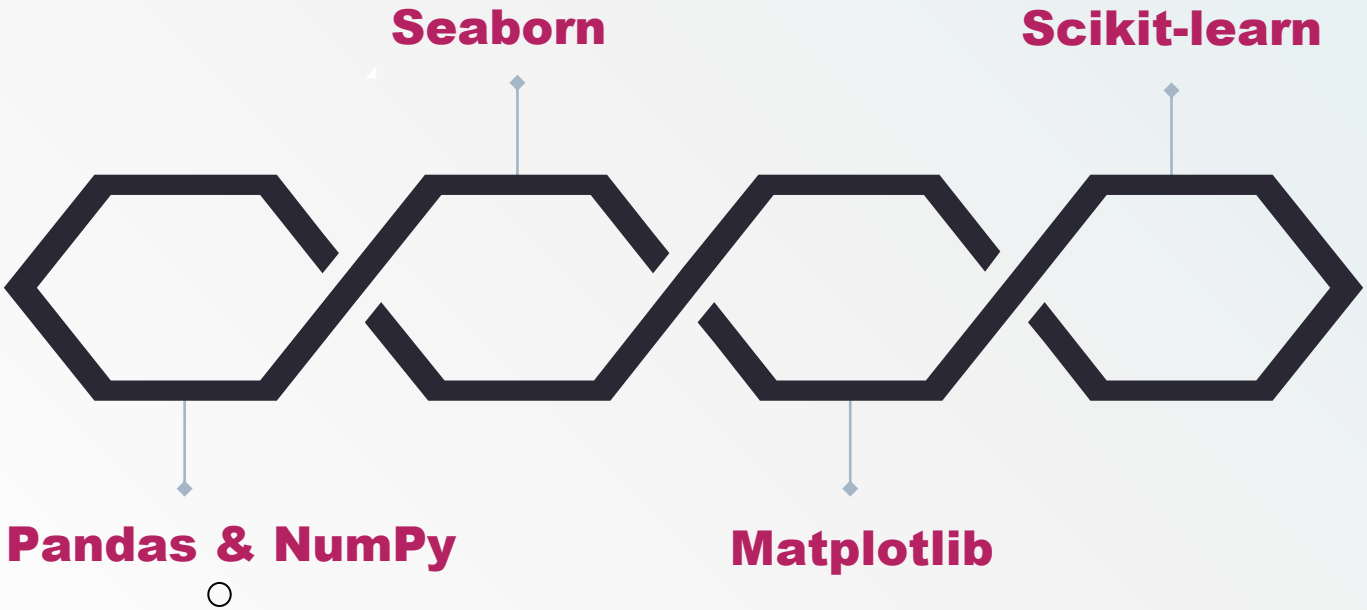


Accuracy: 97%
Precision: 100%
Recall: 94%
F1 score: 97%

Accuracy: 100%
Precision: 100%
Recall: 100%
F1 score: 100%

Accuracy: 100%
Precision: 100%
Recall: 100%
F1 score: 100%

Tools used



Conclusion

In conclusion, adapted classification models results satisfies the main goal of this project , in terms of high ability to learn the data pattern thus the precise prediction process.

In future, further steps to manipulate data will be taken , beside trying to use another classifier as ANN

(artificial neural network)

THANK YOU!

Do you have any questions?