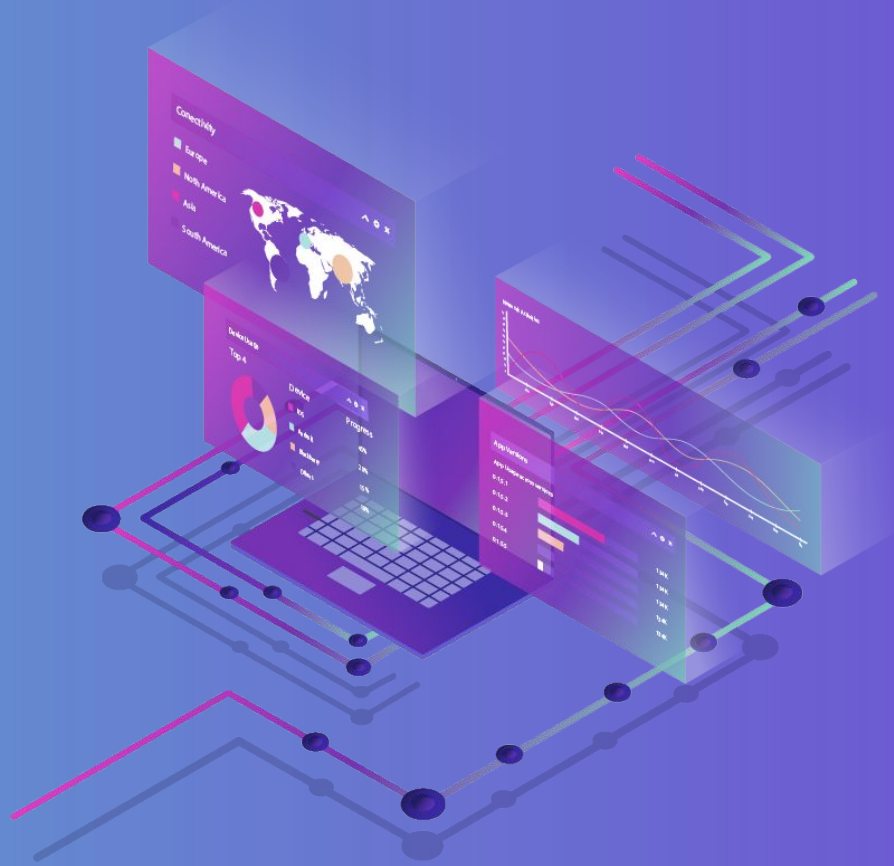
Network Anomaly Detection

Slick Blue Teamers



Team Members

Robinson Bill

Team Lead

Data Understanding, Modelling.

Mwangi Muna David

Member

Data Cleaning, EDA, Modelling.

Wanjiku Githu

Member

Data Cleaning, EDA, Modelling.

Annette Ngao

Member

Data Understanding, Modelling, CRISP DM Report.

Context

Anomaly-based network intrusion detection refers to finding exceptional or nonconforming patterns in network traffic data compared to normal behavior.With new types of attacks appearing continually, developing flexible and adaptive security oriented approaches is a severe challenge.

Market Relevance

- To maintain the principles of network security i.e, confidentiality, integrity and availability of our network and data.
- Protect client data and to protect computers from harmful spyware.
- Keep shared data secure from Industrial Espionage.

Objectives

Our goal is to create an anomaly detection model to detect a cyber attack based on the UNSW-NB 15 dataset. Our Specific objectives are:

- To understand the criteria for an anomaly.
- To give insights on the frequency & types of attacks.
- To provide recommendations for deployment.

Data Understanding

The UNSW-NB 15 dataset was collected by the Australian Centre for Cyber Security (ACCS). We merged 4 datasets, totalling to 2540043 records and 49 attributes with the following data types: 28 int64, 12 float64 & 9 object features.

Binary label: 0 for normal and 1 for attacks

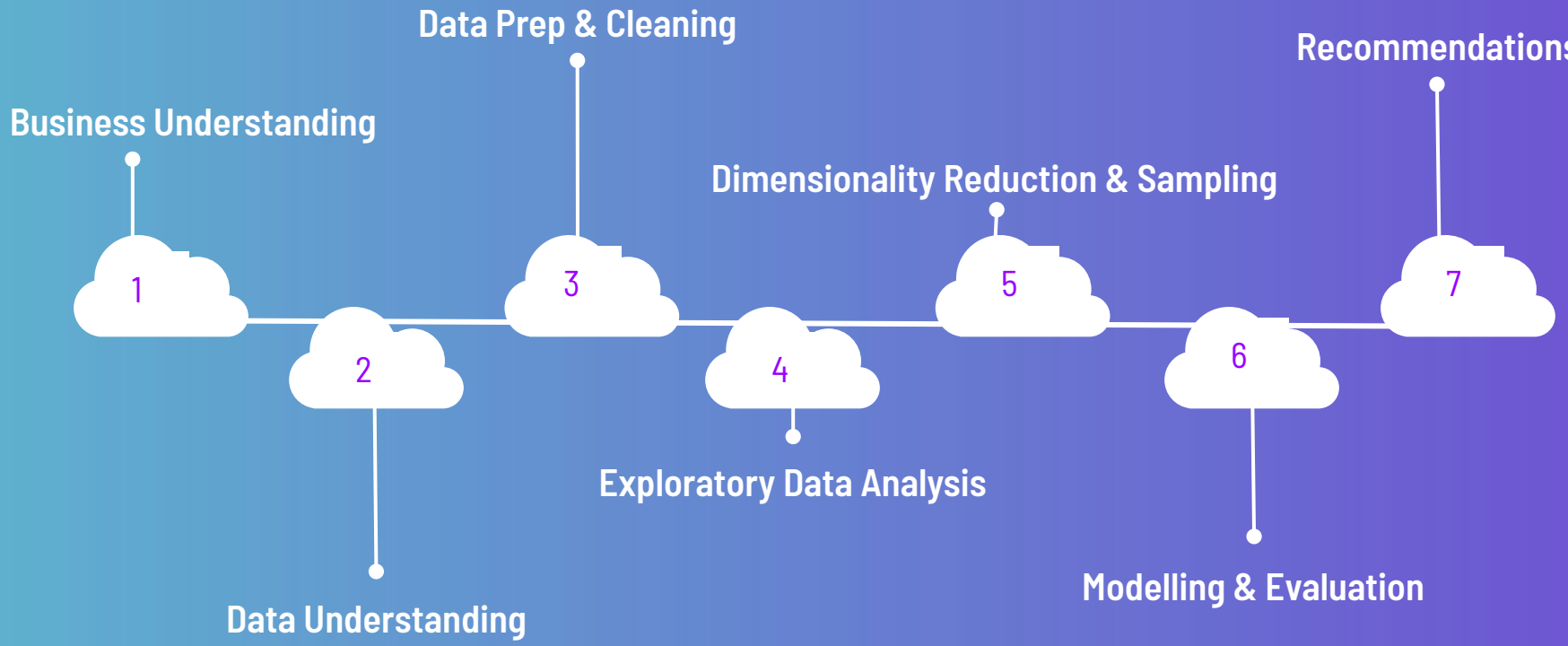
Terminologies Used

- **Generic Attack:** Type of password attack
- **Exploit** : Code/software that exploits flaws in OS and applications.
- **Fuzzers :** an automated process for finding errors in a program by feeding different data permutations into the program to find vulnerabilities

Terminologies Cont'd

- **DNS :** Domain Name System is the phonebook of the Internet
- **TCP/UDP:** Protocols for transmission of files
- **IPS/IDS:** Network intrusion and detection systems

Scope


Business Understanding
1
Data Understanding
2
Data Prep & Cleaning
3
Exploratory Data Analysis
4
Dimensionality Reduction & Sampling
5
Modelling & Evaluation
6
Recommendations
7

Tools


statsmodels


SciPy


NumPy


matplotlib


pandas


+ableau


scikit
learn

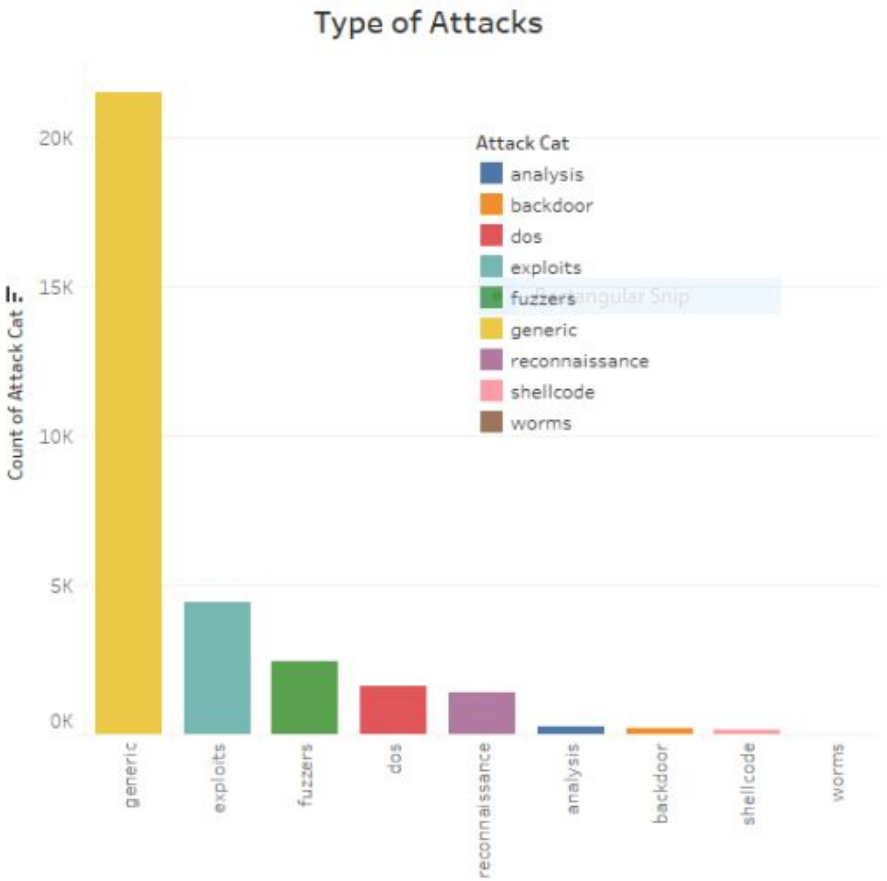

SMOTE

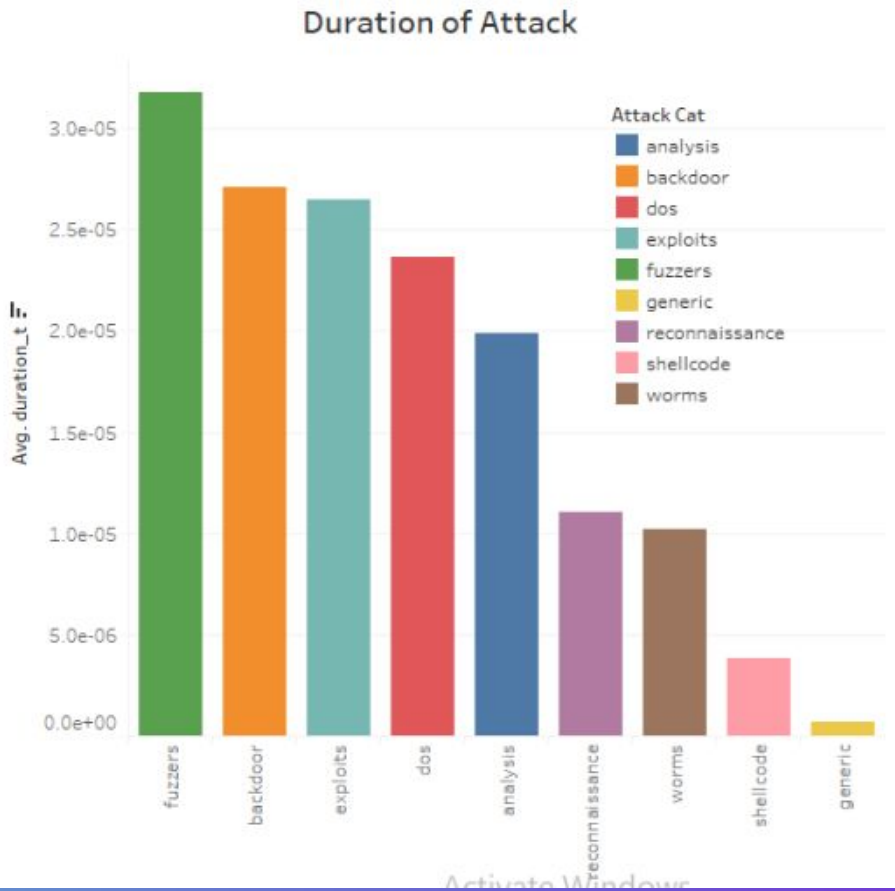

seaborn

Exploratory Data Analysis

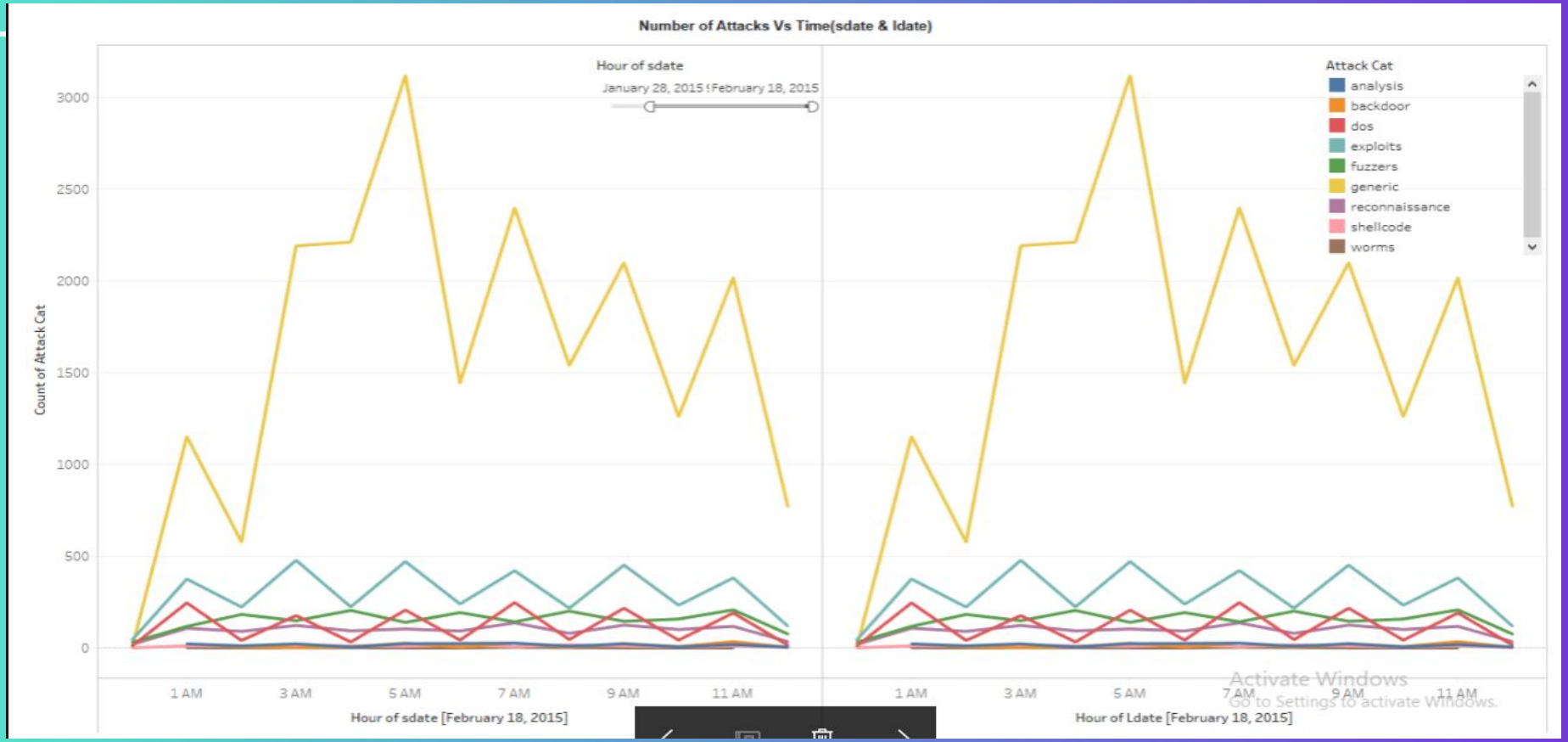
Tableau Visualisation Link

Attack Categories & Duration of Attack

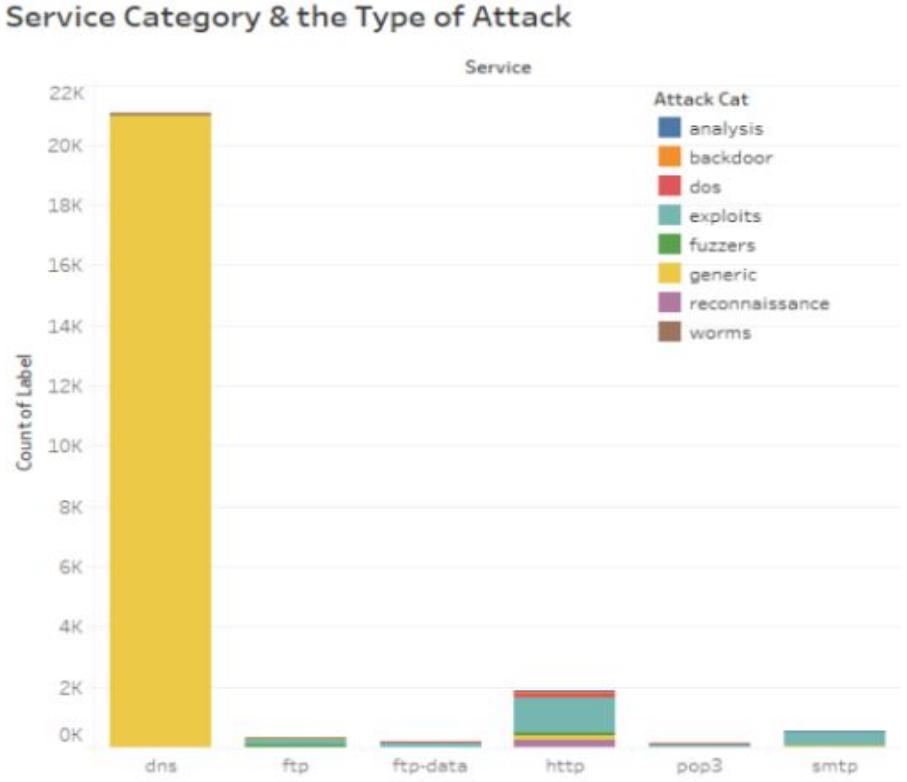


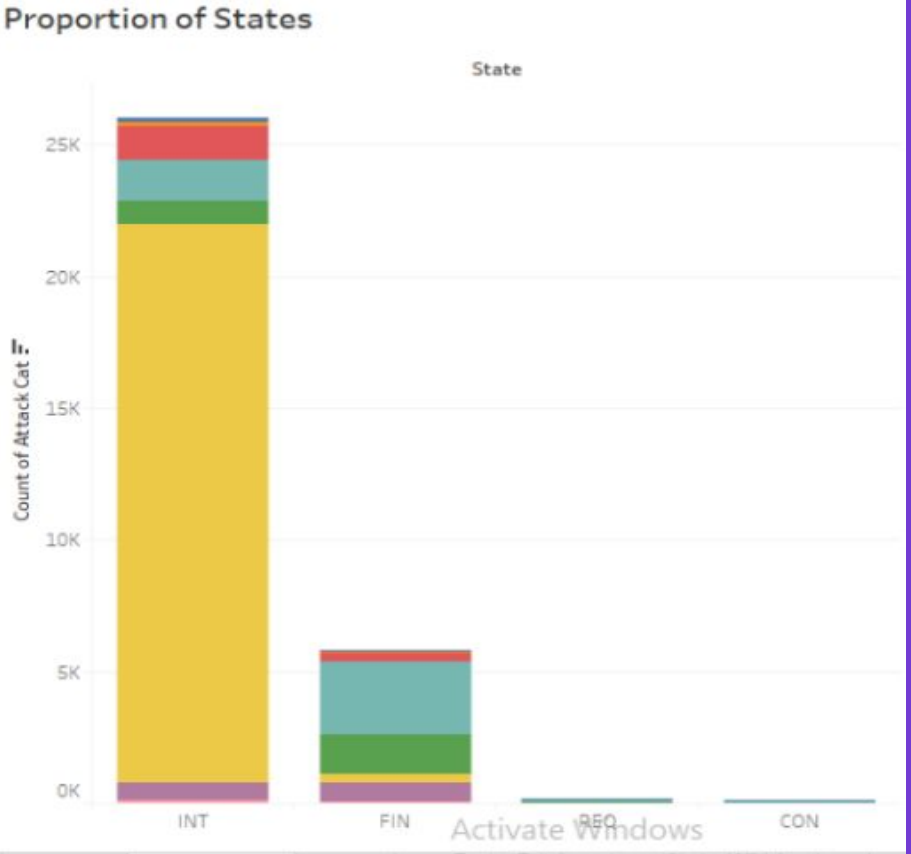


Number of Attacks Vs. Time

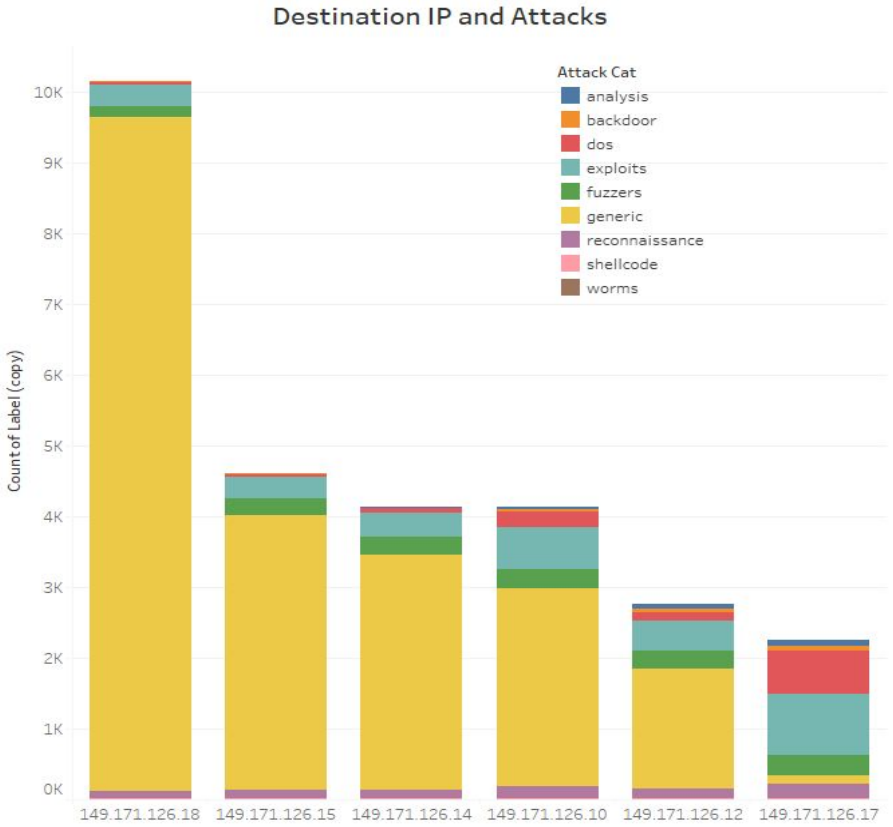


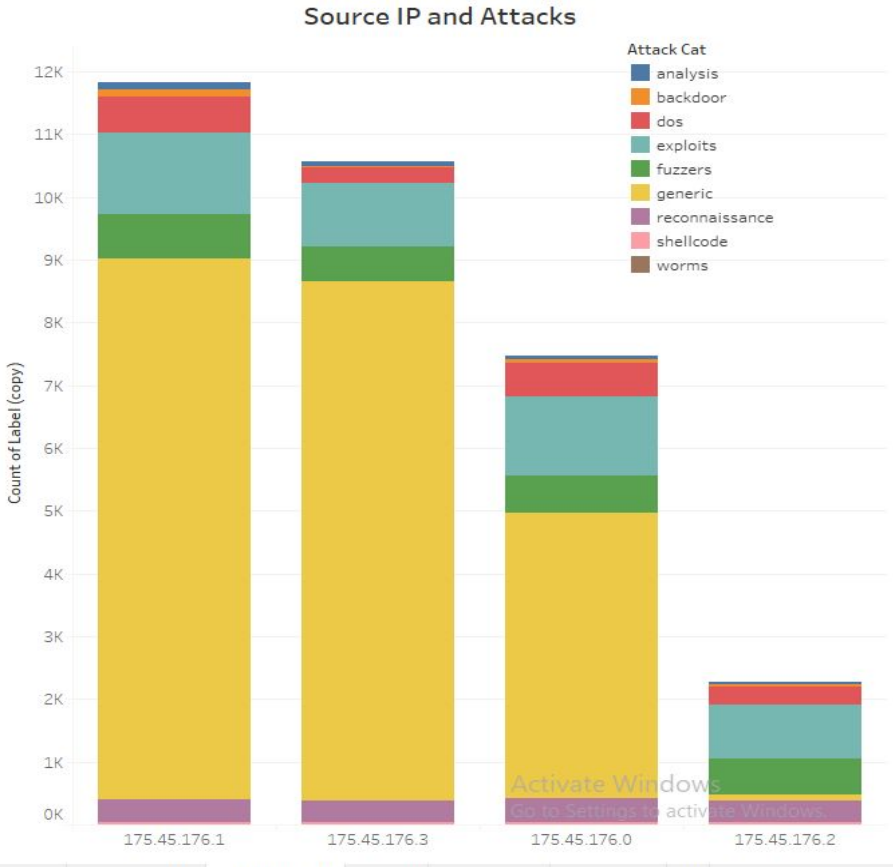
Service Category & Type of Attack





Attacks Vs Source/Destination IP Addresses

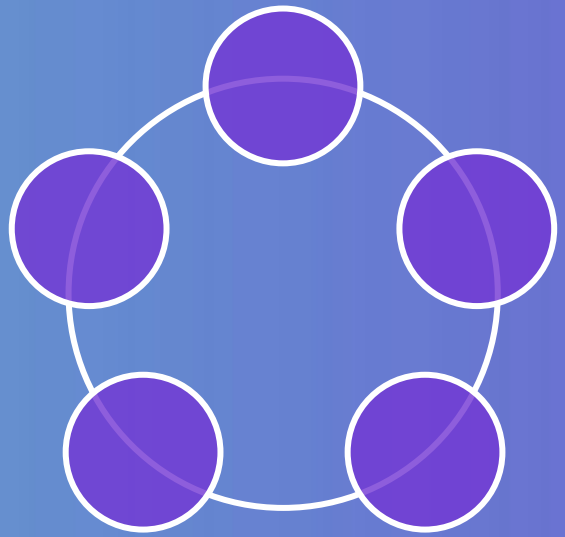




Statistical & Machine Learning Techniques Used

Filling missing values in binary features with mode

Supervised Learning Algorithms



Multicollinearity checks using VIF

Synthetic Minority Over-sampling Technique (SMOTE)

Stratified Sampling

Modelling

1

Logistic Regression.

F1 Score 0.9395

2

Naive Bayes

F1 Score 0.94

3

Random Forest

F1 Score 0.9587

4

Gradient Boosting

F1 Score 0.9881

Recommendations

- Generic,exploits and fuzzers attacks were the most so the business should come up with methods to mitigate these attacks.
- The network admin should secure the vulnerabilities in the DNS & flag IP Addresses 175.45.176.1 whose origin was from Potong-gang District in N. Korea
- Attacks mostly happen at peak hours, at 5 am. The network admin should vigilant during these hours or employ intelligent systems e.g IPS,IDS

Future Improvements

01 Sourcing for unseen data and test performance of our models

02 Hypothesis Testing on whether traffic originating from N.Korea are attacks.

03 Use Recurrent Neural Networks which works well with large datasets.

04 Deploying our model in control systems e.g.firewall & antivirus tools

Challenges

One

The ROC curves were not smooth because our label was binary.We tried to plot the probabilities to see if the sharp corners will turn into a curve but it didn't change

Two

Understanding the attributes specific to networking so as to gain more insight.

Three

Our data was imbalanced but it was corrected using SMOTE.

Muchos Gracias!

Any Questions?