ANOMALIES DETECTION IN NETWORK TRAFFIC USING MACHINE LEARNING

Adil Ergazin, Nurkhan Zaulanbay, CS-2115N

Outline

- Introduction
- Main part
- Methodology
- Conclusion

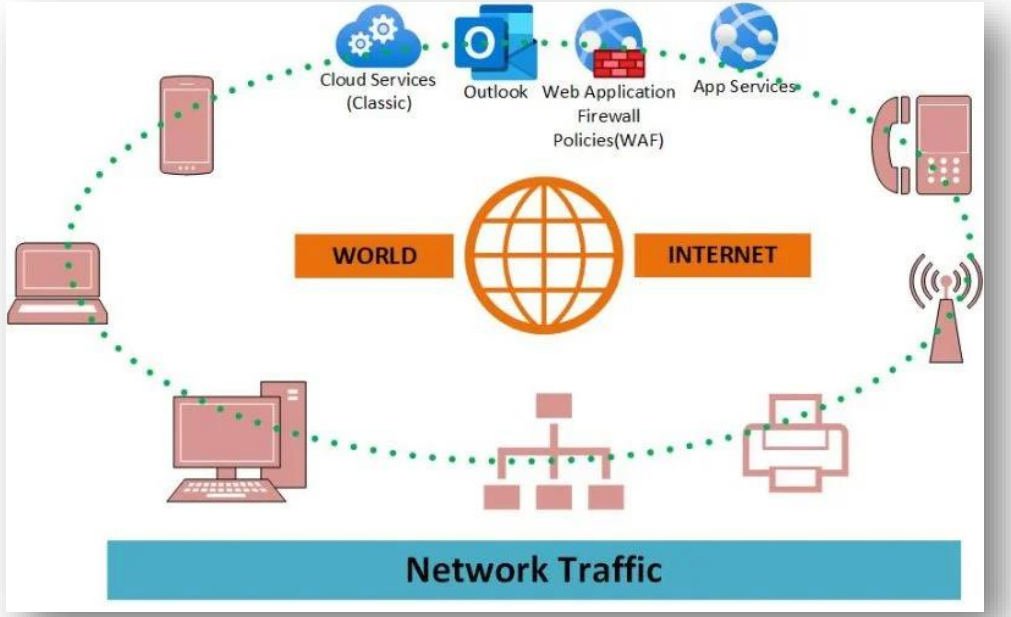
Introduction

Network traffic

- It refers to the movement of amount of data across a computer network in real-time.

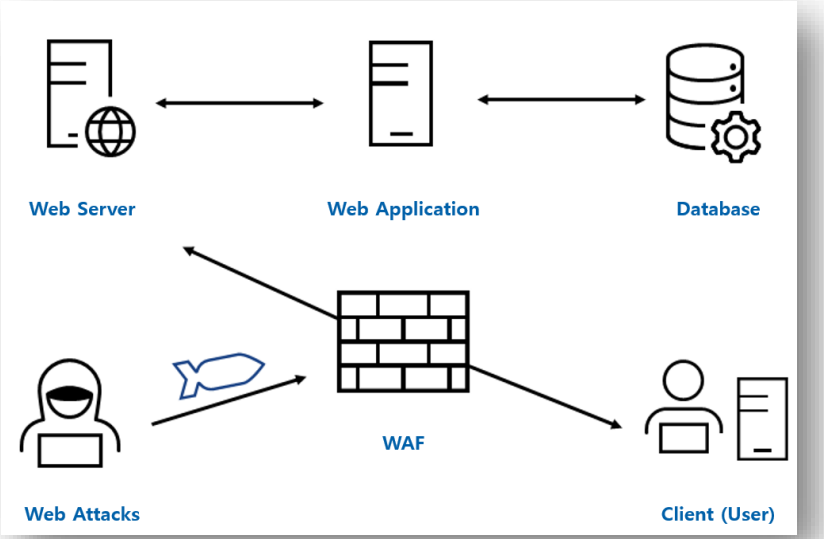
Anomalies in network traffic

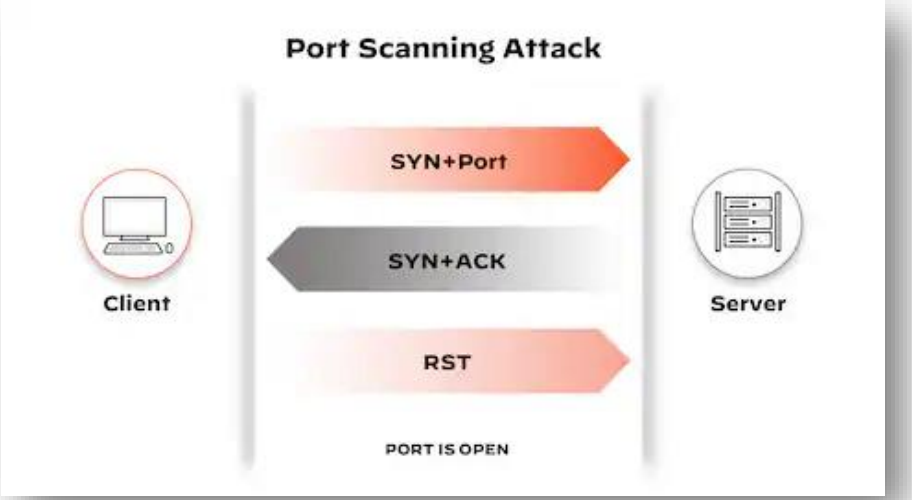
- They are abnormal or unexpected activities in network flow

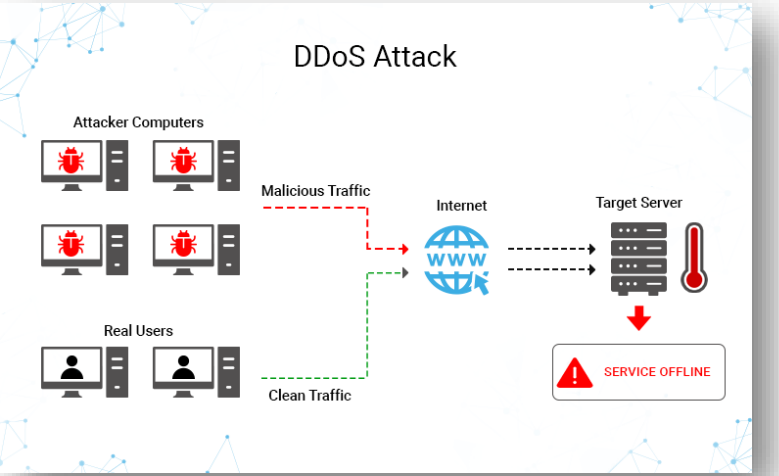


Introduction

Types of anomalies in network traffic:	DDoS - a cybercrime in which the attacker floods a server with internet traffic to prevent users from accessing connected online services and sites.
	Port Scan - a technique that enables threat actors to find server vulnerabilities. Ports enable devices to recognize different kinds of traffic: webpages, emails, instant messages, etc.
	Web attacks – unauthorized actions on the digital assets within an organizational network. Malicious parties usually execute network attacks to alter, destroy, or steal private data.

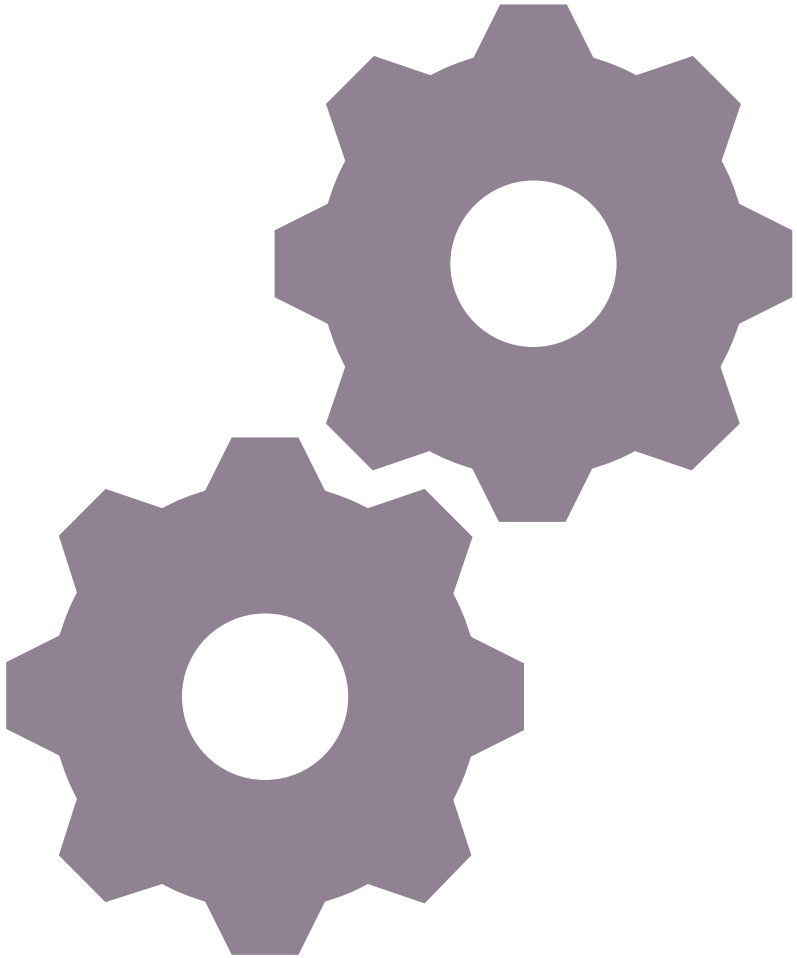

Web Server
Web Application
Database
WAF
Web Attacks
Client (User)


Port Scanning Attack
SYN+Port
SYN+ACK
RST
Client
Server
PORT IS OPEN


DDoS Attack
Attacker Computers
Malicious Traffic
Internet
Target Server
Real Users
Clean Traffic
SERVICE OFFLINE

Main part

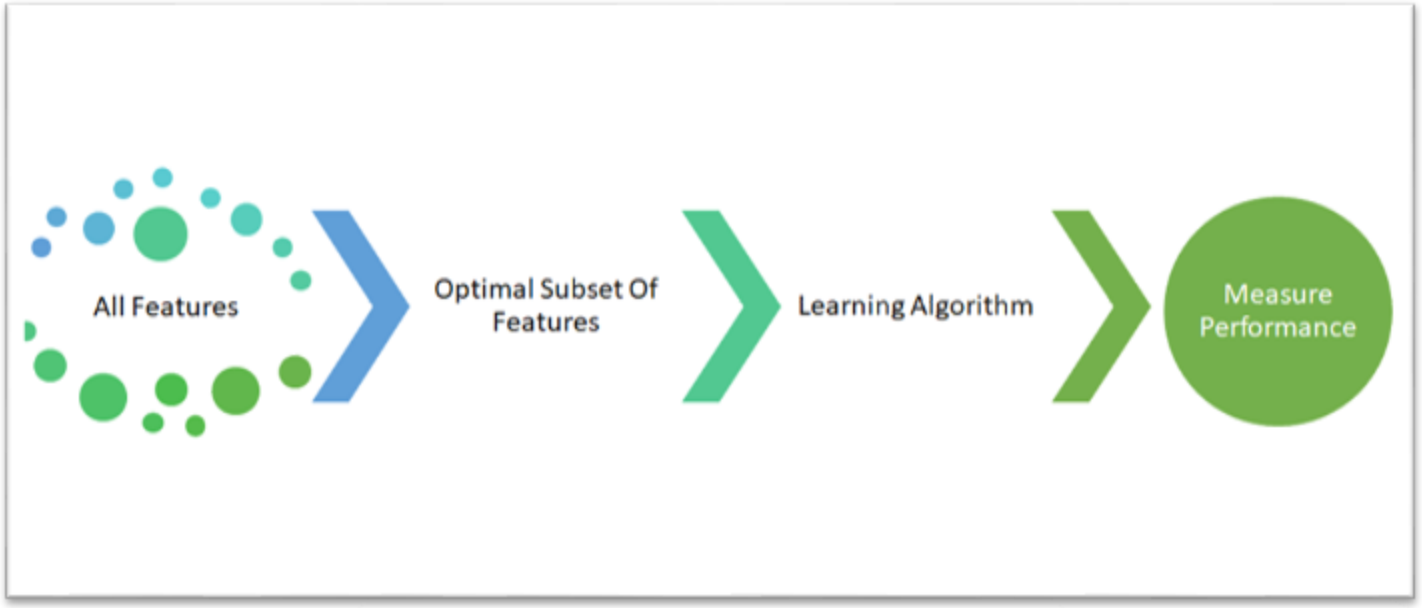
- Feature selection
- Anomalies
  detection algorithms and model

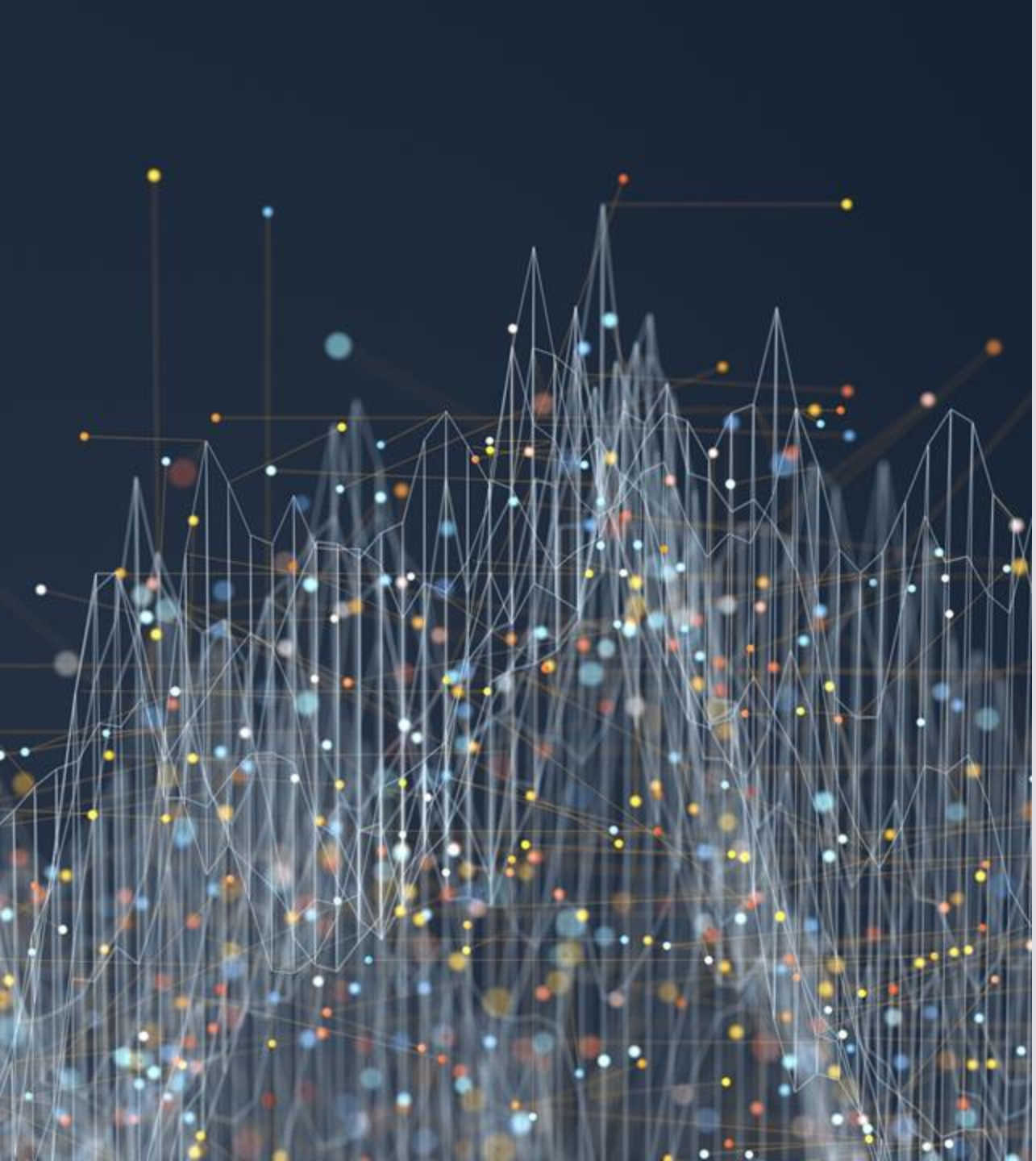


Main part

Feature selection

- Weighted importance
- 5 significant features
- Improve the prediction model



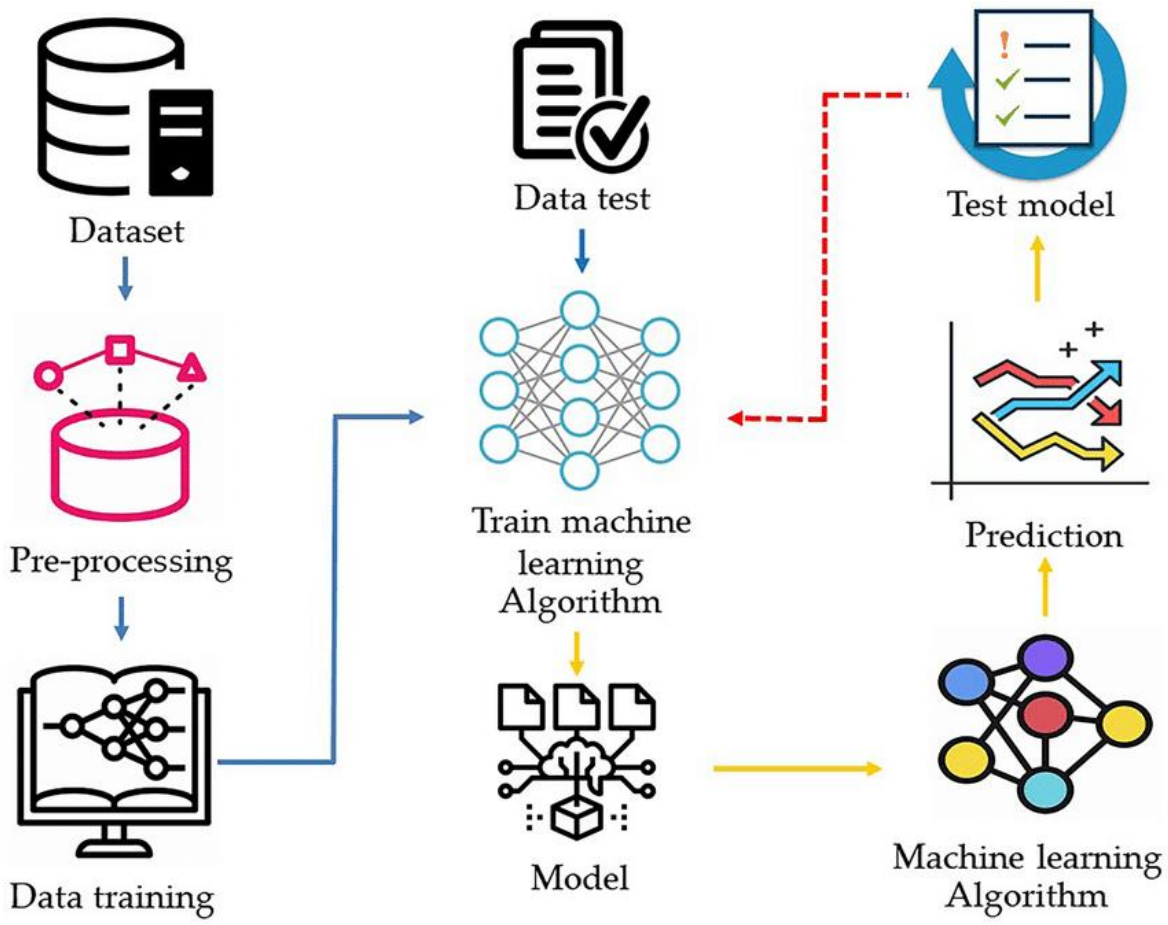


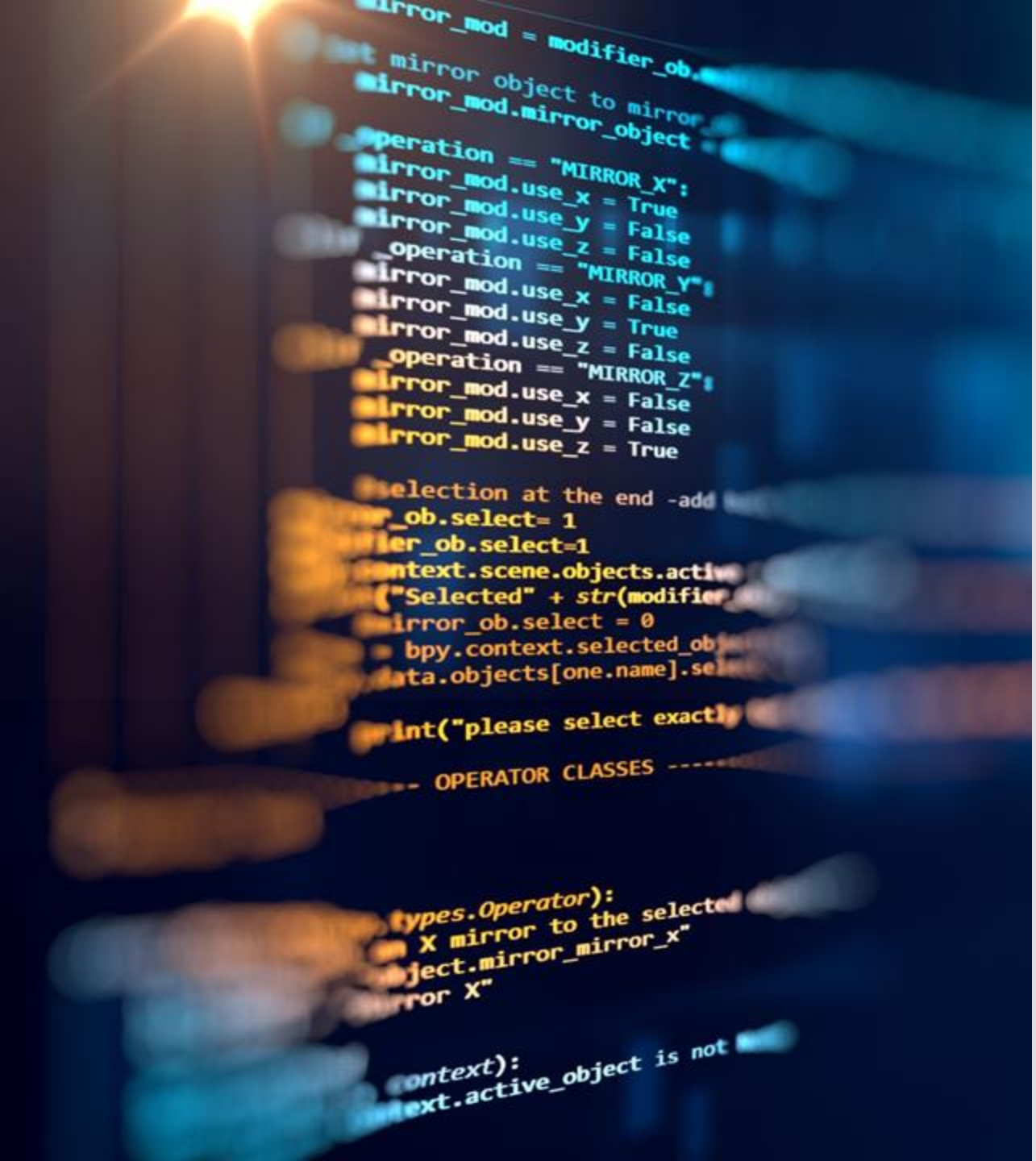
Main part

Anomalies detection algorithms and model

- Adaboost algorithm
- Random Forest
- ID3 (Iterative Dichotomiser 3)
- K-nearest neighbors
- Multilayer Perception

ANOMALIES DETECTION ALGORITHMS AND MODEL





Methodology

- Selecting the dataset
- Pre-processing
- Attack filtering
- Feature selection
- Implementation of machine learning

Methodology

Selecting the dataset

- CICIDS2017 dataset – dataset which includes the most common network traffic attacks and commends the real-world data Wireshark recorded pcap in CSV format files.

Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv

Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX.csv

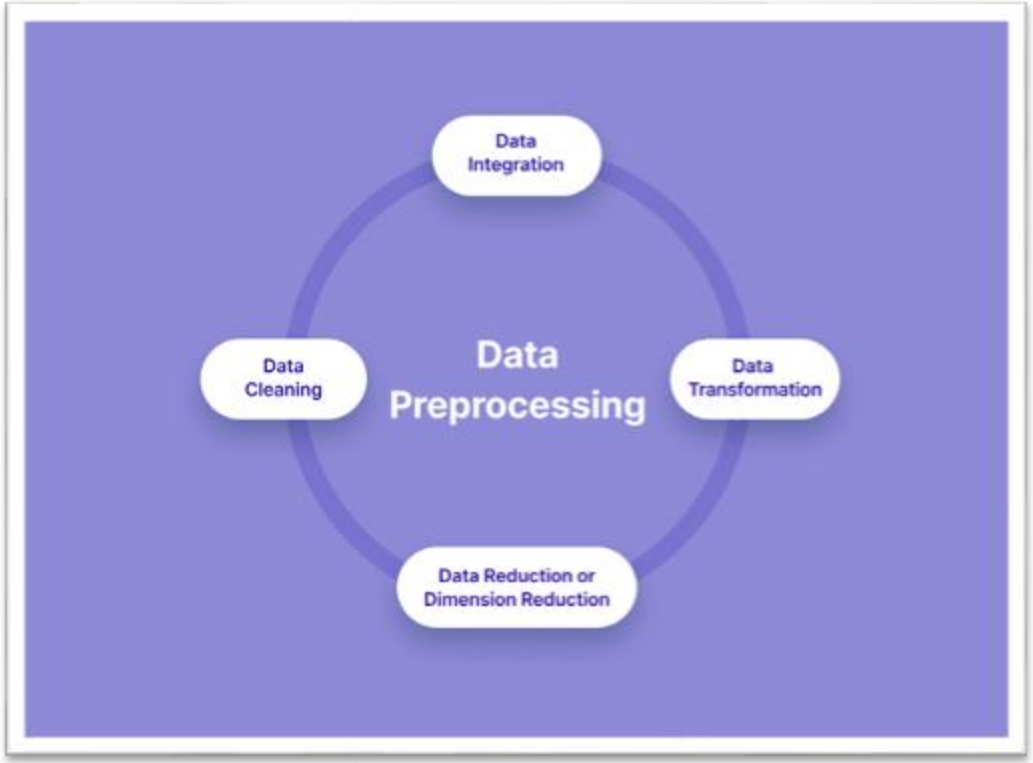
Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv

Flow ID	Source IP	Source Port	Destination IP	Destination Port	Protocol	Timestamp	Flow Duration	Total Fwd Packets	Total Backward Packets
192.168.10.16-199.244.48.55-41936-443-6	192.168.10.16	41936	199.244.48.55	443	6	7/7/2017 3:30	143347	47	60
192.168.10.16-54.210.195.63-42970-80-6	192.168.10.16	42970	54.210.195.63	80	6	7/7/2017 3:30	50905	1	1
192.168.10.16-199.244.48.55-41944-443-6	192.168.10.16	41944	199.244.48.55	443	6	7/7/2017 3:30	143899	46	58
192.168.10.3-192.168.10.17-53-12886-17	192.168.10.17	12886	192.168.10.3	53	17	7/7/2017 3:30	313	2	2
192.168.10.16-199.244.48.55-41942-443-6	192.168.10.16	41942	199.244.48.55	443	6	7/7/2017 3:30	142605	45	58
192.168.10.3-192.168.10.17-53-33063-17	192.168.10.17	33063	192.168.10.3	53	17	7/7/2017 3:30	253	2	2
192.168.10.16-199.244.48.55-41940-443-6	192.168.10.16	41940	199.244.48.55	443	6	7/7/2017 3:30	142499	46	53
192.168.10.16-199.244.48.55-41938-443-6	192.168.10.16	41938	199.244.48.55	443	6	7/7/2017 3:30	23828	27	31
192.168.10.16-199.244.48.55-41946-443-6	192.168.10.16	41946	199.244.48.55	443	6	7/7/2017 3:30	119090	23	28
192.168.10.25-17.253.14.125-123-123-17	192.168.10.25	123	17.253.14.125	123	17	7/7/2017 3:30	63021198	2	2
192.168.10.16-199.244.48.55-41946-443-6	192.168.10.16	41946	199.244.48.55	443	6	7/7/2017 3:30	23841	26	31
192.168.10.3-192.168.10.9-53-65431-17	192.168.10.9	65431	192.168.10.3	53	17	7/7/2017 3:30	227	2	2

Methodology

Pre-processing

- Data integration
- Data cleaning
- Data transformation
- Data reduction or dimension Reduction




```
The pre-processing phase of the  Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX  file is completed.
```

```
The pre-processing phase of the  Friday-WorkingHours-Afternoon-DDos.pcap_ISCX  file is completed.
```

```
The pre-processing phase of the  Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX  file is completed.
```

Methodology

Attack filtering

- Within each file are 30% attack and 70% benign registry

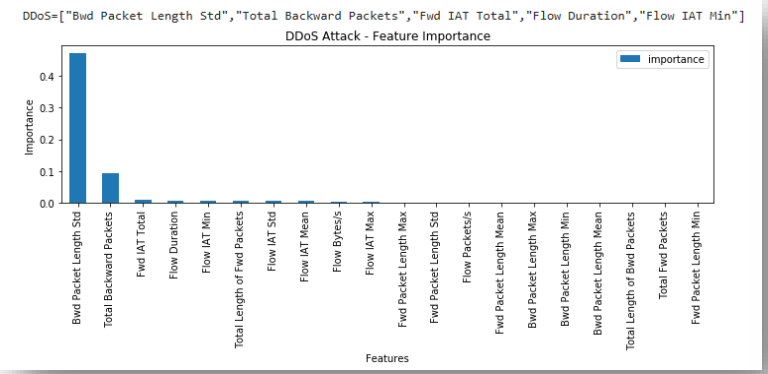
```
DDoS file is completed
 attack:41835
 benign:99398
```

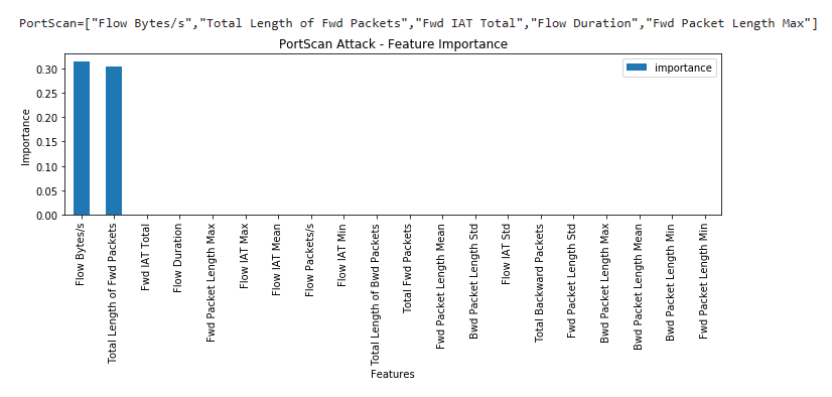
```
Web Attack
 attack:21
 benign:49
```

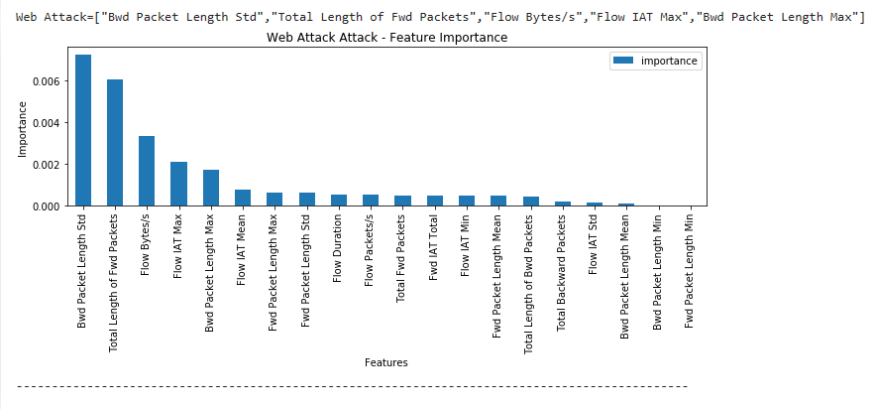
```
PortScan file is completed
 attack:158930
 benign:393748
```

Methodology

- Feature selection – 5 relevant and particular features were selected to each attack to compare them and chooses the best one



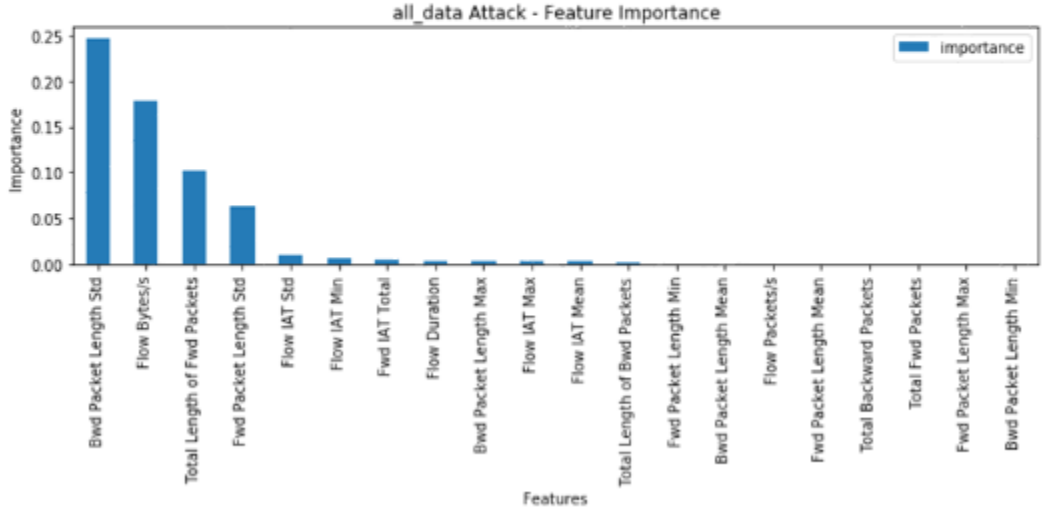




FEATURE SELECTION

```
all_data=["Bwd Packet Length Std","Flow Bytes/s","Total Length of Fwd Packets","Fwd Packet Length Std","Flow IAT Std"]
-----------------------------------------------------------------------------------------------------------------

mission accomplished!
Total operation time: =  25929.417772769928 secomds
```



Methodology

Implementation of machine learning

- Adaboost algorithm
- Random Forest
- ID3 (Iterative Dichotomiser 3)
- K-nearest neighbors

DDoS	Random Forest	0.96	0.96	0.96	0.96	0.4187
DDoS	ID3	0.96	0.97	0.96	0.96	0.1891
DDoS	AdaBoost	0.96	0.96	0.96	0.96	2.6724
DDoS	MLP	0.78	0.8	0.78	0.76	3.4024
DDoS	Nearest Neighbors	0.92	0.93	0.92	0.92	1.3234

Web Attack	Random Forest	0.97	0.97	0.97	0.97	0.0317
Web Attack	ID3	0.97	0.97	0.97	0.97	0.0097
Web Attack	AdaBoost	0.97	0.97	0.97	0.97	0.1716
Web Attack	MLP	0.64	0.63	0.64	0.6	0.1182
Web Attack	Nearest Neighbors	0.93	0.94	0.93	0.93	0.0174

Conclusion

In conclusion, the findings from our analysis and implementation of various methods to handle anomalies, Web attacks, DDoS attacks, and Port Scan attacks in network traffic have underscored the criticality of prioritizing safety measures. Throughout our investigation, we focused on three specific attack types, but it is essential to recognize that the landscape of cyber threats is continuously evolving, and we can anticipate encountering a broader array of attacks in the future.