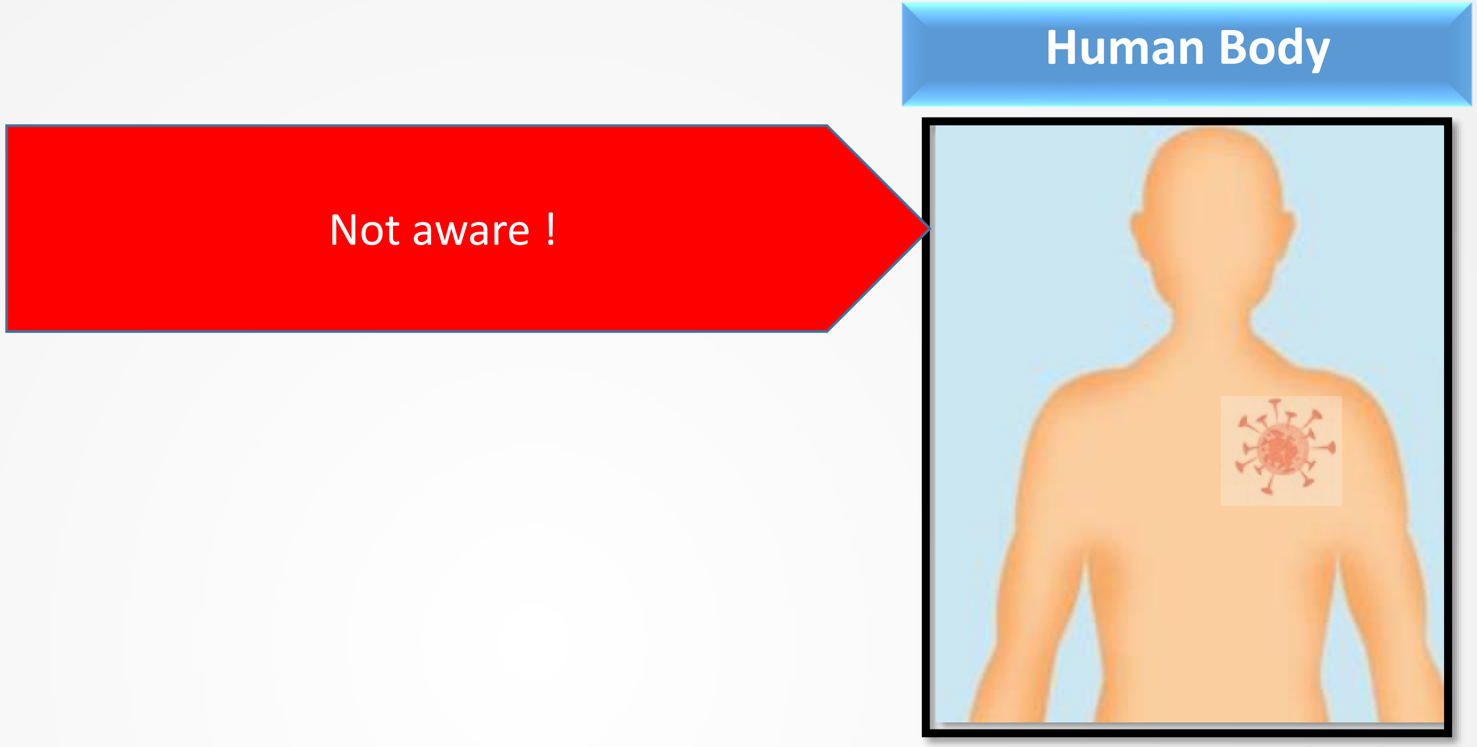
Project Title:

The Efficiency of Machine Learning Techniques in Minimizing Dwell Time of Cyberattacks

Meshal AL-Anazi

18 November 2021

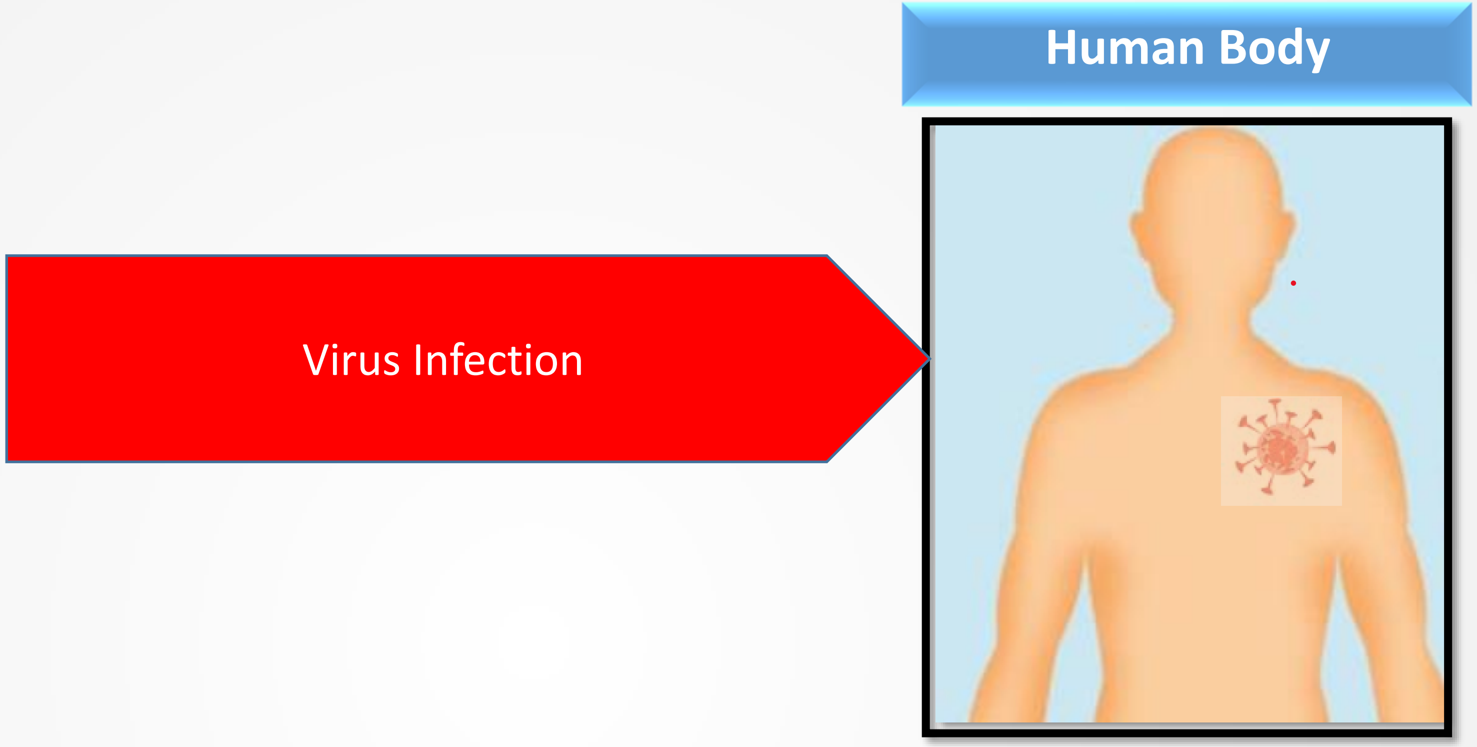
Motivation



Human Body

Not aware !

Motivation

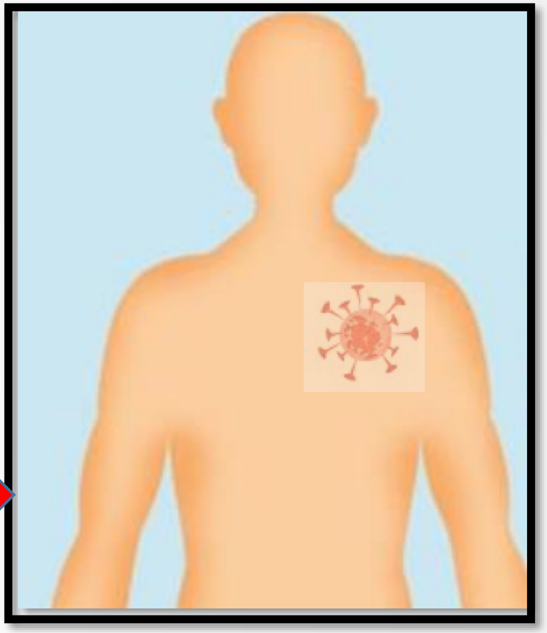


Human Body

Virus Infection

Motivation

Human Body

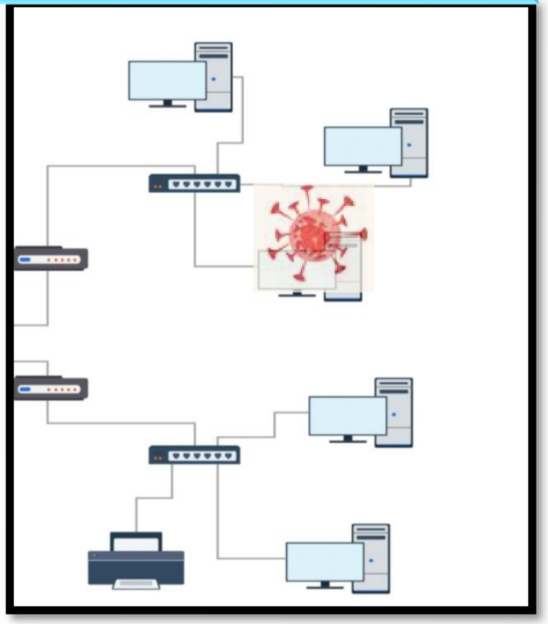


The virus can hide for long period before symptoms appear

Motivation

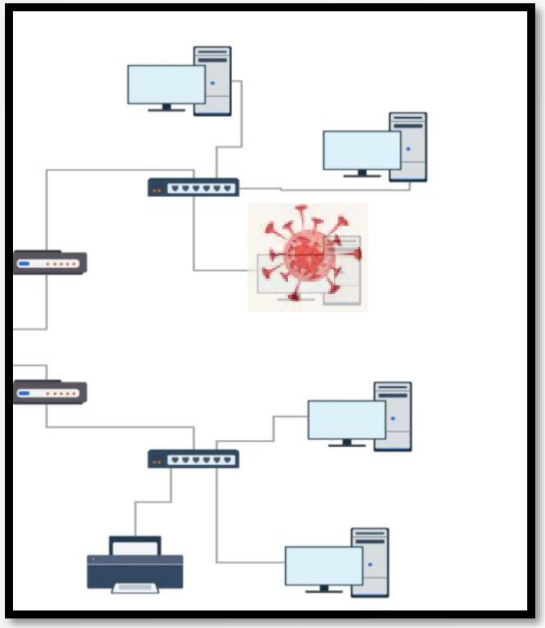
The cyber space has no exception

Computer Network



Motivation

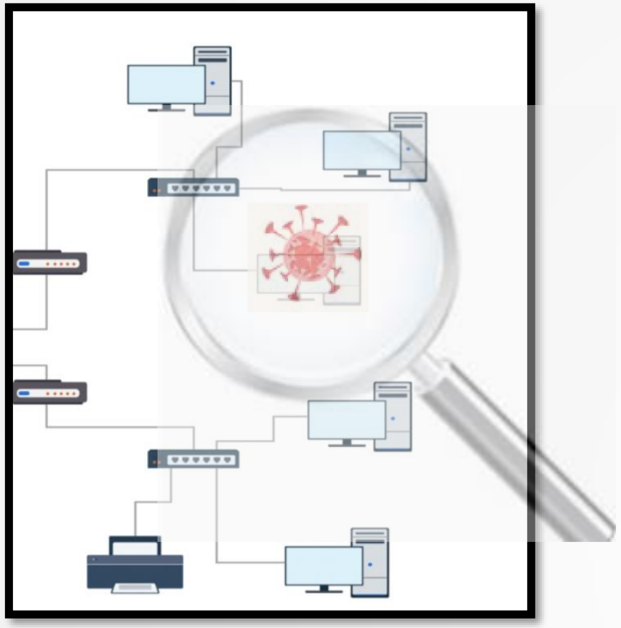
Computer Network



Virus Infection Start

Motivation

Computer Network

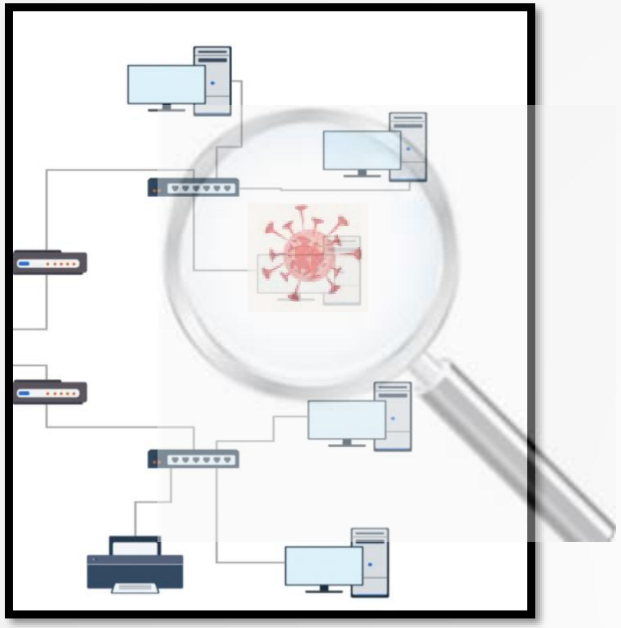


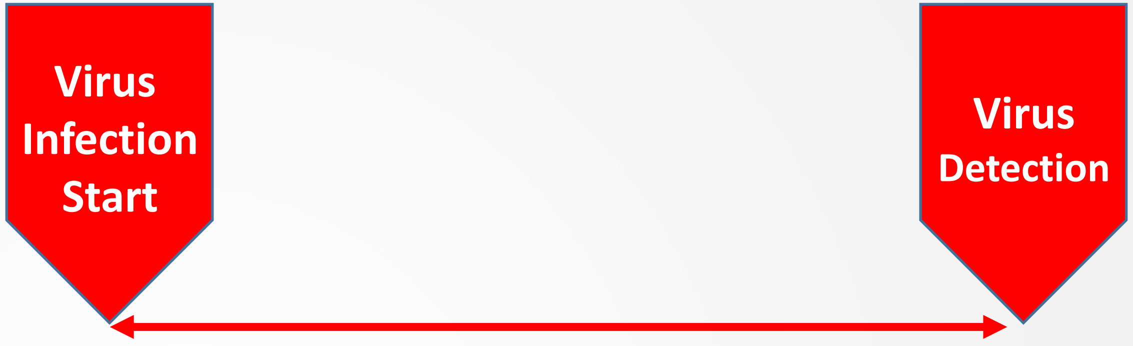
Virus Infection Start

Virus Detection

Motivation

Computer Network





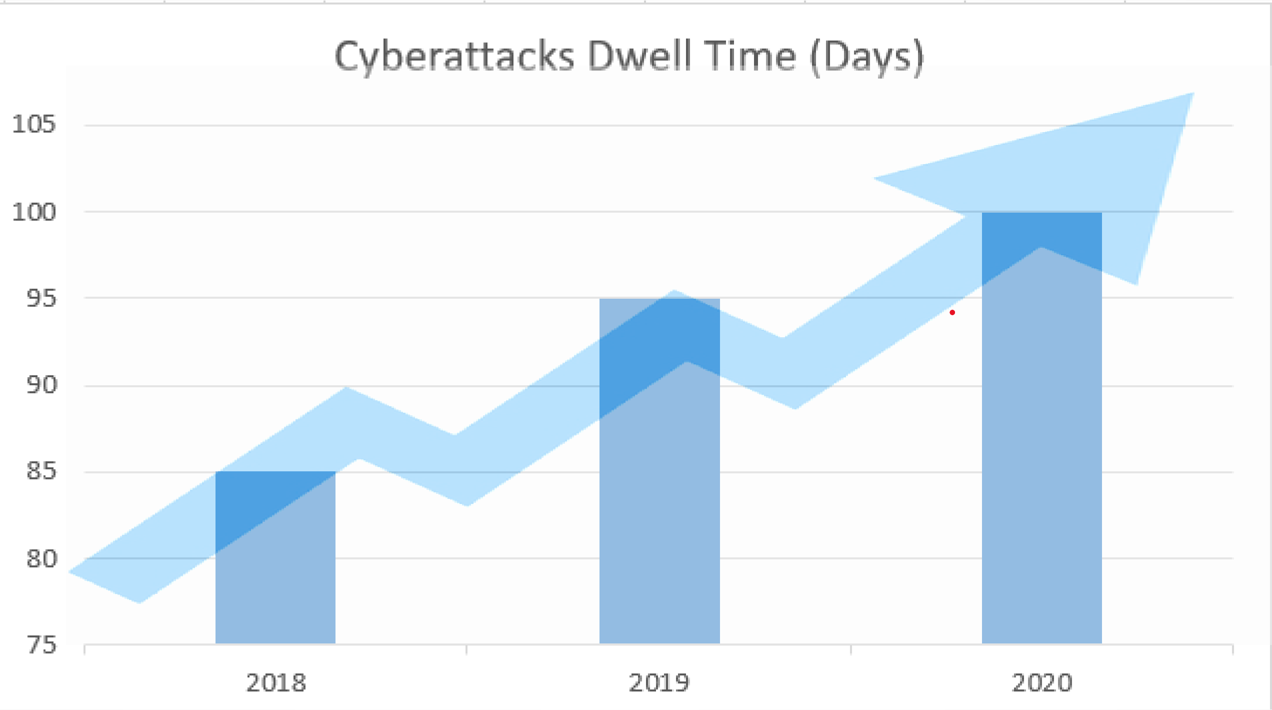
Virus Infection Start

Virus Detection

Dwell Time of Cyberattacks

Motivation

Recent researches show that the average cyberattacks Dwell time is **growing**



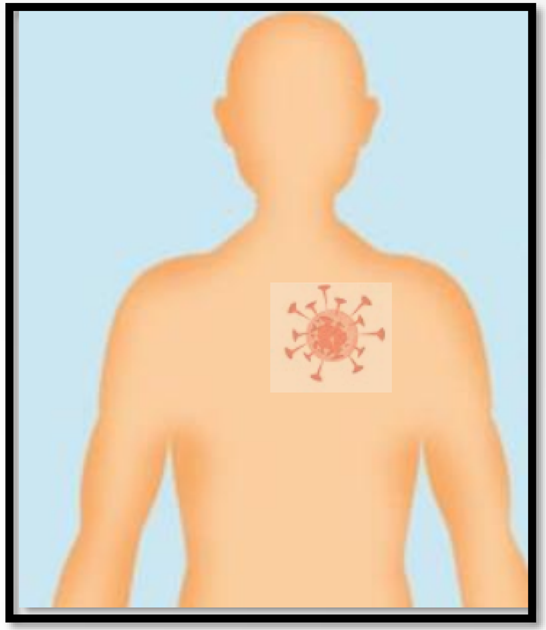
The Project Goal

To **detect** the Virus early → **Minimize** Cyberattacks Dwell Time

Datasets

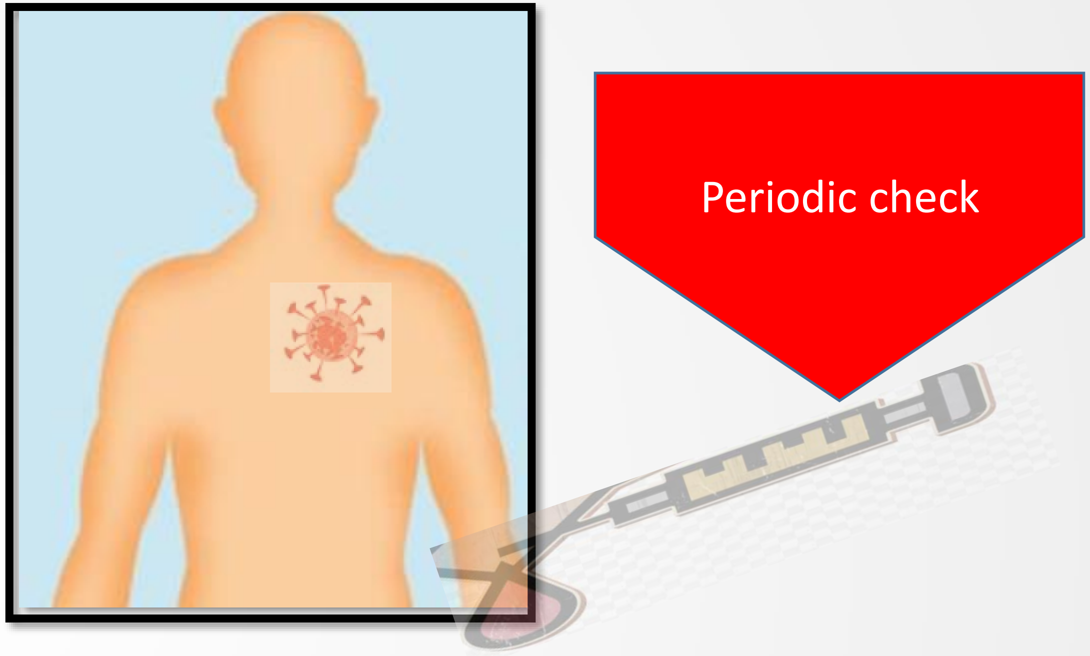
What is the recommendation to detect the virus early ?

Human Body



Datasets

Human Body

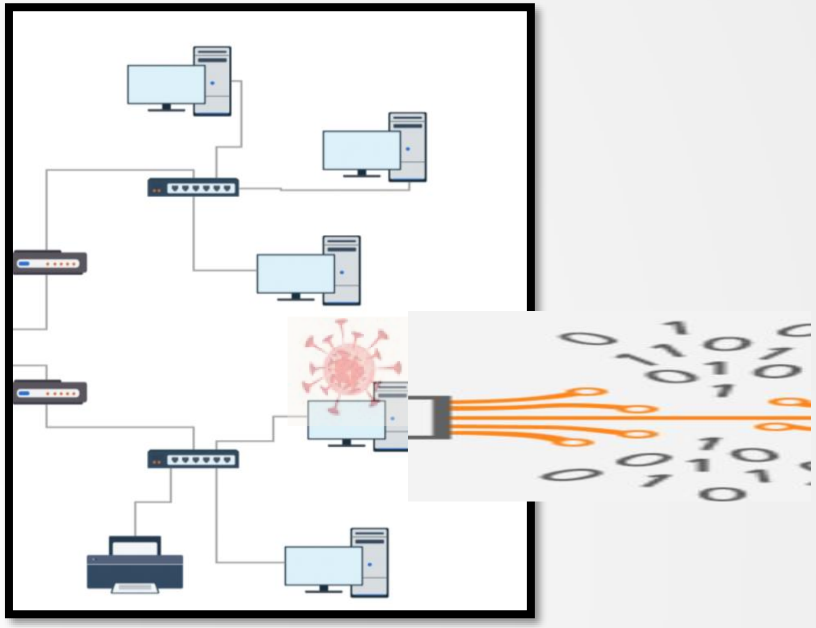


Periodic check

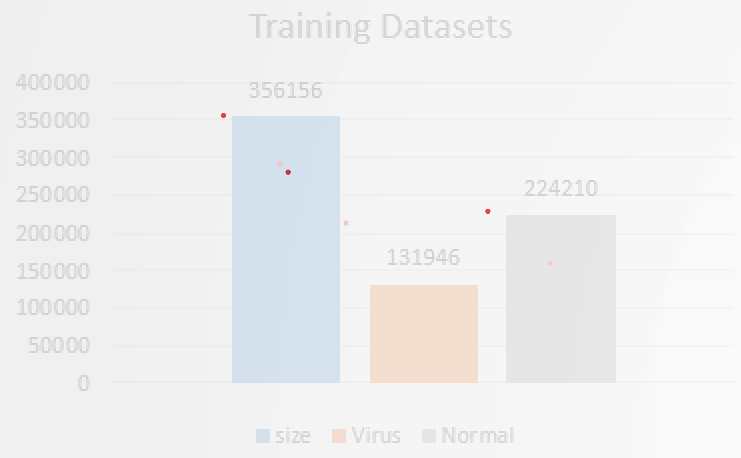
Datasets

The periodic check recommendation is valid in cyber space

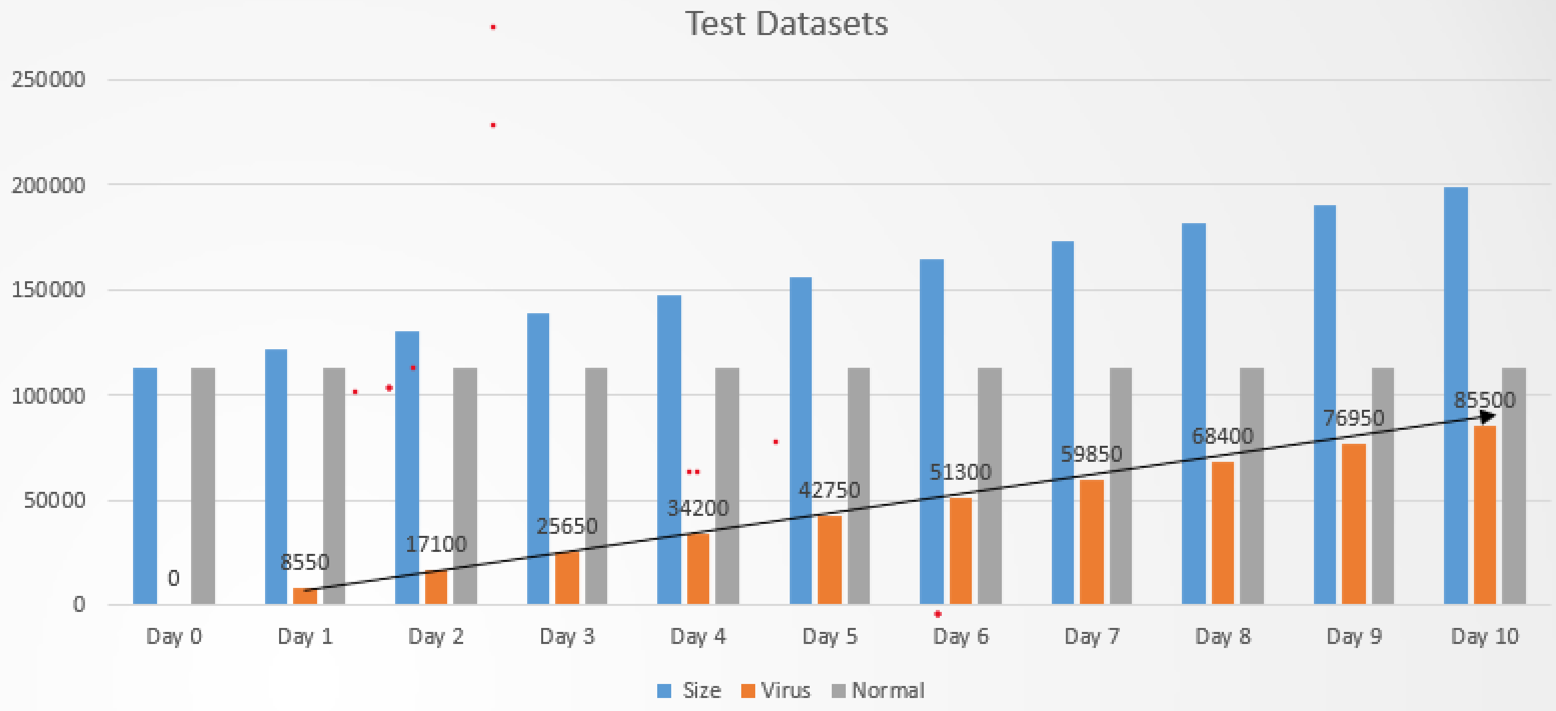
Computer Network



Datasets

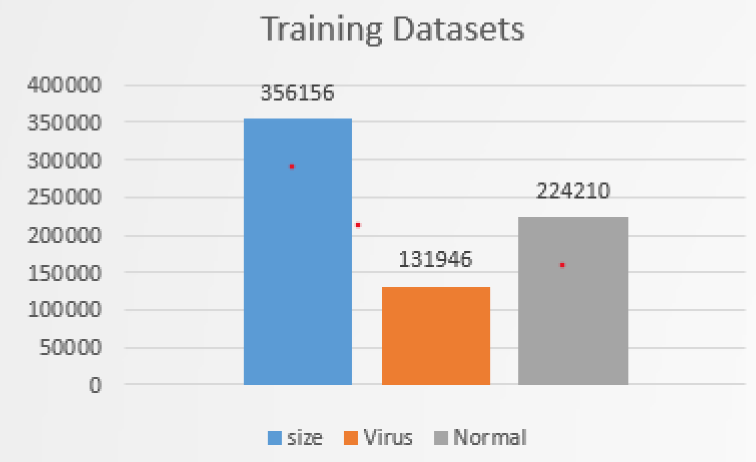


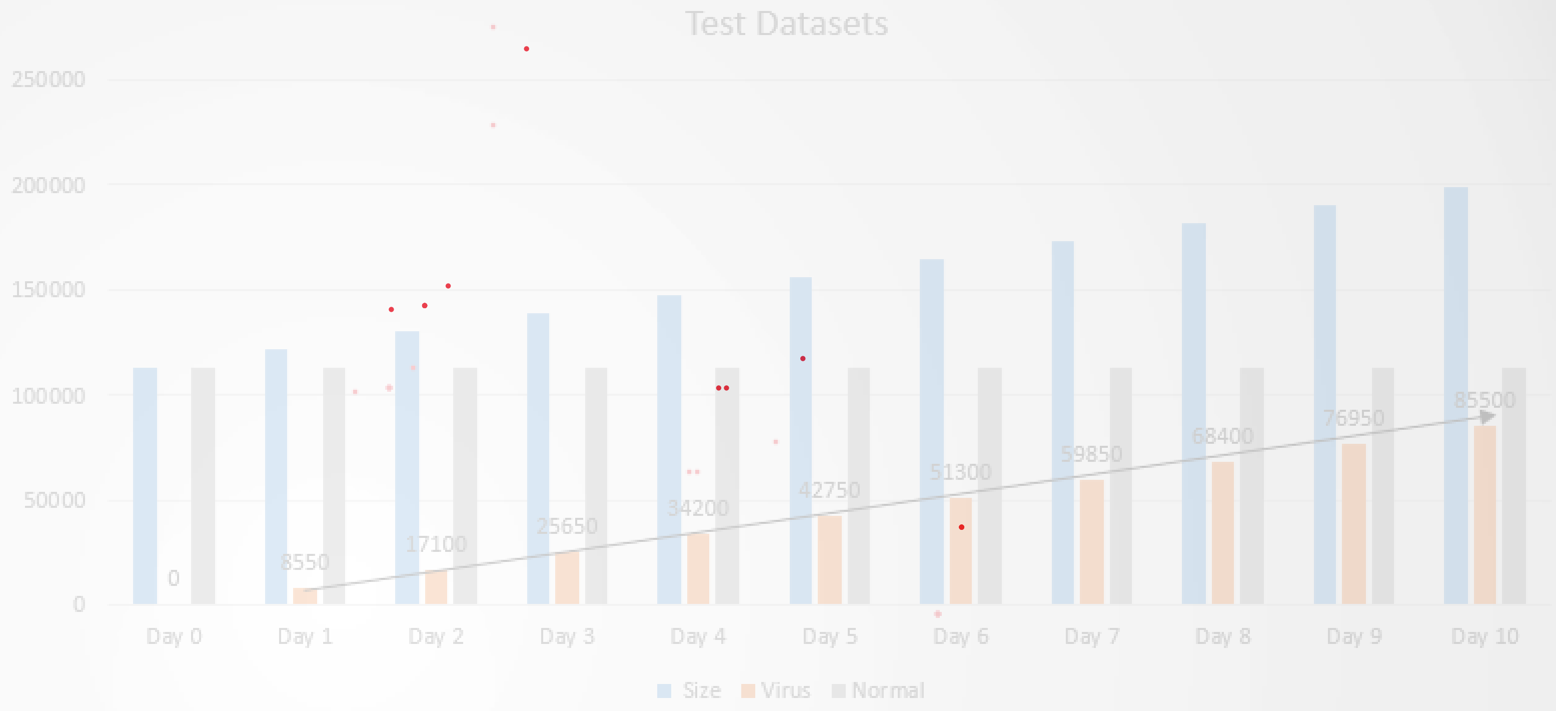
Training Datasets
400000
350000
300000
250000
200000
150000
100000
50000
0
356156
131946
224210
size
Virus
Normal



Test Datasets
250000
200000
150000
100000
50000
0
0
8550
17100
25650
34200
42750
51300
59850
68400
76950
85500
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
Size
Virus
Normal

Datasets


Training Datasets
400000
350000
300000
250000
200000
150000
100000
50000
0
356156
131946
224210
size
Virus
Normal


Test Datasets
250000
200000
150000
100000
50000
0
0
8550
17100
25650
34200
42750
51300
59850
68400
76950
85500
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
Size
Virus
Normal

Tools

ML Algorithm

Logistic Regression

Decision Tree

Balanced Random Forest

Model Validation & Evaluation

scikit learn

Imbalanced learn

Visualization

matplotlib

seaborn

Excel

Data Analysis

pandas

NumPy

Tools

ML Algorithm

Logistic Regression

Decision Tree

Balanced Random Forest

Model Validation & Evaluation

scikit learn

Imbalanced learn

Visualization

matplotlib Version 3.4.3

seaborn

Excel

Data Analysis

pandas

NumPy

Tools

ML Algorithm

Logistic Regression

Decision Tree

Balanced Random Forest

Model Validation & Evaluation

scikit learn

Imbalanced learn

Visualization

matplotlib

seaborn

Excel

Data Analysis

pandas

NumPy

Tools

ML Algorithm

Logistic Regression

Decision Tree

Balanced Random Forest

Model Validation & Evaluation

scikit learn

Imbalanced learn

Visualization

matplotlib

Version 3.4.3

seaborn

Excel

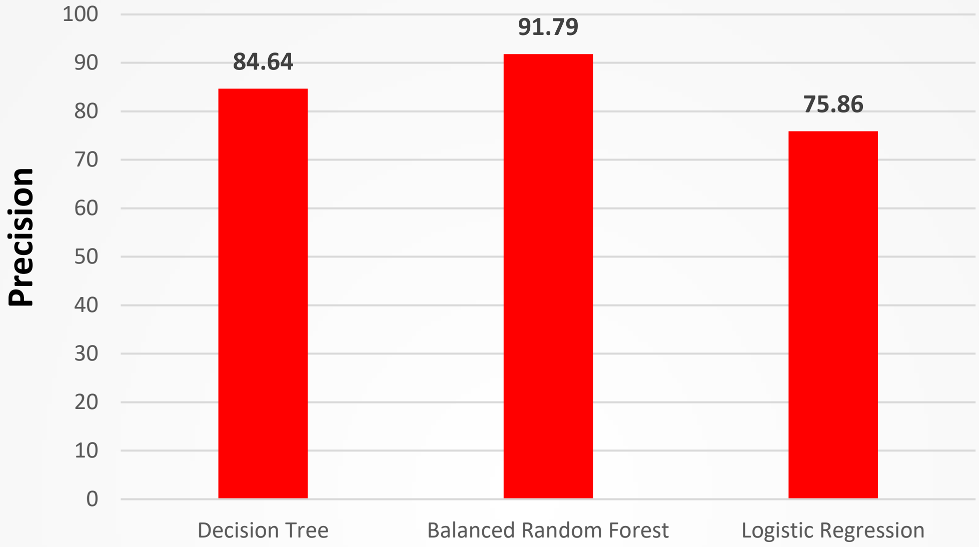
Data Analysis

pandas

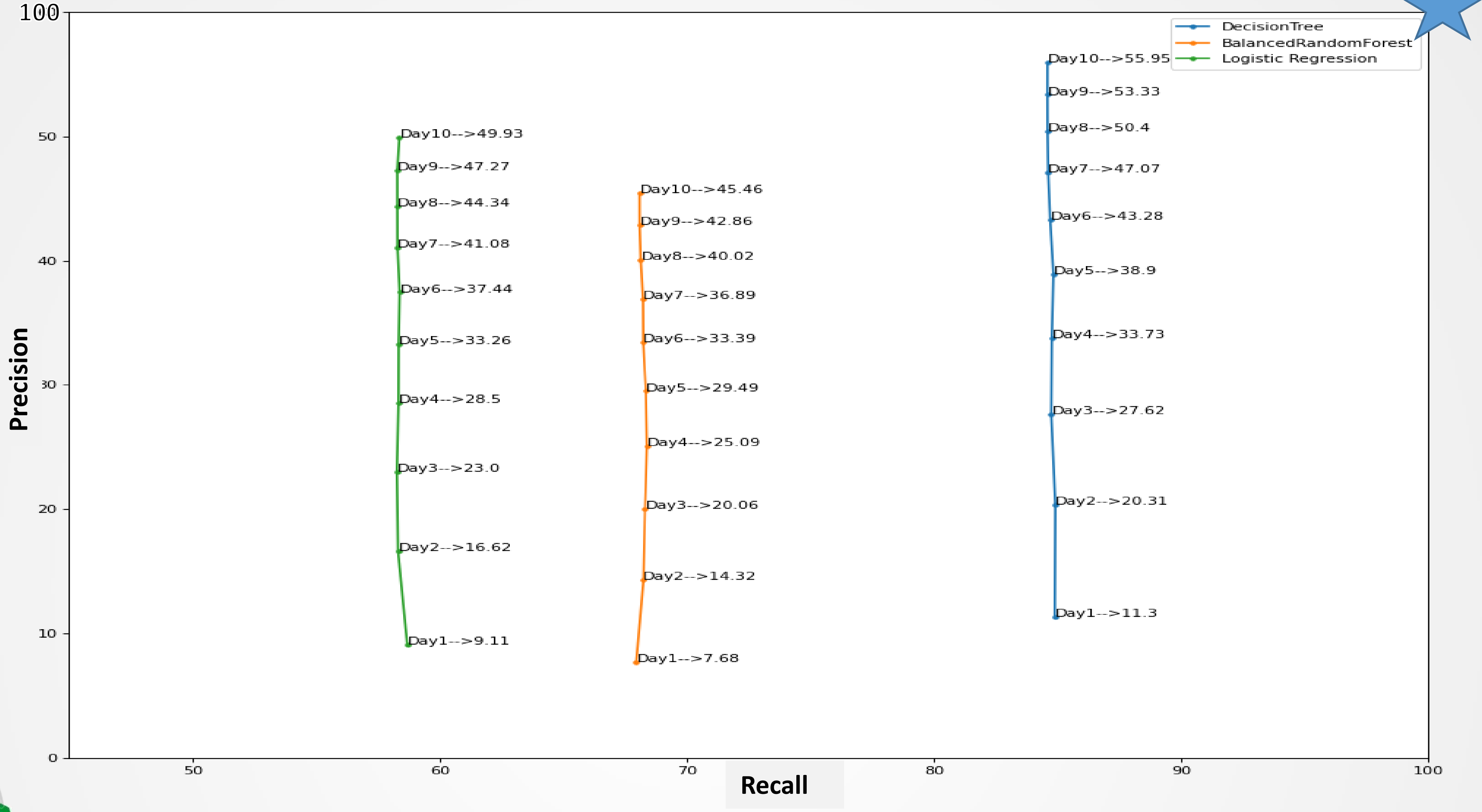
NumPy

Model Validation & Selection

The Best Models are selected by Tuning the hyper parameters using Random Search & cross validation (K = 3)



The Best Models Evaluation



FP = 0

FN = 0

The Efficiency of Machine Learning Techniques in Minimizing Dwell Time of Cyberattacks

Conclusion

All the best models in the project are not performing well in detecting the cyberattack early

Conclusion

All the best models in the project are not performing well in detecting the cyberattack early

which not help in achieving the project goal: Minimizing the Dwell Time of Cyberattack

Future Work

Improve the model performance through feature extraction , Feature selection ,and build the models using other algorithms (eg: XGboost)

Thank You