Anomaly Detection Using ML

Area 54

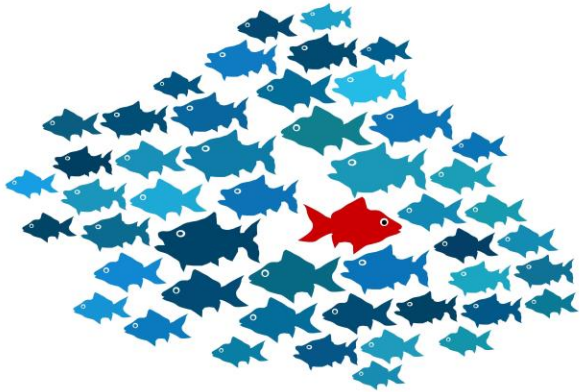
Athithya Jayadevan, Kinaar Desai, Krish Avvari, Sonia Reddy Kolli, Paige Godvin

Problem Statement

- Cyber attacks are becoming more prevalent and more sophisticated
  - Very difficult to detect
    - 6 months = average time to detect breach
- Business value
  - Costs of online crime predicted to reach $6 trillion by 2021
  - $2.4 million = average cost of malware attack
  - Global anomaly detection market estimated to double over the next 5 years to $4.45 billion

Solution Description

- Incorporating ML/AI
- Implementing model to detect network anomalies
- Relieving reliance on manpower
- Increase visibility
- Helps prevent any kind of fraud and threat to a business process
- Potential to detect previously unknown attacks (zero-day)



How it Works?

- Network communication datasets are collected from available source
  - (Kaggle, GitHub)
- Data is processed and stored as DataFrames and 1-D arrays for the algorithms.
- ML algorithms are trained with the DataFrames and various parameters
  - average accuracy of 95%
  - FFN is observed to have an average accuracy of 97%

Naive Bayes

```
------------*Naive Bayes CONFUSION MATRIX*-------------------
[[2137  186]
 [ 300 2416]]


              precision    recall  f1-score   support

           0       0.88      0.92      0.90      2323
           1       0.93      0.89      0.91      2716

    accuracy                           0.90      5039
   macro avg       0.90      0.90      0.90      5039
weighted avg       0.90      0.90      0.90      5039
```

Support Vector Machines

```
------------*SVM CONFUSION MATRIX*--------------------
[[2000  349]
 [   1 2689]]
              precision    recall  f1-score   support

           0       1.00      0.85      0.92      2349
           1       0.89      1.00      0.94      2690

    accuracy                           0.93      5039
   macro avg       0.94      0.93      0.93      5039
weighted avg       0.94      0.93      0.93      5039
```

K-Nearest Neighbors

```
------------*KNN CONFUSION MATRIX*--------------------
[[2311   12]
 [  17 2699]]


              precision    recall  f1-score   support

           0       0.99      0.99      0.99      2323
           1       1.00      0.99      0.99      2716

    accuracy                           0.99      5039
   macro avg       0.99      0.99      0.99      5039
weighted avg       0.99      0.99      0.99      5039
```

RandomForest

```
------------*RandomForest CONFUSION MATRIX*--------------------
[[2316    7]
 [   6 2710]]


              precision    recall  f1-score   support

           0       1.00      1.00      1.00      2323
           1       1.00      1.00      1.00      2716

    accuracy                           1.00      5039
   macro avg       1.00      1.00      1.00      5039
weighted avg       1.00      1.00      1.00      5039
```

Decision Trees

```
------------*Descision Tree CONFUSION MATRIX*-------------------
[[2315    8]
 [   9 2707]]


              precision    recall  f1-score   support

           0       1.00      1.00      1.00      2323
           1       1.00      1.00      1.00      2716

    accuracy                           1.00      5039
   macro avg       1.00      1.00      1.00      5039
weighted avg       1.00      1.00      1.00      5039
```

Feed-Forward Neural Networks

```
------------*FeedFordward Neural Network CONFUSION MATRIX*----------------
[[ 6339  3096]
 [    9 10709]]


              precision    recall  f1-score   support

           0       1.00      0.67      0.80      9435
           1       0.78      1.00      0.87     10718

    accuracy                           0.85     20153
   macro avg       0.89      0.84      0.84     20153
weighted avg       0.88      0.85      0.84     20153
```

Next Steps

- Can be used for learning tool for Knowledge-Based Systems.
- Can be made largely scalable using REST.

Teamwork

- Athithya Jayadevan
  - Provided in-depth knowledge of AI and primary support
- Krish Avvari
  - Generated .csv files to compare each algorithm
- Kinaar Desai
  - Sklearn Classifiers and Confusion Matrix Creation
- Paige Godvin
  - Performed research on data sets and prepared presentation
- Sonia Reddy Kolli
  - Prepared presentation and researched on anomaly detection

Questions


CISCO