---

title: "CyberAmongus"

author: "Shawn"

date: "2024-10-11"

The Standard Features

Unauthorized-Logons

File Activity to sensitive files

Emails to outside domains

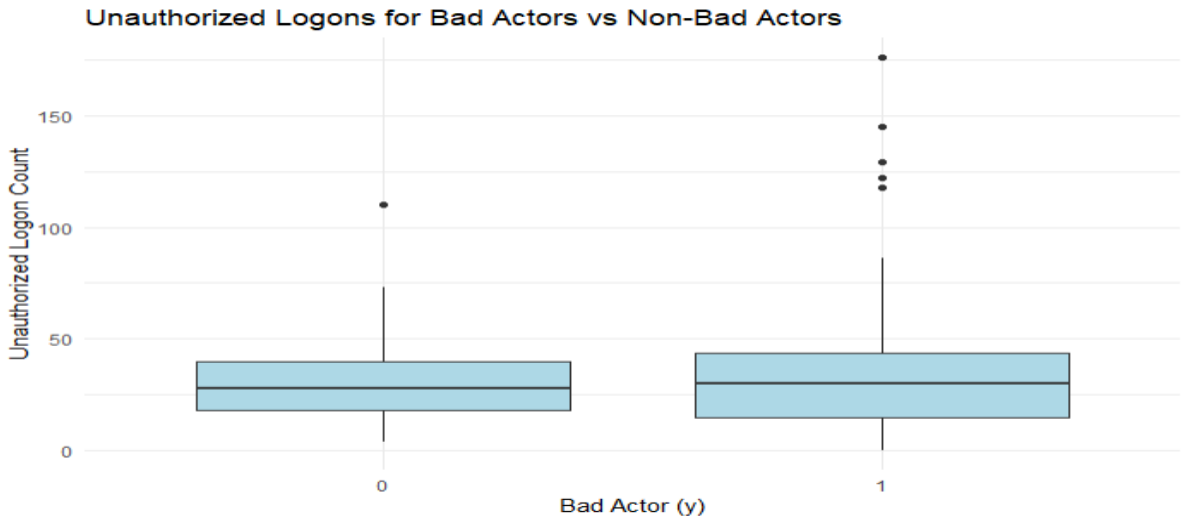
Email sentiment

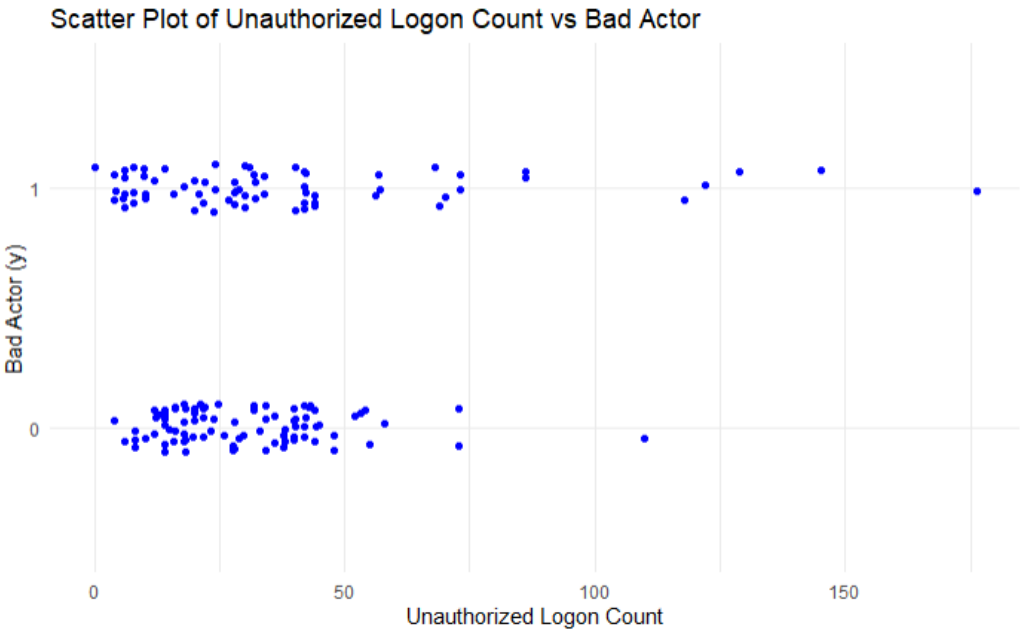
Outside working hours Logons (not built)

External Device Connections (not built)

Feature 1, Unauthorized Logons

In this section I first read in the file access data and logon data and also make the essential Features data-frame which will hold all of the constructed features. Then I made the feature, Unauthorized Logons which I did by checking to see if an employee was part of the employee list when he logged onto the company's server. I grouped by user and counted the times that user committed an unauthorized logon. I saved the feature and visualized the results:



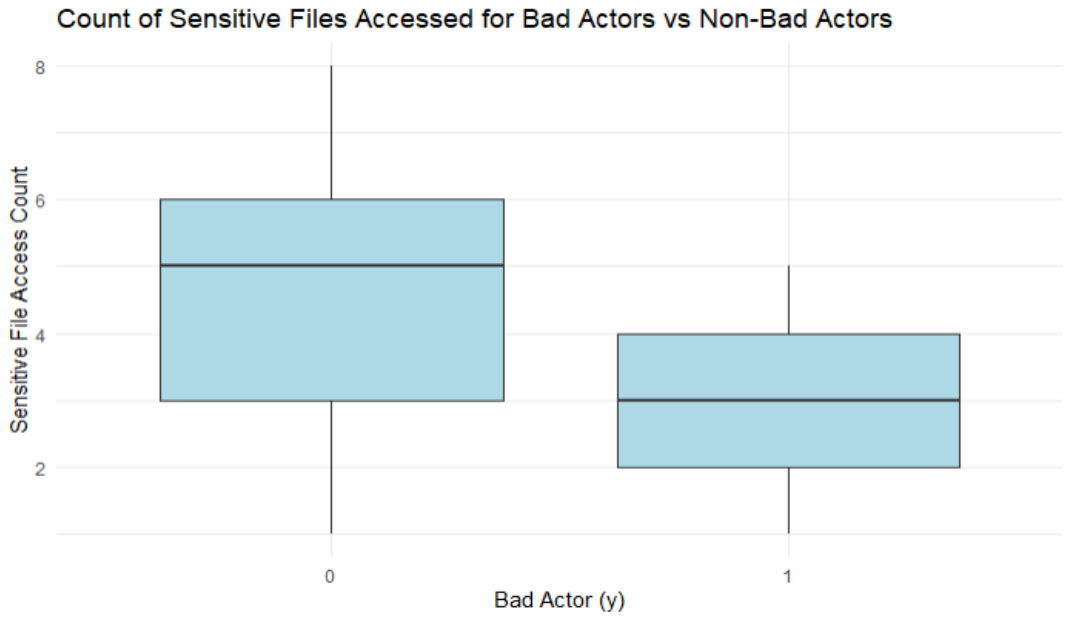


The 0 indicates a benign user while the 1 indicates that a targeted bad actor was included in the results. As you can tell, unauthorized access alone is not enough to determine if the user is a bad actor or not.

Feature 2, Sensitive File Access

In this chapter I load access to files. I define sensitive files as being a file type: ("docx", "pdf", "xlsx", "exe") and containing words such as: ("confidential", "secret", "financial", "strategy"). The results were that about 25% of all sensitive accessed files were by malicious users. Now the feature would not be very good if it just flagged a user for accessing sensitive files, so I created a baseline for each user based on the

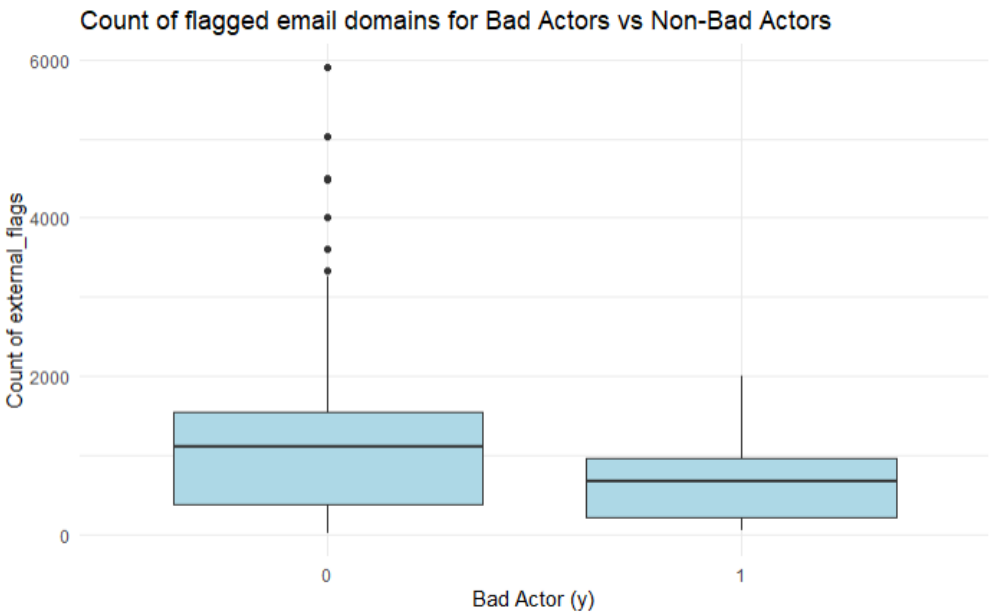
median number of sensitive files they accessed per month and flagged them if in a month they accessed more than their median. Then I counted the flags and added the count of each flagged access to the feature data. Below is the results for the number of access counts per user that was flagged separated based on malicious vs benign users.



Feature 3, Emails

This section has two parts. The first is for the number of emails involving external domains and the second is for email sentiment.

In the first part I wanted to flag individuals who sent and received emails to members outside of the organization whos email did not include another member of the company. Since that combination could make for a red flag.



The results of this section show a significant number of outliers and a higher trend for malicious users to engage in this feature compared to the benign users.

For part two of emails, I was looking for email sentiment. I used the In-built sentiment lexicons in tidytext to sum up a user's email conversation as either positive or negative then counted the number of negative emails and saved that as a feature. I didn't visualize this section, but all 70 malicious users were found to send negative emails.

Supervised Model Learning

For building the model that would determine if a user was malicious or not, I used the random forest model as well as the caret package to split and balance the dataset. I balanced by down-sampling and used 80% of the dataset for training.

Here is the result of my trained model:

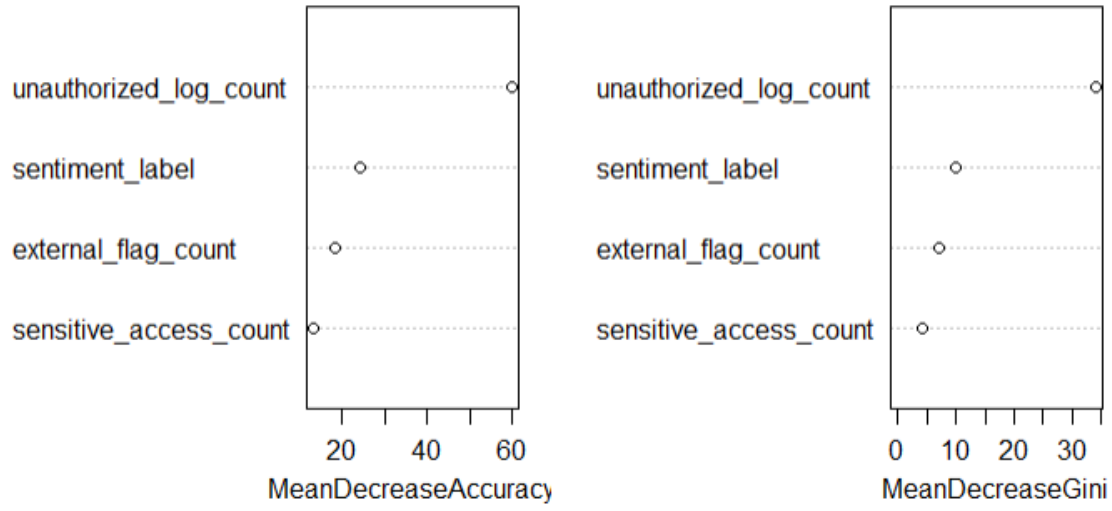
```
Call:
 randomForest(formula = Class ~ ., data = train_data, importance = TRUE,      ntree = 500)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 6.25%
Confusion matrix:
   0  1 class.error
0 51  5  0.08928571
1  2 54  0.03571429
```

Here is the result of my tested model:

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 13  0
         1  1 14

               Accuracy : 0.9643
                 95% CI : (0.8165, 0.9991)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 1.08e-07

                  Kappa : 0.9286

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9286
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.9333
             Prevalence : 0.5000
         Detection Rate : 0.4643
   Detection Prevalence : 0.4643
      Balanced Accuracy : 0.9643

       'Positive' Class : 0
```

I like that the results favored safety, 1 false positive and no false negatives. The model will perform between 81-99% on any run. The goal moving forward will be to add psychometric features and see if the accuracy improves.

Here are the importances of the features.



This is the same as seen in the previous experiments done using machine learning with the CERT-dataset, but the scientific advancement of this paper will be to see where the psychometric features fall on the above charts.

The Psychometric Features

Psychometric-Behavioral Risk:
1
risk_psycho_behavior: A composite risk score that combines high-risk personality traits (e.g., high Neuroticism or low Conscientiousness) with observed risky behaviors (e.g., **unauthorized logons** or **accessing sensitive files**).

Personality Traits and Off-Hour Activity:
0
psycho_extro_offhour: Users with high Extroversion who **log in outside of** working hours may be more likely to collaborate or socialize, which could suggest either legitimate or risky activity.
psycho_neuro_offhour: Users with high Neuroticism **logging in during off-hours** might indicate stress-driven behavior, which could signal potential risks.

Conscientiousness and Device/File Usage:
0
psycho_consc_device_freq: Users with low Conscientiousness could be flagged if they frequently **connect external devices**, as this could indicate less cautious behavior.
psycho_consc_file_freq: Similarly, low Conscientiousness and high **sensitive file** usage might indicate carelessness or intent to exfiltrate data.

Agreeableness and Social Engineering:
1
psycho_agreeableness_email: Users with high Agreeableness might be more susceptible to phishing or social engineering attacks, so their email activity can be monitored for potential risks (e.g., high interaction with **unknown external domains**).
Psycho_email_sentiment: Users with negative email sentiment who are also high in neuroticism will be flagged twice as often.

Neuroticism and High-Risk Activity Clustering:
1
psycho_neuro_risk_cluster: Users with high Neuroticism who d*emonstrate clusters of high-risk activities* (e.g.,