Phishing Website Detection

Alex Webster[1], Ryan Saweczko[1], Ciaran Byles-Ho[1], Allen Zhong[1]

1. QMIND, Queen's University

Have you ever received a text message about a bank account you don't have, or an email saying an account has been hacked but it isn't from the company? Those are Phishing Scams and attackers target individuals and entire companies. In the corporate world, this issue accounts for 90% of data breaches where the average breach costs 3.86 million dollars[1]. Phishing attacks give a link to a 'fake' website trying to act like it was created by a certain group. The goal of this work is to detect these websites by using the URL and create a RESTful API around the model for ease of application development. We developed highly correlated features from the number of subdomain names to individual word parsing based on statistical analysis. From our feature engineering, we quickly experimented with models and evaluation methods such as confusion matrices (figure 1) and f1-score (as our dataset is skewed). Our best performing models were Deep Neural Networks and Random Forests to achieve an accuracy of 91% with most errors as False Negatives. Building off this work, a chrome-extension has been created to automatically send links when browsing to our RESTful API for prediction. These results prove the validity of using URLs for Phishing Website detection. Thus, future work shall include further feature engineering and dataset development as the Phishing URL domain changes.

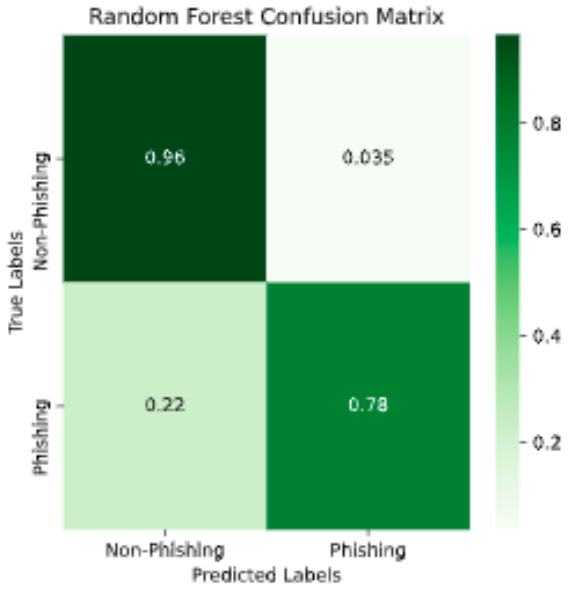

Figure 1 – Confusion matrix for Random Forest Predictor

1. https://retruster.com/blog/2019-phishing-and-email-fraud-statistics.html