A Machine Learning Approach to Phishing Detection

Kendra Maggiore
Texas State University
firstauthor@i1.org

Jon Pugh
Texas State University
JPugh90@gmail.com

James Knepper
Texas State University
secondauthor@i2.org

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

Email is a widely used and effective means of electronic communication, and it is a commonly accepted method of contact between entities regardless of the specific field. Since email is such a primary correspondence in the modern world, there are those who wish to exploit email service and its users. A common exploitation technique of email services is known as Phishing. Phishing is a major threat to today's email communication. In a Phishing attempt, a malicious entity will send an email to a user that appears to be legitimate. The purpose of the email is to solicit the user to provide personal or sensitive information, often by containing a URL to an unsecure or malicious website.

2. Existing Solutions

[2] is an example of a phishing email detection machine learning algorithm that utilizes Rapid Miner. We plan to use a similar feature set but we will implement our algorithm with Python to provide greater flexibility in development. It also uses a dataset that was created in 2007. We will use a dataset available from [3] that was created in 2018. If time allows we will implement TF-IDF as discussed in [1] as an additional feature.

3. Preliminary Plan

- Convert latest .mbox from [3] to a .csv containing feature set
- Design and implement Logistic Regression algorithm
- Design and implement SVM algorithm
- Implement TF-IDF as an additional feature

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

4. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books.

References

[1] S. K. Harikrishnan NB, Vinayakumar R. A machine learning approach towards phishing email detection. Technical report, Center for Computational Engineering and Networking(CEN), 2018.

[2] D. Ocampo. https://github.com/diegoocampoh/machinelearningphishing, 2017.

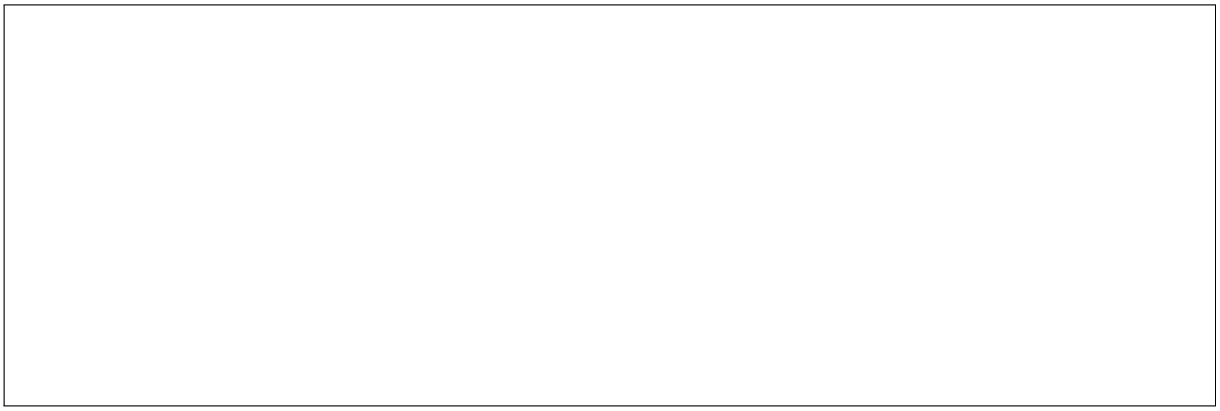

Figure 1. Example of a short caption, which should be centered.