Phishing Website Detection Using URL-Based Features

A Machine Learning Approach to Combat Fraudulent Links

Ayrton Joseph DiPina  
 Data Science  
 Institution/University Name  
 Boston, Massachusetts, USA  
 dipinaa@wit.edu

ABSTRACT

Provide one paragraph abstract for your report.

∗Article Title Footnote needs to be captured as Title Note

†Author Footnote to be captured as Author Note

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*WOODSTOCK’18, June, 2018, El Paso, Texas USA*

© 2018 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...$15.00

https://doi.org/10.1145/1234567890

KEYWORDS

Insert 3-5 keywords for your project.

1 Introduction

On July 19th, 2024, a faulty software update from CrowdStrike, a reputable cybersecurity firm, caused a global IT outage. The update contained a defect that caused Windows computers to crash, rendering the computers unable to restart. This incident affected about 8.5 million worldwide devices. This incident impacted airports, banks, and multiple other businesses. At the time of this global panic, I was a cybersecurity intern at a water company. I was researching what was happening to see if this faulty update led to our systems being down. After confirming that CrowdStrike did not affect our computers, I noticed many different links that did not look official. Then I realized that malicious users were creating fake links to try and socially engineer employees that are panicking to try and fix their computers that were impacted by CrowdStrike, acting as CrowdStrike Support.

The proliferation of online phishing attacks has become a critical cybersecurity concern, as these deceptive techniques exploit users into divulging sensitive information such as passwords, credit card details, or personal identification numbers. Phishing detection, therefore, plays a vital role in securing internet users and organizations from financial loss and data breaches. This project, which focuses on identifying phishing websites using URL-based features, is a significant step towards enhancing cybersecurity. It offers a lightweight and real-time approach that can augment broader cybersecurity measures, providing reassurance and confidence in the fight against phishing.

2 Data

In this part, you should introduce your datasets.

2.1 Source of dataset

Where did you download it? Is it a credible source? When were the datasets generated? How were the datasets generated by the creator? If you create the datasets, how did you generate it?

Example: xxxx

2.2 Characters of the datasets

What’s the format and size of the datasets? What parameters/columns/rows/character and their units are included in this dataset. Use a table to explain this is recommended. Did you clean the data or convert any unit in the dataset? If so, what’s the formula/rule did you apply? Did you combine any datasets? If so, how do you combine them? Did you create any new category for analysis in the datasets? If so, what and how do you create?

3 Methodology

In this part, you should give an introduction of the methods/model. First, what’s the method/model. What’s the assumption of this method/model. What’s the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extra work did you adjust to make the results better. If you have multiple methods, feel free to use subsection 3.1, 3.2, 3.3, … to separate them.

3.1 Heading Level 2

3.2 Heading Level 2

…

Example format: The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

4 Results

In this part, you need to select a reasonable way to deliver the result of your topic. For example, equation or numerical results, or visualization of your result. You also need to provide a clear explanation of all results and how to understand the results. If there exist any unexpected results, please explain why or possible cause of this special result. You can use subsection 4.1, 4.2, … to separate your results.

4.1 Heading Level 2

Example format: In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

5 Discussion

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the feasible suggestions of future work to revise/improve your result.

6 Conclusion

In this part, you should summarize your project. What important results did you find for your topic and what’s the effect of this result on the real-world?

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here.

REFERENCES

Use the following ACM Reference format for your citation

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK’18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: <https://doi.org/>10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

<https://techcrunch.com/2024/07/19/faulty-crowdstrike-update-causes-major-global-it-outage-taking-out-banks-airlines-and-businesses-globally/?utm_source=chatgpt.com>

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00