Paper Title:

A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques

Paper Link:

https://ieeexplore.ieee.org/document/9795286

1 Summary

1.1 Motivation

This research intends to apply natural language processing (NLP) to a thorough examination of phishing email detection techniques. The growing threat of phishing assaults and the requirement for efficient detection methods serve as the driving forces behind this. The main theory centers on using natural language processing (NLP) approaches to improve both the efficacy and accuracy of phishing detection.

1.2 Contribution

The contribution of the research is a systematic literature review that covers 100 studies from 2006 to 2022 and identifies trends, techniques, datasets, standards for evaluation, and instruments for phishing email detection. As a useful resource for scholars and practitioners, it compiles data on prominent datasets, machine learning algorithms, and natural language processing techniques.

1.3 Methodology

The methodology involves a systematic literature review, systematically addressing 11 research questions. The authors meticulously analyze studies, categorizing findings related to NLP techniques, machine learning algorithms, datasets, evaluation criteria, and tools employed in phishing email detection. This methodological approach ensures a comprehensive and organized presentation of the research landscape.

1.4 Conclusion

The importance of supervised methods, new developments in deep learning, and the necessity of more research into unsupervised and semi-supervised models are all highlighted in the conclusion. It highlights the need for updated methods, such as deep learning strategies like LSTM and CNN, and the increased interest in phishing detection, especially after 2019.

2 Limitations

2.1 First Limitation/Critique

The scope of the literature review is severely limited when important databases like Web of Science, Scopus, ERIC, and ProQuest are excluded, and pertinent research may go unnoticed. Future research must incorporate a larger database due to this reliance on sparse sources in order to have a more thorough and objective grasp of the area. Increasing the amount of data available will reduce the possibility of overlooking important discoveries and improve the analysis by fostering interdisciplinary collaborations. The creation of comprehensive knowledge depends on making use of all the information that is accessible.

2.2 Second Limitation/Critique (15%)

Understudied Arabic phishing email topics expose non-English speaking populations to internet threats. Due to its rarity, Arabic NLP is lacking in this crucial area, which makes it difficult to spot problems and puts many people's digital safety at risk. To close this gap and guarantee inclusive and thorough internet security for everyone, research on Arabic natural language processing must be given top priority.

3. Synthesis

In addition to the results, the synthesis sparks debate on future directions and useful uses. It supports cutting-edge instruments to strengthen phishing detection, such as social honeypots and customized recommendation algorithms. It advocates for a more comprehensive knowledge and strong defenses against developing cyber threats by utilizing semantic analysis and endorsing a holistic strategy that takes attacker views into account.