Sentence Level Analysis Model for Phishing Detection

Lindah Sawe¹, Joyce Gikandi¹, John Kamau¹, and David Njuguna¹

¹Mount Kenya University

September 11, 2023

Abstract

Phishing emails have experienced a rapid surge in cyber threats globally, especially following the emergence of the COVID-19 pandemic. This form of attack has led to substantial financial losses for numerous organizations. Although various models have been constructed to differentiate legitimate emails from phishing attempts, attackers continuously employ novel strategies to manipulate their targets into falling victim to their schemes. This form of attack has led to substantial financial losses for numerous organizations. While efforts are ongoing to create phishing detection models, their current level of accuracy and speed in identifying phishing emails is less than satisfactory. Additionally, there has been a concerning rise in the frequency of phished emails recently. Consequently, there is a pressing need for more efficient and high-performing phishing detection models to mitigate the adverse impact of such fraudulent messages. In the context of this research, a comprehensive analysis is conducted on both components of an email message – namely, the email header and body. Sentence-level characteristics are extracted and leveraged in the construction of a new phishing detection model. This model utilizes K Nearest Neighbor (KNN) introducing the novel dimension of sentence-level analysis. Established datasets from Kaggle was employed to train and validate the model. The evaluation of this model’s effectiveness relies on key performance metrics including accuracy of 0.97, precision, recall, and F1-measure.

Hosted file

PAPER SUBMISSION.docx available at https://authorea.com/users/661940/articles/664823-sentence-level-analysis-model-for-phishing-detection