Phishing email detection through linguistic patterns and sentiment analysis

Student: Daniel Pedrinho Nº 107378

Supervisor: João Rafael Duarte de Almeida

Co-supervisors: Sérgio Guilherme Aleixo de Matos

Summary: With the wide usage of e-mail as a communication tool, phishing attacks have become increasingly common and sophisticated. This dissertation aims to explore **the use of linguistic patterns** and **sentiment analysis to detect phishing emails**. By analyzing the emotional tone and language used in emails, we may be able to identify potential phishing attempts and improve email security.

Methodology Overview

- Dataset Preparation:
  - No ideal, fully-labeled dataset exists for both phishing and emotional classification tasks
  - Generate phishing and non-phishing email datasets using an LLM, using an existing dataset as example
  - Add emotion annotations to enrich label diversity and create a more representative dataset.
- Initial Model Pipeline
  - Create combined dataset, load it and perform text normalization
  - Split 80/20 Train/Test sets and vectorize them with TF-IDF
  - Train unsupervised clustering models
  - Evaluate Results
  - Preliminary Results:
    - Phishing F1 Score: 0.4501
    - Emotion F1 Score: 0.0532

Objectives / Work done / Results

- Objectives:
  - Develop the initial model pipeline and refine it overtime
  - Monitor the annotation progress
  - Continuous development of document

- Results:
  - Ready-To-Run pipeline prepared
  - Dataset processing and normalization developed
  - Preliminary results

Next Steps / Challenges / Bottlenecks

- Next Steps:
  - Keep annotating the dataset
  - Run initial pipeline with the “full” dataset
  - Expand pipeline with Transformers models

- Challenges:
  - Annotation process is lengthy