

Phishing email detection through linguistic patterns and sentiment analysis

Student: Daniel Pedrinho N° 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.

Objectives / Work done / Results

❖ Read:

- “Streamlining the Analysis of Phishing Emails Using Artificial Intelligence”

❖ Understand:

- How email and .eml works
- How prevalent phishing is
- What security measures exist
- Machine Learning applications in phishing
- How sentiment analysis can improve phishing detection

❖ Results:

- State of the art begins taking shape

Next Steps / Challenges / Bottlenecks

❖ Next Steps:

- Use LLM to generate phishing emails based on an existing annotated dataset
- Annotate newly generated dataset using Doccano
- Use dataset to generate pre-annotated phishing emails

❖ Possible challenges:

- Manual annotation with students

