Phishing email detection through linguistic patterns and sentiment analysis

Student: Daniel Pedrinho Nº 107378

Supervisor: João Rafael Duarte de Almeida

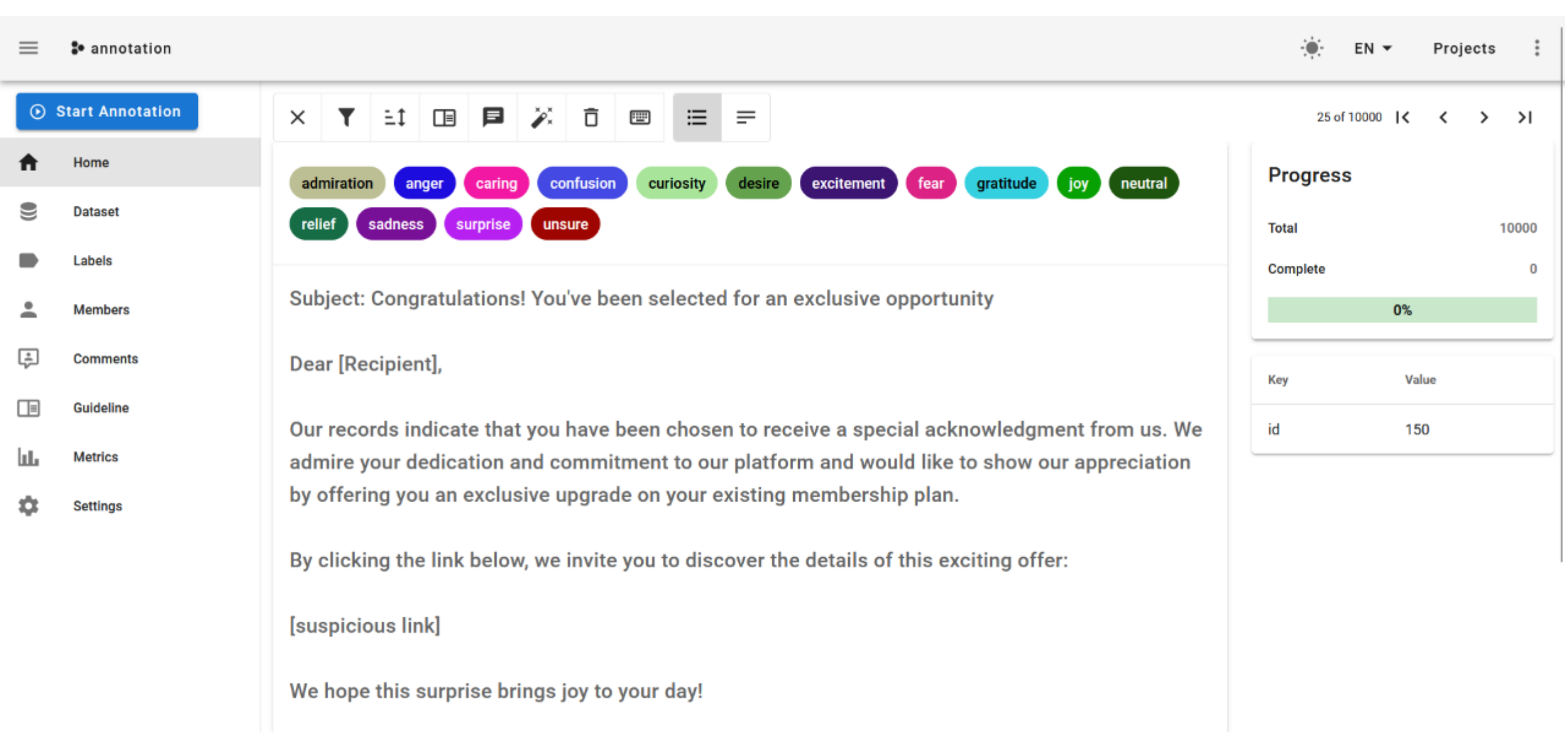
Co-supervisors: Sérgio Guilherme Aleixo de Matos

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.

Objectives / Work done / Results

- Objectives:
  - 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

- Results:
  - Generated dataset with 10,000 emails, evenly split across emotions
  - Doccano plataform deployed and ready, with annotation guidelines
  - User creation + Email automation script ready


annotation
EN
Projects
Start Annotation
Home
Dataset
Labels
Members
Comments
Guideline
Metrics
Settings
25 of 10000
admiration
anger
caring
confusion
curiosity
desire
excitement
fear
gratitude
joy
neutral
relief
sadness
surprise
unsure
Subject: Congratulations! You've been selected for an exclusive opportunity
Dear [Recipient],
Our records indicate that you have been chosen to receive a special acknowledgment from us. We admire your dedication and commitment to our platform and would like to show our appreciation by offering you an exclusive upgrade on your existing membership plan.
By clicking the link below, we invite you to discover the details of this exciting offer:
[suspicious link]
We hope this surprise brings joy to your day!
Progress
Total
10000
Complete
0
0%
Key
Value
id
150

Next Steps / Challenges / Bottlenecks

- Next Steps:
  - Annotate the dataset
  - Use the annotated dataset to generate non-phishing emails

- Bottlenecks:
  - Annotation speed of volunteers