

Email classification

Objective:

Build an AI model to classify emails into different categories based on the content of the emails

Data:

Please download the "A subset of about 1700 labelled email messages" from

https://bailando.berkeley.edu/enron_email.html

Procedure:

- Pre-process and clean the data
- Build a model to categorize them into the following classes
 - a) Company Business, Strategy, etc. (elaborate in Section 3 [Topics])
 - b) Purely Personal
 - c) Personal but in professional context (e.g., it was good working with you)
 - d) Logistic Arrangements (meeting scheduling, technical support, etc)
 - e) Employment arrangements (job seeking, hiring, recommendations, etc)
 - f) Document editing/checking (collaboration)
- Train and test the model and report the performance

Instructions:

- Please use Python for the implementation
- You may use 80% of data for training and 20% for validation
- You may use opensource models, packages or libraries if needed.