Email Classification Using Natural Language Processing

Andrew Ferreira, Michael VanDusen

January 27, 2022

Project

The objective of this project is to help a person to classify an incoming email.
We will use different datasets with a similar data structure to create the data pipeline, which can be applied to different purposes later.

Critical path

Data cleaning — Due Date: Jan 31

The processing of transformation the YELP! dataset into the structure we need: review and classification

Exploratory Data Analysis — Due Date: Feb 4

Understand the data and it's distribution

Defining the model — Due Date: Feb 11

The bag-of-words model VS N-gram word model

Create the data pipeline (ETL) — Due Date: Feb 18

Create the pipeline that might work with other similar projects

Paper — Due Date: Feb 25

write the mid-term paper