COMP3162
Spam Classification Competition
Apr 7, 2024 11:55PM

Description

You will be competing in a spam classification kaggle [competition.](#)You have been given a
dataset on emails and you are to train a model that predicts from the content of an email
whether that email is spam or not spam. In this assignment your models will be judged based on
the [F1 score](#) which is better for imbalanced datasets.

Complete the following

- Extract data
- Do the necessary data preprocessing and feature engineering
- Pick any one of the various supervised learning models you would have learnt of to train
  your predictive model on the data in **train.csv**.
- Use the predictive model to predict the data in **test.csv**
- Convert predictions to csv and submit. The format of the csv can be seen in the
  sample_submission.csv file.
- You must submit your predictions to kaggle.

What to Submit
- Ipynb notebook file.
- Submission csv file.
- Screenshot of place on leaderboard