Week starting 24/10/22 Status Report

What I achieved

* Temporarily fixed the email parsing script
* Setup a data frame to hold all tokenized messages/ lemma/ and predictions
* Split up the email into training and testing data
* Created pipelines for each classification method
* Trained and tested each method and stored results in the dataframe
* Compared the results of each method, rf was the best

Time Spent - ~ 9 hours

Issues

I worked around the issue with parsing through the emails, It is a temporary fix as it does not store and “to” and “from” details from the emails

I am not sure if the NB and SVM classifications are setup correctly as the results were not great