CA02 – ReadMe – David Rider

The following file was made to see if we could train a model to determine when emails were spam or not. We would then get a comparison to the truth and determine how well our model actually worked.

This was done through first creating a dictionary with the 3000 most common alpha-numeric words. This was followed by creating another function that extracts a feature matrix and train labels from a dataset.

Finally, we used our functions to use the Gaussian model on our datasets and determine our accuracy scores.