Lab 4

Exercise 1

1. Caps, selected, over, additive
2. My conclusion does not change no matter the seed or how many folds you rerun it with.
3. In r
4. The selected model is the best model. This model is the most accurate out of the four different models. The caps model is the least accurate of all the models. The next model with the lowest accuracy is the additive model. With the over model being the second most accurate.

The over model is the second most accurate however, the problem that it creates is due to the fact it can overestimate how many of the emails are considered to be spam. If some emails are marked spam that aren’t spam, that creates a problem as the email would never be seen no matter the importance. The caps model is the least accurate model because it only deals with emails with caps, however not all spam includes this and also some important email headlines are put in all caps so you would miss important information. The third most accurate model is the additive model due to the fact that it adds up each spam and reports the final amount of spam emails. However, this one can miss more spam than the other two ahead of it so that is why it ranked here. People do not want to have their inboxes filled with spam while their actual emails could still be marked as spam.