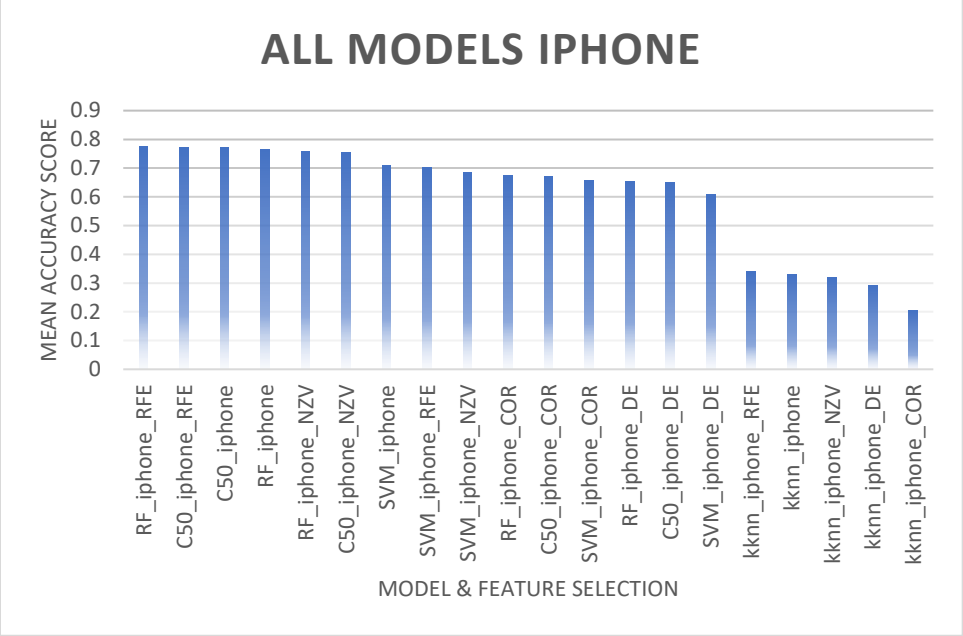
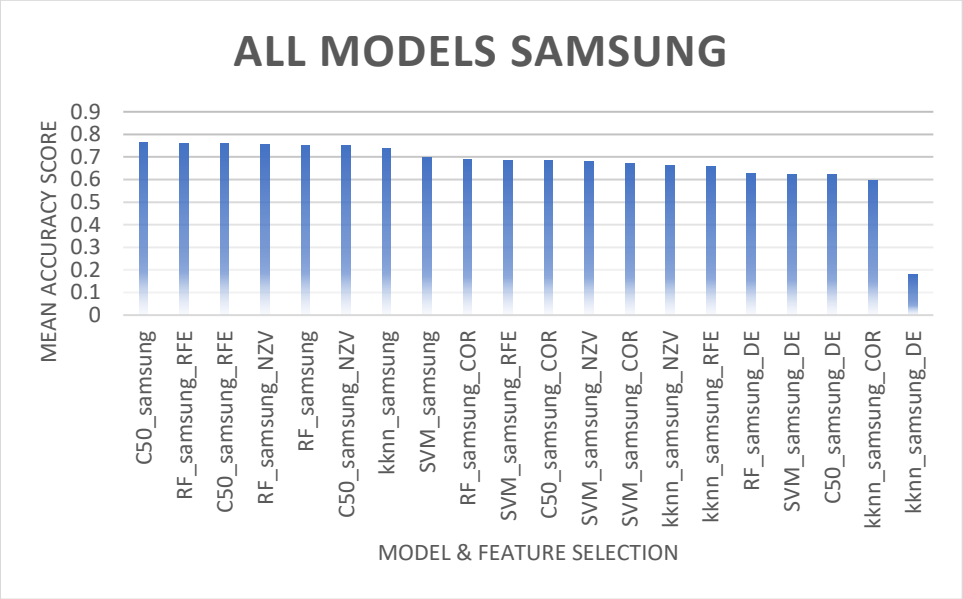
Lessons Learned

The iPhone model selected was Random Forest modelling using Recursive Feature Elimination for its feature selection, this yielded an accuracy score of 77.5% when testing it against with the raw sentimental data matrix. For Samsung the model selected was C50 using the original data with no feature selection, this yielded an accuracy score of 76.6%

The reason for selecting these combinations of classifier and feature selection is because they yielded the best mean accuracy score out of them all (using resampling), this is depicted in the charts below:





Marcelo Jimenez

The most difficult part of the project was making sure all the models ran without any problems. They each take a while to run, so running them a lot of times after fixing errors would make me waste a lot of time. Therefore, I decided to do both iPhone and Samsung at the same time, being very careful as to which data goes where, with what model, etc. It is easy to lose track using loops, so I decide to write every model and feature selection data frame instead of making a loop that could fail and make me waste time.

In the future, I will create the loops as I believe after a certain number of times running the algorithm, I am confident that they wont fail. This would allow me to considerably shorten the code and make it neater. I think making both phones at the same time actually helped with time management so I would do it again, being very careful when writing the code.

I really liked the entire project, it really put me to the test in all the modelling and feature selection algorithms using R. In the future, I would try to experiment more with other models and see if there's something better out there.

Thanks for a great course! Loved it!