DS 315 Final Report - Max Higgins

This project aimed to use data from cognitive performance research to predict reaction time. The dataset we used had over 80,000 rows, so I wasn't worried about not having enough data. I would analyze linear and non-linear models to try and achieve this goal and identify a correlation between the features and reaction time.

I felt confident that the model would be able to accurately predict reaction time values. Midway through the project, I started to realize that this expectation might've been false. Even though the model never was able to accurately predict reaction time, I still gained a lot of insights and understanding for Machine Learning.

After preprocessing the data and engineering a few new data columns, I was encountering a low correlation between almost all the columns. This was the first sign that the model might be weak. Despite this warning, I continued to try and train the model to predict reaction time scores.

After trying a few linear models, such as linear regression and polynomial regression, I was still encountering poor results from the model. The coefficients were near-zero and the intercept was high, two big signs that the model wasn't trained well. The root mean squared error (RMSE) was also very high, nearing 115 for both of the linear models.

After a bit of studying, I realized that a non-linear model might fit the training set the best, so I used a decision tree regression to try and get a lower RMSE. I did manage to achieve a lower RMSE of 47. Because the model was non-linear, this meant that I had multiple RMSE, and the highest was 111. Overall, the model had an initial RMSE of 115, the same as the linear models, so the non-linear model didn't improve much from the linear models.

However, this doesn't mean that the findings were useless or wrong, rather that the features from the dataset cannot be used to find reaction time scores. While I could've taken the early warnings as a sign to switch datasets to something that is more associated with reaction time, I'm glad I stuck with the original dataset, even though the finding couldn't predict reaction time, it's just as useful to know that the findings *cannot* predict reaction time.

If I did want this project to accurately predict reaction time, I could've switched my dataset over to something that is associated with reaction time, such as physical performance, sleep statistics, and even cognitive load.