I felt like I had a pretty good method for the last project so I went with the same. I started with exploration where I looked at the dataset and saw if there was anything I could see from visualizations or if there were a lot of missing values. I found that there were no missing values and only one column had a few zeros. Other than that, I could not find that much from exploration so I had to go into testing and selection somewhat blind. I played around with some different steps for feature engineering which only slightly improved the accuracy. I also used gridsearchcv but for the bigger datasets that began to be really slow and didn’t help out that much. In the end I used a function to replace the zeros in the one column with the mean of the column, I used the sklearn polynomial function to create 4th degree features, and I scaled down the data using the standard scaler. Unfortunately I couldn’t get big increases in the performance and if I had more time I would do more feature engineering and try dropping columns to reduce collinearity /do feature selection.