Chad Six

The goal of this project is to make a predictive model that can classify which round a player will get drafted or not get drafted by an NFL team according to position, height, weight, along with many other factors.

After importing the combine dataset, I had to drop some insignificant/redundant columns from the dataframe. The dataframe had a lot of 0.0 values, so I first checked to see which columns had a high percentage of 0.0 values and dropped those columns from the dataframe. In the position column, I dropped rows with values P, K, and LS, since I have no interest in analyzing those positions. Position OC and C, both are position center, so I replaced OC to be C, therefore creating centers to only be in class C. I had to do the same with position NT and DT, replace NT as DT, so defensive tackles are only classified as DT. Safeties were classified as FS and SS, I replaced both to be classified only as S. Replaced all 0 values with NaN, so calculations will exclude NaN values. Some players in the dataframe had the same name, so had to remove duplicate player names by assigning \_dp to the duplicate player name. For instance, there were two players named Brandon Moore, after removing duplicates, there is now one player named Brandon Moore and the other Brandon Moore\_dp. One issue with this dataframe is that it had nan column round for every player in year 2015. This dataframe is called nflnan.

Luckily, I found another dataset that had the rounds that all players from 2015 were drafted in, so I had to import this dataset and do a left merge with nflnan. This new dataframe is called draft2. Now nflnan has players that were drafted AND undrafted by NFL teams, so will have more players than dataframe draft2, since draft2 ONLY has players that were drafted, hence the left join with nflnan. Prior to the merge, I had to make sure there were no duplicate names in dataframe draft2 and replace duplicates with \_dp at the end of their name. I had to make sure there were no duplicate names in both nflnan and draft2 dataframes, in order to have the same number of rows as nflnan, after doing the merge. I used players name and year for the keys of the merge, which is another reason I couldn’t have duplicate names. Nflnan had 4924 rows, draft2 had 4332 rows and after merging both, had 4924 rows, success! The merge dataframe is called newdf. The newdf dataframe now has values in column round for players in year 2015, success! I then dropped column year and other insignificant columns. I still have plenty of NaN values and plan on dealing with those when some machine learning algorithms on a later date. I briefly explored the data for outliers, with box-plots, but plan on doing more later.