Can you be drafted for the NFL? Albert Men

The NFL Scouting Combine is an annual event occurring in February in Indianapolis, where college players around the U.S. compete in mental and physical competitions in front of NFL coaches, managers, and scouts. This is a chance for these football prospects to display their athleticism in hopes of getting drafted onto a professional football team. Gathering this data from https://www.kaggle.com/savvastj/nfl-combine-data, my goal is to predict whether a player can get drafted onto an NFL team, given some general descriptions and their performance at the NFL Scouting Combine. In addition, the information I've gathered can help assess what competition coaches emphasize more and how many factors coaches tend to take into account to help players improve and have a greater chance of being drafted.

I decided to select these features for my model: (1) Player Height (in.), (2) Player Weight (lbs.), (3) Forty Yard Dash Time (seconds), (4) Vertical Jump Height (in.), (5) Broad Jump Length (in.), (6) 3-Cone Drill Run Time (seconds), (7) Twenty Yard Shuttle Times (seconds), and (8) Number of Bench Reps. Features 1 and 2 are descriptions of the player. Features 3, 6, 7 measure the player's agility. Features 4, 5, and 8 measure the player's strength. Before starting my analysis, I removed 568 observations with missing values for any of the features, resulting in 1702 players' data. Next, I removed all data for the year 2018 because at the time this data was collected, the NFL Draft had not occurred. Thus, all of the players in the year 2018 were marked as not drafted, even if they had incredible athletic talent. Lastly, I decided to include a column to act as a binary classifier, where 1 represents the player was drafted, and 0 represents the player was not drafted.

Looking at **Figure 1**, we can observe the counts of the players that were and were not drafted each year. We can see that as time progresses, there is an increasing amount of players who participate in the NFL Combine, and although each year there were more people drafted, there was also an increasing amount of people not drafted. The highest proportion of people drafted was in 2007 with a rate of 88.57%. In contrast, the lowest proportion of people drafted was in 2016 with a rate of 67.85%. In more recent years, there is a steady decline in the proportion of people drafted. This information can tell us that although more players are interested in competing in the NFL, it is less likely for them to be drafted as there are a limited amount of positions on each team.

Next, I completed a principal component analysis (PCA) with the 8 features described above. I used two models, one with and without standard scaling. **Figure 2** shows the cumulative variance of these two models. Looking at the non-scaled model, it suggests that one component explains 96.78% of the variance. This is mostly due to the weight column having values in the range of 166-370 (lbs.), whereas the height contains values in the range of 65-80 (in.) and the other columns that measure the player strength and agility with even smaller variability. After applying standard scaling, we can see that now it is more difficult to reduce the dimensionality of our data, as the most significant component makes up only 70.13% of the variance.

Lastly, I created a logistic regression model with standard scaling to figure out which features had the most significant impact on determining whether a player gets drafted. The model yielded a score of 79.22%, in comparison to the same model without standard scaling that had a score of 79.58%. Although without standard scaling had a slightly better score, I decided to include standard scaling as the coefficients were more comparable. From **Figure 3**, we can see that the Forty Yard Dash Time has the most negative impact with a coefficient of -0.954, which makes sense as the faster the player is, the more athletic they possibly are and the likelier they get drafted. In fact, features 3, 6, 7 all correspond to agility and have negative coefficients, which suggests that a faster player has a better chance of getting drafted. A player's weight has the most positive impact on their draft chance with a coefficient of 1.215. I thought a surprising feature was that the broad jump had a negative coefficient, however, the coefficient for that feature was -0.135 which is comparably small and can suggest that there might not be a correlation between that feature and the chances of a player getting drafted.

In conclusion, my analysis shows that as time progresses, the proportion of players getting drafted decreases as the sport is getting more popular with a limited amount of positions to be filled. The most significant factors when determining whether a player gets drafted is their 40-yard dash and vertical jump performance, as well as their weight.

Special thanks to Professor Tyler Caraza-Harter for his guidance!

Figure 1: Yearly Report of Players Draft Results

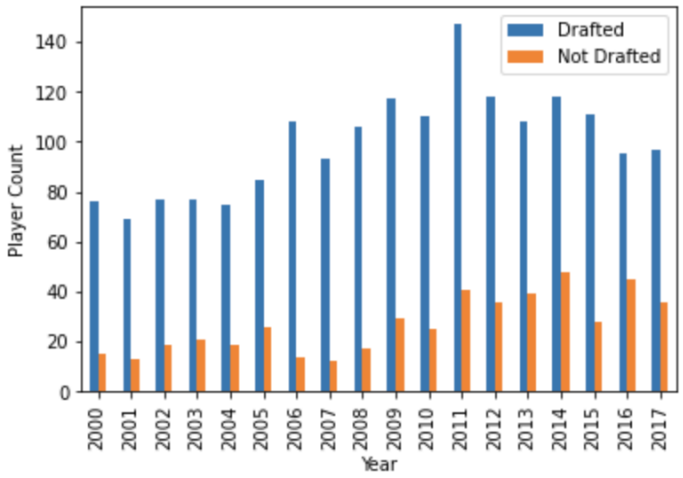

Figure 2: Cumulative Principal Components

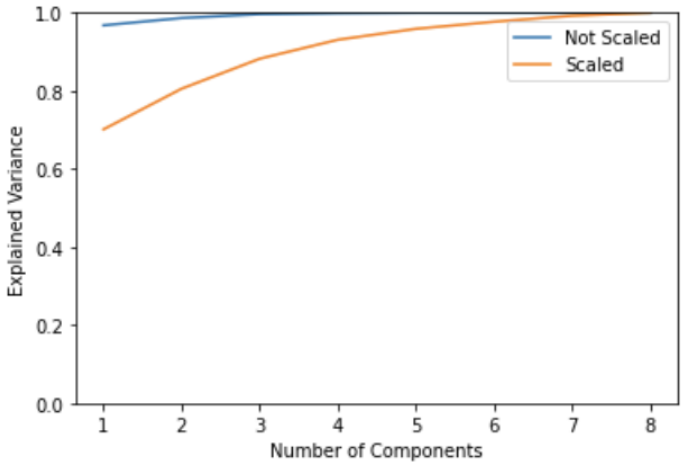

Figure 3: Logistic Regression Coefficients