

Julian Estrada  
83226902

**Final Project Proposal**

For this final project, I have decided to combine the data science principals we have learned in class this semester with my love for football. I've always been very interested in the data and analytics side of football, especially during this season of the year, in which NFL teams are looking towards the upcoming collegiate class of athletes to improve their rosters. During this time, graduating (or those leaving school early) college players are put through a series of evaluations, physical as well as mental (through interviews and meetings), to determine if they are worthy of being selected in the NFL Draft in April. My goal with this project is to create a machine learning algorithm that will determine exactly that: which players are likely to have positive returns for their future franchise. In specific, I will be focusing on the Wide Receiver position, as I myself am a collegiate Wide Receiver here at UBC. Wide Receiver is a notoriously difficult position to evaluate and project, as much of college production is influenced by external sources. For example, Receiver A and Receiver B may be two identical athletes, but Receiver A benefits from better Quarterback play, or playing in a system that passes more often, or playing weaker competition, and so forth. Thus, there is some subjectivity at play. I think this is where my knowledge of the position comes in handy: I can use this subjectivity to create more 'standard' data.

Ultimately, the end goal is to predict whether individual rookie Wide Receivers will be successful or not in the NFL. I have outlined a rough way to do this, of course many changes are in order.

I will need to combine NFL combine data with college statistics. The NFL combine is an evaluation event that tests athletes for different variables (we will only be focusing on the physical portion as the interview process is incredibly subjective and not publicly available, also it is not as important anyways). Such as 40 yard dash times, bench press repetitions, vertical jump, board jump, and so forth.

In terms of college statistics, I will need to find a way to 'standardize' this data. For example, receiving yards is, for many, the most important statistic a Wide Receiver can produce. Intuitively, this statistics measures how much yardage a Wide Receiver provided for a team. However, this is heavily influenced by the scheme of an offense. For example, team A may pass more often than team B, inflating the yardage for Wide Receivers on team A. Thus, I can standardize this data by dividing receiving yards by 30 passing attempts as an example.

Next, I will need to evaluate current NFL Wide Receivers so I can have a classifier to predict future receivers. This part requires a high degree of subjectiveness. I think I have enough knowledge of the position, paired with resources that rate players, to do this effectively. Some resources that rate NFL players are EA Sports' Madden NFL video game, Pro Football Focus, and Pro Football Reference.s From this, I can classify players as 'highly successful', 'successful', 'moderately successful', and so forth (as an example).

I will create a model using K-nearest neighbours that will attempt to accurately predict the category an individual rookie Wide Receiver will fit into. As well, I will be able to see which variables are the most correlated with success through training the classifier.

I have attached a few pages of notes that go into further detail on my rough sketch for the final project, complete with an example in a separate pdf.

Some issues of note:

1. it is difficult to determine career success when later round picks are at a disadvantage: many late round picks are not afforded the opportunities of higher selected players and thus have short, less successful careers.

2. it is difficult to evaluate success when Pro's have varying career lengths. For example, player A may have been a superior producer than player B for 5 seasons, but retired after those 5 years, whereas player B continued to produce for another 5 seasons. Who enjoyed a more 'successful' career? It is hard to say.

Potential workarounds:

1. Filter the data to only include players drafted in the first 3(ish) rounds. This would remove some very successful pro's from the model, however.

2. Only base success off a player's first 4 seasons. When an NFL team drafts a player, they are given the rights to this player for 4 seasons. After this, the player becomes a free agent and is eligible to sign with other teams, or resign with his current team of course. This is a key juncture for pro's, as those who have enjoyed success thus far are often resigned by the team, whereas those who have been disappointments are often not resigned. Thus, it is almost always true that a successful career sees valuable production within the first 4 seasons.