ISyE 6740 – Spring 2024
Project Proposal

Team Member Names: Akira Taniguchi (Team 264)

Project Title: Predicting the Success of NFL Wide Receivers Based on Their College Careers Using Classification Models.

Problem Statement

The National Football League (NFL) and the National Collegiate Athletic Association (NCAA) are the most popular American football leagues in the USA, a sport that is deeply ingrained in the country's culture and history but is relatively unknown abroad. We will refer to American football as football for the sake of brevity. A survey by Gallup in December 2023 revealed that 41% of US adults named football their favorite sport to watch. This is backed up by the fact that the NFL is the most watched professional US sports league and 19 of the 20 most-watched broadcasts of all time in the US are the Super Bowl, the annual league championship game. Meanwhile, college football is the biggest crowd puller in all global sports where no fewer than 19 university teams had bigger average home attendances in 2023 than that of the NFL, which has the highest average paying attendance out of all professional sports leagues in the world.

Unlike other professional sports leagues, the NFL does not have a minor league where players can be developed and trained until deemed ready to compete at the highest level. College football fills this role as the second highest tier of football competition in the US. The overwhelming majority of NFL players compete in the NCAA Division I before being selected in the NFL draft or being signed as undrafted free agent. Despite this, only around 1.6% of NCAA college football players are given the opportunity to play in the NFL. Even with these insurmountable odds of playing in the NFL, getting drafted doesn't guarantee a successful career in the NFL. Based on a study of 1996-2016 draft picks, only about 8% of drafted players make a difference in the NFL beyond replacement value, and around 30% see significant play time or make contributions to their teams.

Football is often regarded as one of the most complicated team sports since every player on the field has a distinct role and specialized set of skills. I will be focusing on the wide receiver position on the offense that specializes in catching the ball. Wide receivers, unlike the other primary scoring position, the running back, must earn a target by the quarterback to complete a reception. Meanwhile, whether the running back gets to run the ball is usually predetermined by the play

call. In this project, I am interested in predicting whether an NFL wide receiver will have a productive season based on data from their college careers. I will create a threshold on what constitutes a successful season for a WR in the NFL and explore the various classification models we learned in class to see if we can accurately make predictions about the success of a wide receiver in the NFL.

Methodology

The goal of this project is to determine if we can accurately predict whether a college wide receiver will have a successful season in the NFL. First, we will define what a "successful" season in the NFL as a wide receiver means. There are certain metrics such as how many receiving yards and the number of touchdowns a wide receiver has in a season that can be used to label a successful season. We can also take the metrics of successful wide receivers from the past and use them as a baseline for what we will consider a successful season. Some criteria to determine which past players were successful can be whether they won awards throughout their career for their achievements on the field, and if they've been invited to the NFL Pro Bowl or inducted into the NFL Hall of Fame. I will consider using and comparing several different labels for this classification problem.

The dataset I will be using to conduct this experiment will primarily come from Pro Football Focus (PFF), a sports analytics company that focuses on NFL and NCAA Division 1 football analysis. Advanced data on college football players began in 2014, thus this project will examine the careers of wide receivers that have played college football from 2014-2022 and NFL wide receivers from 2015-2023. I will only be considering players that appear in both datasets for the classification problem. The datasets contain 41 features, and I will consider implementing more datasets that contain features not available in the PFF dataset. These can include physical traits such as age, height, weight, wingspan and information from draft results when college football players are selected to play in the NFL.

Once I have successfully labelled the data I will begin feature selection for the classification models. I will look at the various feature selection methods we've learned in class, as well as dimensionality reduction techniques to choose what features I will use for the models. After I have decided on what features/parameters to use for the models, I will build several classification models that we have learned in class. These include but are not limited to Naïve Bayes, Logistic Regression, SVM, Neural Networks, K Nearest Neighbors. I am open to incorporating any new models or techniques that we will continue to learn in class as well. After conducting classification with the various models, I will use various methods to evaluate the results of each model. These include measuring accuracy, constructing a confusion matrix and the ROC and AUC curves.