Fantasy Football Machine Learning Project

CS 425 - Team 4

December 5th, 2021

Jared Staman

Trevor Adams

Trent Florey

Lohith Gangula

**Introduction**

For this project, we wanted to use real world data and apply it to machine learning models that we have learned in class. Because we wanted to work with data that is easily accessible along with something we find interesting, we decided to work with fantasy football data. However, there are already a multitude of fantasy football models out there that people have created, so we wanted to make sure we did something that was unique. Most of the models out there use statistics that directly correlate with fantasy scoring (yards, touchdowns, etc.) to create a predictor for a player’s next year.

We decided to look at statistics that are not a part of fantasy football scoring. Specifically, we are interested in finding what statistics have a higher correlation to a better fantasy score, and if we can use these statistics to create a relatively accurate predictor. To do this, we created four ‘.csv’ files, with each file having data for a different position: quarterback (QB), running back (RB), wide receiver (WR), and tight end (TE). We created a linear regression model and a support vector machine to train and test our data.

**Data Processing**

There are four data files: QBs.csv, RBs.csv, WRs.csv and TEs.csv. The data for these files were gathered from [www.pro-football-reference.com](http://www.pro-football-reference.com) and [www.fantasyfootballdatapros.com](http://www.fantasyfootballdatapros.com). The ‘pro-football-reference’ site gave most of the data on the statistics used for the models’ features. The ‘fantasyfootballdatapros’ site gave the fantasy scores, which is the target data for the models. The process of gathering the data was lengthy since it involved matching data from two different sites. The data scrubbing also took time, as statistics that came from players who played less than fifteen games in the season were removed to increase prediction accuracy. Below lists the statistics used for each position.

Quarterbacks: age, pass completion percentage, yards per attempt, times sacked and NFL rating.

Running backs: rushing attempts, first down rushes, longest rush, rushing yards per attempt,

number of pass targets, receptions, first down receptions and reception yards per target.

Wide receivers: targets, first downs, yards before catch, yards after catch, average depth of target, broken tackles, drop percentage, interceptions and quarterback rating.

Tight ends: snap percentage, targets, target share, catch percentage, 40-yard dash time

**ML Methods**

(describe linear regression and svm)

**Results**

(put heatmaps of each position)

(put graphs from both linear regression and svm models)