

# Term Project Proposal: Predicting NFL player scores in fantasy football PPR leagues using machine learning models

Joshua Gonsalves

December 14, 2025

## 1 Introduction and Motivation

Fantasy football is a pastime that many Americans enjoy, where you 'draft' a team of National Football League (NFL) players who score points based on their weekly game performance. Each week, you are matched up against another person's team, and whichever team's point total is higher wins for that given week.

One of the most important skills to have as a team manager is the ability to predict how well a player will perform for their upcoming game, in order to decide whether to start them or not. In this project, I would like to train a regression model to accurately predict an NFL player's fantasy points for a given game using data about their past performances, opponent defense statistics, and any other relevant factors.

## 2 Dataset

The dataset I will use for this model comes from the PyPI package `nfl_data_py`, which contains play-by-play data, weekly and seasonal data, rosters, win totals, and schedules among other things. The data it contains ranges over 1999 - present. I think that I will begin by training a model on a single position first, because different positions score points in different ways. As such, for each player at some position, there is approximately 5 - 10 seasons ( 16 weeks per season) of weekly stats. Across the quarterback position, there are probably 5000 games worth of data. Other position with more players per game will have more data. Features of the data include: player data like age and team, opponent defense data, player stats like touchdowns or yards, and game metadata like whether it was a home game. The main label for this model will be "fantasy points in PPR leagues", where points are awarded normally and for each reception.

The reason I chose this dataset is because I am personally interested. I am participating in a fantasy league, and would like to have a better way of determining who to start on a given week.

### 3 Methods to be Used

I will be focusing on linear regression techniques. I chose this algorithm because fantasy points are a continuous variable. I also think that I will use gradient descent with momentum to speedup training, and maybe experiment with a neural networks to see if they have a significant advantage over a single layer one.

### 4 Preprocessing

I think that the main preprocessing step that I will need to take is to use some sort of averaging over a player's most recent games, so that the model does not equally weigh games from 5 years ago as ones played two weeks ago. Another preprocessing step is taking care of null values in the case of injuries or bye weeks. Finally, I will need to correctly split the data into the train and test portions, which I think I will do by years.

### 5 Hyperparameters

For linear regression, I will not be using centered data, so I will need to fit intercepts. Furthermore, if I am to use neural networks, then I will likely use the ReLU activation function because of its non-vanishing gradient. Also, it has been recommended to be a good starting point for neural networks. Finally, I will start with some number of layers but test different configurations to determine which is best.

### 6 Performance Metrics

- Mean absolute error: Average prediction deviation in predicted vs actual fantasy points for a given game.
- Root Mean Squared Error: In order to penalize larger errors more.