

# Progress Report – WeHateFantasyFootball

Christian Gerber (cgerber3), Ido Tamir (itamir2)

## 1) Which tasks have been completed?

Initial coding research and implementation has been completed. We setup our project repo and set up our initial data streams and initial sentiment analysis tool.

First, we setup the fantasy football data from the ESPN API (Application Program Interface). We are pulling data for each player listed in the fantasy football database and the points stored for each week. We will be able to compare the total amount of points and projected points to the sentiment of each player.

Second, we got the text retrieval for Reddit researched and initially implemented. We found a tool called PRAW, which uses Reddit's development tools to retrieve reddit posts and the comments on them. It works surprisingly well, and as we look through the comments retrieved; we can see how the sentiment analysis will work well with the data we have.

Third, we got the initial sentiment analysis set up. We are using a library called NLTK (Natural Language Toolkit) which is suited for online discourse, and we have tested it with a few sentences.

## 2) Which tasks are pending?

The main tasks we have left are integrating all the components of the project as well as constructing the final report/presentation. We can fetch player scores, crawl through reddit, and calculate the sentiment on a particular document but now we need to combine all 3 tasks to get the sentiment for players in the league and look at their respective score. In addition, we need to store the scores as well as the sentiment of the discourse for each player in an efficient way so we can quickly retrieve and update those statistics as we keep exploring. Lastly, we will need to explore the data we generate which will involve creating visualizations and calculating regressions to measure how predictive certain measures are.

## 3) Are you facing any challenges?

One challenge that we are going to face is that we have a master player list of names of the football players, but we are going to have to match the names with the comments of the players. Players often have nicknames (such as Marquise Brown being called "Hollywood"), which we will have to deal with. We are wondering if we are going to have to use a relevance algorithm to determine if the comment/post is about a certain player.

We are wondering if another challenge will be training the sentiment analysis to deal with football terms. We are unsure if the sentiment analysis tool will do a good job with that. It will require further research.