## What to Include & What is Required?

* “Academic-like” paper with dense text inline figures, no direct dense code cells within the paper.
* Code files related to the project are included in the repository files but are self-contained. Detailed instructions are provided accordingly. No need to submit these files with the progress report, they should be in your team’s GitHub which graders have access to.
* ~~Paper should include a URL link to your project repository on GitHub in the heading~~
* Describe in depth the novelties of your approach and your initial discoveries/insights/experiments, etc. and the analysis that is still to be done to conclude your analysis
* ~~Necessary background information/framing of the problem~~
* ~~Include an overview of the problem in general as well as your planned approach (it is ok if this approach changes later in the project as you learn more information)~~
* Walkthrough your data cleaning process end-to-end
* What type of models have you used so far? Which ones remain?
* If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn’t work is also encouraged.
* Include key visuals in line with text, but always be sure to include labels, axes, captions, legends, and most importantly context!
* ~~Literature survey/references of at least 2-3 sources~~
* Works cited section.

**2020 NFL Draft Fan Sentiment Insights**

MGT 6203 Group Project Progress Report

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-79

**TEAM 79:**

Hugh Eugene Hoagland

Raju Ahmed

Pablo Francisco Ramos Soszna

Alexander Martin Stetzer

Nicholas Weist

Table of Contents

[What to Include & What is Required? 1](#_Toc149577978)

**Background:**

Due to Covid-19, the 2020 NFL draft was fully remote. In lieu of attending, fans were instead encouraged to take to the NFL and associated team subreddits to voice their opinions on their teams’ draft picks. These comments and reactions, spanning the entirety of the draft, received a sentiment score ranging from -1 (extremely negative) to 1 (extremely positive). We’re curious if NFL teams might be able to harness this wisdom of crowds to identify better draft picks.

We intend to identify patterns, trends, and sentiment shifts among fans in response to draft picks, with the

aim of assisting GMs in making more informed player selection decisions based on the sentiments expressed by the fanbase. We can then compare fan sentiment, actual draft position, and on-field results over the past three seasons to assess if the teams can learn anything from the NFL fan base.

Can fan sentiment provide any measurable benefit to the General Managers in improving draft selection as measured by a player’s success in the league?

**Approach:**

We intend to investigate whether fan sentiment can provide any additional insight into a player’s early success beyond what is understood by the General Manager. To do this we will first have to see how predictive draft position is of future success, and then see if fan sentiment offers any predictive power that deviates from the pick offered by the General Manager. We intend to investigate using a variety of models we’ve learned in this program including Linear Regression, Decision Trees, Time Series Analysis, Clustering, Classification, and Principal Component Analysis.

Linear Regression, Clustering, and Classification will help us determine the nature of the relationship between fan sentiment and player performance. Time Series Analysis can help us determine if fan sentiment is that a player was drafted at the wrong time, for example a good player who is considered to be drafted too late. Principal Component Analysis can help make sure that we are using the appropriate variables.

To prepare the data for this we will make sure that our player and sentiment scores accurately reflect what they are trying to measure. For the player score, fantasy points help measure success across different positions, but we will need to develop a more comprehensive way to measure players that do not play or do not stay in the league. We will also need to sift through the sentiment data to make sure that we’re capturing the true sentiment of the comment. Since the data is from social media and the sentiment score is based purely on the words in the comments, we will have to develop a sentiment score that accounts for what is happening in the draft, and for reactions like pictures or sarcasm that could skew the sentiment results.

With the data cleaned we will follow the process of training the various models, validating them, and testing to determine which model has the best predictive ability as measured by on field performance.

**Sources:**

There are two types of datasets from Kaggle. One, picks.csv, showing the name, timestamp, round, and team of the draft picks. The others, round\_n.csv, show the Reddit comments, username, team affiliation, timestamp, comment, and corresponding sentiment score that were posted during the draft. We hope to use these together to get a sense of how various draft picks were received by the fanbase to measure against later performance.

A white background with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

One potential way to evaluate ultimate player performance across different positions is to leverage the scoring system from fantasy football.

A screenshot of a computer

Description automatically generated

The pro-football-reference data will give us some insight into how players ultimately performed, whether or not they were on the field to generate fantasy points. It will also allow us to look at player performance from other draft years to provide context to how well they played their position against the general population of NFL talent.

A screenshot of a computer

Description automatically generated

**Works Cited:**