Monitoring Public Sentiment of NFL Draft Picks via Machine Learning Techniques

Jonathan Wiseman

Elizabethtown College

(717)857-4069

wisemanj@etown.edu

**ABSTRACT**

Sentiment analysis is a topic in natural language processing that seeks to automatically extract positive and negative polarity from text data. Its applications are diverse, ranging from marketing and sales to forum moderation to gauging public opinion. One particularly interesting application area can be found in professional sports: fans share a huge volume of opinions, predictions, and reactions online that can be used to gauge public opinion on specific entities. This paper explores the application of machine learning based sentiment analysis on a hand-labeled social media dataset focused on reacting to National Football League draft picks. The resulting model provides information for more detailed analysis, including attitude towards drafted players, comparison between fan reactions and on-field performance, and comparison between drafted players based on the language used to describe them. Additionally, a labeled dataset for sentiment analysis on professional football will be created for further use.

# INTRODUCTION

The National Football League is the world’s most profitable sports league, achieving over $13 billion in revenue in 2017 alone [#]. The league continues to expand and evolve, drawing in viewers from around the world and regularly updating safety standards, rules and regulations, and even team locations. Accompanying this expansion has been an interest in applying analytics to data generated by NFL players, coaches, and outside observers. In 2019, the NFL hosted its inaugural Big Data Bowl, challenging college and independent teams to make use of its databases to generate valuable insights about the game and its players [#]. The spirit of the Big Data Bowl reflects a growing interest in using the techniques of data analysis and machine learning to generate insights that stretch across many different sports areas.

The application areas of data analysis on professional football are diverse, ranging from a Sabermetrics-like approach to predicting game and player performances to suggesting rule and safety changes to market analysis of commercial placement and fan engagement. A large media empire has grown up around professional football, with injury reports, game predictions, and assorted player and coach news providing constant coverage on all aspects of the game.

Sentiment analysis is a field spanning the disciplines of natural language processing, machine learning, information retrieval, and text mining [#]. Its primary aim is the automatic extraction of standpoint, view, and mood of an author [#]. The most common use of sentiment analysis is determining the polarity (i.e. positive or negative) of a particular sample of text. This can be of particular use in marketing research, where companies seek to gauge public opinion of their products; other application areas include monitoring of online forums, automatically assessing product reviews, and as additional information for search engines [#].

Sentiment analysis can be performed in one of two ways. The first is a grammatical approach based on the linguistic features of text, such as descripting adjectives and adverbs, negation words (i.e. “not”), intensifiers (i.e. “very,” “extremely”), case, and tense [#]. This approach involves the creation of a carefully crafted lexicon that accurately captures the sentiment of words specific or important to a particular domain; for example, a lexicon crafted for determining sentiment in sports articles would have to assign sentiment to words like “interception” and “fumble.” The second approach involves the use of machine learning algorithms to create models that can predict the sentiment of a given text.

A general problem with sentiment analysis is its inability to generalize across multiple domains; for example, a lexicon or model crafted for use in the movie reviews domain will not generalize well to the sports domain [#]. This makes the crafting of specific lexicons time-consuming and requiring a significant amount of domain knowledge. The machine learning approach runs into similar problems: supervised classification requires carefully labeled datasets, which are often not publicly available or are based on implicit ratings (for example, movie and product reviews are standardized on a five “star” scale that gives text data implicit ratings). Either approach requires a significant investment in either crafting a lexicon or acquiring a significant dataset that captures the nuances of a given application field.

Any potential use of sentiment analysis on NFL articles must be performed with a specific goal in mind. Using news articles to predict the outcome of NFL games is problematic. For one, most articles are not specific to one aspect of the game: there are injury reports; news and updates on trades, signings and draft prospects; articles about players’ personal lives; and news about retired players and coaches that are no longer active in the game. Each of these areas requires a specific lexicon, and it is doubtful that each is useful in predicting the outcome of a specific game. Secondly, each article deals with multiple players and topics, such that extracting entity-based sentiment is difficult. For example, one sentence in an article might deal with an offensive and a defensive player. This makes sentiment analysis difficult, since phrasal extraction is a difficult area of natural language processing [#]; additionally, this requires a model that is capable of orienting sentiment-bearing words to specific players (i.e. an interception is bad for an offensive player but good for a defensive player).

It is clear that any sentiment model based on football text must be directed and purposeful. One particularly useful application area is determining public sentiment around NFL draft picks. The NFL draft is an annual event in which college football players are selected by professional teams for short-term “rookie” contracts [#]; it is the primary mechanism by which college talent enters the NFL. This task is useful for several key reasons. For one, high-valued draft picks (i.e. those selected in the early rounds of the draft) are expected to be polished, capable players. Although rookie contracts are generally inexpensive compared to those for veteran players [#], teams wish to avoid selecting players whose draft stock does not translate well into actual on-field performance. Creating a model to process text data related to draft picks is a useful analytics tool for gauging expert and public opinion towards a player’s potential. Secondly, gauging sentiment towards a player is useful from a marketing perspective. The off-field (and sometimes on-field) actions of a player influence fans’ perspectives of players and their willingness to engage with the franchises to which they belong. For example, the impact of on-field protests by NFL players such as Colin Kaepernick on NFL revenues is examined in [#]; for an example of a player’s actions harming team reputation, see the example of Antonio Brown in [#].

To gauge public reactions to NFL draft picks, I propose *DraftSense*, a machine learning approach to sentiment analysis on text relating to draft picks after they are made. The key design goals of DraftSense are:

**Comprehensive:** the ability to collect a large volume of data

**Specific:** collecting data specific to NFL draft picks

**Accurate:** accurately predict sentiment to summarize the public’s reactions to NFL draft picks

# BACKGROUND

## Related Work

### Aggregate Forecasting

It has been consistently observed that aggregating a number of individual forecasts performs better over time than relying on a single forecast [#].

In [#], this principle was applied in the sphere of politics and international events by the Good Judgment Project and tested over time in the U.S. Intelligence Advanced Research Projects Activity’s Aggregative Contingent Estimation program [#]. The team made probabilistic judgements about specific events (i.e. Greece leaving the Eurozone) by framing them as yes-no questions and presenting them to a poll of 2400 Americans from myriad demographics and professions [#]. The team employed various aggregation techniques ranging from simple averaging to log-odds extremizing of weighted averages [#]. Overall, their methods outperformed U.S. intelligence community predictions by about 30%, even when intelligence officials were given access to classified material [#].

The work carried out by the Good Judgment Project presents several interesting findings. Chief among these is the idea that combining individual predictions (as biased and perhaps ill-informed as they may be) outperforms the singular opinion of an expert. This means that many opinions of perhaps lower quality can be used to obtain a fairly reasonable predictor of future events. It also suggests that social media, where opinions are clear and abundant, might be able to provide a good source of material for making predictions.

Secondly, the attempt to quantify the outcome of events as binary allows one to frame problems as questions of classification. This brings complex events into the realm of prediction, ignoring any potential nuance in favor of a quantifiable outcome. For evaluating NFL draft picks, the question now becomes simple: was the choice to draft player X a good choice?

Finally, the Good Judgment Project utilized a number of different aggregation methods. This makes it possible to break *DraftSense* into two distinct components: one for analyzing sentiment and one for aggregating predictions.

However, the work presented in [#] suffers from a few drawbacks that limit its overall effectiveness. Its primary weakness is its reliance on polls to produce predictions. Sending out a poll for every question that needs answering can be time consuming, expensive, and produce biased results. Here, the volume of data available on the Internet to be collected by *DraftSense* can help increase speed and scale. Rather than waiting for thousands of individual polls to be answered and returned, *DraftSense* can scrape social media posts for predictions focused on specific players.

### Sentiment Analysis on NFL Data

There are two existing projects utilizing sentiment analysis to make predictions on NFL games: *Lydia*, developed by Hong and Skiena in [#], and the work of Sinha et al. in [#].

In *Lydia*, a lexical approach to sentiment analysis was applied to text data from news, blog, and other web sources in order to produce a betting paradigm for NFL games. The favorability of a team is derived from its daily positive and negative mentions in the text dataset [#]. Utilizing sentiment alone, the authors achieved 60% prediction accuracy for the 2006-2008 seasons [#]. The authors found that combining sentiment, statistical performance prior to games, and home field advantage produced the most robust model; however, the authors note that the sentiment model only produced significant improvements over the second half of the NFL season [#].

*Lydia* offers a generic framework for how social media data can be used to predict world-world events. The production of raw positive and negative mentions, and their aggregation, is a simple and intuitive approach to deriving general feeling towards a team. However, it suffers from being far too general for practical use. For example, there is no filtering performed on any of the data being scraped. That means that injury reports, coach news, historical articles (ex. recapping last week’s game), and more are all included in the raw counts. Secondly, *Lydia* was not developed specifically for analyzing sentiment in sports (and not specifically for American football). Its lexicon-based analysis of sports articles is thus questionable. Finally, *Lydia* was used as part of a prediction related to betting lines. This significantly hampers its scope: rather than predicting game outcomes themselves, *Lydia* is used to predict when to bet against the odds.

*DraftSense* improves upon these limitations by scraping comments from player-specific Reddit threads. For example, all of the comments scraped from the Patrick Mahomes thread are related to Patrick Mahomes and his selection by the Kansas City Chiefs; thus, there is no extraneous information included. Secondly, *DraftSense* is trained on a dataset specifically focused on football. Finally, *DraftSense* avoids predicting betting lines and instead focuses on evaluating public opinion.

The work of Sinha et al. in [#] represents another significant inspiration for *DraftSense*. Here, the authors utilize a simple lexicon-based sentiment analysis on Tweets to predict game outcomes. Their approach follows the general logic of *DraftSense*: aggregating social media opinions to produce a forecast. Additionally, the authors combine their text analytics with traditional game statistics (such as a team’s win/loss record) to increase accuracy, much like the designers of *Lydia*.

However, the authors of [#] make no attempt to increase scale or speed. There is no component to automatically collect and analyze data, with Tweets needing hand-labeling for effective analysis. One of the major goals of *DraftSense* is its training on a comment dataset so as to be able to automatically analyze a huge volume of data at high speeds. Finally, the authors make no further use of their text data beyond attempting to beat the bookies over/under line. There is no attempt to track sentiment, trending topics, or compare teams over the course of the season. *DraftSense* will utilize sentence embeddings in order to provide direct comparisons of player similarity. This has applications beyond prediction, such as visualization of the language used to discuss a player or attempting to find a correlation between a player’s attributes (i.e. a good arm or fast run speed) and their performance in the league after their draft.

# REFERENCES