Tweets to Touchdowns: Predicting NFL Achievement from Social Media Optimism

Rohan Erasala, Ian McCulloh

***Abstract*—** The NFL Draft is a chance for every NFL team to select their next superstar. As a result, teams heavily invest in scouting, and millions of fans partake in the online discourse surrounding the draft. This paper investigates the potential correlations between positive sentiment in individual draft selection threads from the subreddit r/NFL and if this data can be used to make successful player recommendations. It is hypothesized that there will be limited correlations and nonviable recommendations made from these threads. The hypothesis is tested using sentiment analysis of draft thread comments and analyzing correlation and precision at *k* of top scores. The results indicate weak correlations between the percentage of positive comments in a draft selection thread and a player's approximate value, but potentially viable recommendations from looking at players whose draft selection threads have the highest percentage of positive comments.

***Keywords*—** National Football League, NFL, NFL Draft, sentiment analysis, Reddit, social media, NLP, sentiment analysis

I. Introduction

THE National Football League (NFL) is the most popular sports league in the United States of America, followed by over 50% of all sports fans in the United States of America today [1]. Every year, millions of fans also tune into the NFL Draft - a process that allows teams to select their future stars from the eligible pool of college football players [2]. The NFL Draft has seven rounds, and each of the 32 NFL teams has one pick per round. The NFL Draft represents many teams' best opportunity to acquire young players who contribute to the team, increase ticket sales, and increase overall team success. Accordingly, teams have advanced scouting departments that spend hundreds of thousands of dollars conducting private workouts, tests, visits, and research into what players will be successful in the NFL [3], [4], [5].

The rise of sports betting in the National Football League (NFL) has created financial interest in the NFL draft. This phenomenon can be traced back to the legal shifts that occurred in the United States following the Supreme Court's decision in Murphy v. National Collegiate Athletic Association in 2018, which overturned the Professional and Amateur Sports Protection Act of 1992 (PASPA). This landmark ruling effectively allowed states to legalize sports betting, leading to a rapid expansion of legal sports gambling across the nation. The NFL, once a staunch opponent of sports betting, adapted to this new reality by forming partnerships with betting companies and integrating betting content into its broadcasts and digital platforms. This integration has not only altered the way fans engage with the sport but has also opened new revenue streams for the league. However, it has also raised concerns about the potential impact on the integrity of the game and the well-being of its players and fans [6].

In the context of the NFL, the proliferation of sports betting has had multifaceted implications. On one hand, it has increased fan engagement, as betting allows fans to have a financial stake in the outcomes of games, thus intensifying their overall experience. On the other hand, it has raised ethical and regulatory questions, particularly regarding the safeguarding of the sport's integrity and the prevention of gambling-related harm among vulnerable populations. Additionally, the NFL's embrace of sports betting has necessitated the development of new policies and educational programs aimed at ensuring responsible gambling practices among both players and fans [7]. The ongoing evolution of this relationship between the NFL and sports betting presents a dynamic area of study, with implications for sports management, marketing, and policy.

The intersection of NFL fantasy football and its impact on sports betting represents a compelling area of study within the broader context of sports economics and consumer behavior. Fantasy football, a game where participants act as owners to build a team that competes against others based on statistical performances of real NFL players, has seen exponential growth over the past two decades. This growth has been paralleled by a significant increase in sports betting, especially following the legalization of sports gambling in many U.S. states. The convergence of fantasy football and sports betting has resulted in a unique dynamic where the two industries both compete and complement each other [8].

From a behavioral standpoint, fantasy football has been shown to influence betting patterns among its participants. According to research by Drayer and Rascher [9], individuals engaged in fantasy sports are more likely to participate in sports betting. This correlation can be attributed to the similar skill sets used in both activities, such as statistical analysis and player performance tracking. Moreover, fantasy sports have been credited with enhancing fan engagement and viewership, indirectly influencing betting behaviors by increasing knowledge and interest in specific games or players. This enhanced engagement can lead to a deeper understanding of the sport, potentially giving bettors an edge in their wagering decisions. However, this synergy also raises concerns regarding problem gambling and the need for robust regulatory frameworks to ensure responsible gambling practices [10].

Given the popularity of the NFL, the rise in sports betting, and the importance of the NFL Draft, millions of people engage in online conversations about future NFL players and guessing who will be successful. On the day of the first round of the 2020 draft, X (formerly Twitter) received 4 million posts/tweets related to the draft [11]. A few minutes on any social media platform such as X, Reddit, or Facebook will show someone how many people voice their opinions on team and player performance and how many strongly believe those opinions. These posts garner thousands of views and interactions, and these numbers increase every day. The goal of this paper is to understand if the sentiment of online public opinion has value when projecting the success of NFL players. This is investigated with the following research questions:

• RQ1: Are there correlations between the positivity in social media's reaction to a draft selection and player performance?

• RQ2: Can we use social media's reaction to a draft selection as a recommendation of a future asset to a team?

It is hypothesized that there will be limited correlations between the positivity of social media's reaction to a draft selection and player performance, and recommendations based on these reactions will be poor.

II. Background

A. Related Work

There have been several studies done to predict player and team performance in sports. Lutz [12] investigated the prediction of NFL player performance specifically for fantasy football using statistical techniques to analyze player data. He offers a robust model for forecasting future performance based on historical data. Others [13] offer a unique research angle by considering the influence of team context on individual player performance. Their method provides a more nuanced understanding of player performance within the context of team dynamics. King [14] focuses on the quarterback position to analyze decision-making patterns and performance. His work is particularly relevant for understanding the complexities of predicting performance in key positions. Lyons et al [15] found that past performance was a better prediction of NFL player performance than physical ability. Unfortunately, these models do not perform well for new players entering the NFL, since they all require historic data. Furthermore, they do not necessarily benefit from crowd wisdom that may be able to perceive additional context that has not otherwise been captured in data.

A study by Howe [16] highlights the democratization of data collection through crowdsourcing, emphasizing its potential to gather vast and varied inputs that traditional methods might overlook. This inclusive approach not only enriches the dataset but also fosters a more holistic understanding of the subject matter. Similarly, Brabham [17] discusses the application of crowdsourcing in problem-solving, where the aggregation of diverse viewpoints can lead to innovative solutions and the identification of previously unnoticed features.

The potential of social media as a platform for crowdsourcing predictive insights into sports player performance has garnered increasing interest in academic circles. Social media platforms, with their vast user bases and diverse demographic profiles, offer a rich, real-time source of data and opinions that can be harnessed for predictive analytics in sports. A study by Clavio and Kian [18] highlights the depth and breadth of sports-related discussions on social media platforms, suggesting their utility as a crowdsourcing tool for gauging public sentiment and opinions about player performances. Similarly, Hambrick [19] notes the propensity of sports fans to share detailed observations and analyses on players and games, which, when aggregated, could provide valuable insights for predictive modeling. This is further supported by Xu et al. [20], who demonstrate the feasibility of using social media data to predict outcomes in sports events, underscoring its potential in assessing individual player performances as well. Reed et al [21] found that social media networks of discourse between NBA players were a useful proxy for team chemistry and correlated with performance on the court.

Several researchers have investigated the potential of valuing social media opinions in sports. Wiseman [22] created a machine-learning model called DraftSense, aimed at monitoring public sentiment around a draft pick using comments from Reddit. Wiseman was able to achieve 84% accuracy with a support vector machine model trained to categorize comments as positive, negative, or jokes. This attests to viability of understanding sentiment from discussions about the NFL Draft. Silva [23] conducted a study that analyzed the semantic similarity between comments in live game threads of soccer matches and match commentary from reporters, as well as player ratings from the matches played. Silva also conducted sentiment analysis of comments in game threads to see what conclusions could be drawn. They found that comments from Reddit were semantically similar to match commentary and that positive sentiment in comments was correlated with the team that was currently performing well. Sinha et. al [24] studied the relationship between social media output and NFL results using posts from X and statistics from the games. The researchers trained a logistic regression classifier with tweets before, during, and after the game. With this model, they predicted the winner of a game with 63.8% accuracy during the 2012 NFL season. Both of these studies endorse the value of public opinion when projecting player and team performance. This study aims to build on these analyses by investigating the viability of using sentiment analysis of reactions to draft selections to project player performance and recommend successful players.

B. Sentiment Analysis

Sentiment analysis is the process of classifying emotions, opinions, and sentiments in textual data [25], [26]. The goal of sentiment analysis is to provide a way for computers to understand the human emotion behind public opinion, perceptions, and sociological insights. Pieces of text are classified into different categories such as ”negative”, ”positive”, ”neutral”, ”angry”, ”sad”, etc. as specified by the model and training data.

C. Approximate Value

One of the challenges of analyzing player performance in a team sport is the difficulty of measuring a player's individual impact. This is especially difficult in football because for every position different statistics are measured (ex. defensive players make tackles, but offensive players gain yards). To address this, the performance metric for players in this study is the approximate value statistic [27]. This statistic, invented by Pro Football Reference founder Doug Drinen, aims to quantify the seasonal value of a player regardless of position by using a unique combination of awards, individual statistics, and team success [27]. This is one of the few publicly available statistics that allows comparison of football players across positions, making it the ideal choice for this study.

III. Dataset

This work leverages the use of data available on Reddit, an online community platform with millions of users. All comments were collected from the r/NFL subreddit, a subreddit (forum for a specific topic on Reddit) dedicated to discussion of events in the NFL with over 6 million users. Moderators of r/NFL have created threads for users to discuss and react to draft selections since 2017. The dataset for this study consists of comments from these threads for first-round draft selections from 2017 and first and second-round draft selections from 2018 - 2021. This represents 289 different players/threads and a total of 40,522 comments. The data was collected using the Python Reddit API Wrapper (PRAW), a Python package that allows access to Reddit's API. All data was collected from October 3rd, 2023 to December 4th, 2023.

This work also leverages the use of data available on Stathead Football, a reference site that tracks professional and collegiate football statistics. All data related to NFL players/draftees including name, draft year, position, and approximate value was gathered using this website. Data was aggregated on November 30th, 2023.

IV. Methods

This study used sentiment analysis to analyze the comment dataset. After cleaning, stemming, lemmatizing, and processing each comment, it was passed into a pre-trained language model from HuggingFace, BERTweet [28]. This model was selected based on the extensive size of the training corpus (over 900 million tweets in total) and because it was trained on data from social media (Twitter, formerly X) [11], [24]. Each comment was assigned one of three labels: negative, neutral, or positive, as well as a floating-point score representing the confidence in the label. The number of comments in each category (negative, neutral, or positive) in each thread was recorded in the data table. The percentage of positive comments in a player's draft selection thread was measured against that player's approximate value to determine potential correlations, and the precision at the top k (k = 5, 10, 15, 25, 50, 100) threads with the most positive comments was measured to determine possible recommendations. Some players receive more attention on social media than others. It is therefore possible that only players with a sufficiently high volume of discourse will present enough data for prediction.

V. Results

A. Sentiment and Player Performance Correlations

Correlation was determined by plotting the percentage of positive comments in each thread against the performance metric (approximate value). From this, visual and quantitative observations were made from the resulting trend line and R-squared value of the trend line. Below are the results of this investigation.

Positive sentiment was plotted against a player's approximate value per game that he played to control for player injuries and other factors that may have limited a player's ability to spend time on the field. Below are the results of this investigation.

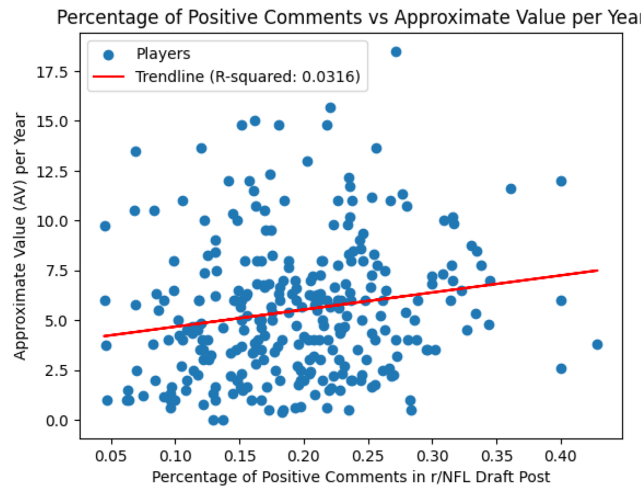

Fig. 1 Percentage of positive comments in each player's r/NFL draft selection thread against each player's approximate value (AV) per year

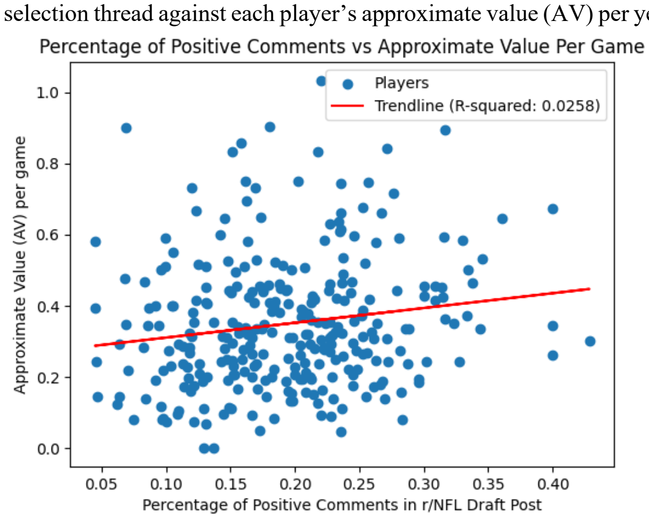

Fig. 2. Percentage of positive comments in each player's r/NFL draft selection thread vs each player's approximate value (AV) per game.

The R-squared value for both trendlines in these graphs is < (0.05), indicating an extremely weak correlation between the percentage of positive comments in a player's draft thread and the player's approximate value. These results are consistent with the hypothesis that the correlation between positive comment sentiment in draft threads and player performance is minimal. This lack of correlation is readily explained by the fact that most people on social media are not professional football players, nor do they have experience with professional scouting. They also do not have the same access to player interviews, workouts, film, coaches, and other resources that give valuable information about players.

B. Sentiment as a Recommendation of Player Performance

Potential recommendations of players were aggregated by analyzing the precision at *k* for the top *k* draft thread comment percentages. Recommendations were made according to each player's approximate value per year compared to the median approximate value for a player in the dataset. If the player's approximate value was higher than the median approximate value of that group, it was labeled as a good recommendation. Therefore, the expected precision at *k* for all measures is 0.5. Below is the result of this investigation for first-round selection (R1), second-round selections (R2), and all selections (R1 + R2).

Given that all measures are above or at 0.5, this indicates that the top *k* recommendations from positive comment sentiment could be a viable recommendation system for successful players on an NFL team, with higher confidence from looking at the top 5-10 recommendations. There are multiple possible explanations for this observation. It can be argued that if social media users react overwhelmingly positively to a draft selection, there may be some tangible ability users can identify in the games that they watch. This could also mean that players are more likely to be successful in the NFL if they have the support and belief of social media users.

TABLE I
PRECISION AT K FOR R1, R2, AND R1 + R2 SELECTIONS

Measures	**R1**	**R2**	**R1+R2**
P@5	0.80	0.60	0.60
P@10	0.80	0.70	0.70
P@25	0.64	0.56	0.72
P@50	0.62	0.50	0.60
P@100	0.54	0.52	0.59

VI. CONCLUSION

The goal of this paper was to investigate potential correlations between positive comment sentiment in draft selection threads from r/NFL and player performance, as well as understand if any recommendations can be made from the analysis of positive comment sentiment. The hypothesis was that there are limited correlations between positive comment sentiment in draft threads and player performance and that there would not be viable recommendations. This hypothesis was confirmed for the first research question, but not confirmed for the second research question. The results of the investigation into potential correlations yielded extremely weak R-squared values (< 0.05) when plotted against each other. We can safely conclude that this sentiment analysis showed no correlation with player performance. This indicates that the collective opinion of social media is no more accurate than most. In the investigation into viable recommendations, valuable recommendations were identified by looking at the top 5-10 highest percentages of positive comments in a player's draft selection thread. Possible explanations may be that players that have the belief of social media users perform better, or that a strong collective public opinion has value when projecting player performance.

This paper only focused on raw sentiment analysis of r/NFL draft selection threads. Future work that builds on this study could focus on developing advanced natural language processing tools that extract greater meaning from comments than raw sentiment. It is important to remember that the careers of many players in this study are not over, and there are many players who may not currently be in the best situation to succeed. Future work could continue to monitor these players' careers and analyze how correlations and recommendations may change over time.

By focusing on the top 25% of R1 + R2 discussion threads, precision is higher than chance and may provide an effective measure for estimating rookie performance. Many draftees do not have enough discussion to provide any predictive value. Perhaps combining social media related measures with other measures of physical performance will allow even more powerful composite measures. The fact that social media discourse provides value in estimating NFL player performance suggests that there is overlooked value in the wisdom of the fan base.

REFERENCES

[1] Richter, F. (2022, September 6). Infographic: Americans love the NFL, but change is looming. Statista Daily Data. https://www.statista.com/chart/15869/favrotie-sports-league/

[2] Lewis,J.(2023,April28).NFLdraftviewershipreboundsfromlastyear. Sports Media Watch. https://www.sportsmediawatch.com/2023/04/nfl-draft-viewership-increase-espn-abc-nfl-network-first-round/

[3] Kraeutler, M. J., Carver, T. J., Belk, J. W., & McCarty, E. C. (2018). What is the value of a national football league draft pick? An analysis based on changes made in the collective bargaining agreement. The Journal of Strength & Conditioning Research, 32(6), 1656-1661.

[4] Johnson, R. (2018, April 12). How much does it cost a team to scout a top NFL prospect?. SBNation.com. https://www.sbnation.com/college-football/2018/4/12/17193092/cost-to-scout-an-nfl-prospect

[5] Rizzuti, A. (2023, July 12). Bryce Young remains amongst NFL's top 10 jersey sellers through June. USA Today. https://pantherswire.usatoday.com/lists/nfl-jersey-sales-top-10-june-bryce-young-tom-brady-patrick-mahomes-aaron-rodgers/

[6] Lopez-Gonzalez, H., & Griffiths, M. D. (2018). Betting, forex trading, and fantasy gaming sponsorships—a responsible marketing inquiry into the 'gamblification' of English football. International Journal of Mental Health and Addiction, 16(2), 404-419.

[7] Humphreys, B. R., & Perez, L. (2020). Sports betting, sports bettors and sports gambling policy. Sport Management Review, 23(1), 1-7.

[8] Martin, R. J., & Nelson, S. (2020). Fantasy sports, real money: Exploration of the relationship between fantasy sports participation and sports betting. Journal of Gambling Studies, 36(1), 193-209.

[9] Drayer, J., & Rascher, D. A. (2019). Fantasy football's potential influence on NFL fan engagement and sports betting decisions. Sport Management Review, 22(3), 350-362.

[10] Weiss, M. B., & Loeb, S. E. (2021). The impact of fantasy sports on television viewership and online sports betting. Journal of Media Economics, 34(1), 26-45.

[11] Cohen, D. (2020, April 24). Sports twitter awakens during round 1 of the 2020 NFL draft. Adweek. https://www.adweek.com/performance-marketing/sports-twitter-awakens-during-round-1-of-the-2020-nfl-draft/

[12] Lutz, R. (2015). Fantasy football prediction. arXiv preprint arXiv:1505.06918.

[13] Whiting, S. W., & Maynes, T. D. (2016). Selecting team players: Considering the impact of contextual performance and workplace

deviance on selection decisions in the National Football League. Journal of Applied Psychology, 101(4), 484.
[14] King, N. A. (2017). Predicting a Quarterback's Fantasy Football Point Output for Daily Fantasy Sports Using Statistical Models (Doctoral dissertation). Arlington, TX: University of Texas at Arlington
[15] Lyons, B. D., Hoffman, B. J., Michel, J. W., & Williams, K. J. (2011). On the predictive efficiency of past performance and physical ability: The case of the National Football League. Human Performance, 24(2), 158-172.
[16] Howe, J. (2006). The rise of crowdsourcing. Wired magazine, 14(6), 1-4.
[17] Brabham, D. C. (2008). Crowdsourcing as a model for problem s
[18] Clavio, G., & Kian, E. M. (2010). Uses and gratifications of a retired female athlete's Twitter followers. International Journal of Sport Communication, 3(4), 485-500.
[19] Hambrick, M. E. (2017). Social media in sports: A narrative review of research on social media in sport and physical activity. The Physical Educator, 74(3), 439.
[20] Xu, Z., Frankwick, G. L., & Ramirez, E. (2019). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. Journal of Business Research, 94, 61-74.
[21] Reed, A., Piorkowski, J., & McCulloh, I. (2018). Correlating NBA team network centrality measures with game performance. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1291-1294). IEEE.
[22] Wiseman, J. (2020). Monitoring Public Sentiment of NFL Draft Picks via Machine Learning Techniques.
[23] Silva, G. H. D. S. (2021). Football players performance analysis and formal/informal media: Sentiment analysis and semantic similarity (Master’s thesis).
[24] Sinha, S., Dyer, C., Gimpel, K., & Smith, N. A. (2013). Predicting the NFL using Twitter. arXiv preprint arXiv:1310.6998.
[25] Perez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., & Mart ́ınez, M. V. (2023). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks.
[26] Taboada, M. (2016). Sentiment analysis: An overview from linguistics. Annual Review of Linguistics, 2, 325-347.
[27] Approximate value. Pro Football Reference. (n.d.). https://www.pro-football-reference.com/about/approximate value.htm
[28] Nguyen, D., Vu, T., & Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 9–14). Association for Computational Linguistics.