2023

Football By the Numbers: A Look Into Sports Analytics Currently Used in the National Football League

Christopher Schoborg
University of Central Florida

FOOTBALL BY THE NUMBERS: A LOOK INTO SPORTS ANALYTICS CURRENTLY USED IN THE NATIONAL FOOTBALL LEAGUE

by

CHRIS SCHOBORG
B.S. Florida State University, 2018

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Statistics and Data Science
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2023

ABSTRACT

Sports analytics is a fast-growing field of analytics. In particular, sports analytics with a focus on National Football League (NFL). In this thesis, we will review many articles on football analytics to have an in-depth understanding of the current stat of football analytics. In addition, we can learn from past research to identify interesting research direction to advance sports analytics with a focus on football analytics. In this thesis, we have carefully examined all current analytical results in the following fields: current state of football analytics, analytics regarding the draft, analytics for wide receivers as well as offensive linemen, analytics on other offensive positions, and we have identified the following research direction: the need for a scale rating system that is equal of all positions but unique to expectations of that position especially when it comes to wide receivers and offensive linemen. Lastly, we lay the groundwork for future work, which will make use of the following statistical learning algorithms: logistic regression, XG Boost, decision trees, and time series, to analyze the NFL data, both tracking data from the first six weeks of the 2020 season as well as play by play data from 1999 to 2022 to introduce these new algorithms to sports analytics community.

Key Words: Sports Analytics, NFL Football League, Data Science, Statistical Learning, Machine Learning, and AI.

ACKNOWLEDGMENTS

I would like to thank my committee members, Chung-Ching Wang, Ph.D., Liqiang Ni, Ph.D., and C. Keith Harrison, Ed.D. Dr. Wang, my gratitude for all of your help in getting this thesis put together cannot be fully expressed on paper. You have helped me greatly from the initial idea for this thesis all the way through. Dr. Ni, your advice and guidance early on as we were discussing the possibilities of things to consider have helped shape this thesis. Dr. Harrison, you have truly been an inspiration to me. Without your class and your guidance, this thesis would not have been possible; I am eternally grateful for all the mentorship that you continue to provide to me.

I would also like to thank my parents, Terry and Faye; the two of you first introduced me to sports and helped develop my love of football with all those Sundays watching games with the two of you. To Ann and Dave, the two of your support throughout my journey have been amazing. I cannot thank you enough for all that you have done along this journey to help keep me on task and sane. Finally, to my wife, Lisa, you are my everything; your unyielding support through the tough times has been recognized and appreciated. Your sacrifice of making the four-hour drive from home to Orlando to provide me more time for my studies and writing was always seen by me, even if I did not always acknowledge it. You are the rock that I leaned on, the gentle hand supporting me when I needed it, and the big kick in the butt to remotivate me at those times I was not sure I was going to get to this moment.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I: INTRODUCTION TO SPORTS ANALYTICS

In 2011 Brad Pitt helped introduce the United States to a concept many sports fans had known many years before, Moneyball. The movie Moneyball was based on a real story of how Billy Beane used data and statistics to develop a method for identifying what players may be undervalued and, therefore, could be signed to the Oakland Athletics baseball team and make them a successful team while keeping their payroll low as possible. Thanks largely to the movie, Moneyball is one of the most popular movies on sports analytics. However, those who understand statistics and love sports had used sports analytics for years before Billy Beane developed Moneyball. This master's thesis will look into the current literature and methods in sports analytics with a significant emphasis on the National Football League, NFL, and the sports analytics written about in journals for this particular sport.

This chapter will introduce the world of sports analytics. First, a definition of what is meant by sports analytics for this thesis (Section 1.1); then, the major contributors to the world of sports analytics will be introduced (Section 1.2). Lastly, why sports analytics is essential and what we hope to learn about sports, and how these techniques could be utilized in areas outside of sports (Section 1.3).

1.1 Sports Analytics Defined

A ChatGTP search for "sports analytics definition" will get the following response "Sports analytics is the use of data and analytics to better understand and improve the performance of athletes, teams, and leagues. Through the use of data, analytics can provide insights into a team's performance, help identify areas for improvement, create competitive advantages, and improve the fan experience.". However, we can also use Google search to get multiple responses. Indeed.com defines sports analytics as "Sports analytics is the analysis of sports data, including components of sports such as player performance, business operations, and recruitment" (Indeed.com, 2023). For this thesis, we will define sports analytics simply as using numbers and statistics to produce a method for describing or making a prediction in a particular sport. While talking on a panel at the 2023 Sloan Conference, Dr. Broadie of Columbia University said, "sports analytics can be divided into one of three buckets: the first is performance, the second in-game decision, the third improving performance" (Baldwin et al., 2023). It has been suggested that a fourth of storytelling could also be lumped into performance. Let us look at the three buckets and describe some of the work.

The first and probably oldest is performance; we have been describing performance in simple terms for almost as long as sports have been played. We calculate batting averages and free throw percentages to more advanced statistics such as baseball's wins above replacement, WAR (Baumer et al., 2015) which looks at how the performance of one player relates to other players that could potentially play that position on the same team. Next is looking at game decisions, which are a little more advanced such as when to foul in late-game situations in basketball (Annis, 2006). All sports have multiple papers written on game decisions. The last is improving performance, mainly using case studies to determine if a change in a factor can improve performance. One example is seeing if the Ketogenic diet can improve physical performance in endurance sports (Phinney, 2004).

1.2 Major Contributors in Sports Analytics Literature

Trying to identify important literature in sports analytics would prove to be difficult at best. Just looking at journals available through the University of Central Florida's Library, nine academic journals have sports analytics in their title. So instead of determining what is essential literature, this thesis will explore some of the publishing of the more prominent names in sports analytics. The authors listed below have been identified by using www.researchrabittapp.com as well as google scholar h-index. It includes the authors who had a large number of publications as well as a high h-index. The first such author is Jim Albert, who has been cited over 2,000 times since 2018, according to Google Scholar. One of his more cited papers is his work evaluating hitting streaks in baseball (Albert, 1993). The next author is Benjamin C Alamar, who has created two sports analytics journals and written three different books on the subject of sports analytics (B. Alamar, 2013). The last for this thesis is Nick Elam, who single created a way of ending basketball games that awards consistency throughout an entire game known as Elam Ending, which is used in many tournaments worldwide for determining the winner, including the NBA All-Star Game (Elam, 2020). Countless others are doing great work in sports analytics. Also, many sports analysts in the corporate world are doing great work but not publishing in academic journals. This thesis will introduce some people publishing in sports analytics, specifically in football and the NFL.

1.3 Why Sports Analytics Matter

Sports analytics uses of sports are well-known and understood. This chapter has already talked about some essential examples of sports, and the rest of this thesis will delve more into sports analytics and how they are used to improve sports. However, what can be learned from sports analytics for those with no interest in sports? The answer is a lot, and there are many cases where sports analytics has helped the world outside of sports.

The first example of this is auto racing. Auto racing uses sports analytics to make their race cars faster, turn better, more fuel efficient, and safer. They then use the results to evaluate whether their hypothesis is correct by implementing it into their race cars. Then they utilize the new configuration to collect data and test their hypothesis. Then once changes have been shown to improve their race cars, these changes get transferred into the production vehicles made for the general public to purchase in their cars. Some examples are seat belts, the safety cage under your car's interior, power steering, anti-lock brakes, and push-button starting systems(Deaton, 2008).

Many sports teams use business analytics to determine what price to charge for tickets based on many factors, including the opponent they are playing, which is called dynamic pricing. Dynamic pricing and how to determine the price dependent on variables has now spread to restaurants, movie theaters, and the performing arts.

Some bestselling self-help books are made possible by the use of sports analytics. The New York Times bestseller "Nudge" (Thaler & Sunstein, 2009) resulted from the authors' work in sports psychology and translating what they learned to make athletes better into a self-help book to help everyday people. As you read this, you may even be benefiting or utilizing a popular advancement that came from a need for sports analytics teams to gather data from their athletes. That is the wrist-based heart rate and activity tracker that is built into many wristwatches these days from companies like Apple, Garmin, and Fitbit.

Even the world of business and economics has had improvements from sports. One example is work done looking at team scouting and a return on investment in the National Hockey League, NHL, which concludes with how these lessons can be translated to business and hiring professionals when attempting to make decisions about whom to hire for a company's open positions(M. E. Schuckers & Argeris, 2015).

Now that a working definition of sports analytics has been established, an introduction to some of the key researchers in sports analytics, and an explanation of how different industries are improving from sports analytics, the rest of this thesis will look at the literature currently out there on sports analytics in the world of football specifically in the National Football League, NFL. In Chapter 2, we will do a broad look at the current state of football analytics, Chapter 3 will discuss the existing literature about analytics around the NFL draft, Chapter 4 will delve into analytics around the specific position of wide receiver, Chapter 5 will delve into offensive linemen and literature currently available about that position, Chapter 6 will talk about analytics for other offensive positions, and finally in Chapter 7 recommendations of possible future work for football analytics will be discussed.

CHAPTER II: CURRENT STATUS OF FOOTBALL ANALYTICS FOR THE NATIONAL FOOTBALL LEAGUE

As discussed in the previous chapter, there is a lot of analytics being done all over sports, and as improvements to data collection methods are made, the data that are available to analyze are increasing both in size as well as complexity. This is true also for the National Football League, NFL, and the analytics currently being done and written about. This paper will break down publications for analytics conducted on NFL data into subcategories that we will call pre and post-NextGen stats. In 2016, the NFL started placing two RFID chips in every player's shoulder pads, and in 2017 one was placed in the ball for the purposes of tracking players and the ball electronically. This entered the new era that the NFL calls the NextGen stats era. However, as of the writing of this paper, this data is not publicly available, and each team has limited availability to the data they can request.

The other important thing to note is that are lots of non-academic, non-peer-reviewed information and data analysis in the NFL. I will take a moment and talk about these and what they are doing briefly, but the major purpose of this paper is to do a review of the literature that has been peer-reviewed and published regarding NFL data analytics. The first company is the NFL itself; as mentioned above, the NFL has player and ball tracking data available to it and creates many charts, graphs, and advanced statistical categories. While the statistics are available on their website (NFL, 2023a), the raw data and information on how these statistics are calculated are kept from public access. The Next Gen Stats is powered by and supported by AWS, which also has its own advertising and statistics on its site (Amazon Web Services, 2023); however, data and processes are kept secret as well. In 2019, the NFL started providing some of this player-tracking data available to the public in an effort to crowd-source analytics. This has become known as the Big Data Bowl (NFL, 2023b), and participation has grown every year since its inception. For example, in 2022, they provided all player tracking data for all punting and kicking plays from the regular seasons from 2018 – 2020. The other company with popular analytics being done is Pro Football Reference. They have created a statistic they call approximate value, which provides information on how it is calculated on their site (Palmer et al., 2021). This is a good start on player rating systems, which we will discuss in more detail in later chapters.

As for the current literature in NFL analytics, these can be broken down into two main categories, those that deal with using limited availability of player tracking data which I will review in section 2.2, and those using other available data such as play-by-play or other data sources which I will cover first in section 2.1.

2.1 Current Analytics Using Traditional Data

One of the first papers written about football analytics dates back to 1971 (V. Carter & Machol, 1971). In this paper, the authors use data from the first half of games played in the 1969 season to conduct their analysis. They develop an expected points value for first-down plays and make a commentary about teams calling time-outs to stop the clock when trailing by seven points or less. For their paper, they devised that given any play could result in any one of 103 possible scenarios, score a touchdown, score a field goal, defense records a safety against them, the opponent gets a touchdown from interception or fumble recovery, or the opponent gets the ball on turnover and is on one of the other 99-yard lines. They then split the 99 possible outcomes into ten-yard chunks to determine a probability. They then established an expected point value using the equation of $E(X) = \sum X_i P(X_i)$ to determine an overall expected points for each of the ten-yard blocks. Carter and Machol then use this newly generated expected point value to create an argument about calling timeouts in late-game situations. In many of the current football analyses being published, the expected point value system established by Carter and Machol is still used today to perform even more advanced analyses.

The remainder of the literature can be grouped into six main categories. These categories will be reviewed together in the following subsections. The groups and subsection numbers are: 2.1.1 Play-calling strategy, 2.1.2 Pre-play analysis or play call prediction, 2.1.3 Score or winner of game predictions, 2.1.4 Fourth Down analysis (should teams be going for it on fourth down more), better ranking of draft prospects (Chapter 3), or individual position analytics (Chapters 4-6).

2.1.1 Play-calling strategy

The first area of literature for review is those papers that deal with a play-calling strategy. Play-calling strategy can be seen as analyzing a certain situation during a football game and determining what play should be called next. The first thing to note about all of these papers and analyses of this type, in general, is that specific play-calling predictions cannot be made, as each individual coach has their own playbook and creates plays that are specific to their team's strengths and weaknesses. Instead, predictions in this category deal with should a team call a running or passing play (B. C. Alamar, 2006; Boronico & Newbert, 1999; Jordan et al., 2009), should a team run a play vs. kick a field goal in overtime (Sahi & Shubik, 1988), or what to do on fourth down which will be discussed in more detail in section 2.1.4.

When looking at determining if a team should call a run or pass play, there are several factors that participate in the decision-making process. Several factors, such as down and distance; score of the game; time remaining; yard line; and game theory itself, comes into play when a coach makes the play call for the next play.

Looking at down and distance as a factor, both Alamar, as well as Boronico and Newbert, establish this in their analysis and have specific situations they are looking at. Alamar limits his play call analysis to only plays that are first and ten with between 60 and 80 yards (the team's own 20-yard-line to 40-yard-line) for their analysis. Whereas Boronico and Newbert limit their analysis to only plays that are first down and goal-to-go.

If we look at score of the game, time remaining, position on the field of the offense, and yard line, these three can be grouped into a term referred to as game situation. Coaches make their play-calling decisions with the game situation in mind. For example, if a team was ahead by 14 points with less than 2 minutes to go in the game, the probability of the coach calling a running play is increased even if it is third down and 15. Whereas a team down by seven on their own 20-yard line with first-and-ten, the probability of them calling a long pass play in the fourth quarter is increased. This is why game situation decisions are important, and analysis on improving decision-making based on the game situation is useful. This is the main focus of Jordan et al.; they focus on developing models based on game situations to make their decisions. Sahi and Shubik also have game situations as the main focus of their paper; however, they specifically look at overtime or sudden death situations after crossing mid-field and if a team should then kick a field goal or run a play for the next play.

Lastly is game theory in general. There are many papers written on game theory in general, and I will not be discussing these papers in this review. All the above-mentioned papers utilize game theory as part of their decision-making process. However, only Boronico and Newbert make game theory a large part of their paper and discuss more in-depth game theories and their application to football. One important principle in game theory states that if an action gains value, a player may actually use the action less because the opponent also knows the action has gained value and will adjust their decisions based on this fact. In football terms, a team may choose to pass less in obvious passing plays because they know the defense is going to prepare to stop pass plays first. This principle is one of the main drivers in Jordan et al.'s modeling adjustments. They use risk factors and risk tolerance and user risk surveys to develop what they term a certainty equivalent, or how certain it is that a team will call a given play type (i.e., pass or run).

Whenever modeling is used, there must be some way of evaluating the model for its accuracy and effectiveness. For the purposes of the papers in this section, the target variable the authors use varies as to what they determine as the best variable to quantify success. Jordan et al. model an entire game, and as such final score is the factor that determines the overall performance of their model. Sahi and Shubik utilize an expected point value similar to that developed by Carter et al. versus the probability of making a field goal from the same yard line to evaluate if a team should kick a field goal or run another play in sudden death situations. Boronico and Newbert utilize the same expected point value with looking at run or pass where, including the fact that passing plays have more negative potential than running plays (i.e., interception, incomplete, loss of yards vs. loss of yards, fumble for results from pass vs. run respectively). Alamar utilizes gaining four yards on a play as a model of a successful play versus an unsuccessful play on first down. This is based on what is a common belief in football that gaining four yards or more on first down is good. Alamar uses expected yards gained for each play type to model the fact that, in the opinion of his paper, teams should pass more on first downs than they do.

As for modeling itself, all four papers use very simplistic models, such as probability models or logistic regression models, to make their predictions and analysis. Many of the advanced analytics that will be discussed later in this chapter, however, build off the models built by these four papers, and all remaining papers cite at least one of the four aforementioned papers in their own research.

2.1.2 Pre-play analysis

Of all the analyses that have been written about in academic literature, the pre-play analysis would most likely be the area that NFL coaches would be most interested in. The ability to know before the ball had been snapped what play a team was going to run would provide a great help to the defense and their ability to stop the offense. Likewise, the offense would like to know what type of defense was called so that the weakness of that defensive play could be exploited. When in it comes to predicting plays, there are multiple variables that are available before the ball is snapped that can be used to predict the play. There are also multiple models that can be used to predict a play as well as how in detail, a play prediction is going to be made. That being said, again, every team has its own playbook with similar plays to all other teams, but formations, names, and other minor details will change from team to team, so predicting the exact play called and subsequently ran by the offense is almost impossible without knowledge of the team's actual offensive playbook.

When it comes to the literature currently out there about pre-play analysis and predicting play, the most common prediction is just simply run or pass (R. E. Baker & Kwartler, 2015; Fernandes et al., 2020; P. Lee et al., 2016; Ötting, 2021). Knowing if a team was going to run the ball or pass the ball would change how a defensive coach may call the next play in order to best stop the play for the lowest possible gain. It is argued by Fernandes et al. that incorrectly predicting a pass when the offensive team had called a running play versus incorrectly predicting a run when the play called was a pass play. The argument is based on the stated fact that a defense expecting pass is better setup to stop a run then a defense expecting run would be setup to stop a pass play. Heiny and Blevins took their prediction a step further and predicted one of five different outcomes for their play prediction (Heiny & Blevins, 2011). The five outcomes they predict are: short pass; medium pass; long pass; run; and scramble. It is important to note that Heiny and Blevins make no notes about distinguishing between scramble verses designed quarterback run.

When it came to the data used by the literature, there were different levels of amount and type of data used. Ötting used the most extensive data set utilizing all regular season play-by-play data from 2009 to 2018 (289,191 observations). Baker and Kwartler used the largest spread of years of data as they used the past 13 seasons, 2001 season to 2014 season; however, they limited their analysis to only looking at offensive plays for Cleveland Browns and Pittsburgh Steelers (23,310 observations). Fernandes et al. used all regular season data from 2013 season to 2016 season (130,344 observations) for their analysis, while Lee et al. used all regular season data from 2012 to 2014 (101,753 observations); they also included madden ratings for every player that was used in these games to help in the modeling they did. The smallest data set was that used by Heiny and Blevins, who only used 2005 regular season offensive data for the Atlanta Falcons, which is only a total of 988 total observations. While just having a larger data set does not necessarily guarantee better accuracy in the models, it is interesting to note that the smallest data set was also the only paper to go beyond just predicting run vs. pass.

Fernandes et al. and Lee et al. did not limit their modeling to a single type of modeling to determine final prediction and instead each used four models to evaluate which method was best. Fernandes et al. however stated that they did the complex modeling first and in the end wanted an easily interpretable and usable model and as such ended with decision trees for their final model; however, Lee et al. used training and test accuracy for determining which model to use as their final model. Both papers utilized a random forest with Lee et al.'s reported prediction accuracy slightly higher than that of Fernandes et al. (75.1% vs 74.7%). However, the other three models, and reported accuracy for Lee et al. was logistic regression (73.7%); Linear Discriminant Analysis, LDA (72.7%); and Gradient Boosting Machine, GBM (75.7%). The other models, and reported accuracy, for Fernandes et al. was CART (73.3%); K-nearest neighbors, KNN (71.3%); and Neural network (75.3%). In the end, Fernandes et al. reduced their independent variables to allow for a quickly utilized neural network as their final model (75.1% accuracy) while Lee et al ended with the GBM as their final model. Baker and Kwartler used logistic regression to build their model and had reported accuracy of 66.4% for Cleveland Browns and reported accuracy of 66.9% for the Pittsburgh Steelers. Ötting utilized a Hidden Markov Model, HMM, to evaluate a play based on looking at each game as an independent time series and using HMM with time series techniques of prior knowledge to affect the prediction. They produced a model for overall prediction using all teams as well as splitting the dataset in such a way that each team had its own model built and accuracy calculated. Ötting utilized AIC as the criteria for selecting covariates and number of states for their models and had an overall reported accuracy of 71.6% with the highest accuracy by team being for the prediction done for New York Giants at a reported accuracy of 90.5%. Heiny and Blevins used the discriminant procedure in SAS, PROC DISCRIM, however they do not report what options they selected so specific type of discriminant analysis is not reported the reported accuracy they had was 40.38%. Off all the models reported on the highest prediction amongst all these papers was the one done using GBM by Lee et al. However, mostly due to papers intent the easiest to implement during a game the final decision tree used by Fernandes et al. with an accuracy of 75.1% is the one that it is the opinion of this author would be the best model to be considered by NFL teams.

2.1.3 Score/Winner prediction

While the ability to predict what play will be ran next may be most important to coaches. What fans and gamblers want to know is who is going to win. Gambling and statistics have a long history, it is a common belief that the first use of probability theory was done in the 1600s to solve a gambling dispute. While this paper does not endorse gambling in any manner, knowing who is going to win, and by how much in case of point spread, is important aspect of football analytics as an estimated 100 billion dollars are bet each year on NFL Football each year (Molter, 2023). When it comes to the academic literature about winning or score predictions there are several important papers written on the subject (R. D. Baker & McHale, 2013; Collins et al., 2017; Lock & Nettleton, 2014; Quenzel & Shea, 2016; Roumani, 2022). The papers vary in what they are predicting exactly, data sources, and model selection types.

When looking at what each paper was trying to predict, and reported accuracy of each, Baker and McHale for their paper looked at forecasting the exact score of a game given multiple variables (they reported an accuracy of 63.6% when picking winner based on score of each side). As an example of modeling techniques used Baker and McHale used a continuous time Markov model. The first step in this modeling technique is to create what is known as the hazard value for each type of scoring play possible on any given play. This hazard is calculated for home and away. The equations for home and away respectively are:

$$\lambda_{ijk} = b_h^{\delta} exp\left\{\left(\sum_{l=1}^{p} \beta_l x_l\right) + c_k + c_h + \gamma_1 i + \gamma_2 j\right\}$$

$$\rho_{ijk} = b_h^{\delta} exp\left\{\left(\sum_{l=1}^{p} \beta_l x_l\right) + c_k + +\gamma_1 i + \gamma_2 j\right\}$$

Where γ_1, γ_2 are for current score of home and away respectively. Then the probability of either team scoring on any given play is calculated by the authors using the following equation:

$$Q = \frac{\gamma_k\{\exp(-a) - \exp(-b)\}}{b - a}$$

Where $\gamma_k$is the hazard of the home team scoring from a nil-nil state and b is the sum of the ten hazards of scoring on the final state of play and a is the sum of the ten hazards of scoring from a nil-nil state. Collins et al.(reported accuracy of 64%), and Quenzel and Shea (reported accuracy of 58%), all made predictions about who would win a regular season game, similar to these Roumani also looked at predicting the winner, however limited his predictions to just the Super Bowl (reported accuracy of 84%). Lock and Nettleton instead looked at creating a statistic that they called win probability that can be calculated at any given time in a game before the next play is ran. All of these papers had varying reported accuracy and part of the reason for the variability is in the prediction itself.

The other factor that could affect the above accuracies is the data used for each. All used regular season data however the number of observations varied greatly due to qualifying data used as well as years of data. The largest data set was that used by Collins et al. who used play by play data for all regular season games for all teams from 2002 to 2012 (over 450,000 observations), Lock and Nettleton used a similar data set of all regular season games from 2003 to 2012 (430,168 observations), Baker and McHale used game data from all regular season games from 2001 to 2008 (2128 observations), Roumani only used regular season totals for each team for all 32 teams from 2002 to 2019 (576 observations), the smallest data set belong to

Quenzel and Shea were only looking at games that were tied at half time so while they had the largest range of years, 1994 to 2012, there data sources were limited (429 observations).

The last item that could greatly affect the overall accuracy of a model is the model selection and model type itself. Roumani was the only paper to write about multiple methods to make their predictions. In the paper by Roumani they first utilized the C4.5 algorithm, which is a specific decision tree builder that can be used with packages in both R and python, this algorithm starts with step one which is to calculate an initial entropy for the sample distributions using the equation:

$$I(T) = \sum_{i=1}^{k} P_i \log_2(P_i)$$

Where P_i is the probability distribution for a specific category and T is the training sample. The second step is to calculate entropy which follows the equation of:

$$I(A,T) = \sum_{i}^{m} \frac{|T_i|}{|T|} I(T_i)$$

Where A is the selected attribute you are considering for classification. After that has been calculated you need to determine gain information for A using $G(A,T) = I(T) - I(A,T)$. The fourth step is then to calculate the split information and information gain ratio using the following equations:

$$Split(A,T) = -\sum_{i}^{m} \frac{|T|}{|T_i|} \log_2 \frac{|T|}{|T_I|}$$

$$G_R = \frac{G(A,T)}{Split(A,T)}$$

The last step is to select the node with the maximum information. Then steps two through four are repeated for every level until such a time that every leaf node would belong to single category. Roumani then also utilized Artificial Neural Networks, ANNs, and Random Forests. Also, because they had data from all 32 teams and yet only two teams play in the Super Bowl and only one actually wins the Super Bowl they utilized Synthetic minority oversampling technique, SMOTE, which is used when you have imbalanced data. In the end, they determine that the random forest with SMOTE sampling gave them the best accuracy. Likewise Lock and Nettleton also utilized random forest to estimate a win probability for a given scenario before each play. Baker and McHale utilized a point process modeling technique to predict the exact score of a game, although they never report an accuracy of predicting exact score they instead state that predicting exact score was more about picking winner of a game or which side of point spread to bet on. As well as above stated accuracy of 63.6% of picking winner straight up they also reported an accuracy of 66.9% when picking which side of point spread a person should bet on. Quenzel and Shea utilized a simple logistic regression to determine what variables from first half of a game seem to indicate which team is going to eventually win a game when it is tied at half time. Collins et al. also utilized a simple logistic regression as their model choice when predicting win probability for a team given previous statistics of them and their opponent.

2.1.4 Fourth Down Analysis

One of the frequent discussions held on sports talk shows regarding football is if a coach should have "gone-for-it" on fourth down at a certain point in the game. The term go-for-it is in reference to how NFL football rules are set up. Each time a team gains possession of the ball they have four plays to gain a first down, ten yards, or score. The offense has the option to punt the ball on fourth down which gives the ball to the other team but in a worse situation then had the original team failed to get a first down on their fourth down. The critique can be that a head coach should have been more aggressive and went for it, or that the coach that failed to convert a first down when they "go for it." A few papers have been written in other disciplines most notably using the NFL and going for it on fourth down to look at if business firms maximize potential (Romer, 2006). This idea is not new. And in fact the first paper published on this exact question, should a team try to get a first down instead of kicking it (punt or field goal) was written in 1978 (V. Carter & Machol, 1978). The other main paper on this subject for this review was published in 2019 (Yam & Lopez, 2019).

Carter and Machol utilized their combined knowledge of football, at the time of writing the paper Carter was a quarterback for the Cincinnati Bengals and mathematical systems, Robert Machol was a professor of systems at Northwestern University, to develop an optimal strategy for if NFL teams should try to convert a first down or not on fourth down based on location on the field as well as score in the game. The crux of their work was based on once a team had crossed mid field and was being faced with a fourth and "short," although later in the paper most of the focus is on fourth and one yard to go. Utilizing their previously discussed modeling to calculate expected points given a situation and the change in expected points should they try to convert or kick it. They find that the probability of getting a first down on fourth and one was 0.715 or 71.5% chance of converting. They use this as well as difference between expected points if they convert or not to determine that unless a team was faced with greater than 4 yards to first down or had personnel that would suggest their kicker is above average and skill players less than average, they should attempt to achieve a first down as it would increase their chances of winning.

Yam and Lopez revisited this question and use Carter and Machol's paper as the crux of their position. They use data from the NFL regular season from 2004 to 2017 and modern computer and machine learning techniques to revisit this recommendation made by Carter and Machol. In the paper, Yam and Lopez evaluate only plays that a team should have gone for it or did go for it on fourth down. They then use the change in win probability using techniques very similar to how above-mentioned paper by Lock and Nettleton created in 2014. The model chosen is a Rubin Causal Model. This model according to the paper first introducing the model (Imbens & Rubin, 2010) is used evaluate 'potential outcomes' to develop probabilistic models. It is the final recommendation of this paper that maybe teams are underestimating value of going for it on fourth down based on expected points gained, or teams may be overestimating the value of going for it based on win probability and then number of wins in a given season a team could gain.

2.2 Current Analytics Using Player Tracking Data

As previously mentioned, player tracking data is currently private data that only small amounts of the data have been released to the public, with most of the data released has been for the purposes of the Big Data Bowl crowd sourcing competition. However, there have been a few academic articles written using the data provided. Most of the publicly available material using player tracking data is the winners from the previous year's Big Data Bowl and are not published in academic journals and are instead available through the NFL's big data bowl (NFL, 2023b) site, or other sources such as Kaggle (NFL Big Data Bowl 19, 2019; NFL Big Data Bowl 20, 2020; NFL Big Data Bowl 21, 2021; NFL Big Data Bowl 22, 2022).

There are many papers written on analytics using the player tracking data. As the data released by the NFL for the Big Data Bowl has been set up to answer certain questions many of the papers are dealing with specific positions or situations. We will look at those papers in depth when dealing with the positions in later chapters. For now we will just mention that there are papers written that deal with receiver route identification (Chu et al., 2020; Kinney, 2020), completion probability, evaluating quarterbacks, and evaluating pass defenses as a whole. I will review most of these papers in later chapters when I compare and contrast them with non-player tracking analytics for specific positions.

The papers that have been written in academic journals utilizing player tracking data that do not fit in other chapters of this paper are based revisiting a previously published topic about if teams should go for the first down more often than they do on fourth down (Lopez, 2020) while the other looks at how the player tracking data can used to create continuous-time within-play evaluations as well estimating expected yards gained by a ball carrier given any instantaneous data point (Yurko et al., 2020), or completion probability (Deshpande & Evans, 2020).

In section 2.1, we discussed the non-player tracking publications that talked about if teams should go for it more on fourth down utilizing traditionally available data. Lopez address the subject of if teams should be more aggressive and try for first downs more on fourth down then they currently do. This is a departure from Carter et al. or others work where they used play by play data in which a fourth down is just marked as an integer based on distance to go rounded to nearest yard. For example, fourth and 1 yard to go actually means the ball is anywhere from fractions of an inch to 71.9 inches from the line to gain first down. Lopez uses a generalized additive model (GAM) to build success rates for did a team get a first down given the precise distance needed. Lopez surmised that while the data does suggest that if a team had less than 0.7 yards, 25 inches, the team should try for converting a fourth down instead of punting or kicking a field goal as the success rate and win probability differences between converting or not converting were minimal. While given a greater distance than 0.7 yards to go the teams are better off punting or kicking a field goal because success rate is lower such that it has greater impact on win probability.

Yurko et al. as well as Deshpande et al., utilized the data from the first Big Data Bowl which was the first six weeks of player tracking data from the 2017 NFL regular season. As well as this data they also gathered play by play data from the same games using the nflscapR package in R. Both papers develop a model for determining an expected (hypothetical) completion probability, Yurko et al., simply refer to it as expected completion probability whereas Deshpande et al. refer to it as an expected hypothetical completion probability. While both papers predict a completion probability Deshpande used both regular logistic regression, Bayesian logistic regression, Bayesian Additive Regression Trees (BART), and Markov Chain Monte Carlo for their modeling. It was their conclusion that the MCMC model yielded the best results. Yurko et al., utilizes an intercept only model which they used as their baseline, they then experimented with LASSO, XGBoost, Feedforward Neural Network, and finally Long short-term memory model that they determined was the best model to use because it supported high dimensions, it was nonlinear, and accounted for changes in time.

Where they differ is that Yurko et al. went beyond this calculation and also went on to calculate expected yards gained from ball carriers' current position on the field as well as positions of the other players on the field. Yurko et al. also calculate the ball carriers instantaneous speed based solely on change in distance from previous observation. As with completion probability they continued to use an intercept only model, LASSO, XGBoost, Feedforward Neural Network, and finally Long short-term memory model. They then utilized similar modeling to calculate a quarterback decision model that rates if a quarterback made the correct decision based on player tracking data. Next, they looked at probability of each eligible receiver's probability of being targeted. All of these analytics can be used to help better evaluate players on offense.

CHAPTER III: FOOTBALL ANALYTICS ON EVALUATING THE NATIONAL FOOTBALL LEAGUE DRAFT

This chapter will be devoted to reviewing the current literature that is related to the National Football League (NFL) draft (the draft). When it comes to current analytics being done regarding the draft, there are two main areas that the literature can be broken into; they are ranking draft prospects and predicting the success of draft picks in the NFL. It is important to note that there are many current sports sites out there on the internet that talk about the draft, and each of these sites has personnel that they refer to as draft analysts who provide their own rankings and mock drafts, for example, on espn.com you can see Mel Kiper's mock draft for the 2023 draft (Kiper, 2023). The analysts develop these draft rankings and mock drafts themselves, and no information is given about what factors and analyses are used to form these lists. As such, they will not be discussed going forward in this literature review as this paper is devoted explicitly to academic literature regarding the draft. The remainder of this chapter will be divided into three sub-sections; 3.1 ranking draft prospects, 3.2 predicting performance in the NFL of draft picks, 3.3 future work, and improvements that can be made on the current status.

3.1 Ranking Draft Prospects

When it comes to selecting talent in a draft model, all current major US sports have a draft (National Football League, National Basketball Association, Women's National Basketball Association, Major League Baseball, and National Hockey League). However, the NFL draft is unique in the fact that position is significant, and the talent pool does not have a similar statistic for all positions. For example, in basketball, all players record points, assists, steals, blocks, rebounds, and there is some variation in amounts per game of each of these statistics depending on which of five positions a player plays. In baseball, you have hitting statistics (such as batting average and home runs) and pitching statistics (such as earned runs and wins), and most non-pitchers can play multiple positions, and as such, positions are really divided into pitchers and hitters. In hockey, you have again two groups of statistics you have goalie statistics (such as goals against and save percentage) as well as non-goalie statistics (such as goals and assists). For the non-goalies, the positions can be broken down into forwards and defensive men, giving three main positions to be evaluated.

When it comes to the NFL, however, even getting generalized as possible, you have ten unique positions (quarterback, running back, wide receiver, tight end, offensive line, defensive line, linebacker, defensive back, kicker, and punter). Each of these groups has a primary objective and statistics that they are held to. For example, the quarterback has a statistic called quarterback rating (QBR) (NFL.com, 2023); however, no other position on the field has such a rating system. The other offensive positions (running back, wide receiver, tight end, offensive line) do not accumulate the same statistics as the quarterback, so they cannot be judged on the same rating system. Then when looking at the defensive positions (defensive line, linebacker, defensive back), they do not accumulate any of the same statistics during the game as the offensive players. Also, the statistics they do get, the three groups are held to different expectations (for example, a defensive lineman may never record an interception in their career but still be looked at as a great defensive lineman). While the kicker and punter perform tasks on the field that no other players will regularly perform (punts or field goals) With, this uniqueness in player positions and statistics is why evaluating the draft can be a challenge even with the most advanced modeling and large data availability. Also, every team has its own list of needed positions in the draft based on what they view as the weakest positions are their individual team.

It is important to note that sometimes a team will draft what they deem the best available player regardless of position.

Before diving into the literature about the draft, it is essential to understand a couple of terms that will be used by the literature and their meaning. The first is the NFL Combine (NFL, 2020); this event is held over the course of a week in the month of February. Draft-eligible players that the combine player selection committee determines have a high chance of being drafted and are extended an invitation to the combine. At the combine, players are run through several mental and physical evaluations that NFL team scouts are watching and recording data in an effort to help evaluate the draft prospect. At the combine, one of the evaluations done is players administered the Wonderlic test, which is meant to judge intelligence as well as decision-making speed. Another term important to understand is Bowl Championship Series (BCS) (Martinez, 2010); this term is used to signify the top echelon of college football. BCS schools are from one of six conferences (the Atlantic Coast Conference (ACC), the American Athletic Conference (American), the Big Ten Conference (Big Ten), the Big 12 Conference (Big 12), the Pac-12 Conference (Pac-12), and Southeastern Conference (SEC) conferences). These conferences have more money devoted to football and more resources that are given to their football players. Many of the analysis done in the literature has a dummy variable for if a player is coming from a BCS school or not.

When it comes to the literature, the first item that needs to be determined is a system to evaluate the data as an outcome. For this section, the outcome is the draft position relative to all the other players being considered. Also, the question of what data to use for the evaluation. The draft has been held since 1936 (Football operations, 2023) and has changed drastically since the first years. Lyons et al. did their evaluation for players selected in the draft between 2002 and

2004 (Lyons et al., 2009), Treme and Allen used players selected in the draft between 2001 and 2006 (Treme & Allen, 2009), Pitts and Evans collected data on players selected in the draft between 2002 and 2012 (Pitts & Evans, 2018), Dhar had his paper on players selected in the draft between 1999 and 2010 (Dhar, n.d.), Mulholland and Jensen extended it two years and used players selected in the draft between 1999 and 2013, McKenzie evaluated players selected in the draft between 1999 and 2013, whereas Berri and Simmons used players chosen in the draft between 1970 and 2007. One thing to note is that several papers started in 1999. The reason for this is because all got some of their data from the combine results page, which has available to the public for free all information about combines from 1999 to the present day (NFL Combine Results, 2022). While there are several factors in determining what day to go back to, understanding that increasing the pool of eligible drafts to evaluate may, in most cases, improve accuracy and effectiveness of any models built.

The next question is what players to select for evaluation. Only Lyons et al. included all positions in their assessment; all others reduced the player pool in some way. McKenzie reduced the player pool to only quarterbacks, wide receivers, and running backs. Beri and Simmons, as well as Pitts and Evans, limited their evaluations to only quarterback. Dhar, as well as Treme and Allen, both reduced their player pool to only wide receivers. While still, Mulholland and Jensen limited it to only Tight ends. It is important to note that since the NFL draft is all positions in one pool, limiting the positions evaluated for a draft ranking does create a gap in their evaluation by leaving out one or more draft-eligible positions.

Once the decision of which years of the draft as well as which positions had been made, the researchers then had to determine what pre-draft day data they were going to consider for their evaluation and ranking. Lyons et al. and Pitts and Evans limited their assessment to only looking at the Wonderlic scores (Cowen Partners, n.d.). The rest of the papers all used various statistics from both the player's college football playing careers as well as data from the combine. The two papers that limited their pre-draft day data to just the Wonderlic were both looking at the effectiveness of using the Wonderlic to predict or rank draft order.

An evaluation of modeling techniques produces much of the same model types with varying reasons for selecting the significant variables. McKenzie's paper is the most in-depth as far as modeling techniques; in the paper, they use four different standalone models (Naïve Bayes, Logistic Regression, Multiplayer perceptron, Radial basis functions, RBF, network) as well as the paper looks at combining other models in a multilayer modified genetic algorithm. While this is the most in-depth analysis as far as types of models, they also have the smallest data set, and better results would have been achieved if they used a larger dataset. Lyons et al. used bivariate correlations between Wonderlic and draft position for their evaluation of the effectiveness of Wonderlic. All other papers previously mentioned above use an ordinary least squares regression as one or their only modeling technique. Pitts and Evans, along with OLS regression, also did a two-stage least squares model. Mulholland and Jensen, as well as Dhar, not only used regression but also used recursive partitioning trees (CART). The biggest limitation with linear regression is that in order for the results to be accurate, we must assume four assumptions (Linearity, equal variance, independence, and normality); none of these papers do an excellent job of addressing how they make sure that all assumptions are met. The other limitation found in these papers' modeling is that, with the exception of Dhar and McKenzie, none of the other papers make mention of testing and training sets or validation methods of any kind.

As for the results of the models, all papers in this section are trying to make a ranking system that would tell NFL teams which players should be drafted above others based on pre-draft statistics.

3.2 Player Performance in the NFL-Based Draft Analytics

The other form of prediction is using draft data in conjunction with draft position to attempt to quantify how good a player is going to be in the NFL. After all, this is what all teams really want to know is how well the player they are drafting is going to perform. There are a few unmeasurable that can affect a player and how well they are going to perform once they get to the NFL, such as injuries or how a player will react when they have access to the amounts of money they start earning in the NFL.

When looking at the current literature, the first area of focus that all papers need to decide is how they are going to measure performance in the NFL. The reason for this is that, with the exception of quarterback, there is no single measure that explains how well a player is performing that can be used to quantify a player's performance. The website pro football reference has created a score they call approximate value that is intended to provide a scoring system for all players and their performance (Drinen, 2008); the limitation of this is that each player is rated off their team's offensive or defensive production, and so if you have a great player on a lousy team that individuals play may not stand out on the approximate value. When it comes to the literature, most of the papers use multiple dependent variables to judge performance. Several of the papers (Lyons et al., 2009; Mulholland & Jensen, 2016; Pitts & Evans, 2018; Treme & Allen, 2009) use games started as a dependent variable, while games started is a decent measure of NFL performance it is also arbitrary as it is based on the first play of the game for either the offense or defense. For example, if a team decides to use five wide receivers on their first offensive play, then they will not have a tight end or running back get credited for games started, even if the player comes in the next play and plays the entire game. An example of this is Isaiah Pacheco for the 2022 Kansas City Chiefs. For the season he played in all seventeen games and is listed as the number one running back on their depth chart (Kansas City Chiefs, n.d.), yet he has only eleven games started. The next similar statistic for measuring player performance is games played (Boulier et al., 2010; Kuzmits & Adams, 2008; M. Schuckers, 2011; Wolfson et al., 2011, 2017).

This gets rid of the limitations of games started; however, it is still limited in the fact that it does not allow comparisons between players of the same position on separate teams. Another statistic that relates only loosely to actual accumulated statistics is Pro Bowl appearances. Schuckers uses Pro Bowl appearances, while the Pro Bowl is, in theory, the best players at every position, the drawback of using this as a predictor is that players get credited for a Pro Bowl appearance if they are voted into it or if they are a replacement for players that choose not to participate for different reasons. The other difficulty with this is that Pro Bowl selection is based on fan voting, player voting, and coach voting, each getting one-third of the weight; thus, it has some level of a popularity contest. The last non-statistic that is used as a dependent variable is salary; Kuzmits and Adams use salary as a dependent variable. The most significant error with using this is salary is highly correlated with position and where the player was drafted, so using salary as a dependent variable is influenced by where they were drafted, which the authors use as a predictor. Finally, Boulier et al. use the number of years in the league as a dependent variable to measure performance.

Once we get passed discussing non-statistics as a dependent variable of performance, we can then look at statistics that are generated and recorded while playing football. Treme and

Allen use receptions in a rookie year as the other dependent variable for evaluating the performance of wide receivers. Kuzmits and Adams use QBR, average yards per reception, or average yards per carry for the first three years of the player's career as their dependent variables. Boulier et al. use QBR, total receiving yards, and passes thrown over the course of the player's career to evaluate performance. Berri and Simmons use QBR over the course of the quarterback's entire career as their dependent variable for performance as well. Using actual statistics to compare players of the same position is a logical evaluation; however, limiting to less than an entire career may create some problems with the predictions of success. The most prominent example of this is in 2005 the Green Bay Packers drafted Aaron Rodgers 24th overall (Drinen, n.d.), and he sat on the bench for his first three years as the Packers had Hall of Famer Bret Favre at quarterback. (Bret Favre was inducted into the Pro Football Hall of Fame in 2016) (*Pro Football Hall of Fame's Class of 2016 Announced | Pro Football Hall of Fame Official Site*, n.d.). Since then, Aaron Rodgers has won a Superbowl and four league Most Valuable Player (MVP) awards; thus, while looking at his first three years, he would have been looked at as drafted too high for his performance or would not have lived up to the expected performance in his first three years based on where he was drafted.

The last area of performance measure used by the papers is individually created score values. These score values are well described in the individual papers and are the author's ideas of how to measure performance. Some of the papers use the same name for their dependent variable, but it is calculated differently. For example, both Berri and Simmons, as well as Quinn et al., create a statistic they call QB score (Quinn et al., n.d.). However, the statistic is calculated differently. These individually designed scores are a good measure of player performance, yet only the paper by Eicken and the career average value (CarAV) is a score that can be used to compare players from different positions (Eicken, 2017). This CarAV value, however, is limited in the fact that it assigns equal weight to each of the offensive linemen of a team per game played.

After selecting what is going to be the dependent variable, the next decision to be discussed is the position that is evaluated by the papers. Eicken, Schuckers, and Lyons et al. did not limit their analysis to specific positions; instead, they assessed all positions. This is the most appropriate for evaluating the effectiveness of the draft because all positions can and are drafted. McKenzie, as well as Kuzmits and Adams, limited their analysis to only quarterbacks, wide receivers, and running backs. This again leaves out many of the potential draft-eligible players in their evaluations. Boulier et al. reduced their evaluations to only wide receivers and quarterbacks. The interesting thing to note is that they state in their title that it is about the passing game in the NFL; however, they do not include tight ends. Mulholland and Jensen decided to limit their analysis to only tight ends. The author Bless limited their evaluation to the position of running backs (Blees, 2011). Both Dhar and Treme limited the review of performance to only wide receivers. The most popular single position to be evaluated was the quarterback. Wolfson et al., Quinn et al., Pitts and Evans, as well as Berri and Simmons all did their papers on just assessing the performance of the quarterback. While, again, there is nothing wrong with limiting the analysis to only selected positions; however, it does create a limitation in their analysis.

The next item of discussion is the modeling methodology itself. Lyons et al. and Kuzmits and Adams used a correlation matrix to determine relationships between their dependent variables and their independent variables. Correlations are good for showing if there is a relationship between the dependent variables and the independent variables, but they do not allow you to rank players; it limits your analysis to only state what values teams should be looking at when it comes to drafting players. All the rest of the papers mentioned before in this chapter used some form of regression as their first analysis. Quinn et al. used only a logistic regression, while McKenzie started with logistic regression as the first type of regression used. McKenzie then continued to include Naïve Bayes, Multiplayer Perceptron, RBF, Multilayer, modified genetic networks, as well as many different combinations of these modeling techniques to determine performance. The remainder of the papers all started with ordinary least squares, OLS, and regression as the first model for evaluating performance. Regression was the only model used to assess performance in the paper authored by Treme, Eicken, Bless, Boulier, et al., and Berri and Simmons. As an example, the linear regression equation used by Boulier et al. to determine number of years played by a drafted wide receiver is:

$$Years = 9.910 - \sum_{i=1}^{9} C_i R_i$$

Where C_i is constant based on rank of where the player in question was drafted. R_i is value of one if player was drafted at position equal to i and zero otherwise. Regression is a good modeling technique, and the most significant limitation is that we must make assumptions made earlier as well, as there are always debates about which variables should be included as well as what should be excluded. Dhar, as well as Mulholland and Jensen, used the CART modeling for the other model technique they used with their papers. Decision trees are a way of making sure that normality assumptions are met. Both articles by Wolfson et al., mentioned above, included a negative binomial and zero-inflated negative binomial regression models beyond the OLS. These two methods are suitable when the data is over-dispersed count data but assume the conditional means are not equal to the conditional variances. Pitts and Evans included in their analysis a two-stage least squares regression as a second model. The two-stage least squares regression is a good choice when the dependent variable's error terms are correlated with the independent variables, which is the case many times with sports data. Schuckers expanded on the OLS regression by doing a non-parametric as well as a locally estimated scatterplot smoothing (LOESS) modeling. The use of non-parametric as well as LOESS is a great idea to avoid needing the assumption of linearity that OLS requires.

3.3 Future work for draft evaluation

When it comes to predicting draft order or predicting the performance of players in the NFL once they are drafted, there are many areas of improvement that can be made. The first is getting more robust datasets. The data collected during the combine is available going back to 1999 through 2022. Also, statistics for college football that include all box scores of every game going back to 1960 can be found at sports-reference.com (Sports Reference LLC, 2023); these box scores can provide more variables that may be helpful in more in-depth analysis. For example, getting player tracking data for all college athletes would allow work to be done looking at how players are interacting on the field and showing trends about players that may improve future predictions. However, at the time of this writing there, it is not known if any teams are yet using this tracking data like the NFL is currently doing. Also, simply increasing the dataset size by including more years of analysis could help improve the accuracy of predictions.

Other than improving the data that is available to analyze, different and more robust modeling techniques can be implemented as improvement is made in computational ability. For example, a few of the papers used CART for the decision-making process. A newer, more effective method called XGboost is now available, which has been shown in many applications to be an improvement in accuracy over CARTs (Chen & Guestrin, 2016). Moreover, when looking at draft players, there could be more work done with the pre-draft information available. Such as possibly a time series analysis of college statistics to determine if a player is trending up or trending down, or remaining consistent may provide better insight into how to evaluate the potential draft selection of players.

The most extensive area of future work that can be done would be the development of a universal player scoring system. Right now, the average value created by pro football reference seems to be the best starting point, but there are many limitations to the value, and it is not an easy-to-understand scale; currently, there is not a listed maximum score that can be achieved as the score is based largely on the yards produced during a game and while there is a finite amount of time in a game there has yet to be a game in which every single play ran produces a touchdown of 99 yards in length which would be the theoretical maximum amount of yards that could be gained in the finite time. There should be a system where numbers are meaningful, applied to a particular player's position, and yet on the same scale as all other positions, such as the scale for the total QBR, which is from zero to 100 (Katz & Burke, 2016). This could be done by using weighted logistic regression, linear regression, or clustering performance statistics into groups using a clustering method such as k-means or k-nearest neighbors. There may also be a way to a multiple model selection process.

CHAPTER IV: FOOTBALL ANALYTICS FOCUSED ON WIDE RECEIVERS

When it comes to the NFL and the positions on the field, the quarterback is usually the most popular player and usually the highest-paid position for the team. However, in recent years one particular position is becoming more and more popular as well as more important to their team. If we look at the average salary for starting wide receivers over a 20-year period, the average salary for starting wide receivers has grown from 2.6 million in 2003 to 4.0 million in 2022 (Sportrac, 2023a), whereas the average salary for starting running back has changed from 2.1 million to 2.5 million in the same time span. If we also look at printed media, we see books being written by wide receivers, such as Keyshawn Johnson's book "Just Give Me the Damn Ball (Johnson & Smith, 2009) or even Cris Carter's book "Going Deep: How wide receivers became the most compelling figures in pro sports" (C. Carter & Chadiha, 2013). A way to determine player popularity with fans is by looking at jersey sales. Out of the top 50 jersey sales for 2022, 19 were quarterbacks, while 12 were wide receivers, and no other position was more than five representatives (NFL Players Assoc., 2022). It is also shown in the academic literature, looking at published articles by the single position being focused on the quarterback is the top, with wide receiver being second. In this chapter, I am going to focus on the literature that has been written about wide receivers in the NFL. The rest of this chapter will be broken down into three sub-sections. The first (4.1) will be literature that is written about the wide receiver specifically. The second (4.2) will be literature that is written with some focus on wide receivers as well as other positions. Lastly in 4.3 I will discuss future areas of research that could be done regarding the wide receivers.

4.1 Literature focused solely on wide receivers.

Looking at the academic literature about wide receivers in the NFL, there are two main focused areas of sports analytics. The first is about the drafting of wide receivers, both pre-draft ranking (Treme & Allen, 2009) and using draft data to predict success in their first few years (Dhar, n.d.). The next area is about route recognition or prediction of which route the receiver is going to run. It should be noted that there are many papers evaluating injuries or the economics of the wide receiver, but those are beyond the focus of this literature review. Chapter 3 had a very in-depth breakdown of the literature written by both Dhar, as well as Treme and Allen. Instead, this section will review the literature on route recognition.

The importance of recognizing what routes wide receivers are running has long been a part of the study of information. Currently, a team will have a person whose job is to watch game film and record what route each player ran on every passing play and then make a note of the time stamp as well as other data about the situation and provide that to the coaches after they are done. This process is very time-consuming, so creating a way of identifying routes with computers would significantly save time as well as possibly increase the types of future analysis that could be done. When it comes to route recognition, there are two different data types that are used. The first is the limited access to NFL player tracking data (Chu et al., 2020; Kinney, 2020) or using video film analysis (N. Lee & Kitani, 2016). In order to understand route recognition, it is important to note that there are only nine different routes that wide receivers will run down the field from the line of scrimmage. These routes are shown in figure 1. When conducting their video analysis Lee and Kitani used actual game footage to predict what route a wide receiver was going to run as well as predict where the defender covering them would be. They used a Gaussian process regression to build their model. Then once they had their model for the wide receiver, they used Markov chains and a Markov decision process to evaluate where most likely the defender would be in reaction to that route. Overall, they had successful results with an average error rate of 8.51 when using the Kullback-Leibler divergence to measure the difference between two probability distributions. The limitations of this study are that they only used a limited dataset; they only had 20 videos of passing plays from the same team with a maximum of four wide receivers running routes in any given play.

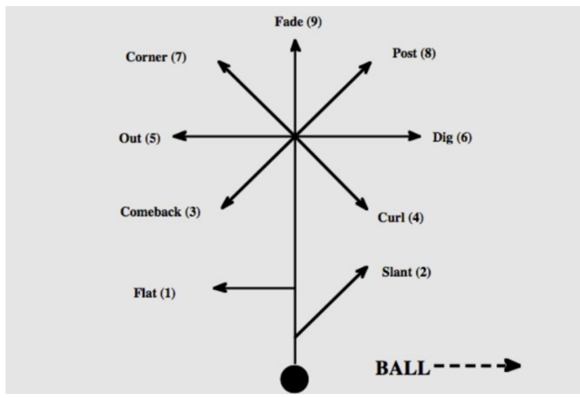

Figure 1 Route tree for wide receivers Chu et al., 2020)

Using player tracking data gives a more accurate positioning of a given player on the field as well as the player's direction and speed. However, at the time of writing, access to this data is limited for the NFL. Currently, the NFL will release parts of the player tracking data as part of the annual NFL Big Data Bowl, which is an open competition the NFL holds each year. The most robust dataset they have released to date is from the 2017 NFL season, in which they provided the player tracking data for every team for every rushing play from the first six weeks of the NFL season, 91 games in total. This is the data used by Kinney as well as Chu et al. for their articles. Both papers use machine learning techniques to identify what route a wide receiver was running on the play. Chu et al. used a Gaussian mixture method referred to as the Expectation-Maximization (EM) algorithm to handle the latent variables needed. The EM algorithm uses a multi-step process that starts with calculating expected probability using this equation:

$$\pi_{ik} = \alpha_k \frac{\prod_{j=i}^{m_i} \mathcal{N}\left(\boldsymbol{y}_{ij} \middle| \boldsymbol{T}_{ij}\boldsymbol{\theta}_k, \sigma_k^2 \boldsymbol{I}\right)}{\sum_{k=1}^{K} \prod_{j=1}^{m_i} \mathcal{N}\left(\boldsymbol{y}_{ij} \middle| \boldsymbol{T}_{ij}\boldsymbol{\theta}_k, \sigma_k^2 \boldsymbol{I}\right)}$$

For all i curves. Then the maximization step to update the values for $\boldsymbol{\theta}_k, \sigma_k^2, \alpha_k$ are found by maximizing the log-likelihood. Starting with $\hat{\alpha}_k$ this is calculated using equation of:

$$\hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^{n} \pi_{ik}$$

Where $\hat{\alpha}_k$ is the mean posterior probability that the i^{th} curve was generated from cluster k. Then using theπ_{ik} obtained from the expectation step then using weighted least squares we can update $\widehat{\boldsymbol{\theta}}_k$, and $\hat{\sigma}_k^2$ are updated using the following two equations:

$$\widehat{\boldsymbol{\theta}}_k = (\boldsymbol{T}^T \boldsymbol{W}_k \boldsymbol{T})^{-1} \boldsymbol{T}^T \boldsymbol{W}_k \boldsymbol{Y}$$

$$\hat{\sigma}_k^2 = \frac{1}{\sum_{i=1}^{n} \pi_{ik}} \left(\boldsymbol{Y} - \boldsymbol{T}\widehat{\boldsymbol{\theta}}_k\right)^T \boldsymbol{W}_k \left(\boldsymbol{Y} - \boldsymbol{T}\widehat{\boldsymbol{\theta}}_k\right)$$

The previous E and M steps were repeated until $\frac{(L_{new} - L_{old})}{l_{old}} < 1 * 10^{-3}$. Where L_{new}, L_{old} are the log-likelihood before E step, old, and after M step, new. Kinney used a grid search method and included Euclidean distance from the actual route and predicted location for each of the nine routes. Both papers presented above 78% accuracy for their predictions which is a good start for analysis. It is expected that had more information been available, then the accuracy could have been improved through the modeling.

4.2 Literature with Sections Devoted to Wide Receivers

When it comes to the academic literature that has specific sections for wide receivers, there are different areas and analytics being done. The first is those that are making predictions based on the NFL Draft, either pre-draft ranking or predicting NFL success based on draft data (Kuzmits & Adams, 2008; Lyons et al., 2009; McKenzie, 2015; M. Schuckers, 2011). These articles were reviewed in detail in chapter 3, so no further review will be done here. With the limited literature outside of draft analysis, instead of comparing and contrasting similar topic articles, this section will rather talk about each article individually, including some of the limitations of their work.

As well as draft predictions, two articles from chapter 3 need to be further evaluated as they develop a statistic that they use that should have a further in-depth look at them when it comes to the wide receiver. Eicken developed a statistic that was referred to as CarAV or the approximate value of a player throughout the player's career (Eicken, 2017). This value for the training set was obtained by using pro-football reference career approximate value and dividing by the number of years the player played. For the independent variables, they utilized several pre-draft statistics, including draft age, combined results, as well as a single value (PosColStats) that was intended to be a combination of all the player's college statistics to make a score that was similar for all positions. No info on how the PosColStats was calculated, just a statement that it was supposed to be a representation of numerous position-based statistics in college to create a level playing field statistically. Without information on how this score was calculated, we cannot access the validity of it for scoring a wide receiver, or any other position for that matter.

Boulier et al. used draft data and career statistics to develop two separate censored regressions (Boulier et al., 2010). The first was developing a model to predict the number of years that a wide receiver will play based on their draft position. They also develop a censored regression to predict the number of career-receiving yards that a player will have again based on the draft position. These models were developed as a method to evaluate if executives making decisions about whom to draft are making the correct choices. The limitation of this is that it is based solely on draft position, and there have been many wide receivers that have been drafted high and not produced on the field. An excellent example of this is the Detroit Lions, between 2003 and 2007, drafted a wide receiver with their first-round picks four separate times (FF Today, 2022). All of these picks were picked earlier than pick 11, and only one of the four had enough career receiving yards to be in the top 250 all-time career yards. Also, two of the four lasted less than five seasons of career playing. There may have been other statistics available before the draft that may have shown that these players were going to have different career receiving yards or lengths of career.

The following paper was done by Deshpande and Evans, which aimed to develop an expected theoretical completion probability (Deshpande & Evans, 2020). For their approach, they used the player tracking data released by the NFL for the Big Data Bowl in 2019. For the model, they used logistic regression to develop completion probability. They used the word hypothetical because they evaluated all wide receivers, tight ends, and running backs that were on the field for each passing play, even though there is only one player who is the target of the pass. The authors do a good job developing a model predicting where defenders may have been able to be in relation to the offensive player in question, even though once a pass has left, the quarterbacks hand the defensive players will leave the person or area they were guarding and converge on the player with the ball. After the simple linear regression, they also did a model with Bayesian logistic regression as well as Bayesian Additive Regression Trees (BART). The limitation of their model is that they could have worked to develop a more sophisticated modeling technique. The authors' results suggest that the BART model had a lower mean square error, but as with all tree methods deciding which variables are most helpful in predictions is not always easy or straightforward. Further work could be done on this by attempting to do a more complex and computationally costly method. Some suggested methods may be employing a deep learning model with multiple hidden layers may have produced better results. A few suggestions may have been using a generalized additive model.

The last paper is one that was first introduced in Chapter 2. This is the paper by Yurko et al. that is about creating a nflWAR or wins-above replacement for the NFL (Yurko et al., 2019). This paper is essential because it develops a universal statistical measure to compare players from different positions. Their approach was first to estimate the value of every play through expected points, as discussed in chapter 2. The next step is the estimate the effect of a player on play value added. This is done through a multilevel, or mixed-effects, model that looks at different statistics within the course of the play. Examples of this include passing plays, the air yards of the pass, the yards after catch (YAC), the side of the field, as well as the player involved in throwing the pass, most of which is a quarterback, and the targeted player for the pass, most of which is a wide receiver. They use these statistics to determine an individual player effect which is based on averages for position groups. A similar methodology is also applied to running plays. Once they have these models that determine the player value-added, they then use a similar technique that was first used in baseball to develop WAR based on other players on the roster (Baumer et al., 2015). This method uses a regression model and the statistics of all players on the roster to determine how much value is added by that player over potential other roster players put in the same position. One of the first limitations in nflWAR is in the data available itself. The play-by-play data does not let you know all twenty-two players on the field for that play. So, while they intend it to be used for all positions and all players, it is only the skill players on offense (quarterback, wide receiver, running back, tight end) as well as the player(s) on defense that were involved in the play (recording a tackle or sack, forcing a fumble, or getting an interception). As such, their nflWAR is only helpful for a limited number of players. Also, as with original baseball WAR, the nflWAR suffers from the uncertainty associated with estimations, as well as a lack of reproducibility due to fact that identical scenarios with different players rarely happens over the course of a game or even season.

4.3 Future work

All of the papers in this chapter have great modeling techniques and provide advancement in the ability to evaluate wide receivers in the NFL. However, there are still lots of areas of improvement that can be made.

The first is utilizing more data, currently through the use of nflscapR discussed earlier play-by-play data is now available from 1999 all the way through 2022 regular and post-season. Increasing the number of observations that are available will usually result in more accurate models with a reduced chance of overfitting the data. Also, there have now been six NFL Big Data Bowls in which the NFL has released player tracking data. Using these six datasets gives us more player-tracking data to determine which players were on the field.

Also, all of the wide receiver information is based solely on pass receptions. If we are going to create a system to evaluate wide receivers, we need to make sure we are creating a method to account for the quarterback. There needs to be a method where we can compare receivers who may have different talent levels at the starting end of the pass. With the addition of player tracking data, more information can be used to develop a better rating system. One example is including accountability for a good wide receiver, always drawing a double team from the defensive side and, as such, creating more openings for the other wide receivers on the team to get open and catch passes. Also, determining separation space from defenders as a measure of the ability to get "open" could be determined from the player tracking data. While this does not show up on the play-by-play data using the player tracking data and spatial statistics, one could determine how often a player is double-teamed. Also, receivers such as Deebo Samuel are being used more and more in running plays and increasing their value to the team; in 2021, Samuel had six receiving touchdowns and eight rushing touchdowns (Pro-Football-Reference, 2022). Currently, there is no statistic that includes this as one rating. For example, Yurko et al.'s nflWAR would give Samuel a WAR score for receiving and a WAR score for rushing but two distinct values. Also, when not involved in running the ball, wide receivers are expected to block for the rusher. Receivers who block effectively on running plays increase their value to the team and help prevent possibly letting them know a play is more likely a run play by substituting out an ineffective blocker during run plays. Developing a method that takes the currently available player tracking data and play-by-play data would help create a single score that would show actual value added to the team by having that player on your team as well as having them on the field.

This could be as simple as creating a linear regression or possibly utilizing more advanced unsupervised machine learning techniques such as clustering or K-nearest neighbors to overall group performance. Deep learning could also be explored with the large amounts of data that are available; however, the drawback of deep learning is that trying to interpret or explain decision processes can be a challenge.

Lastly, when looking at drafting wide receivers, there could be more work done with the pre-draft information available. Such as possibly, a time series analysis of college statistics to determine if a player is perhaps trending up or trending down, or remaining consistent may provide better insight into how to evaluate the potential draft selection of wide receivers. Chapter 7 will look at more in-depth future work that can be done with wide receivers as a better description of types of modeling and techniques that can be done.

CHAPTER V: FOOTBALL ANALYTICS FOCUSED ON OFFENSIVE LINEMEN

Chapter 4 of this paper discussed how wide receivers are fast becoming the second most popular and essential skill position in the NFL. The least talked about by fans, the lowest media attention, and, subsequently, the lowest amount of analytics being done currently is for what is most likely to be the most crucial part of a team's offensive success, the offensive linemen. What makes the offensive linemen so important is the fact that without a good offensive line, there will not be any open areas or holes for the running backs to run through, nor is there enough time for a quarterback to find an open receiver and complete a pass. Yet the only time an offensive lineman has any public attention to them during the game is when they commit a penalty. However, to NFL teams and coaches, the importance of good offensive linemen is known. While quarterbacks are the highest-paid positions on the field, the second-highest average salary by position is the left tackle (Sportrac, 2023b).

With the importance of linemen in the NFL, one would expect many academic articles to be devoted to them, however at the time of this writing; there were only three articles that can be found that talk about offensive linemen specifically (B. C. Alamar & Weinstein-Gould, 2008; B. Alamar & Goldner, 2011; Byanna & Klabjan, 2016). One of the most prominent reasons for the lack of analysis is the lack of statistics; even previously mentioned pro-football reference average value creates a value for the entire offensive linemen or basis individual average value only on games played, games started, and, all pro ap team selection (Drinen, 2008). Offensive linemen do not generate yards themselves; they do not show up in the box score or play-by-play data, except for penalties; there is not even public availability of which plays they are on the field. The player tracking data that the NFL has collected since 2017 may be the secret to creating a system that provides a statistic for NFL linemen. Still, as discussed in previous chapters, access to this data is limited. The remainder of this chapter is composed of two sections. The first section (5.1) will be a review of the three papers mentioned above, and the other section (5.2) will be about future research and future work in areas of analytics regarding offensive linemen.

5.1 Review of Three Papers Related to Offensive Linemen

The first focus will be on the paper by Alamar and, Weinstein-Gould. This paper evaluates an individual lineman's effectiveness in giving the quarterback the most time possible in the area directly behind the offensive linemen to throw a pass, more familiarly known as the time in the pocket (TIP). TIP is more accurately the time from the snap of the ball to the time at which one of three different outcomes happens. The three outcomes are that the quarterback releases the pass, runs outside the area behind the lineman, or is tackled for a loss, a sack. The authors calculate each linemen's success rate by calculating two probabilities using logistic regression. The first is a success probability of an individual lineman based on TIP. The equation for their logistic regression was:

$$\Pr\left(s_{ij} = 1\right) = \frac{\exp(C_i + \beta_i TIP_j)}{1 + \exp(C_i + \beta_i TIP_j)} + e_{ij}$$

Where s is 1, i represents a specific player, and j represents a given play, C_i is a constant for player i, and, TIP is the time in pocket on play j. The second was a probability that a pass would be completed based on many factors, including down and, distance to first down, as well as if a lineman successfully maintained their block for the entire TIP. The logistic regression for this was:

$$\Pr\left(Comp_{t,j}=1\right)=\frac{\exp\left(C_t+\beta_t S_{t,j}+\theta_t X_{t,i}\right)}{1+\exp(C_t+\beta_t S_{t,j}+\theta_t X_{t,i})}+e_i$$

Where $Comp_{t,j}$ is equal to 1 if the pass on play j for team t was completed. C_t is a constant for team t, $s_{t,j}=\begin{cases}1 & if\ s_{i,j}=1\forall i\\ 0 & otherwise\end{cases}$ and, $X_{t,i}$ is a vector of variables for the play. The authors do not specifically state how either constant, C_i or C_t was calculated; however, they provide a value for one player and one team and seem to indicate that it is just a coefficient calculated by the logistic regression equation from their training data, and the value is then used in their test data when testing the equation.

To do this, the authors watched the entire first three games of the 2007 season for seven teams. They limited their analysis to only passing plays and excluded passing plays where a penalty was called. They used stopwatches to create the TIP, and each author recorded it separately; they compared the results of the recorded TIP with a 0.98 correlation between the two times. Next, the authors document if each lineman was successful in maintaining the block they started for the entire time; this was a binary variable of successful or not. Lastly, they used the logistic regression to equations to devise an estimated probability of success value for each lineman in the game based solely on TIP and a constant value calculated for each individual team based on the success value of that team's offensive line from the first logistic equations.

The first limitation of this paper is that their data and statistics are very time-consuming and not something that could be quickly developed as a universal method, as each team has a unique team constant that needs to be calculated for evaluating individual players. Limiting to only passing plays where no penalty was called reduces the total number of plays that could be analyzed. Also, by limiting it to passing plays effectiveness of a player as a run blocker is not measured, and run blocking may be more critical to a particular team. Lastly, limiting to seven teams over the course of three games creates smaller samples which may have influenced the results.

The paper by Alamar and Goldner was focused on improving the model just discussed for projecting the probability of completing a pass and attaching that to each individual offensive line position for 22 different teams during the 2010 season. Then utilizing that to estimate the total number of yards contributed to the teams passing game. For the data, they used play-by-play data from passing plays during the 2010 NFL season. Again the authors individually calculated TIP as well as also recorded down, distance to first down or goal line, and, if the offense was in a formation known as shotgun; shotgun formation is where the quarterback lines up about five yards behind the center and, the ball is sent through the air from the center to the quarterback. In the use of the play-by-play data, the authors were able to get other statistics that were useful for calculating success. The authors then noticed that not all linemen have failed before the ball is thrown or the play is otherwise over. Because of this, they chose to use a shared frailty random effects model as opposed to a simple regression model. However, they do not publish their equation or the coefficients for the variables. Once the probabilities estimated for each individual lineman were calculated, they then created a logit regression of completion. Lastly, they used this result to calculate the number of yards the player contributed to the teams passing game under the assumption that the player in question had played all 16 regular season games.

Again, the limitations of this article start with the data set would be not considering all 32 teams in one season. Also, not recording which linemen were in the game for each play when calculating TIP makes assumptions, and their prediction of yards added is more of a prediction of how many yards that particular offensive line position added to the teams passing yards. Lastly, no consideration is made of performance for rushing plays, thus preventing an accurate estimation of a player's ability overall as an offensive lineman.

The last paper was the one done by Byanna and, Klabjan, in which the authors are doing a more economic theory evaluation of the offensive linemen. This paper attempts to identify which linemen are overvalued, paid above average market value to peers at the same position, or undervalued, not paid average market value. In order to accomplish this, the authors first create a linear regression to determine characteristics pertaining to the salary of offensive linemen. For this step, the authors used play-by-play data for the 2013 and 2014 NFL seasons. They only analyzed linemen that appeared in at least one game during that time. They included penalties and descriptive variables (such as age, position, and number of years in the league, among others); they also used play-by-play rushing and passing statistics, including which side of the field the play happened. They also used which side of field play happened to calculate a differential statistic of if the play had happened to the side of the field the player usually plays based on position. Once they had all of their data, the authors used a stepwise regression model to help reduce independent variables. Also, they decided to exclude any player that was still on the first contract, which is also known as a rookie contract, or any player that had not been a free agent. Then they cluster players based on similar performance statistics using k means cluster analysis. They used cluster selection from one to twenty as the potential number of clusters to consider in conjunction with the Krzanowski-Lai statistic for determining the correct number of clusters to group players in; they decided to create seven clusters for the 133 unique players. Lastly, they compared each player's salary to the average other players in the same cluster and determined if there was a significant difference between the salary they received. In the end, they identified two players as being undervalued and two players overvalued.

As with other papers in this section, there are limitations present in the study. The first noticeable limitation is the lack of being able to determine how many plays each player was involved in during a game they appeared in, only that they played at least one snap. Also, when using the differential statistics calculation, no consideration was given to the fact that many running players will result in an offensive lineman "pulling," which is where they go from one side of the offensive line to the other to help block for the running back. Also, while a player may be listed as a left tackle, there may be situations in which the player actually switches to the other side or, in this case, plays right tackle. This happens on NFL teams for many reasons, including but not limited to injuries. Jeff Saturday was listed as center; however, he once said that he played every position on the offensive line at least one game during his career. Also, some of the clusters had less than twenty players in that cluster and, yet they used a t-test to assess if a salary was significantly different; the t-test assumes that the data is normally distributed and, with a sample size of less than twenty, there is no guarantee that this assumption is met.

Taking all three papers as a collection of analysis on offensive linemen into consideration, all papers did an excellent job of achieving what metric they intended to analyze and had reproducible and data-driven models that could be used to quantify the value of an individual offensive lineman in the NFL. However, all three also had significant limitations that could limit the useability of their model. As well as two of the papers only limited their analysis to passing plays by using information available through nflscrapR (Yurko et al., 2019); in 2010, one of the years of data considered, there were 19,159 passing plays and 14,1418 running plays, so while there are more passing plays during the season, it is still only 57% of the plays from scrimmage.

5.2 Future research

It is important to note that in September of 2022, the NFL started its sixth Big Data Bowl competition, with winners going to be announced at the NFL Draft in late April 2023. This Big Data Bowl was to develop statistics devoted to offensive and defensive linemen on passing plays using player tracking data. This competition may produce better statistics that create a sound ranking system for offensive linemen, but that has not yet been released to the public. However, again this is only based on passing plays, so even this will not be an encompassing rating system of a lineman's ability to do their entire job.

The most significant need would be to determine a repeatable, measurable value for offensive linemen that encompasses all aspects of the job, including run blocking, pass blocking, and special teams blocking (blocking on punts, field goals, and extra points). This could be done if access to the player tracking data was made available to the public. However, looking at which data is open to the public using play-by-play data along with looking at recording factors during the game may be a method of developing a ranking system that is based on video analysis. There are multiple methods available for video analysis techniques that could be used to evaluate an offensive lineman's performance during a game. However, even looking at just one season could be timely as there are around 45,000 plays in a single regular season of NFL games and, as such, is not feasible due to time constraints. Instead, future research using publicly available tracking data may be to utilize player tracking data during passing plays and the ability of a lineman to hold their block successfully for the entire TIP, then extrapolate this to running plays using play-by-play data and estimate if a lineman was successful in blocking the player they were most likely responsible for blocking for the amount of time for the player running the ball to get past that block point. This could be done by several statical model recommendations on at least one model will be discussed in chapter 7 of this paper.

CHAPTER VI: FOOTBALL ANALYTICS FOCUSED ON OTHER OFFENSIVE POSITIONS

As discussed in previous chapters, there is not many analytics currently published in academic journals that focus on positions other than the quarterback. Chapter 4 addressed the analytics for wide receivers, chapter 5 discussed offensive linemen publications, and this chapter will focus on three papers that were written that focus on other offensive positions and special teams, specifically placekickers (Pasteur & Cunningham-Rhoads, 2014), section 6.1, running backs (Blees, 2011), section 6.2, and tight ends (Mulholland & Jensen, 2014), section 6.3. While the papers by Blees and Mulholland & Jensen were first discussed in chapter 3 in relation to the draft, this chapter will delve further into the analytics that the papers used to develop the authors' system of rating the position.

6.1 Positional Analysis of Field Goal Kickers

When it comes to placekickers in the NFL, there may not be a more binary position in the eyes of fans. Place kickers are often placed in a position where at the end of the game, they are asked to kick a field goal which will decide the winner of the game. If the kicker makes the kick, then the team wins, fans celebrate, and yet most of the glory goes to the coach and quarterback. Whereas, if the kicker misses the field goal and his team loses, then the fans and, much time, the media all blame the kicker for the loss. Add to that the fact that during the regular season place, kickers are the only position on the field that routinely gets fired and replaced during the season by teams without an injury; it is easy to see why kickers are mostly overlooked when it comes to analytics. However, Pasteur and Cunningham-Rhoads decided to write their paper looking at kickers and producing an expectation-based metric to determine the probability of success of a field goal based on many factors.

For their data, the authors use all field goals attempted during the regular season and playoffs of the 2008, 2009, and 2010 NFL seasons. This gave them exactly 3000 field goal attempts to analyze, of which 82.5% of these field goals were successful. The authors decided to start by computing correlation coefficients between a binary field goal success variable and the other 20 variables they had in their data set. Of all these variables, distance had the highest correlation (r=-0.39), and the next was crosswind speed (r=-0.06).

The authors started with a simple logistic regression model based solely on the distance of the kick. Their final model equation is:

$$P(success) = \frac{1}{1 + e^{-5.8409+0.1078d}}$$

Where *d* is the distance of the field goal attempt, this gives the probability of success for all possible field goal distances (17 yards to 110 yards) between 98.22% and 0.24%. Although it is important to note that the longest field goal attempted was 76 yards (Austro, 2023), and the longest field goal made was 66 yards (Buller-Russ, 2023). The next step the authors performed was to look at which stadiums would be the easiest and hardest to kick in. This was done by taking the expected field goal percentage of all field goal kicks done by visiting kickers in a particular stadium and comparing it to the actual field goal percentage, they used only visiting kickers as home team kickers will account for about half the field goal attempts in a given season and the authors were using percentages based on "average" NFL kicker and not accounting for the specific kicker's ability.

The next step was to develop a logistic model using all data variables. Then using Mean Square Error as the deciding factor and five-fold cross-validation, they performed a variable selection to determine which variables were most important in determining the expected probability of success of a field goal. The factors they identified were: distance (yards), temperature (degrees Fahrenheit), the square of total wind plus crosswind speeds ($miles^2/hour^2$), Defense quality (points per game allowed), as well as binary variables for a fatigued kicker (more than five attempts in a game was 1), playing Denver (Denver is a very high altitude stadium and as such need to account for the altitude effect on the kick), if the game was during the playoff (playoffs mater because in theory, the pressure of performing in the playoffs is higher for all players), and if the kicker was the original kicker from the start of the season or a replacement kicker. The authors' final logistic model was:

$$P(FG\ Made) = \frac{1}{1 + e^{-(4.9769 + \sum c_i x_i)}}$$

Where c_i is the coefficient for each variable and x_i is the input variable. The last step the authors did was to create a metric for individual kickers, which they called points above replacement kicker (PARK), as well as the average distance of all made field goals by that individual kicker. Unlike other above replacement statistics mentioned previously, the PARK metric compares the kicker with all kickers that were tagged as a replacement kicker, meaning that the kicker did not start the season on the team's roster but instead compared to other players on the current roster. This is because most NFL teams only carry one kicker on their roster; therefore, a comparison to other kickers on the roster could not be made.

Limitations of the study are due to the fact that using logistic regression, their model would overestimate the probability of success for very long field goals. Also, there were limited attempts in certain conditions, so there is no proof that the metric was actually necessary or not; one example is that they had only 6% of all attempts had any recorded precipitation during the game, with less than 1% of all kicks having a precipitation rate of greater than 0.1 inches per hour. The other is that there was not any consideration for extra points. Extra points attempts are basically a 33-yard field goal try and so adding extra points as a 33-yard field goal attempt may improve their model.

Future work could be done to include extra points, looking at more seasons of data, and finally trying different modeling techniques, for example, K nearest neighbor or support vector machines. Both of these models could possibly provide a better method for classifying the success or not of field goal kicks. Lastly, developing a rating system that is not based on the above replacement may provide a more intuitive and easier-to-understand metric, such as the total QBR, which is on a zero to 100 scale.

6.2 Positional Analysis of Running Backs

The next paper was first discussed in Chapter 3 and was by Blees. This paper looks at using NFL draft and NFL combine data to predict the performance of a running back in the NFL. This chapter will focus on the performance metric that was created called the RBScore. This model is a simple linear equation of:

$$RBScore = TotalYards - 3(AllPlays) - 30(AllTurnovers) + 18(Touchdowns)$$

This gives a running back score in terms of yards. The total yards are total rushing yards from scrimmage plus total receiving by that running back. AllPlays is multiplied by three as it is expected that every play should gain at least 3 yards, so this metric takes rushing attempts plus passes caught and multiplies by three and thereby reduces the score by the minimum expected yards of a play. All turnovers are just the number of fumbles lost by the running back; the coefficient was stated to be taken from the book "Stumbling on Wins: Two Economists Expose the Pitfalls on the Road to Victory in Professional Sports, Portable Documents" by David Berri and Martin Schmidt in 2010. All value of 18 was determined by pro football reference (Drinen, 2008), which stated that the value of a touchdown scored is the same value as gaining 18 yards.

Overall, the RBScore is more of a prediction of the number of net yards that a running back would add to an offensive production over the course of a season. If you divided it by 17 (because there are 17 games in a regular NFL season per team), you would get the total predicted number of yards to be added by that running back to the offense for that game.

The major limitation of this RBScore is that there are not any bounds for the score, thereby limiting the ability to possibly determine how it matches with other positions and comparing running backs with different positions. Future work could look at how this RBScore could potentially be changed to put it on a set scale with a minimum and maximum value.

6.3 Positional Analysis of Tight Ends

Mulholland and Jensen's paper was also previously discussed in Chapter 3 for determining career success based on pre-draft data. For the ability to predict the success of tight ends in the NFL, they looked at three different models that looked at other pre-draft data to make a prediction based on linear regression. The three models made predictions for NFL Games Started, NFL Career Score, and NFL Career Score per Game.

For games started, they looked at all tight ends that were drafted between 1999 and 2010 and created a linear regression after getting a number of games started in the NFL. The ending of 2010 was so that there were at least three seasons of data for all tight ends. The games started based on NFL statistics and using adjusted R^2 as the determination as to which variables to include. They settled on the following equation.

$$
\begin{aligned}
&NFLGamesStarted \\
&\quad = -5189.42 + 61.08Height - 8.76Weight + 80.79BMI + 2.31broad\ jump \\
&\quad + 10.48BCSDummy + 1.87CollegeYardsPerReception \\
&\quad + 0.18CollegeReceptions - 26.98FinalYearCollegeYdsPercent \\
&\quad + 8.20FinalYearCollegeTDs
\end{aligned}
$$

Where *BCSDummy* is a binary variable indicating if the tight end played college at a BCS school or not, and *FinalYearCollegeYdsPercent* is the percentage of the total receiving yards the tight end had for their career gained during their final season in college (thus potentially isolating tight ends that just had one good college year vs. a good college career). The most significant limitation of this is that starting the game is just determined by who is on the field for the first play from offense, and depending on the team, may be zero, one, or two tight ends; a better estimate of actual performance may have been looking at total games played in which is easily obtained from sites that would report games started. Future work may be to use a more advanced regression equation such as LASSSO, which has been shown to produce more accurate results over standard ordinary least squares modeling(Ranstam & Cook, 2018).

The following equation they created was one for predicting an NFL career score. The equation for calculating NFL Career Score is as follows:

$$NFLCareerScore = CareerReceivingYards + 19.3CareerReceivingTouchdowns$$

With a coefficient of 19.4 based on equivalent yards added with a receiving touchdown. Once they had this score for all tight ends, the authors again used adjusted R^2 to determine the variables to be used. The final linear regression equation is:

$$NFLCareerScore = -4205.68 + 11.98Weight - 1694.69FourtyYardDash + 68.38BroadJump + 464.06BCSDummy + 107.55CollegeYardsPerReception + 14.48CollegeReceptions - 55.30CollegeTDs$$

The peculiar result of this equation is that a lower career score would be achieved with more touchdowns in college if all other variables were held constant, or in other words, scoring more touchdowns in college may mean less production in the NFL.

The last linear regression equation was for NFL Career Score per game started. While this could have been just taking two scores above and diving them, they decided to create a separate equation again using adjusted R^2 for selecting variables; as such, their linear regression score for calculating NFL Career Score per game is:

$$NFLCareerScorePerGame = 6.99 - 12.50FourtyYardDash + 0.44BenchPress + 0.37BroadJump + 6.31BCSDummy + 0.01CollegeYards - 0.35CollegeTDs$$

The most significant limitation again is that there is no maximum or minimum value that could be potential from any of these equations; therefore, no easy way to compare to other positions other than tight end. Also, these equations do not account for the fact that many tight ends in the NFL are actually starting or playing lots of plays for their ability to be good blockers and act as a sixth lineman on some plays.

Future work could be done to find a way to quantify blocking ability, perhaps with the availability of player tracking data, also making sure to account for entire careers and only including tight ends that have had at least one season after being drafted with no game appearances.

CHAPTER VII: FUTURE WORK

Throughout this thesis, we have reviewed many different current works in sports analytics and made suggestions on areas in that future work could potentially provide improvements. In the article "Football's Hilbert Problems," Aaron Schatz identifies the ten biggest areas that need to be addressed by NFL to improve analytics (Schatz, 2005). Of these ten biggest areas, none have been fully addressed, and all pose potential for future work that can be done once the data is available. This chapter is devoted to three of those suggestions in more detail, further explaining possible methods and why this would help. The three areas specifically are creating a rating system for wide receivers that has a maximum and minimum value (Section 7.1), creating an equally scaled rating system for individual offensive linemen (Section 7.2), and looking at creating a better draft ranking system that may help identify undervalued and overvalued players entering the draft (Section 7.3).

7.1 Wide Receiver Rating System

As discussed in chapter 4, wide receivers are becoming more and more critical players when building an offense in the NFL. As such, the need to develop a rating system for wide receivers that has a maximum and minimum value as well as a scale that is easy to understand is vital for decision-makers to use to help evaluate talent and potentially find a player that is going to fit into their offense and help the wide receiver know where he stands amongst the other wide receivers in the NFL. Also, the responsibilities and expectations of wide receivers have increased in recent years. This is seen in expectations to block on running plays, run the ball themselves, block for other wide receivers on screenplays, and run-pass option, RPO, plays. Some of these responsibilities are easy to quantify as they show up in the play-by-play data. However, other abilities that may not show up in play-by-play data need to be used in any attempt to calculate an accurate value of a wide receiver to their team.

Some examples of this are wide receivers who are able to successfully block for a teammate on a screen pass or running play. Also, if a wide receiver is constantly getting double-teamed, this creates an increased likelihood that one of the other receivers would be able to get open, thus adding value to the team. Using play-by-play data that can be easily obtained by using nflscrapR discussed earlier, we can evaluate wide receivers on their abilities that show up as statistics for that player. Also, we can use the player tracking data made publicly available for the Big Data Bowl competition, the first eight weeks of the 2021 season, all passing plays for all teams. Using the player tracking data, the PFF data provided, as well as the play-by-play data, we can develop a more robust rating.

As for modeling techniques, one possibility is to look at variables that are important and calculate percentages of that individuals compared to the maximum recorded. For example, if looking at receiving yards during a single NFL season, you could have the yards the receiver in question got divided by 1,964, which is the current record for most receiving yards in a season (Drinen, 2022). Then take these percentages and average them out by assigning weights based on covariance factors or potentially an R-squared value of how much that variable adds to the overall accuracy of the predictor. However, a better model would probably look to use the fact that there is not currently a rating system to learn from the model and utilize an unsupervised model. Without a deep dive into the data and experimenting with different techniques suggesting the best model is not possible at this time. However, some suggestions as places to start would be to use a k-means clustering method and then determine the scale for each cluster and assign a maximum and minimum score for each cluster; then, based on the likelihood of belonging to that cluster, could give you the score for each person. It is a recommendation to use a standard scale that can be easily interpreted, such as a rating system from zero to one hundred.

There are many potential ways to perform the modeling. The first thing is to consider what data is available and can be used. As of the writing of this thesis, publicly available data is limited to all passing plays from the first eight weeks of the 2020 NFL season, all regular season and playoff play-by-play data starting in the 1999 season through the current season, also pro-football-reference and NFL.com have many statistics including single game, season, and career data. The next step would be to use all available data in conjunction with correlation or other factors to determine which variables are most significant. Then comes the need to select modeling techniques, as there is not currently a rating system available this leads to the need for unsupervised learning models due to the lack of dependent variables. A few models that have been shown in the past to be successful in this situation could potentially be clustering techniques such as K Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), or Expectation-Maximization (EM). Other unsupervised methods could also be tried, such as XGBoost, or Random Forest; both of these are tree-based models that would group the receivers into categories. Most likely XGBoost would be preferred as we can set the depth of tree splits where random forests are usually grown to full depth.

7.2 Offensive Lineman Rating System

As discussed in Chapter 5, offensive linemen present the most challenging position to create a rating system for as they themselves do not have any statistics that are put on them individually. However, offensive linemen deserve some form of rating system, as they are a massive factor in the overall success of the offense. Also, without a universally accepted rating system looking at every player on an equal scale, how do we determine value to help make decisions if a team was in need of a correct tackle and center and had money to sign one free agent having a system that could compare the value with a single number may aid in the team's decision making as to which free agent to sign.

Regarding the responsibilities of offensive linemen in general, the center is also responsible for snapping the ball; there are two different tasks they are asked to do depending on the play. For passing plays, they are asked to protect the quarterback; this is done by stopping defenders who are trying to get to the passer and tackle them before they can throw the pass. A method of looking at the player tracking data could be developed to determine a value for their ability to do this; let's call it stopping force. Using player tracking data, we can get acceleration; utilizing the team's roster, we can get approximate weight for a defender as well as an offensive lineman. Then we can use simple Newtonian physics of force equation $F = ma$ and calculate a difference in force between the moment before acceleration slows and when acceleration reaches a minimum. This stopping force could be used to help rate offensive linemen in a quantitative manner. The other responsibility of linemen is to use their skills to move defenders out of the planned path of the ball carrier on running plays. Again, using player tracking data, we could come up with a binary variable of if the lineman was able to sustain a block for the time between contact and when the ball carrier had passed that specific area of the field. Then as we would have multiple binary variables, we could create a logistic regression for determining the probability of success for each lineman using the logistic regression equation:

$$\pi(\boldsymbol{X}) = \frac{1}{1 + \exp(-\boldsymbol{X}\boldsymbol{\beta})}$$

Where $\pi(\boldsymbol{X})$ is the success probability, $\boldsymbol{\beta}$ is the vector of weights for each, and **X** is the vector of values that are found to be statistically relevant.

Then you could use the given percentage of passing plays vs. running plays for that particular lineman's team to determine a weighting system and put the score on a scale between zero and one hundred again.

7.3 Draft Ranking system

The last area of future work that could be done is to develop a better draft ranking system to help teams better rank prospects. The development of a universal rating system would help in this process. The other idea that may work is to use the fact that currently, the NFL rules prevent a player from being drafted until three years after their high school graduation. Due to this rule, most players wanting to play in the NFL play college football for three years. This fact would lend itself to evaluating the college statistics to help determine a player's ability. In my research for this paper there was not one attempt to look at the college statistics as a time series. Time series analysis removes the independent assumption and can help us look at the 3 to 5 years of playing in college as not independent for a player. This thought is very logical because a player that shows promise and potential early on will get more playing time and more opportunities as their career in college progress. Also, time series evaluations would help identify players who are trending up, staying the same, or possibly even trending down. Using a college statistic such as total touchdowns and time series modeling packages, such as astsa, can be used to get the trend data which would be in the result of an equation such as:

$$Y[t] = T[t] + S[t] + e[t]$$

Where Y[t] is the data, T[t] would be the trend component. If this were close to zero, we would know that a player was consistent for their college career, where a negative number would indicate a tend down, and a positive number would show the player's career was trending up.

Of course, looking at facts, such as seeing if a player was injured at any time, would need to be considered as well. Once we have trend data as well as a rating system should potentially help teams identify players that may have had just one great season or players that have been consistently improving, thus providing the possibility that although statistically, they are not as good as the player with one good year but they are trending up and improving showing that they would most likely be able to continue to improve once they joined an NFL roster.

7.4 Conclusion

Sports analytics is a field that is growing every year with the advancement of computer technology, methodology for analytics, and the data being collected and access to this data. This is true for professional football in the NFL. The player tracking data is becoming more and more advanced and more and more accurate as older stadiums are getting upgrades that allow accuracy in the data; early tracking data could not even track the football in all stadiums. Allowing for access to more and more of this data will help improve analytics in football. The most significant area needed is a universal player rating system that will enable us to have a same scale rating system that is designed for each position group but has the same scale for all other position groups so that the ability to compare players across positions can become more accessible for coaches, front offices, and fans.

REFERENCES

Alamar, B. (2013). *Sports Analytics*. Columbia University Press. https://doi.org/10.7312/alam16292

Alamar, B. C. (2006). The Passing Premium Puzzle. *Journal of Quantitative Analysis in Sports*, *2*(4). https://doi.org/10.2202/1559-0410.1051

Alamar, B. C., & Weinstein-Gould, J. (2008). Isolating the Effect of Individual Linemen on the Passing Game in the National Football League. *Journal of Quantitative Analysis in Sports*, *4*(2). https://doi.org/10.2202/1559-0410.1113

Alamar, B., & Goldner, K. (2011). The Blindside Project: Measuring the Impact of Individual Offensive Linemen. *CHANCE*, *24*(4), 25–29. https://doi.org/10.1080/09332480.2011.10739883

Albert, J. (1993). A Statistical Analysis of Hitting Streaks in Baseball: Comment. *Journal of the American Statistical Association*, *88*(424), 1184–1188. JSTOR. https://doi.org/10.2307/2291255

Amazon Web Services. (2023, January 17). *NFL Next Gen Stats Powered by AWS*. Amazon Web Services. https://aws.amazon.com/sports/nfl/

Annis, D. H. (2006). Optimal End-Game Strategy in Basketball. In *Journal of Quantitative Analysis in Sports* (Vol. 2, Issue 2). https://doi.org/10.2202/1559-0410.1030

Austro, B. (2023, February 17). *Longest missed field goals – Quirky Research*. Quirky Research. https://www.quirkyresearch.com/football-lists/longest-missed-field-goals/

Baker, R. D., & McHale, I. G. (2013). Forecasting exact scores in National Football League games. *International Journal of Forecasting*, *29*(1), 122–130. https://doi.org/10.1016/j.ijforecast.2012.07.002

Baker, R. E., & Kwartler, T. (2015). Sport Analytics: Using Open-Source Logistic Regression Software to Classify Upcoming Play Type in the NFL. *Journal of Applied Sport Management*, *7*(2), 43–58. https://doi.org/10.7290/jasm

Baldwin, A., Broadie, M., Zelezinski, D., & Whaley, S. (2023, March 4). *The New Age of Golf: Players vs the game*. Sloan Sports Analytics Conference, Boston, MA.

Baumer, B. S., Jensen, S. T., & Matthews, G. J. (2015). openWAR: An open source system for evaluating overall player performance in major league baseball. *Journal of Quantitative Analysis in Sports*, *11*(2), 69–84. https://doi.org/10.1515/jqas-2014-0098

Blees, C. (2011). *Running Backs in the NFL Draft and NFL Combine: Can Performance be Predicted?*

Boronico, J. S., & Newbert, S. L. (1999). Play Calling Strategy in American Football: A Game-Theoretic Stochastic Dynamic Programming Approach. *Journal of Sport Management*, *13*(2), 103–113. https://doi.org/10.1123/jsm.13.2.103

Boulier, B. L., Stekler, H. O., Coburn, J., & Rankins, T. (2010). Evaluating National Football League draft choices: The passing game. *International Journal of Forecasting*, *26*(3), 589–605. https://doi.org/10.1016/j.ijforecast.2009.10.009

Buller-Russ, A. (2023, February 12). *10 longest field goals in NFL history*. Sportsnaut. https://sportsnaut.com/longest-field-goal-in-nfl-history/

Byanna, N., & Klabjan, D. (2016). Evaluating the Performance of Offensive Linemen in the NFL. *ArXiv*. https://doi.org/1603.07593

Carter, C., & Chadiha, J. (2013). *Going Deep: How Wide Receivers became the most compelling figures in pro sports.* Hyperion.

Carter, V., & Machol, R. E. (1971). Technical Note—Operations Research on Football. *Operations Research*, *19*(2), 541–544. https://doi.org/10.1287/opre.19.2.541

Carter, V., & Machol, R. E. (1978). Optimal Strategies on Fourth Down. *Management Science*, *24*(16), 1758–1762. https://doi.org/10.1287/mnsc.24.16.1758

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chu, D., Reyers, M., Thomson, J., & Wu, L. Y. (2020). Route identification in the National Football League: An application of model-based curve clustering using the EM algorithm. *Journal of Quantitative Analysis in Sports*, *16*(2), 121–132. https://doi.org/10.1515/jqas-2019-0047

Collins, M., Moir, G., Muresan, J., & Savage, D. (2017). *NFL Data Analytics For A New Era*.

Cowen Partners. (n.d.). *Understanding the wonderlic test*. Cowen Partners. Retrieved February 8, 2023, from https://cowenpartners.com/understanding-the-wonderlic-test/#:~:text=The%20Wonderlic%20Test%20is%20a,can%20think%20and%20solve%20problems.

Deaton, J. P. (2008, August 4). *10 Everyday Car Technologies That Came From Racing*. HowStuffWorks. https://auto.howstuffworks.com/under-the-hood/trends-innovations/top-10-car-tech-from-racing.htm

Deshpande, S. K., & Evans, K. (2020). Expected hypothetical completion probability. *Journal of Quantitative Analysis in Sports*, *16*(2), 85–94. https://doi.org/10.1515/jqas-2019-0050

Dhar, A. (n.d.). *Drafting NFL Wide Receivers: Hit or Miss*.

Drinen, D. (n.d.). *Aaron Rodgers Stats, Height, Weight, Position, Draft, College*. Pro-Football-Reference.Com. Retrieved February 9, 2023, from https://www.pro-football-reference.com/players/R/RodgAa00.htm

Drinen, D. (2008, April 29). *Approximate Value*. Pro Football Reference. https://www.pro-football-reference.com/about/approximate_value.htm

Drinen, D. (2022, February 10). *NFL Receiving Yards Single-Season Leaders*. Pro-Football-Reference.Com. https://www.pro-football-reference.com/leaders/rec_yds_single_season.htm

Eicken, A. von. (2017). *An Analysis of the Efficiency of the NFL Draft*.

Elam, N. (2020). An Examination of the Effectiveness of the Elam Ending at TBT2018 and TBT2019. *Sports Innovation Journal*, *1*, 120–133. https://doi.org/10.18060/23969

Fernandes, C. J., Yakubov, R., Li, Y., Prasad, A. K., & Chan, T. C. Y. (2020). Predicting plays in the National Football League. *Journal of Sports Analytics*, *6*(1), 35–43. https://doi.org/10.3233/JSA-190348

FF Today. (2022, May 1). *NFL Draft Tracker Detroit Lions, 1995 to 2022—FF Today*. https://fftoday.com/nfl/drafttracker.php?o=by_team&TeamID=9020

Football operations. (2023). *The history of the NFL Draft*. NFL Football Operations. https://operations.nfl.com/journey-to-the-nfl/the-nfl-draft/the-history-of-the-draft/#:~:text=At%20a%201935%20league%20meeting,inaugural%20NFL%20Draft%20in%201936.

Hall Of Fame. (n.d.). *Pro Football Hall of Fame's Class of 2016 Announced | Pro Football Hall of Fame Official Site*. Pfhof. Retrieved February 9, 2023, from https://www.profootballhof.com/news/2016/02/pro-football-hall-of-fames-class-of-2016-announced

Heiny, E. L., & Blevins, D. (2011). Predicting the Atlanta Falcons Play-Calling Using Discriminant Analysis. *Journal of Quantitative Analysis in Sports*, *7*(3). https://doi.org/10.2202/1559-0410.1230

Imbens, G. W., & Rubin, D. B. (2010). Rubin Causal Model. In S. N. Durlauf & L. E. Blume (Eds.), *Microeconometrics* (pp. 229–241). Palgrave Macmillan UK. https://doi.org/10.1057/9780230280816_28

Indeed.com. (2023, January 17). *What Is Sports Analytics? (Definition, Importance, and Tips)*. Indeed Career Guide. https://ca.indeed.com/career-advice/finding-a-job/what-is-sports-analytics

Johnson, K., & Smith, S. (2009). *Just Give Me the Damn Ball!: The Fast Times and Hard Knocks of an NFL Rookie*. Grand Central Publishing.

Jordan, J. D., Melouk, S. H., & Perry, M. B. (2009). Optimizing Football Game Play Calling. *Journal of Quantitative Analysis in Sports*, *5*(2). https://doi.org/10.2202/1559-0410.1176

Kansas City Chiefs. (n.d.). *Chiefs Depth Chart*. Kansa City Chiefs. Retrieved February 8, 2023, from https://www.chiefs.com/team/depth-chart

Katz, S., & Burke, B. (2016, September 8). *How is Total QBR calculated? We explain our quarterback rating*. ESPN.Com. https://www.espn.com/blog/statsinfo/post/_/id/123701/how-is-total-qbr-calculated-we-explain-our-quarterback-rating

Kinney, M. (2020). Template matching route classification. *Journal of Quantitative Analysis in Sports*, *16*(2), 133–142. https://doi.org/10.1515/jqas-2019-0051

Kiper, M. (2023, January 25). *2023 NFL mock draft: Mel Kiper's first-round predictions for 31 picks*. ESPN. https://www.espn.com/nfl/insider/draft2023/insider/story/_/id/35514084/2023-nfl-mock-draft-mel-kiper-first-round-predictions-31-picks-best-team-fits-top-prospects

Kuzmits, F. E., & Adams, A. J. (2008). The NFL Combine: Does It Predict Performance in the National Football League? *The Journal of Strength & Conditioning Research*, *22*(6), 1721. https://doi.org/10.1519/JSC.0b013e318185f09d

Lee, N., & Kitani, K. M. (2016). Predicting wide receiver trajectories in american football. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–9.

Lee, P., Chen, R., & Lakshman, V. (2016). *Predicting Offensive Play Types in the National Football League*. http://cs229.stanford.edu/proj2016/report/LeeChenLakshman-PredictingOffensivePlayTypesIntheNFL-report.pdf

Lock, D., & Nettleton, D. (2014). Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, *10*(2). https://doi.org/10.1515/jqas-2013-0100

Lopez, M. J. (2020). Bigger data, better questions, and a return to fourth down behavior: An introduction to a special issue on tracking data in the National football League. *Journal of Quantitative Analysis in Sports*, *16*(2), 73–79. https://doi.org/10.1515/jqas-2020-0057

Lyons, B. D., Hoffman, B. J., & Michel, J. W. (2009). Not Much More Than *g* ? An Examination of the Impact of Intelligence on NFL Performance. *Human Performance*, *22*(3), 225–245. https://doi.org/10.1080/08959280902970401

Martinez, K. L. (2010, February 5). *Meridian Magazine: Sports :NCAA Football Tournament: An Imagined Solution to a Real Problem*.

https://web.archive.org/web/20100205130213/http://meridianmagazine.com/sports/031128ncaa.html

McKenzie, G. (2015). Data analytics in sports: Improving the accuracy of NFL draft selection using supervised learning [M.S., University of Missouri - Columbia]. In *ProQuest Dissertations and Theses* (1842666793). ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/data-analytics-sports-improving-accuracy-nfl/docview/1842666793/se-2?accountid=10003

Molter, M. (2023, January 3). *How Much Do Americans Bet on Sports*. Legal Sports Betting. https://www.legalsportsbetting.com/how-much-money-do-americans-bet-on-sports

Mulholland, J., & Jensen, S. T. (2014). Predicting the draft and career success of tight ends in the National Football League. *Journal of Quantitative Analysis in Sports*, *10*(4). https://doi.org/10.1515/jqas-2013-0134

Mulholland, J., & Jensen, S. T. (2016). Projecting the Draft and NFL Performance of Wide Receiver and Tight End Prospects. *CHANCE*, *29*(4), 24–31. https://doi.org/10.1080/09332480.2016.1263095

NFL. (2020). *How are players selected for the NFL Combine*. Scouting Combine. http://www.nflcombine.net/faq/how-are-players-selected-for-the-nfl-combine/

NFL. (2023a, January 12). *NFL Next Gen Stats Game Center*. NFL Next Gen Stats. https://nextgenstats.nfl.com/stats/game-center

NFL. (2023b, January 21). *NFL Big Data Bowl | NFL Football Operations*. https://operations.nfl.com/gameday/analytics/big-data-bowl/

NFL Big Data Bowl 19. (2019, March 1). *2019 Big Data Bowl Winners Announced | NFL Football Operations*. https://operations.nfl.com/updates/the-game/2019-big-data-bowl-winners-announced/

NFL Big Data Bowl 20. (2020, March 1). *NFL Big Data Bowl 2020*. https://kaggle.com/competitions/nfl-big-data-bowl-2020

NFL Big Data Bowl 21. (2021, March 1). *NFL Big Data Bowl 2021*. https://kaggle.com/competitions/nfl-big-data-bowl-2021

NFL Big Data Bowl 22. (2022, March 1). *NFL Big Data Bowl 2022*. https://kaggle.com/competitions/nfl-big-data-bowl-2022

NFL Combine Results. (2022, April 9). *NFL Combine Results*. NFL Combine. https://nflcombineresults.com/

NFL Players Assoc. (2022, June 5). *Top 50 NFL Player Sales List—March 1, 2022-May 31, 2022*. NFL Players Association. https://nflpa.com/partners/posts/top-50-nfl-player-sales-list-march-1-2022-may-31-2022

NFL.com. (2023, August 6). *NFL Quarterback Rating*. NFL.Com. https://web.archive.org/web/20110814052052/http://www.nfl.com/help/quarterbackrating formula

Ötting, M. (2021). Predicting play calls in the National Football League using hidden Markov models. *IMA Journal of Management Mathematics*, *32*(4), 535–545. https://doi.org/10.1093/imaman/dpab005

Palmer, P., Pullis, K., & Gillette, G. (2021, November 25). *Approximate Value*. Pro Football Reference. https://www.pro-football-reference.com/about/approximate_value.htm

Pasteur, R. D., & Cunningham-Rhoads, K. (2014). An expectation-based metric for NFL field goal kickers. *Journal of Quantitative Analysis in Sports*, *10*(1). https://doi.org/10.1515/jqas-2013-0039

Phinney, S. D. (2004). Ketogenic diets and physical performance. *Nutrition & Metabolism*, *1*(1), 2. https://doi.org/10.1186/1743-7075-1-2

Pitts, J. D., & Evans, B. (2018). Evidence on the importance of cognitive ability tests for NFL quarterbacks: What are the relationships among Wonderlic scores, draft positions and NFL performance outcomes? *Applied Economics*, *50*(27), 2957–2966. https://doi.org/10.1080/00036846.2017.1412081

Pro-Football Reference. (2022). *Deebo Samuel Stats, Height, Weight, Position, Draft, College*. Pro-Football-Reference.Com. https://www.pro-football-reference.com/players/S/SamuDe00.htm

Quenzel, J., & Shea, P. (2016). Predicting the Winner of Tied National Football League Games: Do the Details Matter? *Journal of Sports Economics*, *17*(7), 661–671. https://doi.org/10.1177/1527002514539688

Quinn, K. G., Geier, M., & Berkovitz, A. (n.d.). *Passing on Success? Productivity Outcomes for Quarterbacks Chosen in the 1999-2004 National Football League Player Entry Drafts*.

Ranstam, J., & Cook, J. A. (2018). LASSO regression. *British Journal of Surgery*, *105*(10), 1348. https://doi.org/10.1002/bjs.10895

Romer, D. (2006). Do Firms Maximize? Evidence from Professional Football. *Journal of Political Economy*, *114*(2), 340–365. https://doi.org/10.1086/501171

Roumani, Y. F. (2022). Sports analytics in the NFL: Classifying the winner of the superbowl. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-05063-x

Sahi, S., & Shubik, M. (1988). A model of a sudden-death field-goal football game as a sequential duel. *Mathematical Social Sciences*, *15*(3), 205–215. https://doi.org/10.1016/0165-4896(88)90007-8

Schatz, A. (2005). Football's Hilbert Problems. *Journal of Quantitative Analysis in Sports*, *1*(1). https://doi.org/10.2202/1559-0410.1010

Schuckers, M. (2011). An Alternative to the NFL Draft Pick Value Chart Based upon Player Performance. *Journal of Quantitative Analysis in Sports*, *7*(2). https://doi.org/10.2202/1559-0410.1329

Schuckers, M. E., & Argeris, S. (2015). You can beat the "market": Estimating the return on investment for NHL team scouting. *Journal of Sports Analytics*, *1*(2), 111–119.

Sportrac. (2023a, January 5). *NFL Positional Payrolls*. Spotrac.Com. https://www.spotrac.com/nfl/positional/

Sportrac. (2023b, February). *NFL Positional Payrolls*. Spotrac.Com. https://www.spotrac.com/nfl/positional/

Sports Reference LLC. (2023). *Complete List of College Football National Champions with Seasons*. College Football at Sports-Reference.Com. https://www.sports-reference.com/cfb/years/index.html

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Treme, J., & Allen, S. (2009). Widely Received: Payoffs to Player Attributes in the NFL. *Economics Bulletin*, *29*, 1631–1643.

Wolfson, J., Addona, V., & Schmicker, R. (2017). *Forecasting the Performance of College Prospects Selected in the National Football League Draft*.

Wolfson, J., Addona, V., & Schmicker, R. H. (2011). The Quarterback Prediction Problem: Forecasting the Performance of College Quarterbacks Selected in the NFL Draft. *Journal of Quantitative Analysis in Sports*, *7*(3). https://doi.org/10.2202/1559-0410.1302

Yam, D. R., & Lopez, M. J. (2019). What was lost? A causal estimate of fourth down behavior in the National Football League. *Journal of Sports Analytics*, *5*(3), 153–167. https://doi.org/10.3233/JSA-190294

Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechrinis, K., & Ventura, S. L. (2020). Going deep: Models for continuous-time within-play valuation of game outcomes in American football with tracking data. *Journal of Quantitative Analysis in Sports*, *16*(2), 163–182. https://doi.org/10.1515/jqas-2019-0056

Yurko, R., Ventura, S., & Horowitz, M. (2019). nflWAR: A reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, *15*(3), 163–183. https://doi.org/10.1515/jqas-2018-0010