SMU Data Science Review

Volume 1 | Number 2 Article 6

2018

Predicting Game Day Outcomes in National Football League Games

Josh Klein
Southern Methodist University, jdklein@smu.edu

Anna Frowein
Southern Methodist University, afrowein@smu.edu

Chris Irwin
Southern Methodist University, cmirwin@smu.edu

Predicting Game Day Outcomes in National Football League Games

Christopher Irwin[1], Josh Klein[1], Anna Frowein[1]
Joshua Koen Ph.D.[2]

[1] Master of Science in Data Science, Southern Methodist University,
6425 Boaz Lane, Dallas, TX 75205
{cmirwin, jdklein, afrowein}@smu.edu,
[2] University of Texas at Dalls, 1660 Viceroy Dr, Dallas, TX 75235
joshua.koen@utdallas.edu

Abstract. In this paper, we present a model for predicting the game day outcomes of National Football League games. Three of the most popular models, AccuScore, Madden, and FiveThirtyEight, for generating game day predictions are analyzed for comparison. Player data and outcomes from previous games are used, but we also incorporate several weather factors into our models. Over 1,700 games were incorporated, and 3 logistic regression models with different sets of attributes: (1) all attributes, (2) feature reduction with principal component analysis, and (2) feature reduction using recursive elimination. We also discuss potential ethical issues of using data science techniques by knolwegable individuals to gain an advantage over others lacking training in data science techniques.

1 Introduction

We present a model that predicts game day outcomes, that is the winning and losing team, for games played in the National Football League (NFL). Predictive modeling in sporting events is currently used for a variety of purposes, including reporting, gambling and in fantasy football leagues. Although the available models (see below) perform well at about 74% accuracy, there is room for improvement. In this report, we propose and develop a model that incorporates additional critical factors that are largely omitted from currently available models. Specifically, we incorporate weather statistics as weather conditions potentially play a critical role in the outcome of NFL games. The model is validated using data from over 1,792 NFL games occurring between 2009 and 2016. We will use accuracy to measure the results of our models, due to the fact that the current available models only provide accuracy as the measure for comparison.

The three currently available models comprise AccuScore, EA Sports Madden, and FiveThirtyEight. Each model focuses on different sets of limited amount of attributes to predict NFL game outcomes. AccuScore looks at a combination of players, team composition, weather, coaching and other factors. Madden focuses heavily on the individual players and predicting the outcome of interactions between players based on the players' attributes. In contrast, FiveThirtyEight looks at the historical team performance while ignoring the individual players. Although these models can predict game day outcomes above chance (e.g., 74%), performance of these models is not perfect. We seek to improve performance above these existing models using factors that likely have a strong influence game day outcomes, such as weather, travel time, or the time of day the game is played.

There are potential ethical issues that should be considered regarding the use of predictive modeling. Specifically, the use of predictive models can lead to different parties, individuals or organizations, having unfair advantages in either gambling or fantasy leagues that require a buy in to participate. In particular, individuals or organizations that make use of training in data science and analytics might have an unfair advantage if they were to participate in recreational gambling or fantasy leagues. Although there are regulations as to what information can be used in legal gambling venues (e.g., Las Vegas), these regulations are not applicable to all forms of gambling. For instance, using NFL player statistics for fantasy leagues is not considered illegal under the current regulations. This is partly due to the assumption that those who participate in fantasy leagues have a knowledge of the game and have to use this knowledge in a skillful way in order to be successful in the league. However, Congress and State legislators are starting reviews of current gambling regulations to create a fair and even playing field for those who participate. There are possible questions that arise from the legalization of sports gambling, such as balancing the benefits of an increase in government revenues via taxes against the cost of potential increases in cases of gambling addiction. As data analytic techniques become more mainstream in the population, the use of them in the context of gambiling becomes ever more important as it could result in the costs of gambling legalization drastically outweighing any potential benefits.

2 History and Rules of National Football League

The National Football League (NFL) has evolved since 1869 to include 32 different teams and the rules each year are reviewed and revised as appropriate. Currently, games are 60 minutes longs played in four quarters, teams score points through touchdowns, field goals and safeties.

2.1 History and Current Rules

The NFL started in 1869, with a college game of soccer football between Rutgers and Princeton, using a modified version of the London Football Association rules. After this game and in the years, that followed the game Rugby became more popular than soccer with schools on the east coast and this would form the base of modern day football. In 1876, the first official rules for football were written and Walter Camp, a player and coach who became known as the “Father as American Football,” became involved in the game. The first player to be paid for playing was William Heffelfinger, who was paid $500 to play in a single game in 1892. In the following years, more players started to receive compensation for playing the game and professional teams and owners were developed [1].

While the teams and players started to form the rules also changed as years progressed in the sport. Goal points were changed in 1898 and the forward pass was introduced or legalized in 1906 [1]. Throughout the years the rules have been reviewed and changed; the current simplified rules of game play are as follows[1]:

1. Points:
    a. Touchdown 6 points
    b. Field Goal 3 points
    c. Safety 2 points
    d. After a touchdown: 1 point for a safety or field goal or two points for another touchdown
2. Downs – are played when a ball is put into play and end when the ball is determined to be dead.
3. A team has a series of downs to make a line to gain, which is defined as 10 yards or the goal line whichever is closer, a team has four downs in a series to meet the line to gain, or else a new series is declared for the other team.
4. Game time: games are 60 minutes long broken up into 15-minute quarters, the time is stopped for certain events to allot time to physical game play.

2.2 Changes in the Game

The NFL is an evolving league where rule changes translate into adaptive strategies and players to support those strategies. The earliest evolution of player personnel came in the 1940s when player substitutions were first allowed [26]. This lead to the position specialization we see in football across all levels and leagues. Specific talents were needed for individual positions instead of all players being able to play every position interchangeably. The next evolution happened around the late 1960s after the popularity of the NFL exploded and along with it came increased money from sponsorships, television contracts, merchandising, and packed stadiums [26]. Football players no longer had to work offseason jobs and could make a career out of playing in the league.

With a highly popular, highly profitable product on the field in the NFL, interest began to trickle down to the college, high school, and even youth leagues. Players began to learn strategy and technique younger than ever before and the excitement of playing in front of tens-of-thousands of people created the demand to be the biggest, fastest, and smartest player.

According to data compiled by Craig Booth, a data scientist that focuses on sports analytics, the average weight of players at each position has increased since 1950 [27]. Some positions for example running backs and linebackers haven't seen huge increases and at times show fairly flat growth lines. However, offensive tackles, guards, and centers have shown massive increases in the average weights of the players (Figure 1). Offensive tackles have increased in weight by almost 30% since 1950. In light of the increase in average weights and size, it is widely accepted that today's players are more athletic than those in previous eras.

---

1 For the complete 2017 NFL Rulebook please see https://operations.nfl.com/the-rules/2017-nfl-rulebook/#2017-rule-changes

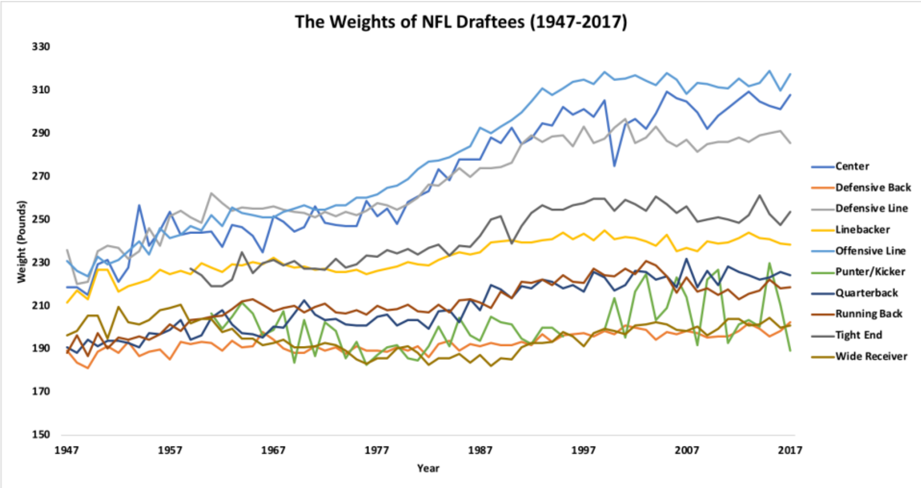

Figure 1: Average weights (lbs) of NFL draftees in each position from 1947-2017.

Given how rule changes and player sizes have developed over the years, comparison of statistics from an individual player in a position (e.g., quarterback) from 1975 to a player in the same position in 2005 is not necessarily commisurate. The models we have explored do not use statistics prior to 2009 and would not be applicable to earlier games given that the characteristics of how football is played have fundamentally changed over time. That same logic can also be used looking towards the future as football continues to evolve. Use of these models in either an academic or business setting would need to be constantly examined and updated to capture the essence of the game at that point of time. We are not advocating the use of a specific window of statistics (e.g., previous 5 years of data), but a realization that statistics from a given year is not necessarily an accurate representation of earlier (or future) years.

3 Gambling and Fantasy Leagues

There are many ways in which sport statistics are used, such as betting websites (e.g., OddsShark[2]) and fantasy leagues (e.g., NFL Fantasy[3]). Congress is currently reviewing the federal gambling regulations to account for fantasy sports and other online betting venues [2]. Given that sports betting is starting to become more main stream acceptable, States and the Federal government are seeing the chance at the potential revenues in legalizing it. At present, the only state that allows for single game betting is Nevada, and other types of regulated betting (e.g.,Indian Gambling Regulatory Act) is limited to only a handful of states [2]. A review of the laws, such as Unlawful Internet Gambling Enforcement Act of 2006, Professional and Amateur Sports Protection Act of 1992 and the Wire Act of 1961, specify a fair playing field for both online and offline sports betting, including fantasy sports leagues [2]. Fantasy leagues are considered at

2 https://www.oddsshark.com/sportsbook-review
3 http://fantasy.nfl.com/

their core a form of sports betting, however, are not as strictly regulated by the laws because they take skill and knowledge from those participating for them to win and not just chance [3].

Several companies are preparing for the potential widespread legalization of sports bettering, including MGM, William Hill, Native American and various state lotteries. These groups are trying to ensure they have substantial input in the development of government regulations resulting from legalization. Moreover, Fantasy Sports websites are likely to vie for a similar level of input as to ensure the growth of their company and by allowing traditional odds-style betting, which would be easily implemented into their websites [2]. However, a potential issue with input from companies the support (and profit from) gambling services is that they may seek to bias regulations to favor increased rates of betting. This, in turn, has the potential to result in using modeling techniques to improve odds of winning and create unfair advantages.

The final change we may see is a peer to peer exchange for online betting, much like a stock exchange. This would allow for every bettor to become their own bookie and take bets with the platform taking a commission fee for using the service [2]. As data and models become available for individual use, it could lead to an increase in illegal gambling or enable those with a gambling addiction to participate in gambling without placing bets themselves.

4 Current Available Models

Betting and simply the pre- and post-event reporting on sports has attracted interest from many sources that proclaim their superior accuracy. With an abundance of options to select from we chose to focus on three of the most popular services that use statistical simulation methods: AccuScore, simulations from the popular EA Sports Madden video game series, and the statistical analysis website FiveThirtyEight.com.

4.1 AccuScore

AccuScore is an online sports betting platform where users pay a fee to access data and information about simulated sporting events. The simulation is done through a proprietary prediction model that provides a winner of each game along with individual player performance for a large variety of sports and sport leagues worldwide. It is the dominant player in the industry to which Nick Bonaddio, the CEO of a competitor called numberFire, proclaimed that his own company would not exist if it wasn't for AccuScore [7].

As a service, AccuScore does not provide the means to bet on any of the sports they track. However, AccuScore delivers data to paid subscribers who can then use betting services to wager money on any major sporting event. Individual player data can also be utilized by users to play fantasy sports through any platform or service. Currently AccuScore provides analytics on games in the NFL, NCAAF, all major soccer leagues (e.g., MLS, LaLiga, and Bundesliga), MLB, NBA, NCAAB, and many major hockey leagues (e.g,. NHL). Paid subscriptions give access to every sport and there is no option to selectively subscribe to just one set of predictions. A 12-month paid in full subscription is $349 and there is an option for a $69 month-to-month plan [5].

AccuScore provides the predictions on some of the most widely used online sports websites, including ESPN, CBS Sports, Yahoo Sports, and The Wall Street Journal. However, the manner which these companies implement AccuScore differs. For example, CBS Sports combines AccuScore predictions with the predictions of their own sports writers as opposed to Yahoo Sports which only uses AccuScore. ESPN offers AccuScore predictions to its paid premium subscribers, but also provides predictions from an unknown source in their free section [6].

Although some details of AccuScore's model are available, specific details of the prediction models are proprietary so as to create competitive advantage with other companies. A high-level overview is provided on their website telling visitors that each game is simulated approximately 10,000 times using "past player performances and relates that data to the coming match: team composition, weather, location, coaching…" [7]. In an interview with SBC News this year, the CEO Tuomas Kanervala also mentioned team cohesion and "qualitative input from our sports analysts who have a detailed information about specific sports and league they are in charge of" [8]. Due to the secrecy of company personnel, be it real or perceived, there is an unknown to the backgrounds of either the statisticians or sports analysts that create the AccuScore predictions for each game.

There are 2 main ways companies and individuals like AccuScore can predict against wins & losses: 1) against the point spread or 2) straight up. If a handicapper is predicting a game straight up, it is simply predicting who wins or loses a game. Point spreads are provided by handicappers and represent the number of points the favored team will beat the other team. To illustrate the difference let's create a hypothetical line for a game between New England and Dallas. The point spread for this game is New England -5.5 points which means that New England is favored to win the game by 5.5 points. In a straight up decision, whichever team wins is the result. Using the point spread, New England would have to beat Dallas by over 5.5 points to be considered the winner.

Based on AccuScore's website, their prediction accuracy was compared to the accuracy of 27 Las Vegas handicappers in a 2010 analysis. AccuScores's accuracy in predicting the outcome of all NFL regular season and post-season games against a point spread[4] was measured at 55%, but was not the highest winning percentage compared to the other 27 handicappers (Figure 2). Twelve handicappers had a higher winning percentage with the best being 64%. However, AccuScore was able to create predictions for 474 games while the Vegas handicappers ranged between 44 and 301 games. AccuScore's value proposition consists of providing a high number of predictions over a wide breadth of sports at a much lower price point compared to the "experts" [9].

[4] Point spread is the number of points by which the stronger is expected to win.

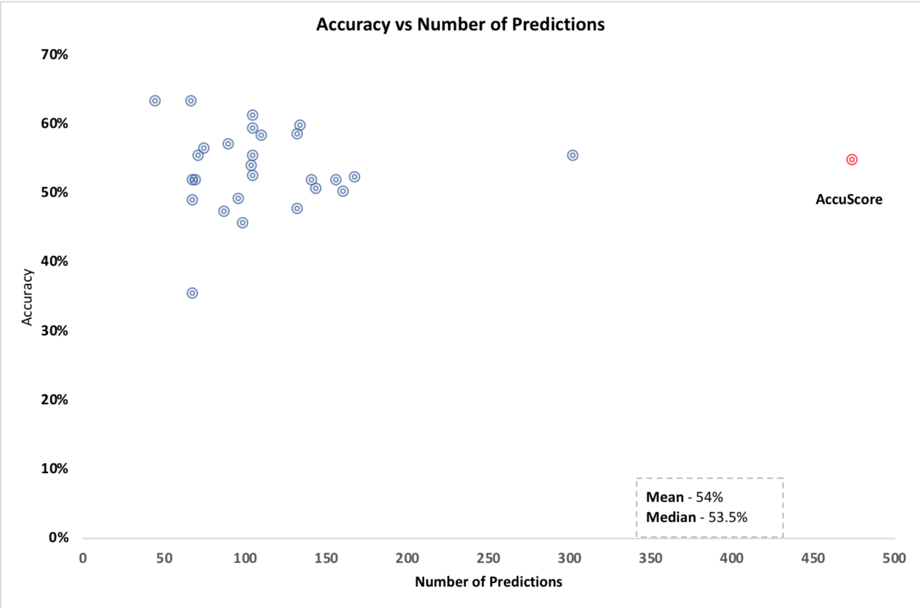

Figure 2: Plot of the AccuScore model performance and individual raters as a function of the number of predictions. Mean and median values refer to the mean and median of the individual raters.

To the casual observer, 55% accuracy may look very ordinary being that it is not too much better than 50/50. This may scare some people into thinking the prediction model used by AccuScore is not very good. Accuracy of any prediction model is relative to the expectations of the user and the nature of the data itself. Given that AccuScore is proprietary, we are unable to independently evaluate the of the results. However, based on AccuScore's own 2010 analysis, the performance of its model is very average when compared to the other 27 handicappers with performance not much above the mean and median of all handicappers in the analysis. Even with the ability to sustain this 55% accuracy over far more games than others, we feel there is opportunity to improve the performance through the addition of more variables and using other statistical techniques.

4.2 Madden

EA Sports is a part of the Electronic Art Inc, a global company headquartered in Redwood, California. Electronic Art Inc is an interactive entertainment software company, that develops and delivers games, content and online-services for internet-connected consoles[5], personal computers and mobile devices [10]. Madden is EA Sports' simulated football video game based on the NFL. Madden by EA Sports runs an artificial intelligence engine to simulate real life game play in the video game.

[5] In Example the Microsoft Xbox or PlayStation PS4

Starting in 1988 for computers running MS-DOS the game has become more sophisticated in its ability to simulate game play [11]. The focus of the creators is not to predict the outcome of future games, but to use current player data to allow for the gamer to participate in the game itself with realistic outcomes. Madden runs a simulated version of each season to predict the seasons outcome with ESPN with a 72% accuracy [12].

Madden's model runs off individual player information and rankings, which for the most part is controlled by a single person, the "Ratings Czar" [13]. The Czar is responsible for updating the players' 43 attributes including speed, strength, stamina, and agility. Some attributes are specific to a position: throw power for quarterbacks, carrying for ball carriers, catching for receivers, and tackle for defenders. Other factors include traits that are thought to capture player tendencies during a game, such as fighting for extra yards or performing in the clutch [14]. These factors as well as others all go into the Madden's game simulator which has been used since 2004 to predict the Super Bowl outcome.

Each year EA Sports uses Madden to run simulations of the Super Bowl without the influence of people playing the game, computer vs computer, which is an experiment in artificial intelligence. This practice has yielded the same result as the actual Super Bowl 9 out of 13 times [15].

The focus of the Madden simulators are the individual players and their Overall Ratings score. This score is a combination of all the attributes and traits, which are controlled and assigned by an individual at EA Sports. Each individual player in the NFL is assigned an Overall Ratings Score, a weighted average of the player's attribute scores. Each of the field positions has their own attribute weighting that is used for the calculation of the Overall Rating Score. Figure 1 below shows the weights given to the different attributes for the various positions.

Table 1: Example of EA Sports Madden weights [13].

Quarterback

Player Attribute	**Weight**
Awareness	20%
Throw Power	20%
Accuracy short	15%
Accuracy mid	15%
Accuracy deep	10%
Play action	8%
Speed	5%
Agility	3%
Throw on run	3%
Acceleration	1%

The systems within Madden that simulate the game play are extremely complex and with the years of additional systems and coding being layered into the game the amount of interaction between them all is unclear [13]. As depicted in Figure 3, albeit it rather simplistic, as two players within in the game start to have an interaction the game will take into consideration the different attributes of each of the players and add a "dice roll"; if when comparing a ball carrier's attributes are ranked higher than a line-backer's tackling attributes, then the "dice roll" would favor the ball carrier getting out of the tackle [13].

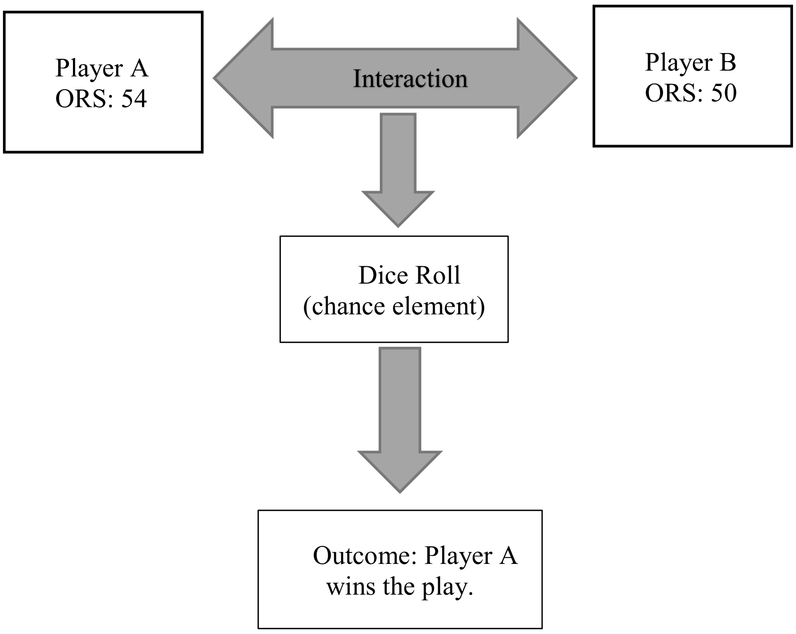

Figure 3: Flowchart of Player interaction in Madden simulator.

As we consider the Madden approach to modeling or predicting a game, what we can deem from it is that the individual players' abilities are going to have the greatest effect on the overall outcome of a model.

4.3 FiveThirtyEight

An additional football ranking systems that can be used to predict possible game outcomes was developed by Nate Silver of FiveThirtyEight.com. This system uses many variables to create an overall score and uses the idea of Elo, which was a ranking system developed by an avid chess player and physicist, Árpád Imre Élö [16]. Élö developed the system to help with the ranking of Chess Masters throughout the world. This same system has been used throughout the world to rank everything from professional sports outcomes to various board and card games.

The model developed by Nate Silver starts by using the final standings of the previous season and resets all teams towards the mean. Unlike several other models which attempt to predict the outcome of NFL football games the FiveThirtyEight model does not take into consideration the individual transactions of a team within a given week or over the entire season. This model does not take into consideration any changes in personnel such as injury or trades, which makes model less susceptible to large swings when predicting wins and losses each week. The ranking system looks at the final scores of the previous games played, when those games occurred and where the games were played. These factors allow the model to be more flexible without trying to account for the impact of a single player or play. Additionally, the model takes into consideration the more important factors and leaves out data points and sources that could help introduce bias in the rankings of a data set that already considered small by most measures.

The FiveThirtyEight model has a mean value of 1,500, which is commonly used to define a team that would have 8 wins and 8 losses. The model takes a team's previous year's results into consideration, as mentioned earlier all teams' previous years point totals are then reverted to the mean using the formula below.

$$((\text{Current Year } Point\ Total - Mean) * (1/3))^{-1} + \text{Current Year } Point\ Total. \quad \textbf{(1)}$$

To create the starting values for a new season the difference from the point total of the previous season is then multiplied by one-third. For example, a team with a 2016-year end total of 1,900 starts the 2017 season with a total of 1,768.

The model also uses a parameter called K. According to FiveThirtyEight the "Elo's K-factor determines how quickly the rating reacts to new game results. It should be set to efficiently account for new data but not overreact to it" [17]. This parameter helps account for and correct the issue of autocorrelation. Autocorrelation according to StatisticSolutions.com is "... a characteristic of data in which the correlation between the values of the same variables is based on related objects. It violates the assumption of instance independence, which underlies most of the conventional models. It generally exists in those types of data-sets in which the data, instead of being randomly selected, is from the same source" [18]. The issue of autocorrelation is particularly important when attempting to predict outcomes for a sport like football, which has only sixteen games per team. The sixteen games are considerably less than other sports which the same model is used for predicting wins and losses. For example, baseball has one hundred and sixty-two games and basketball has eighty-two games a year respectively. K is an important parameter to the model because setting the value too high, the ratings will be very inconsistent. If the value of K is set too low the model will take too long to recognize the importance of one win versus another.

As important as it is for the model to account for autocorrelation by using the value of K, the ability to choose a winner also depends on the overall margin of victory of the game. According to FiveThirtyEight the way the model was originally designed was to compensate for the margin of victory by subtracting more points for a loss than is given for a win [19]. This is great for many of the games played, but attempting to account for the margin of victory can sometimes lead to problems. On average favorites, those who are predicted to win, tend to win more often. Additionally, when they do win, they tend to win by larger margins. Since the model gives more credit for larger wins, this means that favorites tend to have their ratings get inflated over time. To offset this issue the model created a formula to discount the margin of victory.

$$L(ABS(PD)+1) * (2.2/((ELOW-ELOL)*.001+2.2)) . \qquad (2)$$

Where PD stands for Point Differential between the teams, ELOW is for the winning team's Elo rating before game and ELOL is the losing team's Elo rating before game.

Finally, before each week an individual team's win probability can be found by taking the Elo values from both teams in the matchup and subtracting the team you are trying to predict (Team A) from the team they are playing (Team B). The formula below shows how the Elo's from the model can help with predicting the overall win probability of a given game:

$$P(Team\ A) = 1 / (10\text{^}(-(Team\ A-Team\ B)/400) + 1) . \qquad (3)$$

5 Data Source

We collected data for all games played in the NFL since 2009, which gives us over 1,792 games to build a model from. The data collected has a range of data points spanning offensive, defensive and special team's categories. We gathered 55 attributes from the NFL's official website[6]. The NFL's website contains game statistics for most of the games played since 2009. The game information data available for individual games includes identification of the home and away teams and identification the winning and team. Additionally, this data includes team statistics (for both the home and away teams) including: possession time, total number of yards gained, yards gained via passing, rushing and punting, number of penalties, total number of penalty yards, number of first downs and number of turnovers.

Additionally, we expanded our data set to include weather related variables. These data were obtained from the WeatherUnderground.com website, which complies data from over 250,000 weather stations throughout the world. Using the API's provided by the site[7] we added 6 weather related metrics, which correlate to each date, time and locations of the NFL game being played. We included the following metrics: temperature, humidity, precipitation, forecast, wind speed and wind direction. To our knowledge the other models available do not take into consideration the affects weather would have the outcome of a game.

[6] NFL Official website: www.NFL.com

[7] https://www.wunderground.com/weatherstation/overview.asp

6 Model Creation and Selection

To find the best model for predicting wins and losses for the NFL season we chose to build three models[8], all of which used LogisticRegression[9] from the scikit-learn library (version 1.19.1) implemented in Python (version 2.7.14) [20]. Each of the three models we created predicted if the Home Team would win a game (a binary outcome variable of either "yes" or "no"). Our code was adapted from Susan Li's article "Building a Logistic Regression in Python, Step by Step" [21]. The three models we developed included different mixtures of variables that were obtained from using different feature selection techniques. Additionally, all models were validated using ten (10) fold cross validation. Cross validation is commonly used process that splits any data set into different sets in which the model is trained on N-1 datasets, where N is the number of splits (10 in the present analysis) and the resulting model is tested on the withheld data.

The data set used for all three models originally consisted of seventy-one (71) explanatory variables (see Appendix). Once these explanatory variables were imported from a comma delimited file, additional dummy variables were created for following categorical variables WindDir, Condition and TimeOfDay. Additionally, the HomePossTime and AwayPossTime variables were coded as the total seconds of possession for each offensive team and were renamed HomeTeamPossSec and AwayTeamPossSec respectively.

After creating the dummy variables and the new time of possession variables the total data set consisted of a total of one hundred and eighteen (118) explanatory variables. Using this newly created data set each model used a different feature selection technique. The first model that was created used all 118 explanatory variables for the analysis. The second model used traditional Principle Component Analysis (PCA) analysis to find appropriate variables. The final model used a feature selection technique called recursive feature elimination.

The first model created used all one hundred eighteen (118) of the explanatory variables within the data set. This model provided an overall accuracy of approximately 85.3% using ten (10) fold cross validation. Table 2 provides the accuracy of each validation iteration and the overall model.

The second model created used a Principal Components Analysis (PCA) method to determine the numeric explanatory variables. PCA is a method that captures most of the variation of multivariate responses by selecting several linear combinations of the explanatory variables [22]. Figure 4 shows a graph of the variance ratio for the first 20 components of the PCA. These components explained essentially all of the variance in the data (~99.9%).

---

[8] GitHub repository with Code for models: https://github.com/ChrisIrwin5620/CapstoneProject

[9] Scikit-Learn Linear Model Logistic Regrestion: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

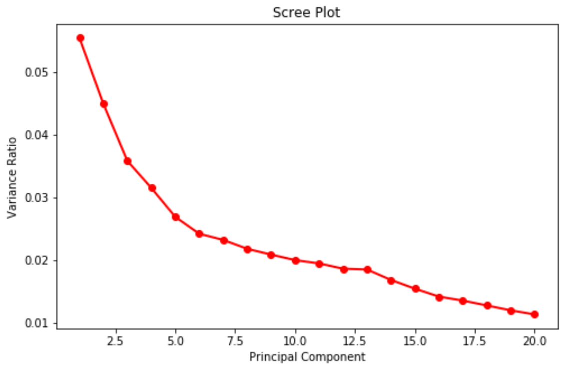

Figure 4: Plot of the variance ratios for the first 20 components obtained from PCA.

Using the results of PCA for feature selection to create another logistic model consisting of thirty-two (32) explanatory variables. The results from the cross validation using the subset of features determined by PCA demonstrate an overall accuracy of the of 70.5% (see Table 2). Importantly, feature reduction using PCA, which resulted in a loss of 86 predictor variables, resulted in an a 19.45% average decrease in model performance..

The final model used recursive feature elimination to eliminate potentially uninformative features. According to documentation is an "external estimator that assigns weights to features (e.g., the coefficients of a linear model), recursive feature elimination (RFE) is to select features by recursively considering smaller and smaller sets of features" [30]. The process works by first training on the initial set of features using an estimator. Once the model is trained, the importance of each feature is obtained through a feature importance attribute. We chosee the ranking function which eliminates the least important features are removed from current set of features, which is then repeated until the desired number of features are eventually reached[10]. We used a Logistic Regression model to train the original features and then selected the one hundred (100) most important features. The recursive model had an accuracy of approximately 85.10% (see Table 2).

Table 2: Accuracy results of proposed models using ten (10) cross validation.

	Full Model	**PCA Feature Reduction**	**Recursive Model**
Average	85.30%	70.61%	85.09%
Iter. 1	85.42%	60.94%	83.85%
Iter. 2	92.18%	63.54%	84.36%
Iter. 3	85.42%	70.83%	82.81%
Iter. 4	79.17%	70.83%	86.46%

[10] http://scikit-learn.org/stable/modules/feature_selection.html

Iter. 5	82.81%	70.31%	87.50%
Iter. 6	84.86%	71.35%	83.33%
Iter. 7	85.94%	70.31%	80.73%
Iter. 8	81.25%	66.15%	82.81%
Iter. 9	80.73%	66.15%	83.33%
Iter. 10	82.29%	70.83%	81.25%

Although similar, the RFE model perfomed numerically worse than the full model. Figure 5 provides the results of 10-fold cross validation for each of the models discussed. From the chart we see that the Recursive Model has a better accuracy rate that the other two models.

Figure 5: Model Accuracy by Run 10-fold Cross Validation

Given it's superior performance, we will use the full model described above that included all features in a final validation test for predicting game day outcomes using data from the 2017 NFL seasons. Additionally, we will compare the accuracy of our model with the prediction accuracy of the existing models.

7 Model Analysis and Results

Two of the three existing models are commercially available (MODELS) for use in predicting outcomes of NFL games. As discussed in the Introduction, the AccuScore, Madden, and FiveThirtyEight models use different variables and different points of view to achieve the same goal. Using the 2017 season as a test set, we found prediction accuracies of 71.4% for Madden [28] and 63% for ELO [29]. Finally, the model we created performed better than these existing models with an an overall success rate 84%. This accuracy level was a 12.6-percentage point increase over the Madden simulation and a 21-percentage point increase over ELO model. Figure 6 shows the accuracy rates of the model presented in the paper compared to the two models commercially available. The AccuScore model discussed earlier did not make accuracy rates for the 2017 NFL season publicly available, and was thus not included in this analysis.

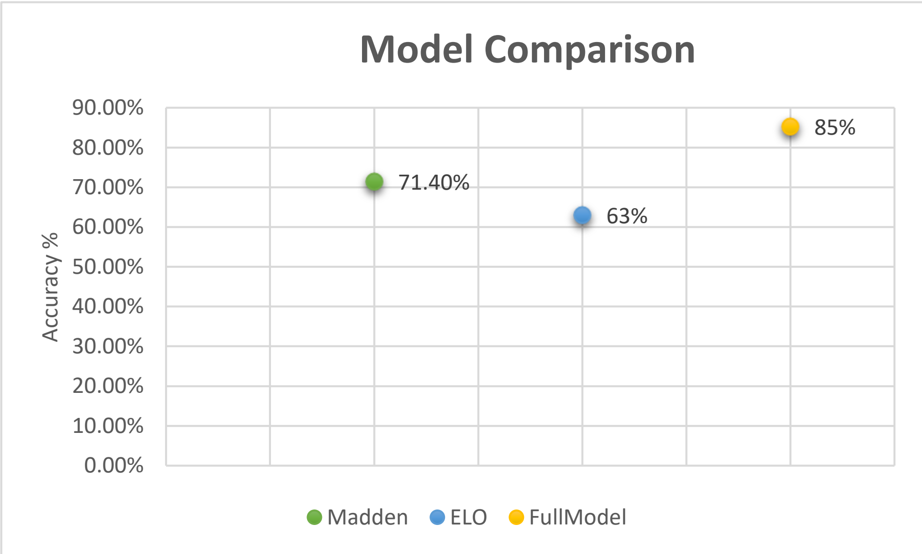

Figure 6: Prediction accuracy for 2017 NFL games from the Madde simulations, ELO model, and the Full Model that was developed in this report.

We believe this increase can be attributed to two main factors. The first is that our model uses a 5-week rolling average to calculate team statistics, this allowed for the model to adjust for situations and the effects of possible player injury, performance slumps, travel schedules and unfavorable team matchups in a much timelier manor. The ELO model specifically penalizes teams that lose games in which they should win, but in turn takes time to adjust the overall teams ELO score after the loss of a significant player. In the case of the Madden simulation, a person would have to manually adjust for such player personal issues and could not take in to account any additional factors such as fatigue or travel. Second the model we have created uses weather data that is captured at the start of the game. These data points are invaluable when determining possible game plan shifts due to the expectation of high winds or inclement weather. ELO does not take any weather into consideration and Madden uses a randomly generated weather setting for each game.

8 Ethics

There are a few ethical questions that arise from modeling game results:

1. Does the use of models promote gambling and is gambling in itself unethical?
2. Does the use of models, if not made public, give some gamblers or fantasy league members an unfair advantage?
3. Is it ethical or right to use the players' personal medical information to increase the accuracy of a model?

To address the first question, we start with reviewing if gambling is unethical or immoral. There is a stigma around the word "gambling" and most people view it as wrong. Churches have declared it a vice, an immoral activity that should be avoided [23]. Religious doctrine states that if a person gambles and wins, they received money without working or earning it through their own abilities. However, there has been a shift I recent times as funding for organizations come from things such as the lottery. Many Catholic churches today sponsor bingo nights with cash prizes. The point of view that gambling as an immoral act is not something we agree with, especially when gambling on sporting events. People watch sports for the suspense and excitement of the game and its outcome. If betting on the outcome of an event can enhance the experience of watching the game, then the person is using their bet as a proxy payment for the additional excitement and entertainment. There is a utility to the wager placed on the outcome - as long as the person can afford to place the bet and lose, then no harm is done to that individual [23]. This leads to the second part of the question which is does the use of a model promote gambling in a way that would cause gambling to become unhealthy? For example, would the use of a model give an individual the confidence to bet money they could not afford to lose because they are counting on the model's outcome.

We look at what causes gambling addiction in the first place. Gambling addictions are considered an impulse control disorder, such as obsessive-compulsive disorder [24]. A person with a problem or a predisposition for a gambling problem could be led to think more obsessively and participate in gambling through either the building, monitoring, or using of a prediction model for sporting events. However, there is another argument to be made that model usage could provide a healthier outlet for a person's addiction as they would be able to focus on the creation or improvement of a model rather than placing bets and losing money. The final thing to consider for the overall ethicality of gambling would be is the degradation of the enjoyment by the majority of participants because a subset of participants is harmed by the activity. We can look to history of several regulations causing more harm than good. Prohibition, for example, saw a rise in illegal alcohol production and consumption. Here the answer is to provide regulations around how and where gambling is allowed, such as casinos and other vendors. A common-sense regulatory approach can control access and provide ways for those with a problem to be recognized and get help.

The next ethical issue is does the use of a model give gamblers or even fantasy league participants an unfair advantage over others. In this we would have to consider the purpose of both these activities: providing entertainment for the participants. We consider the source of the models and the information that is going into them. If all participants have access to the same model results, then no advantage is given to a participant over the others. They all have access to the same information prior to an event. In the case of a participant building their own model, this could be considered using their own skills in pursuit of a goal, much like the athletes themselves who practice in order to become more skilled and stronger than their opponents. In both of these cases no one is using unfair advantage over their own skill and knowledge to aid performance. Finally, we come to the availability of data for the models. If a participant uses data or a model that is not publicly accessible, they hold an unfair advantage which can be likened to insider trading.

Next, we look at player health data if it is ethical to use this information in prediction models for sporting events. The individual players and their abilities have a great impact on a game's outcome along with the other factors we have reviewed. There are laws around the protection of health information, however social media and the internet has made it possible for people to view injury lists and player commentary to help determine if a player is going to play. Now NFL players can actually sell their health data to those who want it through a wearable device from a company called Whoop, Inc. [25]. Whoop, Inc. provides a wristband that players can wear that tracks strain/activity, recovery time, and sleep. This information could be used by players and coaches to determine a strategy for a game, but this information getting out in the open could be used by data scientists to create predictive models with potentially greater accuracy than available to the general public. The availability of such data and its potential to make huge strides in accuracy could create a black market or sorts where players could auction off their own data to the highest bidder. Laws such as HIPAA ban healthcare providers from selling or releasing patient data, but it does not cover and individual willing to release or sell their own health information. An economic opportunity is created for the player to enhance their economic benefit to any entity willing to pay the right price. It would however be unethical to use a player's personal data without their consent, if the data is obtained in a manner that breaks the laws and person's individual privacy the modelers would be in wrong for using the data.

9 Conclusions

In conclusion, the NFL is a constantly shifting environment with the introduction of new players, new techniques and new game plans. These changes make for a more enjoyable watching experience, but do not necessarily translate into something that can be easily predicted. Throughout this paper we have presented 3 different ways, 2 of which are commercially available for use, to predict outcomes of NFL games. Each of the models presented use different variables and different points of view to achieve the same goal. Our model developed here provides an increased prediction accuracy rate compared to the existing models.

The overall strength of the model we have presented is the ability to adapt to a constantly changing environment. To maintain a high overall prediction accuracy, we expect that our model will need periodic updating to track evolution of the game and players. Implementation of the model for public use would be provided as open source software, with the expectation that individuals have the ability to gather the required data for the model and the necessary knowledge and programs to implement the code for the next season and/or game.

References

1. National Football League, "History," 3 December 2017, http://www.nfl.com/history/chronology/1869-1910.
2. Purdum, David and Rodenberg, Ryan "Future of Sports Betting: The Marketplace," ESPN, 04 May 2017, http://www.espn.com/chalk/story/_/id/17892685/the-future-sports-betting-how-sports-betting-legalized-united-states-the-marketplace-look-like.
3. Purdum, David "Congress Reviewing 'obsolete' federal gambling laws, to introduce new legislation," ESPN, 26 October 2016, http://www.espn.com/chalk/story/_/id/17848529/congressional-committee-reviews-federal-gambling-laws-plans-introduce-new-legislation.
4. Burstein, D.: NumberFire: "Moneyball" For Fantasy Leagues. Fast Company (13) (March 2012), https://www.fastcompany.com/1824128/numberfire-moneyball-fantasy-leagues
5. Accuscore Inc. Join Now, https://accuscore.com/join-now
6. Kim, D.: "How Accurate are Yahoo, ESPN and CBS Fantasy Projections?." The New York Times (12) (November 2010), https://fifthdown.blogs.nytimes.com/2010/11/12/how-accurate-are-yahoo-espn-and-cbs-fantasy-projections
7. Accuscore Inc. About Us, http://accuscore.com/about-us/uncategorized/about-us
8. Menmuir, T.: Tuomas Kanervala – Accuscore: Learning Curve…Why prediction is not just for punters. SBC News (29) (August 2017), https://sbcnews.co.uk/latestnews/2017/08/29/tuomas-kanervala-accuscore-learning-curve-prediction-not-just-punters
9. Accuscore Inc. NFL Picks Dominate "Experts", http://accuscore.com/sports-betting-system/nfl-analysis/accuscore-picks-best-vs-experts
10. "Company Information," 3 December 2017, https://www.ea.com/about.
11. Cracknell, Ryan "Madden NFL Covers – A Complete Visual History," The Cardboard Connection, https://www.cardboardconnection.com/madden-nfl-covers-visual-history.
12. McGregor, Todd "Are 'Madden Simulations' sn Accurate Predictor of NFL Games?" Bleacher Report, 09 October 2009, http://bleacherreport.com/articles/269341-are-madden-simulations-an-accurate-predictor-of-nfl-games.
13. Paine, Neil. "How Madden Ratings Are Made The Secret Process That Turns NFL Players Into Digital Gods," FiveThirtyEight, 25 February 2015, https://fivethirtyeight.com/features/madden/#.
14. "Madden Attributes and Ratings Explained," E Athlete, 18 February 2017, http://www.e-athlete.com/madden-guides/madden-attributes-ratings-explained/.
15. Muncy, Julie "How Madden Got to Be so Good at Predicting Super Bowl Winners," Wired, 03 February 2017, https://www.wired.com/2017/02/super-bowl-madden-predictions/.
16. "Arpad Elo and the Elo Rating System," Chess News, 16 December 2007, http://en.chessbase.com/post/arpad-elo-and-the-elo-rating-system.
17. Silver, Nate and Fischer-Baum, Reuben "How We Calculate NBA Elo Ratings," FiveThirtyEight, 21 May 2015, https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/.
18. "Autocorrelation," StatisticsSolutions, 2017, http://www.statisticssolutions.com/autocorrelation/.
19. Silver, Nate "Introducing NFL Elo Ratings," FiveThirtyEight, 04 September 2014, https://fivethirtyeight.com/features/introducing-nfl-elo-ratings/.
20. scikitlearn "Logistic Regression," http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.
21. Li, Susan "Building a Logistic Regression in Python, Step by Step," 28 Sept 2017, https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8.

22. Ramsey, Fred L. and Schafer, Daniel, "The Statistical Sleuth: A Course in Methods of Data Analysis, 3rd Edition." Brooks/Cole, United States, 2013.
23. New Philosopher, "The ethics of gambling," NewPhilospher, 14 Sept 2016, http://www.newphilosopher.com/articles/the-ethics-of-gambling/
24. Bergeson, Boyd "What Causes Gambling Addiction?" 14 Aug 2017, https://www.livestrong.com/article/149406-gambling-addiction-ethical-issues/.
25. Grothaus, Michael "NFL Players can sell their health data to anyone," FastCompany, 25 April 2017, https://www.fastcompany.com/4035608/nfl-players-can-now-sell-their-health-data-to-anyone.
26. "Evolution of the NFL Player," NFL Football Operations, 2018, https://operations.nfl.com/the-players/evolution-of-the-nfl-player/.
27. Fischer-Baum, Reuben "How Have the Weights of NFL Positions Changed Over Time?" Deadspin, 17 March 2014, https://deadspin.com/how-have-the-weights-of-nfl-positions-changed-over-time-1545701731.
28.Owens, Michael, "Madden 18: Our 2017 NFL season simulation results," *Fansided,* September 2017 https://apptrigger.com/2017/09/01/madden-18-2017-nfl-season-simulation-results
29. Gaines, Cork, "One of the computer models picking NFL games has taken a big lead- here is who they like in Week 16," *Business Insiders,* 21 December, 2017, http://www.businessinsider.com/nfl-picks-microsoft-cortana-elo-week-16-2017-12
30. Browenlee, Jason "Feature Selection in Python with Scikit-Learn," Machine Learning Matery, 14, July 2014, https://machinelearningmastery.com/feature-selection-in-python-with-scikit-learn/