Using Poisson Regression to Model National Football League (NFL) Scoring and Exploit Inaccuracies in the Online Betting Market

Mofeed Nagib

12/17/2023

Abstract

In the landscape of NFL score forecasting, existing models often treat the score as a continuous variable, overlooking the inherent patterns of scoring in American football — frequently manifesting as combinations of multiples of 3 and 7. Moreover, these models typically forecast the total points scored by each team, rather than projecting scoring events (e.g., the number of touchdowns and field goals) and summing them to calculate the cumulative predicted score. Given these limitations, we present an innovative Poisson regression that models the discrete nature of NFL scoring. More specifically, we fit a dependent bivariate Poisson regression that jointly models the number of touchdowns and number of field goals, using team, opposing team, season, and home/away status as predictors. The aggregate training dataset in this project was sourced from the nflfastR package and Kaggle, and includes game details, scoring event breakdowns, and betting and spread information spanning 1999 to the present. Our model achieves an accuracy of around 53% in correctly predicting both the spread and the over-under, demonstrating some efficacy in capturing the nuanced dynamics of NFL scoring.

Introduction

The past decade has witnessed a substantial increase in the popularity of online gambling, particularly in the realm of sports betting. Online sportsbooks like DraftKings and FanDuel have risen to prominence in the industry, with the National Football League (NFL) emerging as the predominant league for betting. In an effort to gain a competitive edge, many sports betting enthusiasts have turned to statistical models and advanced analytics to inform their betting decisions. These sophisticated methodologies often involve analyzing player performance, team statistics, historical trends, and various other factors to make more informed predictions. While many of these models have achieved success to a certain extent, they typically rely on certain glaring assumptions that fail to accurately capture the dynamics of play in the NFL.

One of the most common oversights in these models is their treatment of the game score as a continuous response variable. The primary flaw in this approach lies in the fact that NFL scoring is both positive and discrete, introducing a critical discrepancy in accurately reflecting the nature of scoring events in the game. In actuality, scoring in the NFL follows a distinctive pattern, characterized by combinations of multiples of 3 and 7. Moreover, the score of the game is the result of summing together a discrete number of scoring events for a cumulative final score, so forecasting the final score should inherently account for these distinct and structured elements. Therefore, these models would better align with the scoring dynamics in the NFL if they project scoring events and aggregating them appropriately to predict the final score.

Poisson regression is a widely used statistical method for predicting the number of events occurring within a fixed interval. It is particularly well-suited for situations where events are rare and occur independently within a specified time or space. Poisson regression models are founded on the Poisson distribution, which assumes that the events of interest are discrete and that the probability of each event occurring is constant over time or space. In sports analytics, Poisson regression excels at capturing the inherent randomness and infrequency of scoring events. For instance, it has been particularly successful in modeling the sporadic nature of goal-scoring in soccer and hockey. We propose that the assumptions of Poisson regression, designed for modeling stochastic and infrequent scoring occurrences, are applicable and effective in the context of the NFL. This makes it a promising method for modeling touchdowns, field goals, and other scoring events in American football, given its adeptness at capturing inherent variability. Hence, a more in-depth understanding of scoring breakdowns in the NFL is imperative for developing accurate predictive models using Poisson regression.

In the NFL, there are five possible ways to score points: touchdowns, field goals, extra points, two-point conversions, and safeties. The most valuable scoring play is the touchdown, worth six points, which occurs when a player crosses the opponent's goal line with possession of the ball or catches a pass in the end zone. Field goals, worth three points, are achieved by successfully kicking the ball through the opponent's goalposts at varying distances on the field. Following a touchdown, the scoring team can score an extra point by kicking the ball through the uprights in the manner of a field goal, or a two-point conversion by passing or running the ball into the end zone in the manner of a touchdown. Both extra points and two-point conversions are attempted from the two-yard line in the play immediately following a touchdown. Finally, safeties occur when the offensive team is tackled, forced out of bounds, or commits a penalty in their own end zone, awarding the defensive team two points and possession. It is important to note that in 2015, the NFL introduced defensive two-point conversions, allowing defenses to earn two points by returning a conversion attempt to the opposing end zone; however, given their brief history and limited occurrence, we have opted to omit the few games featuring this scoring event from the subsequent analysis.

There are four basic bet types available across nearly every sport: moneylines, spreads, totals, and parlays. Moneylines, the simplest of the four, involve betting on the outright winner of a game, with odds assigned to each team based on their perceived likelihood of winning. Spreads, on the other hand, introduce a handicap to level the playing field between two teams of uneven strength. The favored team must win by more than the specified margin, while the underdog can lose by less than that margin or win outright. Totals, also known as over/under bets, involve predicting whether the combined score of both teams will be over or under a set number determined by the sportsbook. Lastly, parlays are a more complex betting type where bettors combine multiple individual bets into a single wager. While parlays offer the potential for higher payouts, they require all included bets to be correct for the overall bet to win, making them riskier but more rewarding if successful. We utilize predicted spread bet and over/under accuracy to evaluate the quality of our proposed model.

The training dataset for this model was constructed using play-by-play data acquired through the nflfastR package, alongside game betting data sourced from Kaggle. The nflfastR dataset contains NFL play-by-play data spanning from the 1999 season to the present, and contains nearly 400 variables, covering everything from team statistics and player details to drive and series specifics, as well as binary variables indicating each scoring event. The Kaggle betting odds dataset contains simple game details (date, teams, and scores, etc.), and the closing point spread and over/under line for every game, ranging from the 1966 season to the present. In the forthcoming analysis, we leverage the scoring category breakdown data for training our Poisson regression model, while assessing the model's quality with betting odds data to gauge its performance against Vegas.

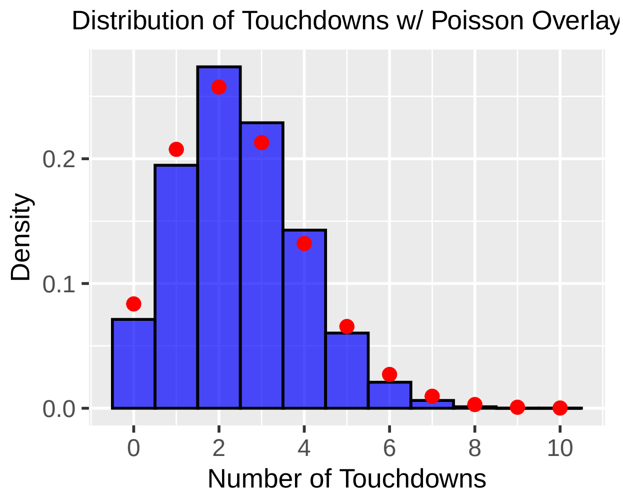
Section 2 of the report examines the data aggregation process, presenting data explorations, visualizations, and a correlation matrix plot illustrating positive and negative correlations among scoring types, highlighting a clear Poisson relationship for each. Section 3 involves fitting two models—a univariate Poisson regression for touchdowns and field goals independently, and a bivariate Poisson regression for their dependent relationship. Section 4 employs our model to simulate game outcomes and subsequently applies these results to the online betting market, evaluating the effectiveness and predictive accuracy of our model in forecasting game outcomes. Finally, in Section 5, we delve into the discoveries made by our model, offer suggestions for refining it, and discuss prospective directions for future research to address some of its identified shortcomings

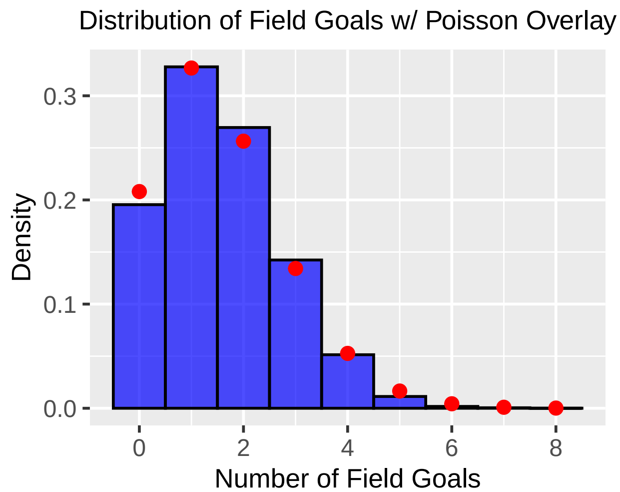
Data exploration and visualization

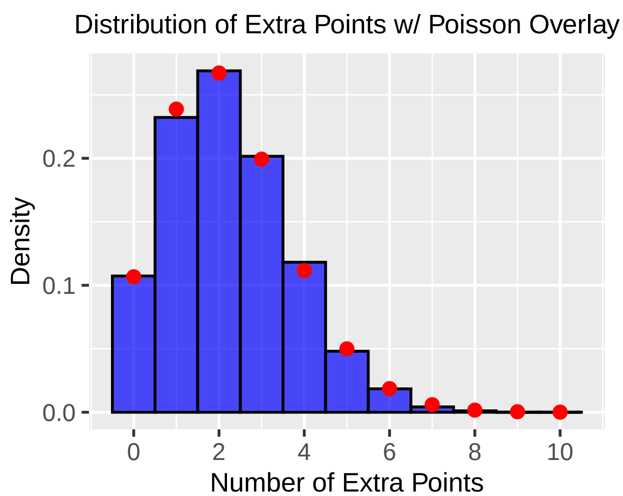
The primary data extraction and processing operations are performed in the `05_get_data.R` script, consolidating all the information that we will need to train and evaluate the model. Within this script, we iterated through the play-by-play data for each game since 1999, tallying the number of touchdowns, field goals, extra points, two-point conversion, and safeties. Following this, we merged that aggregated data with the closing point spread and over/under line data for each respective games. Lastly, we filtered out games where the scoring breakdown did not sum to the original final scores of the game (i.e. games that featured a defensive two-point conversion) or if the betting data is absent or the point spread is 0 (i.e. essentially a moneyline bet) for that particular game. Ultimately, this data preparation culminated in scoring and betting information for 6,583 of the 6,640 games played since 1999, providing a comprehensive foundation for subsequent model training and evaluation. Below is a table that lists all the variables we included along with their corresponding descriptions.

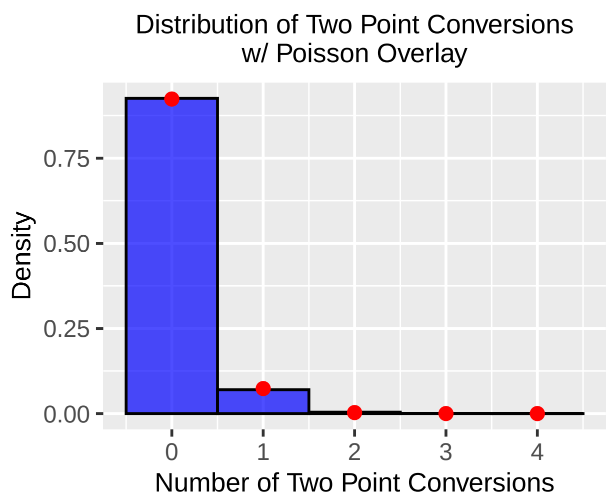
Variable Name	Variable Type	Variable Description
game_id	Categorical	Unique identifier for each game
game_date	Numerical	Date of the game (YYYY-MM-DD)
season	Categorical	Season of the game (YYYY)
home_team	Categorical	Home team name
away_team	Categorical	Away team name
home_final_score	Numerical	Home team's final score
away_final_score	Numerical	Away team's final score
home_TD_count	Numerical	Home team's touchdown count
away_TD_count	Numerical	Away team's touchdown count
home_FG_count	Numerical	Home team's made field goal count
home_FG_attempts	Numerical	Home team's field goal attempt count
away_FG_count	Numerical	Away team's made field goal count
away_FG_attempts	Numerical	Away team's field goal attempt count
home_XP_count	Numerical	Home team's good extra point count
home_XP_attempts	Numerical	Home team's extra point attempt count
away_XP_count	Numerical	Away team's good extra point count
away_XP_attempts	Numerical	Away team's extra point attempt count
home_TWO_PT_count	Numerical	Home team's successful two point conversion count
home_TWO_PT_attempts	Numerical	Home team's two point conversion attempt count
away_TWO_PT_count	Numerical	Away team's successful two point conversion count
away_TWO_PT_attempts	Numerical	Away team's two point conversion attempt count
home_SAFETY_count	Numerical	Home team's safety count
away_SAFETY_count	Numerical	Away team's safety count
team_favorite_id	Categorical	Team favorited to win the game
spread_favorite	Numerical	Expected winning margin for the favored team
over_under_line	Numerical	Expected total combined score of both teams

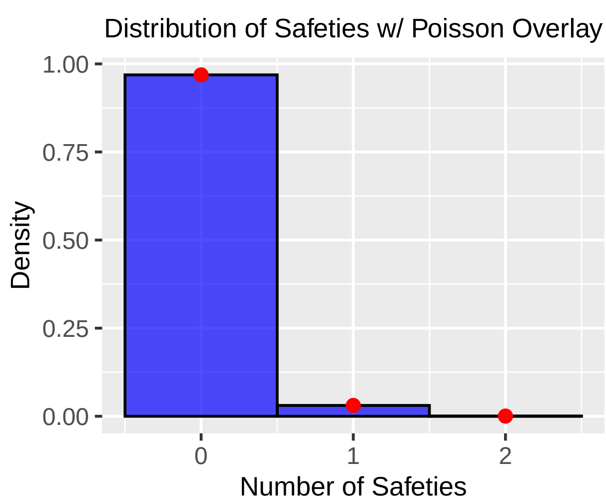
Using this consolidated dataset, we investigated the suitability of a Poisson distribution for modeling each scoring event. For each of the five scoring categories, we consolidated the tallied values for both home and away teams, resulting in two rows per game. One row represents the home team with their tallied scoring event values, and the other row represents the away team with their respective tallied values, totaling 13,166 observations overall. Subsequently, for each of the five scoring categories, we generated a histogram overlaid with a Poisson distribution for their respective mean frequency. Below are the five plots, and while the overlay of the Poisson distribution does not precisely match the histogram distributions, indicating some difference between their mean and variance, they exhibit a notable resemblance in overall patterns. Hence, it can be reasonably asserted that the underlying assumptions of Poisson regression are adequately satisfied.





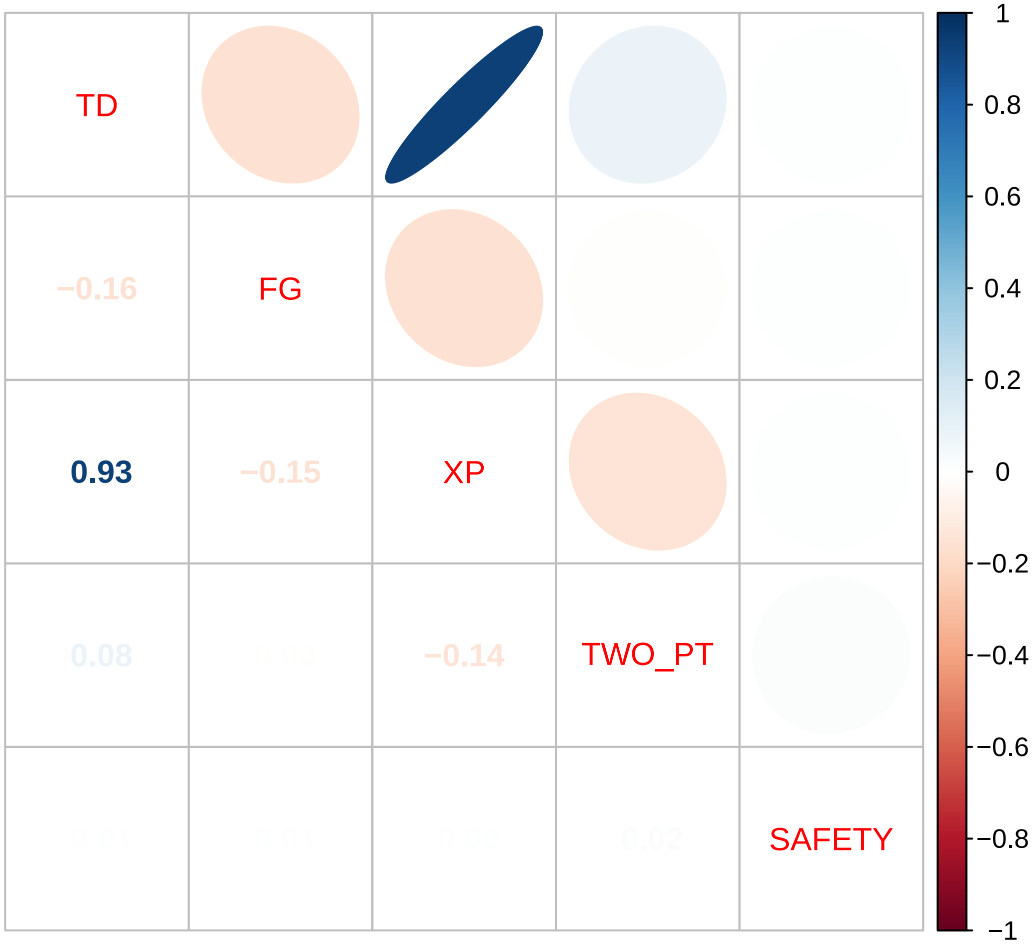






Subsequently, we explored the relationship between each of the five scoring types by plotting the correlation matrix shown below. Given the games' limited duration (60-minute regulation time plus a potential 15-minute overtime) and scoring procedures, it is understandable that certain scoring events may exhibit correlation with one another. For instance, given that each touchdown is followed by either an extra point or a two-point conversion, it makes sense that both conversion types positively correlated with touchdowns.

However, considering that the vast majority of conversion attempts after touchdowns are extra points, it follows that there is a notably stronger positive association between touchdowns and extra points compared to touchdowns and two-point attempts. This also implies that the two conversion types are negatively correlated with one another, as opting for one implies forgoing the other, and vice versa. Furthermore, given the limited game duration and consequent restriction on the number of drives for each team, it stands to reason that touchdowns and field goals are negatively correlated, as more drives resulting in touchdowns reduce the number of possible drives that result in field goals and vice versa. Consequently, a similar negative relationship is expected between extra points and two-point conversions in relation to field goals, considering that both types of conversions directly follow touchdowns. Finally, since safeties are inherently random and unrelated to other scoring events, it follows that they are uncorrelated with the four other scoring types.



Modeling/Analysis

Poisson regression relies on four key assumptions. Firstly, the response variable should follow a Poisson distribution, representing counts per unit of time or space. In our case, this assumption is clearly met as we are specifically modeling the tallies for each of the five NFL scoring categories, which all seemingly follow a Poisson distribution. Secondly, the observations must be independent of each other, ensuring that the occurrence of an event does not influence the likelihood of another. While there are two rows dedicated to each game (one for home team scoring tallies and the other for away team scoring tallies), the tallies for each game are independent, ensuring that the outcome of one game has no bearing on the outcome of another. Thirdly, the mean of the Poisson random variable must equal its variance, emphasizing a crucial property of the Poisson distribution. Based on the histograms plotted above, the distribution of counts for each of the five scoring categories resemble a Poisson distribution, sufficiently satisfying this requirement. Lastly, the linearity assumption dictates that the natural logarithm of the mean rate ($\log(\lambda)$) must form a linear relationship with the predictor variable x. Since all the predictors are categorical variables, the linearity assumption is inherently satisfied, as the levels of the categorical predictor variables are modeled as binary indicators and a straight line always fits two points exactly.

Using the aggregated scoring data, we aimed to develop a multivariate Poisson regression model incorporating all five scoring categories as response variables. However, this proved to be unfeasible due to the sparse literature regarding multivariate Poisson regression and the lack of available programming libraries that support it. After scouring a wide range of R packages and Python libraries, we were only able to source a bivariate Poisson regression model using the bpglm package, meaning that our model was limited to considering only two scoring categories simultaneously. Since touchdowns and field goals stand out as the two most common and valuable scoring events, we opted to regress on them, using number of touchdowns and number of field goals as the response variables. Additionally, in response to this constraint, we also decided to fit separate univariate Poisson regression models for each scoring category independently. While this approach may neglect potential cross-category interactions, it would enable us to capture the distinctive relationships between each scoring category and the predictors. Nevertheless, to ensure comparability with the bivariate Poisson model, we only consider the univariate Poisson models that regress on touchdowns and field goals.

To resummarize, we constructed two models: (1) a pair of univariate Poisson regressions (one regressing on touchdowns and the other on field goals), each fit with the `glm()` function, and (2) a bivariate Poisson regression (regressing on both touchdowns and field goals, fit with the `bpglm()` function. Both models were trained on all 13,166 observations of the dataset and use the same categorical predictors - team, opposing team, season, and home/away status. The team and opposing team variables encompass 32 levels each, representing every NFL team. The season variable consists of 25 levels, corresponding to each season from 1999 through 2023. Additionally, the home/away status variable includes 2 levels, indicating whether the game is played at home or away. Therefore, our Poisson regression model comprises 88 coefficients for each response variable, which is the result of incorporating an intercept term and excluding one level from each of the four predictors (specifically ARI, ARI, 1999, and away). Consequently, we have included the coefficient summaries for both models in the appendix. Tables 1 and 2 refer to the coefficients for the univariate Poisson regression on touchdowns and field goals, respectively, while Table 3 refers to the coefficients for the bivariate Poisson regression on both touchdowns and field goals.

The general form of the univariate model is given by the equations

$$\lambda_{TD} = e^{\beta_{TD\_Intercept} + \beta_{TD\_teamATL} X_{teamATL} + \cdots + \beta_{TD\_opp.teamATL} X_{opp.teamATL} + \cdots + \beta_{TD\_2000} X_{2000} + \cdots + \beta_{TD\_home} X_{home}}$$

and

$$\lambda_{FG} = e^{\beta_{FG\_Intercept} + \beta_{FG\_teamATL} X_{teamATL} + \cdots + \beta_{FG\_opp.teamATL} X_{opp.teamATL} + \cdots + \beta_{FG\_2000} X_{2000} + \cdots + \beta_{home} X_{home}}$$

In these expressions, the β coefficients with the prefix TD correspond to Table 1, while those with the prefix FG correspond to Table 2, and the X inputs are indicator variables that are dependent on specific game details (meaning that at most four of the X inputs can evaluate to 1).

The coefficients in Poisson Regression can be interpreted as the logarithm of the rate ratio or percentage change in the expected count of the dependent variable associated with a one-unit change in the corresponding independent variable, holding all other variables constant. In our case, this one-unit change is relative to the base level that was dropped from each of the predictors (i.e. ARI, ARI, 1999, and away, respectively). For instance, $\beta_{TD\_teamBUF}$ has a coefficient of 0.0610786, meaning that the Buffalo Bills (BUF) would be expected to score $e^{0.0610786} = 1.062982$ times as many touchdowns as the Arizona Cardinals (ARI), had they both played away, in 1999, against the Arizona Cardinals. Although this interpretation does not make much sense, as it is a comparison to ARI playing against ARI, which is impossible, it gives a good metric of which teams is better. Similarly, the interpretation for $\beta_{FG\_teamBAL}$, which has a coefficient value of 0.2429988, is that the Baltimore Ravens (BAL) would be expected to score $e^{0.2429988} = 1.275067$ times as many field goals as the Arizona Cardinals (ARI), had they both played away, in 1999, against the Arizona Cardinals.

Moreover, we can interpret $\beta_{TD\_season2015}$ with a coefficient value of 0.0887177 as saying that the we would expected $e^{0.0887177} = 1.092772$ times as many touchdowns in the 2015 season as in the 1999 season, had the Arizona Cardinals (ARI) played the Arizona Cardinals (ARI). Again, while the interpretation seems illogical, it illustrates the increase in touchdowns score count between the two seasons. Finally, given the $\beta_{FG\_opp.teamJAX}$ coefficient of -0.0270338, it can be interpreted as playing against the Jacksonville Jaguars (JAX) is associated with a $1 - e^{-0.0270338} = 0.0266717 \approx 2.6$ reduction in the expected number of field goals, again, holding all other predictors constant.

In the case of the bivariate Poisson regression, it was quite challenging to interpret exactly what was happening under-the-hood of the `bpglm()` function due to the limited documentation, but from what we gather, the bivariate model is an untruncated, dependent (marginal/conditional) bivariate Poisson model that accounts for the inherent negative correlation between touchdown and field goal counts. Hence, the interpretation of λ_{TD} should be the same as in the univariate model; however, there is not much clarity on the interpretation of λ_{FG}. Based on the functions documentation, the model is fit such that the joint pdf follows the equation

$$g(y_{TD}, y_{FG}) = \frac{e^{-\lambda_{TD}}\,(\lambda_{TD}^{y_{TD}}) \cdot e^{-\lambda_{FG}\cdot y_{TD}}\,(\lambda_{FG}\cdot y_{TD})^{y_{FG}}}{y_{TD}! \cdot y_{FG}!}, \quad y_{TD} = 0, 1, \ldots;\ y_{FG} = 0, 1, \ldots;\ \lambda_{TD}, \lambda_{FG} > 0$$

where the number of touchdown occurrences Y_{TD} in a given interval follows a Poisson distribution with parameter λ_{TD}, and the number of field goal occurrences Y_{FG}, for a given Y_{TD}, is also Poisson with parameter $\lambda_{FG} \cdot y_{TD}$. Hence, we can see that the expected number of field goals scored is dependent on the expected number of touchdowns scored, meaning that this model theoretically takes into account the inherent negative correlation between the two scoring events.

Visualization and interpretation of the results

To assess the effectiveness of our models, we evaluated how well it would perform in the online betting market. In the case of the univariate model, we used the `predict()` function to calculate the λ_{TD} and λ_{FG} for the home and away team in each game of our training dataset (i.e. all 13,166 observations). Using each pair of λ_{TD} and λ_{FG}, we simulated the number of touchdowns (y_{TD}) and field goals (y_{FG}) scored 1,000 times by sampling from their respective Poisson distributions. Subsequently, we computed corresponding final score for each simulation using the generated touchdown and field goal values. For simplicity, we assumed a scoring system of 7 points for each touchdown and 3 points for each field goal. Hence, we now had 1000 simulated final scores based on the Poisson distributions of touchdowns and field goals. Next, we calculated the difference and sum of scores for each of the 1,000 simulation of each game, reducing the dataset back down to 6,583 observations. Using the reported closing point spread and over/under line for each game, we determined the proportion of simulations that surpassed the point spread and over/under line, as well as those that fell below these benchmarks, and "bet" on the higher of the two proportions for each of the two betting types. In other words, we selected the betting strategy that had a greater likelihood based on the simulation outcomes, aligning our bets with the more favorable predicted scenario. In the bivariate model, we conducted similar simulations as in the previous case. However, since the `bpglm` function lacked a `predict()` equivalent, in each of the 1,000 simulations, we additionally sampled a set of beta values from the beta distribution. These beta values were then utilized to compute both λ_{TD} and λ_{FG} for the simulations.

Finally, to evaluate the accuracy of our predictions, we determined how the predicted bets performed (i.e. whether they won or not) against the true outcomes of the 6,583 games. The univariate model consistently predicts both the point spread and over/under with around 51% accuracy, whereas the bivariate model consistently predicts both with approximately 53% accuracy. Hence, both models surpass random guessing, demonstrating their capability to discern some underlying patterns within the data. Moreover, the enhanced accuracy of the bivariate model can be attributed to it's ability to capture the inherent negative correlation between the number of touchdowns and field goals scored by a team. However, it's crucial to approach these results with a dose of skepticism. While the univariate and bivariate models showcase a modest improvement over random guessing, their predictive capabilities may not be as remarkable as they initially seem. Both models only marginally surpassed the 50% threshold, and despite attempts at experimental optimization, there was minimal enhancement in their betting accuracy. Consequently, we began to suspect an error in the fundamental assumption of a Poisson regression model.

To investigate the performance issues observed in our models, we conducted a thorough examination of the dispersion of the data. This analysis revealed a notable overdispersion, challenging the Poisson regression assumption that the response variables exhibit equidispersion. The presence of overdispersion implies that the variability in the data is higher than what the Poisson distribution can adequately capture. This finding prompted a critical reassessment of our modeling approach, leading us to consider alternative methodologies. Given the observed overdispersion, a negative binomial model emerges as a promising candidate for more accurately representing the underlying patterns in the data. Unlike the Poisson model, the negative binomial distribution accommodates varying levels of dispersion, making it better suited to handle the complexities inherent in our dataset. Hence, a future analysis fitting a negative binomial model to our data will likely provide more accurate and reliable results. Additionally, we plan to incorporate more covariates and explore potential sources of heterogeneity to further enhance the model's predictive power.

Conclusions and recommendations

In conclusion, our exploration into NFL score forecasting has revealed promising insights and challenges. The traditional approach of treating scores as continuous variables has limitations in capturing the nuanced dynamics of American football scoring, particularly the distinctive patterns involving multiples of 3 and 7. Our innovative bivariate Poisson regression model, focusing on touchdowns and field goals, has demonstrated an accuracy of around 53% in predicting both the spread and over-under in the online betting market. While this surpasses random guessing, the observed overdispersion in the data calls for a reevaluation of our modeling approach. The Poisson regression's assumption of equidispersion appears violated, prompting consideration of a negative binomial model for a more accurate representation.

Moving forward, we plan to further research incorporating a negative binomial model to address overdispersion. This alternative modeling approach has the potential to enhance predictive accuracy by accommodating the variability present in NFL scoring data. Additionally, incorporating features like player statistics, weather conditions, and team dynamics as additional covariates could introduce valuable insights and improve the overall accuracy of the predictions. Furthermore, considering temporal dynamics, such as trends across different seasons or changes in team strategies over time, will contribute to a more adaptive and robust forecasting model. Moreover, leveraging advanced machine learning techniques such as ensemble models and deep learning architectures might offer additional avenues for improvement.

Appendix (optional)

Table 1: Coefficients Summary for Univariate Poisson Regression on Touchdowns

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.8081080	0.0535156	15.1004080	0.0000000
teamATL	0.0629570	0.0457387	1.3764488	0.1686827
teamBAL	0.0662048	0.0458318	1.4445163	0.1485938
teamBUF	0.0629741	0.0464035	1.3570981	0.1747500
teamCAR	0.0011705	0.0464065	0.0252225	0.9798775
teamCHI	-0.0466207	0.0468529	-0.9950425	0.3197156
teamCIN	0.0652129	0.0464932	1.4026346	0.1607258
teamCLE	-0.1012697	0.0487423	-2.0776561	0.0377410
teamDAL	0.1408778	0.0447949	3.1449543	0.0016611
teamDEN	0.0804472	0.0459237	1.7517562	0.0798157
teamDET	0.0408151	0.0464229	0.8792015	0.3792920
teamGB	0.2467031	0.0434135	5.6826423	0.0000000
teamHOU	-0.0198800	0.0485032	-0.4098690	0.6819020
teamIND	0.2066368	0.0443984	4.6541512	0.0000033
teamJAX	-0.0344542	0.0473474	-0.7276880	0.4668046
teamKC	0.1917401	0.0445327	4.3056015	0.0000167
teamLA	0.0694715	0.0451979	1.5370513	0.1242807
teamLAC	0.1720389	0.0450888	3.8155582	0.0001359
teamLV	0.0069071	0.0469544	0.1471033	0.8830505
teamMIA	0.0156508	0.0470830	0.3324091	0.7395804
teamMIN	0.1302878	0.0448881	2.9024988	0.0037020
teamNE	0.2693701	0.0435607	6.1837821	0.0000000
teamNO	0.2573797	0.0436830	5.8919822	0.0000000
teamNYG	0.0315664	0.0459837	0.6864689	0.4924175
teamNYJ	-0.0869939	0.0480727	-1.8096330	0.0703527
teamPHI	0.1695335	0.0441346	3.8412867	0.0001224
teamPIT	0.1082154	0.0454054	2.3833136	0.0171576
teamSEA	0.1207581	0.0443477	2.7229844	0.0064695
teamSF	0.0483198	0.0454422	1.0633252	0.2876345
teamTB	0.0208203	0.0458819	0.4537813	0.6499863
teamTEN	0.0643691	0.0460056	1.3991563	0.1617661
teamWAS	-0.0389968	0.0470659	-0.8285575	0.4073549
opp.teamATL	-0.0473898	0.0429723	-1.1027995	0.2701142
opp.teamBAL	-0.3345835	0.0466602	-7.1706438	0.0000000
opp.teamBUF	-0.1319285	0.0444557	-2.9676427	0.0030009
opp.teamCAR	-0.1110643	0.0437133	-2.5407446	0.0110617
opp.teamCHI	-0.1579584	0.0441479	-3.5779338	0.0003463
opp.teamCIN	-0.0707445	0.0439457	-1.6098162	0.1074380
opp.teamCLE	-0.0424433	0.0436969	-0.9713103	0.3313938
opp.teamDAL	-0.1173981	0.0438063	-2.6799361	0.0073636
opp.teamDEN	-0.1355157	0.0442911	-3.0596598	0.0022159
opp.teamDET	0.0491868	0.0422817	1.1633121	0.2447029
opp.teamGB	-0.0975387	0.0433237	-2.2513948	0.0243605
opp.teamHOU	-0.0468934	0.0448189	-1.0462858	0.2954291
opp.teamIND	-0.0676369	0.0434691	-1.5559759	0.1197138
opp.teamJAX	-0.0930296	0.0440557	-2.1116378	0.0347175
opp.teamKC	-0.0935534	0.0436908	-2.1412587	0.0322532
opp.teamLA	-0.0560212	0.0426788	-1.3126222	0.1893103

	Estimate	Std. Error	z value	Pr(>|z|)
opp.teamLAC	-0.0845569	0.0439379	-1.9244636	0.0542965
opp.teamLV	-0.0057006	0.0429529	-0.1327172	0.8944170
opp.teamMIA	-0.1211130	0.0443741	-2.7293643	0.0063457
opp.teamMIN	-0.0995553	0.0435507	-2.2859625	0.0222565
opp.teamNE	-0.2176536	0.0448769	-4.8500153	0.0000012
opp.teamNO	-0.0351762	0.0430248	-0.8175789	0.4135977
opp.teamNYG	-0.0854172	0.0432428	-1.9752931	0.0482349
opp.teamNYJ	-0.1426547	0.0443907	-3.2136180	0.0013107
opp.teamPHI	-0.1342091	0.0436351	-3.0757154	0.0021000
opp.teamPIT	-0.2346910	0.0453647	-5.1734241	0.0000002
opp.teamSEA	-0.1646595	0.0437646	-3.7623934	0.0001683
opp.teamSF	-0.0979025	0.0431738	-2.2676361	0.0233514
opp.teamTB	-0.1463033	0.0438281	-3.3381181	0.0008435
opp.teamTEN	-0.0813466	0.0437766	-1.8582228	0.0631374
opp.teamWAS	-0.0767778	0.0433977	-1.7691691	0.0768657
season2000	-0.0098558	0.0412853	-0.2387239	0.8113197
season2001	-0.0344257	0.0413212	-0.8331230	0.4047754
season2002	0.0734992	0.0399371	1.8403740	0.0657133
season2003	0.0038620	0.0406239	0.0950659	0.9242625
season2004	0.0684919	0.0400329	1.7108898	0.0871015
season2005	-0.0196046	0.0410224	-0.4778998	0.6327215
season2006	-0.0125981	0.0407552	-0.3091173	0.7572323
season2007	0.0410003	0.0402309	1.0191236	0.3081443
season2008	0.0435835	0.0401559	1.0853573	0.2777635
season2009	0.0512758	0.0400603	1.2799650	0.2005575
season2010	0.0770417	0.0400112	1.9255025	0.0541665
season2011	0.0557447	0.0400327	1.3924781	0.1637777
season2012	0.0903310	0.0397928	2.2700349	0.0232055
season2013	0.1138086	0.0395445	2.8779892	0.0040022
season2014	0.0825193	0.0397706	2.0748814	0.0379975
season2015	0.0886031	0.0399089	2.2201355	0.0264096
season2016	0.0883712	0.0399286	2.2132321	0.0268816
season2017	0.0266698	0.0403640	0.6607317	0.5087844
season2018	0.1356232	0.0393462	3.4469242	0.0005670
season2019	0.1102870	0.0395899	2.7857388	0.0053406
season2020	0.2068068	0.0386152	5.3555812	0.0000001
season2021	0.1163980	0.0389128	2.9912490	0.0027784
season2022	0.0477193	0.0395279	1.2072318	0.2273429
season2023	0.0321025	0.0399440	0.8036897	0.4215762
home_or_awayhome	0.1192712	0.0110435	10.8001273	0.0000000

Table 2: Coefficients Summary for Univariate Poisson Regression on Field Goals

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.3156620	0.0679205	4.6475230	0.0000034
teamATL	0.1215440	0.0561869	2.1632109	0.0305250
teamBAL	0.2412542	0.0549067	4.3938963	0.0000111
teamBUF	0.0522782	0.0578327	0.9039565	0.3660185
teamCAR	0.0735429	0.0569043	1.2923960	0.1962200
teamCHI	0.0149899	0.0576521	0.2600068	0.7948585
teamCIN	-0.0030182	0.0584768	-0.0516136	0.9588366
teamCLE	-0.0535358	0.0594677	-0.9002494	0.3679875
teamDAL	0.0523231	0.0569570	0.9186425	0.3582826
teamDEN	0.1206521	0.0568602	2.1219093	0.0338454
teamDET	0.0072580	0.0580338	0.1250643	0.9004726
teamGB	0.0129528	0.0569217	0.2275544	0.8199926
teamHOU	0.0712654	0.0590912	1.2060239	0.2278083
teamIND	0.1215210	0.0563408	2.1568930	0.0310140
teamJAX	0.0100368	0.0583005	0.1721567	0.8633143
teamKC	0.0448073	0.0572927	0.7820780	0.4341688
teamLA	0.1086589	0.0557469	1.9491473	0.0512778
teamLAC	0.0545582	0.0576254	0.9467729	0.3437545
teamLV	0.0860002	0.0573361	1.4999312	0.1336322
teamMIA	0.0386782	0.0581619	0.6650093	0.5060445
teamMIN	0.0138411	0.0576768	0.2399777	0.8103476
teamNE	0.1432940	0.0555042	2.5816763	0.0098322
teamNO	0.0266765	0.0573415	0.4652206	0.6417735
teamNYG	0.0424313	0.0570122	0.7442498	0.4567254
teamNYJ	0.1069179	0.0571490	1.8708600	0.0613645
teamPHI	0.0850734	0.0559765	1.5198055	0.1285599
teamPIT	0.1270131	0.0560387	2.2665234	0.0234194
teamSEA	0.0663807	0.0560378	1.1845701	0.2361874
teamSF	0.0741460	0.0562118	1.3190463	0.1871536
teamTB	0.0402989	0.0569875	0.7071541	0.4794707
teamTEN	0.0063799	0.0581054	0.1097987	0.9125690
teamWAS	0.0104186	0.0578715	0.1800300	0.8571290
opp.teamATL	-0.0160689	0.0565255	-0.2842773	0.7761979
opp.teamBAL	0.0266242	0.0560643	0.4748871	0.6348675
opp.teamBUF	-0.0316134	0.0570381	-0.5542513	0.5794069
opp.teamCAR	0.0026409	0.0562954	0.0469123	0.9625831
opp.teamCHI	0.0221048	0.0562027	0.3933043	0.6940948
opp.teamCIN	0.0561843	0.0557221	1.0082960	0.3133124
opp.teamCLE	0.0201428	0.0564174	0.3570323	0.7210677
opp.teamDAL	0.0078747	0.0560651	0.1404557	0.8882999
opp.teamDEN	0.0591280	0.0558113	1.0594264	0.2894057
opp.teamDET	-0.0103697	0.0569698	-0.1820204	0.8555667
opp.teamGB	-0.0454622	0.0564452	-0.8054226	0.4205758
opp.teamHOU	-0.0140688	0.0586509	-0.2398736	0.8104283
opp.teamIND	0.0130820	0.0560153	0.2335437	0.8153393
opp.teamJAX	-0.0300979	0.0571014	-0.5270952	0.5981275
opp.teamKC	-0.0703718	0.0571270	-1.2318491	0.2180054
opp.teamLA	0.0403768	0.0551683	0.7318832	0.4642399
opp.teamLAC	-0.0588326	0.0574691	-1.0237245	0.3059654
opp.teamLV	0.0330397	0.0562600	0.5872670	0.5570244

	Estimate	Std. Error	z value	Pr(>|z|)
opp.teamMIA	0.0088931	0.0564795	0.1574569	0.8748848
opp.teamMIN	0.1083512	0.0549096	1.9732662	0.0484652
opp.teamNE	-0.1556515	0.0577938	-2.6932204	0.0070765
opp.teamNO	0.0210671	0.0557141	0.3781285	0.7053351
opp.teamNYG	0.0070535	0.0559536	0.1260603	0.8996842
opp.teamNYJ	0.0317099	0.0562099	0.5641330	0.5726636
opp.teamPHI	-0.0383898	0.0562342	-0.6826766	0.4948112
opp.teamPIT	-0.0612444	0.0569162	-1.0760443	0.2819074
opp.teamSEA	0.0414178	0.0548098	0.7556634	0.4498510
opp.teamSF	0.0023406	0.0556967	0.0420235	0.9664800
opp.teamTB	-0.0409270	0.0565361	-0.7239087	0.4691218
opp.teamTEN	0.0036194	0.0563278	0.0642557	0.9487666
opp.teamWAS	0.0408252	0.0558193	0.7313805	0.4645468
season2000	-0.0262292	0.0517459	-0.5068845	0.6122359
season2001	-0.0055970	0.0512406	-0.1092294	0.9130205
season2002	-0.0321932	0.0511919	-0.6288730	0.5294322
season2003	-0.0061589	0.0508996	-0.1210003	0.9036908
season2004	-0.0901172	0.0520380	-1.7317563	0.0833170
season2005	0.0166474	0.0507589	0.3279701	0.7429343
season2006	0.0200428	0.0505305	0.3966476	0.6916273
season2007	0.0255785	0.0504836	0.5066690	0.6123871
season2008	0.0932843	0.0495725	1.8817779	0.0598662
season2009	-0.0244205	0.0509819	-0.4790037	0.6319360
season2010	0.0196099	0.0506691	0.3870181	0.6987428
season2011	0.0899263	0.0496224	1.8122131	0.0699533
season2012	0.1053720	0.0495165	2.1280194	0.0333355
season2013	0.1189434	0.0493660	2.4094198	0.0159779
season2014	0.0734705	0.0497975	1.4753849	0.1401091
season2015	0.0956559	0.0497765	1.9217086	0.0546424
season2016	0.1119147	0.0496482	2.2541555	0.0241864
season2017	0.1221024	0.0493356	2.4749361	0.0133260
season2018	0.0486697	0.0501921	0.9696672	0.3322124
season2019	0.0466991	0.0502253	0.9297931	0.3524782
season2020	0.0664181	0.0498073	1.3335025	0.1823670
season2021	0.0773323	0.0490643	1.5761415	0.1149932
season2022	0.1062434	0.0487450	2.1795761	0.0292889
season2023	0.1061959	0.0490655	2.1643692	0.0304360
home_or_awayhome	0.0488650	0.0138620	3.5251113	0.0004233

Table 3: Coefficients Summary for Bivariate Poisson Regression on Touchdowns and Field Goals

	Coefficients	Standard Error	t-value	p-value
Y1:Constant	0.808108	0.048413	16.6919629	0.000000
teamATL	0.062957	0.041378	1.5215090	0.128153
teamBAL	0.066205	0.041462	1.5967633	0.110343
teamBUF	0.062974	0.041979	1.5001310	0.133605
teamCAR	0.001170	0.041982	0.0278691	0.977758
teamCHI	-0.046621	0.042386	-1.0999151	0.271389
teamCIN	0.065213	0.042060	1.5504755	0.121054
teamCLE	-0.101270	0.044095	-2.2966323	0.021655
teamDAL	0.140878	0.040524	3.4764090	0.000510
teamDEN	0.080447	0.041545	1.9363822	0.052842
teamDET	0.040815	0.041997	0.9718551	0.331135
teamGB	0.246703	0.039274	6.2815858	0.000000
teamHOU	-0.019880	0.043879	-0.4530641	0.650508
teamIND	0.206637	0.040165	5.1447031	0.000000
teamJAX	-0.034454	0.042833	-0.8043798	0.421190
teamKC	0.191740	0.040287	4.7593517	0.000002
teamLA	0.069472	0.040888	1.6990804	0.089334
teamLAC	0.172039	0.040790	4.2176759	0.000025
teamLV	0.006907	0.042477	0.1626056	0.870830
teamMIA	0.015651	0.042594	0.3674461	0.713294
teamMIN	0.130288	0.040608	3.2084318	0.001338
teamNE	0.269370	0.039407	6.8355876	0.000000
teamNO	0.257380	0.039518	6.5129814	0.000000
teamNYG	0.031566	0.041599	0.7588163	0.447974
teamNYJ	-0.086994	0.043489	-2.0003679	0.045482
teamPHI	0.169534	0.039926	4.2462055	0.000022
teamPIT	0.108215	0.041076	2.6345068	0.008436
teamSEA	0.120758	0.040119	3.0099953	0.002618
teamSF	0.048320	0.041109	1.1754117	0.239858
teamTB	0.020820	0.041507	0.5016021	0.615952
teamTEN	0.064369	0.041619	1.5466253	0.121979
teamWAS	-0.038997	0.042578	-0.9158955	0.359745
opp.teamATL	-0.047390	0.038875	-1.2190354	0.222855
opp.teamBAL	-0.334584	0.042211	-7.9264647	0.000000
opp.teamBUF	-0.131929	0.040217	-3.2804287	0.001039
opp.teamCAR	-0.111064	0.039545	-2.8085472	0.004984
opp.teamCHI	-0.157958	0.039939	-3.9549813	0.000077
opp.teamCIN	-0.070744	0.039756	-1.7794547	0.075184
opp.teamCLE	-0.042443	0.039531	-1.0736637	0.282985
opp.teamDAL	-0.117398	0.039630	-2.9623518	0.003058
opp.teamDEN	-0.135516	0.040068	-3.3821503	0.000721
opp.teamDET	0.049187	0.038250	1.2859346	0.198494
opp.teamGB	-0.097539	0.039193	-2.4886842	0.012834
opp.teamHOU	-0.046893	0.040546	-1.1565383	0.247473
opp.teamIND	-0.067637	0.039324	-1.7199929	0.085462
opp.teamJAX	-0.093030	0.039855	-2.3342115	0.019601
opp.teamKC	-0.093553	0.039525	-2.3669323	0.017950
opp.teamLA	-0.056021	0.038610	-1.4509454	0.146813

opp.teamLAC	-0.084557	0.039749	-2.1272736	0.033414
opp.teamLV	-0.005701	0.038858	-0.1467137	0.883367
opp.teamMIA	-0.121113	0.040143	-3.0170391	0.002558
opp.teamMIN	-0.099555	0.039398	-2.5269049	0.011519
opp.teamNE	-0.217654	0.040598	-5.3612001	0.000000
opp.teamNO	-0.035176	0.038923	-0.9037330	0.366146
opp.teamNYG	-0.085417	0.039120	-2.1834611	0.029018
opp.teamNYJ	-0.142655	0.040158	-3.5523432	0.000383
opp.teamPHI	-0.134209	0.039475	-3.3998480	0.000676
opp.teamPIT	-0.234691	0.041039	-5.7187310	0.000000
opp.teamSEA	-0.164659	0.039592	-4.1588957	0.000032
opp.teamSF	-0.097903	0.039057	-2.5066697	0.012201
opp.teamTB	-0.146303	0.039649	-3.6899543	0.000225
opp.teamTEN	-0.081347	0.039603	-2.0540616	0.039989
opp.teamWAS	-0.076778	0.039260	-1.9556291	0.050530
season2000	-0.009856	0.037349	-0.2638893	0.791873
season2001	-0.034426	0.037381	-0.9209491	0.357104
season2002	0.073499	0.036129	2.0343491	0.041937
season2003	0.003862	0.036751	0.1050856	0.916310
season2004	0.068492	0.036216	1.8912083	0.058618
season2005	-0.019605	0.037111	-0.5282800	0.597322
season2006	-0.012598	0.036869	-0.3416963	0.732584
season2007	0.041000	0.036395	1.1265284	0.259960
season2008	0.043584	0.036327	1.1997688	0.230258
season2009	0.051276	0.036241	1.4148616	0.157131
season2010	0.077042	0.036196	2.1284672	0.033319
season2011	0.055745	0.036216	1.5392368	0.123770
season2012	0.090331	0.035999	2.5092641	0.012109
season2013	0.113809	0.035774	3.1813328	0.001469
season2014	0.082519	0.035979	2.2935323	0.021831
season2015	0.088603	0.036104	2.4541048	0.014136
season2016	0.088371	0.036121	2.4465270	0.014438
season2017	0.026670	0.036515	0.7303848	0.465177
season2018	0.135623	0.035595	3.8101700	0.000139
season2019	0.110287	0.035815	3.0793522	0.002079
season2020	0.206807	0.034933	5.9201042	0.000000
season2021	0.116398	0.035203	3.3064796	0.000947
season2022	0.047719	0.035759	1.3344613	0.182073
season2023	0.032103	0.036135	0.8884184	0.374345
home_or_awayhome	0.119271	0.009991	11.9378441	0.000000
Y2:Constant	-0.501868	0.064502	-7.7806580	0.000000
teamATL	0.064042	0.053206	1.2036612	0.228743
teamBAL	0.179475	0.051934	3.4558286	0.000550
teamBUF	-0.008946	0.054794	-0.1632661	0.870312
teamCAR	0.071582	0.053896	1.3281505	0.184151
teamCHI	0.060786	0.054547	1.1143784	0.265135
teamCIN	-0.071704	0.055394	-1.2944362	0.195539
teamCLE	0.052227	0.056401	0.9259942	0.354471
teamDAL	-0.089414	0.053961	-1.6570115	0.097541
teamDEN	0.039075	0.053832	0.7258694	0.467933
teamDET	-0.037098	0.054927	-0.6754055	0.499439

teamGB	-0.234388	0.053895	-4.3489749	0.000014
teamHOU	0.098618	0.056026	1.7602185	0.078393
teamIND	-0.078575	0.053319	-1.4736773	0.140592
teamJAX	0.036510	0.055267	0.6606112	0.508869
teamKC	-0.139437	0.054224	-2.5714997	0.010136
teamLA	0.044380	0.052929	0.8384817	0.401774
teamLAC	-0.115088	0.054523	-2.1108156	0.034809
teamLV	0.081092	0.054403	1.4905796	0.136096
teamMIA	0.022314	0.055127	0.4047744	0.685652
teamMIN	-0.109748	0.054577	-2.0108837	0.044356
teamNE	-0.128531	0.052584	-2.4442986	0.014526
teamNO	-0.227444	0.054285	-4.1898130	0.000028
teamNYG	0.008376	0.054022	0.1550479	0.876783
teamNYJ	0.202972	0.054219	3.7435585	0.000182
teamPHI	-0.086232	0.053070	-1.6248728	0.104215
teamPIT	0.023285	0.053048	0.4389421	0.660711
teamSEA	-0.051309	0.053109	-0.9661074	0.334006
teamSF	0.031137	0.053351	0.5836254	0.559488
teamTB	0.030368	0.054016	0.5622038	0.573993
teamTEN	-0.047724	0.054970	-0.8681826	0.385311
teamWAS	0.044285	0.054817	0.8078698	0.419176
opp.teamATL	0.029389	0.053493	0.5493990	0.582735
opp.teamBAL	0.365802	0.053171	6.8797277	0.000000
opp.teamBUF	0.097780	0.054095	1.8075608	0.070698
opp.teamCAR	0.112928	0.053273	2.1197980	0.034042
opp.teamCHI	0.184135	0.053266	3.4568956	0.000548
opp.teamCIN	0.125757	0.052785	2.3824382	0.017214
opp.teamCLE	0.064007	0.053473	1.1969966	0.231337
opp.teamDAL	0.128458	0.053110	2.4187159	0.015588
opp.teamDEN	0.193257	0.052805	3.6598239	0.000253
opp.teamDET	-0.062905	0.053875	-1.1676102	0.242983
opp.teamGB	0.053790	0.053398	1.0073411	0.313790
opp.teamHOU	0.032567	0.055493	0.5868668	0.557308
opp.teamIND	0.081286	0.052990	1.5339875	0.125055
opp.teamJAX	0.068948	0.054036	1.2759642	0.201987
opp.teamKC	0.028305	0.054055	0.5236333	0.600546
opp.teamLA	0.095815	0.052222	1.8347631	0.066562
opp.teamLAC	0.025702	0.054335	0.4730284	0.636208
opp.teamLV	0.038589	0.053194	0.7254390	0.468197
opp.teamMIA	0.125716	0.053595	2.3456666	0.019007
opp.teamMIN	0.213114	0.051973	4.1004752	0.000041
opp.teamNE	0.061615	0.054754	1.1253059	0.260479
opp.teamNO	0.053226	0.052703	1.0099235	0.312544
opp.teamNYG	0.094365	0.052978	1.7812111	0.074904
opp.teamNYJ	0.174635	0.053227	3.2809476	0.001037
opp.teamPHI	0.096269	0.053270	1.8071898	0.070752
opp.teamPIT	0.180889	0.053870	3.3578801	0.000788
opp.teamSEA	0.209726	0.051915	4.0397958	0.000054
opp.teamSF	0.107415	0.052801	2.0343365	0.041938
opp.teamTB	0.096568	0.053571	1.8026171	0.071471
opp.teamTEN	0.086614	0.053393	1.6221977	0.104785

opp.teamWAS	0.114670	0.052909	2.1673061	0.030229
season2000	-0.012219	0.049098	-0.2488696	0.803471
season2001	0.033550	0.048615	0.6901162	0.490125
season2002	-0.104368	0.048561	-2.1492144	0.031636
season2003	0.005748	0.048283	0.1190481	0.905245
season2004	-0.145654	0.049378	-2.9497752	0.003186
season2005	0.049216	0.048170	1.0217148	0.306931
season2006	0.046524	0.047977	0.9697147	0.332209
season2007	-0.001714	0.047962	-0.0357366	0.971493
season2008	0.060782	0.047107	1.2902966	0.196971
season2009	-0.059153	0.048446	-1.2210090	0.222103
season2010	-0.051177	0.048135	-1.0631973	0.287715
season2011	0.045701	0.047170	0.9688573	0.332633
season2012	0.022239	0.047042	0.4727478	0.636411
season2013	0.014436	0.046892	0.3078564	0.758199
season2014	-0.001261	0.047285	-0.0266681	0.978718
season2015	0.009245	0.047241	0.1956987	0.844857
season2016	0.034949	0.047119	0.7417178	0.458269
season2017	0.101477	0.046820	2.1673857	0.030221
season2018	-0.077040	0.047641	-1.6170945	0.105881
season2019	-0.059315	0.047681	-1.2439966	0.213522
season2020	-0.133137	0.047293	-2.8151523	0.004883
season2021	-0.029719	0.046578	-0.6380480	0.523461
season2022	0.068714	0.046283	1.4846488	0.137666
season2023	0.078381	0.046596	1.6821401	0.092562
home_or_awayhome	-0.074271	0.013148	-5.6488439	0.000000

Work Cited

1. Gifford, Matt, and Tuncay Bayrak. "A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression." *Decision Analytics Journal*, vol. 8, Sept. 2023, https://doi.org/10.1016/j.dajour.2023.100296.
2. The Factory of Sadness. "Machine Learning Model for NFL Betting (Model 5.0)." *Medium*, Medium, 15 May 2023, medium.com/@bravenewworld21/machine-learning-model-for-nfl-betting-model-5-0-8e916428c330.
3. AlMuhayfith, Fatimah E., et al. "On bivariate Poisson Regression Models." *Journal of King Saud University - Science*, vol. 28, no. 2, Apr. 2016, pp. 178–189, https://doi.org/10.1016/j.jksus.2015.09.003.
4. Ford, Clay. "Getting Started with Multivariate Multiple Regression." *UVA Library StatLab*, 27 Oct. 2017, library.virginia.edu/data/articles/getting-started-with-multivariate-multiple-regression.
5. https://bmacgtpm.github.io/notes