Algorithmic Projections of Fantasy Football Performance

Samuel Hughes
Khoury College of Computer Science
Northeastern University
hughes.sa@northeastern.edu

Jason Curtis
Khoury College of Computer Science
Northeastern University
curtis.jas@northeastern.edu

Abstract

With over 29 million participants in fantasy football, a virtual game based on real-world National Football League (NFL) player statistics, accurate fantasy point projections are essential. Using supervised learning, we developed open-source projections trained on historical NFL game data from 2020 to 2023. After a robust analysis of different modeling techniques, we selected dense neural networks as our modeling algorithm, yielding a mean absolute error of 4.63 fantasy points. Feature impact was analyzed using Shapley values, highlighting the importance of statistics derived from the betting market along with consensus expert rankings. The transparent, open-source projections provide accurate forecasts of player performance with interpretable results, showing promise for applications in fantasy football. Future work may explore projections as a statistical distribution and correlations between teammates to further refine predictive ability.

1. Introduction

American football is a sport where two teams aim to advance the ball into the opponent's end zone. The game combines physical skill and tactical planning, featuring specialized roles on offense, defense, and special teams. Offensive players work to advance the ball and score, while defensive players focus on stopping the opposing offense, and special teams handle specific tasks such as kicking the football [1]. This complex game offers a rich statistical landscape that has paved the way for fantasy football.

Fantasy football is a virtual game where participants, called fantasy football owners, create teams made up of real-life National Football League (NFL) players, competing based on the weekly performance of these players. Points are awarded for each player's real-world statistics, such as yards gained and touchdowns, contributing to the fantasy team's score in weekly matchups. Fantasy football merges knowledge of the sport with statistical analysis, as participants strategize in selecting the football players that maximize their fantasy team's potential. The game has grown immensely popular, with a player base of over 29 million Americans [2].

In fantasy football scoring, several key statistics are critical for success, as they directly impact a player's scoring potential. As an example, passing yards and passing touchdowns are especially significant for quarterbacks, while rushing yards, receiving yards, and touchdowns are vital for running backs, wide receivers, and tight ends. Appendix Table 1 offers a description of the vital fantasy football statistics along with the corresponding fantasy point values.

Fantasy football owners choose between various NFL players to insert into their lineup in an effort to accumulate a maximum amount of fantasy points. However, choosing between NFL players is not always straightforward. Advice from fantasy football experts can sometimes be conflicting, thus lacking a unanimous source of truth. In addition, other projection algorithms from popular fantasy football sources lack transparency. Because of this, it is difficult to determine the statistical validity of current algorithms. Using a combination of consensus expert rankings, information from the NFL betting markets, and machine learning, we propose open-source projections with clearly reported and validated model performance.

2. Data

In order to utilize supervised learning, a target variable is necessary. In this projection algorithm, the target variable is the amount of fantasy points a player scored in a previous NFL game [3].

To predict the fantasy points scored by NFL players, one crucial source of information is consensus expert rankings. Experts are tasked with ranking players from best to worst based on the amount of fantasy points they believe the player will score. The rankings are then aggregated across numerous experts to create a consensus. These consensus rankings are provided for all skill position players along with a separate ranking within each position grouping

[4]. Figure 1 shows the relationship between the consensus rankings and the target variable. Please note that the rankings are designated from first to worst, thus an ordinal ranking of 1 corresponds to the player that the experts believe will score the most fantasy points.

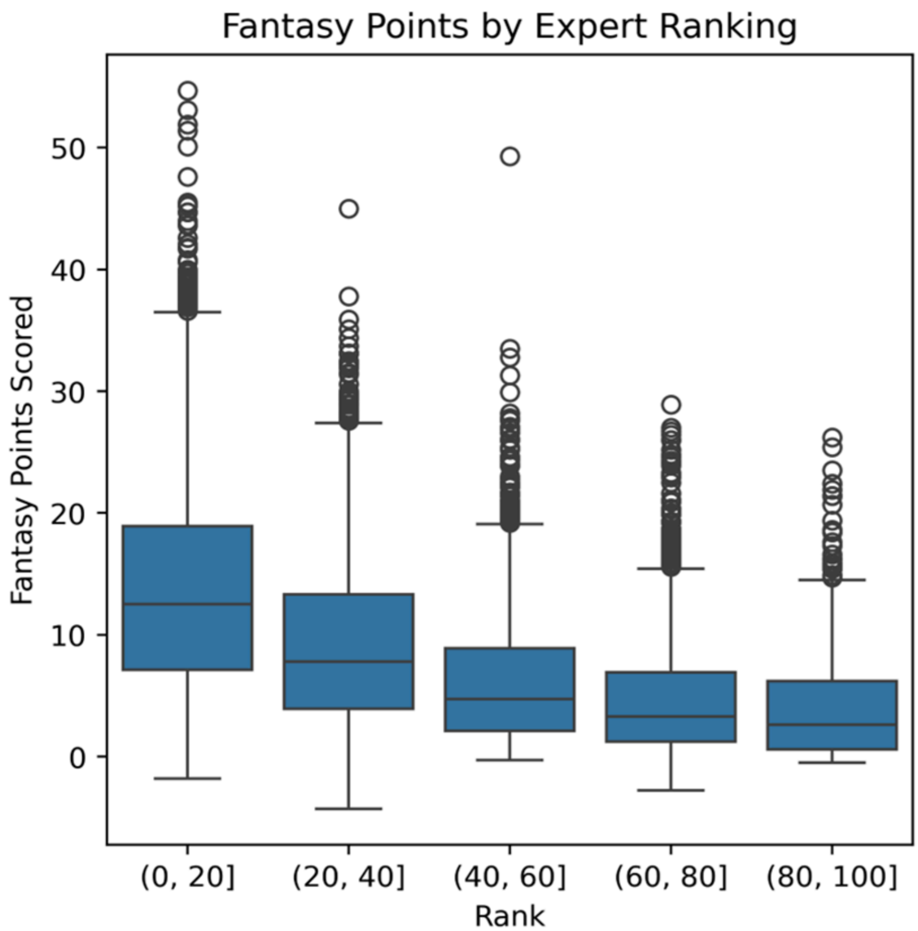

Figure 1: Relationship between consensus expert rankings and actual fantasy points scored

Another dataset source used in our prediction algorithm is betting market information. In countries where sports betting is legal, a person can place a wager on various team and player statistics. These include the total number of points a team will win an NFL game by, or if a player will score a touchdown [5]. We derive the projected NFL player statistics, called prop bets, and thus convert this information to a fantasy point projection strictly from betting market information. Figure 2 visualizes the association between the betting market projection and target variable.

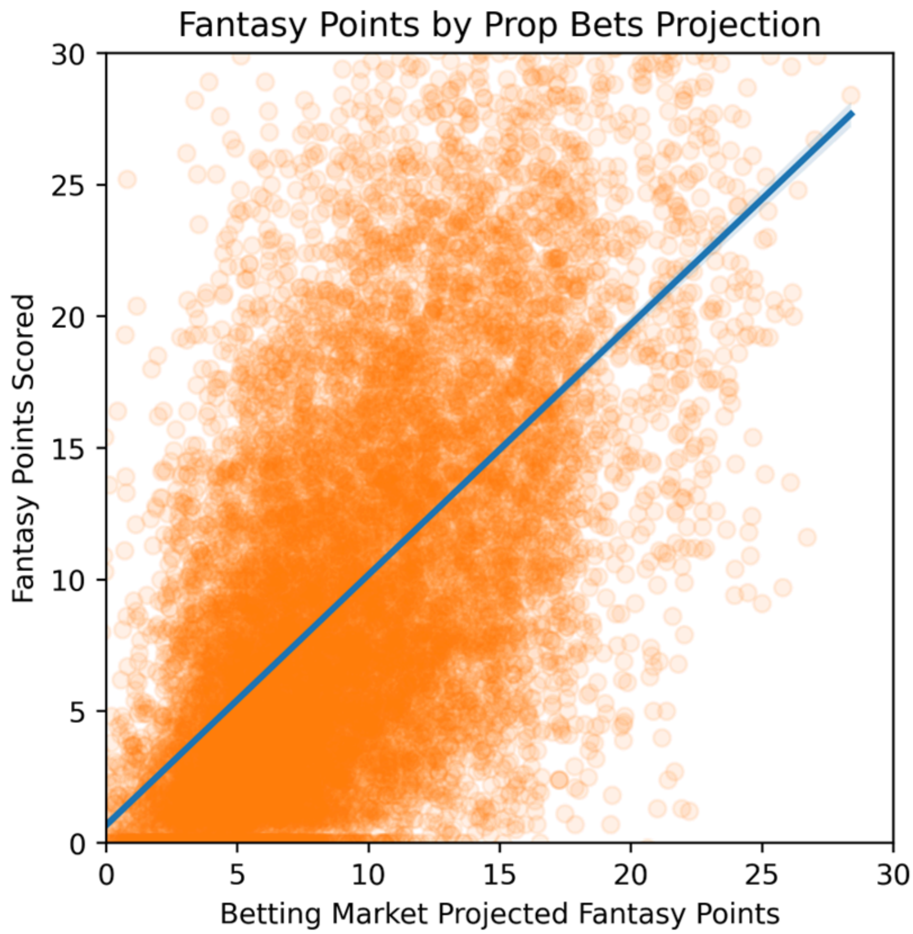

Figure 2: Relationship between betting market projection and actual fantasy points scored

Appendix Table 2 shows a data dictionary of extracted or derived features from the above three sources. We consider player performances in all regular season games from the 2020 NFL season until the 2023 NFL season, excluding the last week of the regular season from each year as it is typically omitted in fantasy football. In addition, to limit the problem space to relevant players, only the top quarterback, top two running backs, top three wide receivers, top tight end, and top kicker are considered. This leaves us with the task of retroactively projecting 2,005 quarterbacks, 4,014 running backs, 6,045 wide receivers, 2,015 tight ends, 2,016 defenses, and 1,946 kickers.

3. Methods

An important element to fantasy football is the positional constraints involved in choosing NFL players. As an example, a quarterback generally cannot be chosen in place of a running back, and vice versa. Because of this, along with the differences in relevant features across some position groupings, we believe that it makes sense to create separate models for each position. With this in mind, all models are benchmarked using identical season-week pairings in a grouped three-fold cross validation. Once the folds are created, any missing data is imputed iteratively using the Bayesian Ridge algorithm. A hyperparameter search is then performed to minimize mean squared error, reporting the best out-of-sample results across the three folds.

In terms of model architecture, we focused on two algorithms: gradient boosting with LightGBM and dense neural networks with TensorFlow. LightGBM (Light Gradient Boosting Machine) is a fast and efficient implementation of gradient boosting developed by Microsoft, designed to handle large datasets and complex models [6]. Dense neural networks consist of layers where each neuron is fully connected to all neurons in the subsequent layer, enabling the model to learn complex patterns in data. TensorFlow provides a flexible framework for constructing, training, and deploying these networks [7].

After careful analysis of statistical performance with considerations for model complexity and simplicity of results, we chose a dense neural network model for each of the positions. The performance of the dense neural network models was generally slightly better than the gradient boosting models, and we preferred to use the same algorithm for all models as it was more practical to analyze the same architecture for each position. In combination, the models had a root mean squared error of 6.03, and a mean absolute error of 4.63. In the context of the domain, a mean absolute error of 4.63 means that, on average, the difference between a player's projection and actual fantasy points scored was 4.63 fantasy points, which is approximately equivalent to 46 rushing or receiving yards. Performance for each individual model is available in Appendix Table 3.

4. Results

To help interpret the results of the projection models, we utilize Shapley values. Shapley values are a mechanism to help interpret the feature impact in advanced machine learning models [8]. A detailed analysis of feature relationships is provided for three key positions of fantasy football.

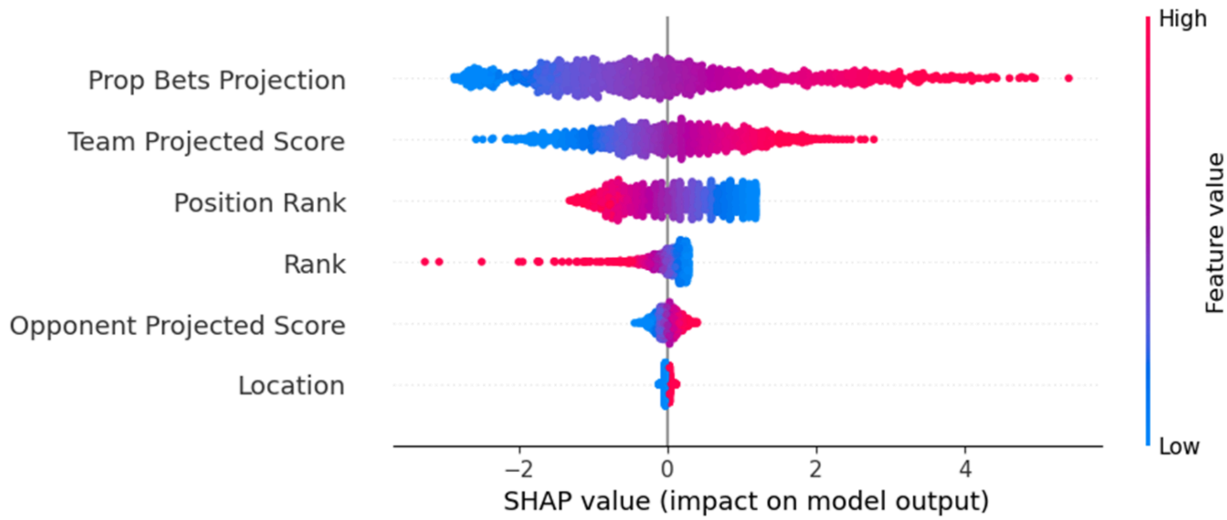

Figure 3: Shapley values of quarterback projections

Figure 3 shows the Shapley values of the features used in the quarterback projections algorithm. From the figure, we see that the prop bets projection derived from the betting market is the most important feature. The model's reliance on betting data suggests that current quarterback projections benefit most from market-driven expectations. In addition, the projected score of an NFL team is key in projecting fantasy points. Quarterbacks are generally considered to be the most impactful offensive position in American football, thus it makes sense that there is a relationship between the amount of points an NFL team is expected to score and the projected fantasy points of their quarterback. A relationship that may be considered counterintuitive is that the model projections generally increase if the opposing team is projected a high score. However, this makes sense in the context of football strategy. Teams that are losing tend to throw the ball more frequently, thus leading to more opportunities for the quarterback to score fantasy points.

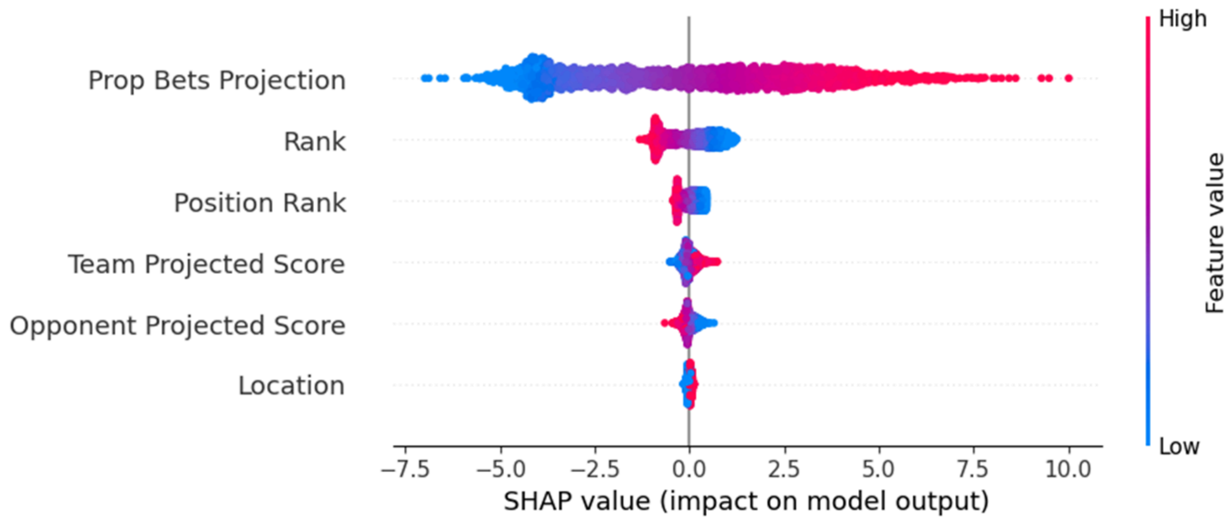

Figure 4: Shapley values of running back projections

Figure 4 shows the Shapley values of the features used in the running back projections algorithm. We see most of the same intuitive relationships as in the quarterback model. However, one deviation is with the projected score of the opponent. Conversely to quarterbacks, running backs are typically projected less as the projected score of the opponent increases. This once again aligns with American football strategy, as teams tend to run the ball more if they have a lead in the football game.

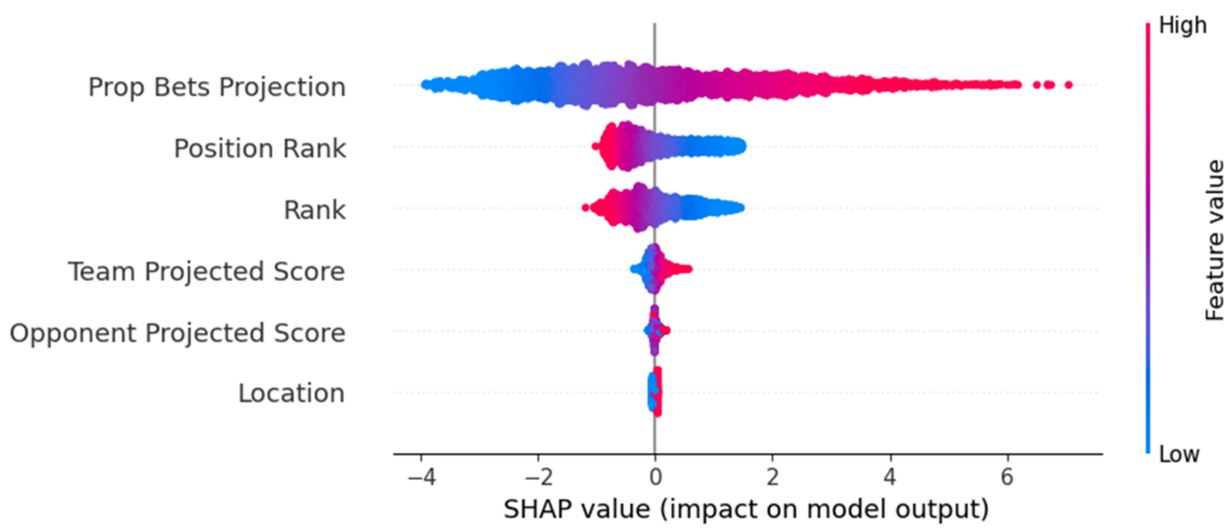

Figure 5: Shapley values of wide receiver projections

Figure 5 shows the Shapley values of the features used in the wide receiver projections algorithm. Once again, the prop bets projection derived from the betting market sticks out as the most impactful feature. In addition, the two types of consensus expert rankings, denoted as position rank and rank in the plot, both have a meaningful impact. It is also noteworthy that the location of whether a team is playing a home or away game seems to be a minimally impactful feature. This is likely due to the confounding nature of the features, where experts and bettors would already take the location of the game into account.

In addition to the above positions, Shapley values were generated for each model. For all Shapley plots, please refer to Figures 3 through 8 in the appendix.

5. Discussion

By integrating consensus expert rankings, NFL betting market data, and machine learning, we presented open-source fantasy football projections with transparent and validated model performance. We found that dense neural networks gave the best mix between predictive ability and practical use, resulting in a mean absolute error of 4.63 fantasy points across the different fantasy football positions. We found that statistics derived from the betting market are generally the driving force behind the model's predictions, followed by consensus rankings from fantasy football experts.

With the conclusion of our study, we believe that there is room for future improvements in the realm of fantasy football projections. Specifically, some scenarios of fantasy football involve selecting the player who is most likely to exceed a certain fantasy point threshold. In this setting, it may make more sense to model the expected distribution of fantasy point outcomes as opposed to a singular point estimate. In addition, as American football is a team sport, performances between players on the same team are somewhat correlated. In future versions, incorporating this interdependence by using teammate projections as an additional feature may improve model performance.

Contributions

Samuel Hughes: project architecture, data scraping, feature engineering, initial modeling, presentation format, report details
Jason Curtis: modeling improvements, SHAP analysis, presentation key takeaways, report format and structure

Repository

All relevant code is available on GitHub:
https://github.com/shughes1000/fantasy-football-projections

References

[1] https://www.nike.com/a/how-to-play-football
[2] https://www.statista.com/topics/10895/fantasy-sports-in-the-us/
[3] https://www.footballguys.com/playerhistoricalstats
[4] https://www.fantasypros.com/nfl/rankings
[5] https://www.bettingpros.com/nfl/odds
[6] https://lightgbm.readthedocs.io/en/stable/
[7] https://www.tensorflow.org/
[8] https://shap.readthedocs.io/

Appendix

Table 1: Vital Fantasy Football Statistics

Statistic	Description	Fantasy Points
Touchdown	A player possesses the ball in the opposing end zone for a touchdown.	6 points per touchdown
Passing Touchdown	A player throws the ball to a teammate, and the teammate scores a touchdown.	4 points per passing touchdown
Rushing Yards	A player carries the ball on a rushing play and runs toward the opposing end zone to gain yardage.	0.1 points per yard (1 point per 10 yards)
Receiving Yards	A player catches the ball and runs toward the opposing end zone to gain yardage.	0.1 points per yard (1 point per 10 yards)

Passing Yards	A player throws the ball forward to a teammate, and the teammate catches the ball and runs toward the opposing end zone to gain yardage.	0.04 points per yard (1 point per 25 yards)
Reception	A player catches a pass.	0.5 points per reception
Interception	A player throws the ball forward, but the opposing team catches the ball.	-2 points per interception
Lost Fumble	A player loses possession of the football, conceding possession to the opposing team.	-2 points per lost fumble
Extra Point Made	On the next play following a touchdown, a player successfully kicks the ball through the uprights.	1 point per extra point made
Extra Point Missed	On the next play following a touchdown, a player unsuccessfully kicks the ball through the uprights.	-1 point per extra point missed
Field Goal Made	A player successfully kicks the ball through the uprights on a normal play.	3 points per field goal made 0-39 yards, 4 points 40-49 yards, 5 points 50-59 yards, 6 points 60+ yards
Field Goal Missed	A player unsuccessfully kicks the ball through the uprights on a normal play.	-1 points per field goal missed

Table 2: Data Dictionary

Feature	**Description**	**Source**
Fantasy Points	The target variable of how many fantasy points a player actually scored in their NFL game.	FootballGuys
Rank	The ordinal expert ranking of all quarterbacks, running	FantasyPros

	backs, wide receivers, and tight ends in a given week, starting from 1 for the best ranked player.	
Position Rank	The ordinal expert ranking of a position grouping in a given week, starting from 1 for the best ranked player of the given position.	FantasyPros
Prop Bets Projection	An estimation of the player's expected fantasy points scored given a compilation of all relevant betting market projections.	BettingPros
Team Projected Score	The projected score of the player's NFL team, derived from the consensus of the betting market.	BettingPros
Opponent Projected Score	The projected score of the player's NFL team's opponent, derived from the consensus of the betting market.	BettingPros
Location	An ordinal variable, denoted as 1 or -1, indicating whether the player's team was designated as home or away, respectively.	BettingPros

Table 3: Model Performance

Position	**Root Mean Squared Error**	**Mean Absolute Error**
Quarterback	7.28	5.80
Running Back	6.32	4.78
Wide Receiver	6.14	4.74
Tight End	5.33	4.05
Defense	5.60	4.32
Kicker	4.56	3.66

Overall	**6.03**	**4.63**

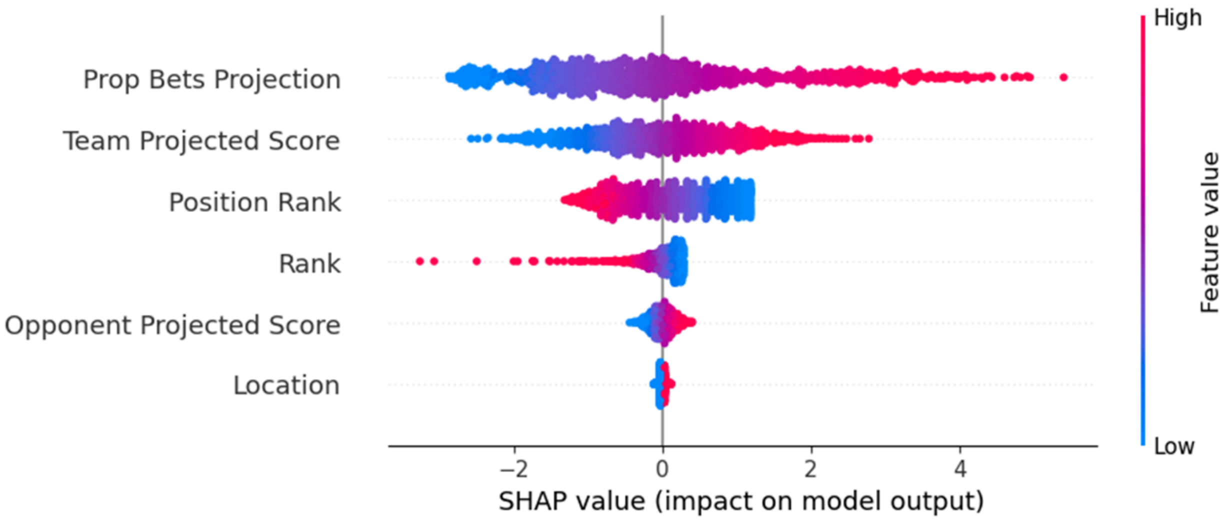

Figure 3: Shapley values of quarterback projections

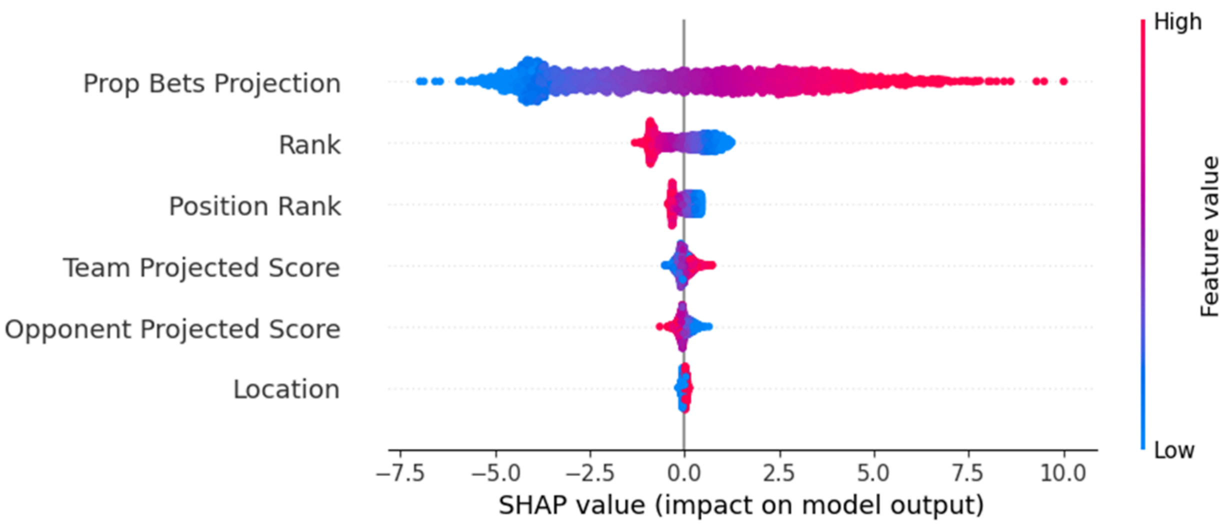

Figure 4: Shapley values of running back projections

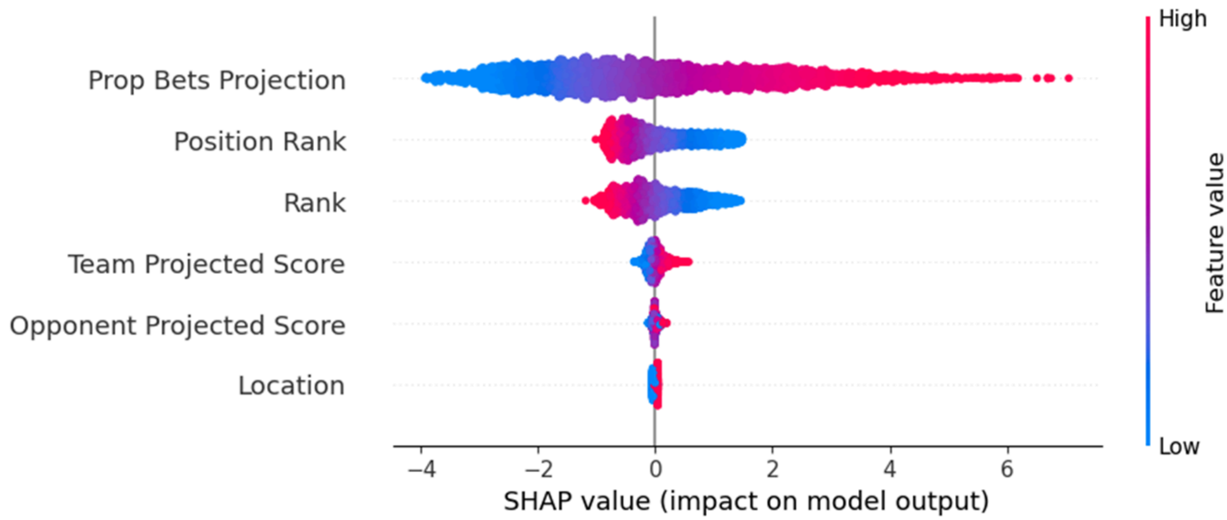

Figure 5: Shapley values of wide receiver projections

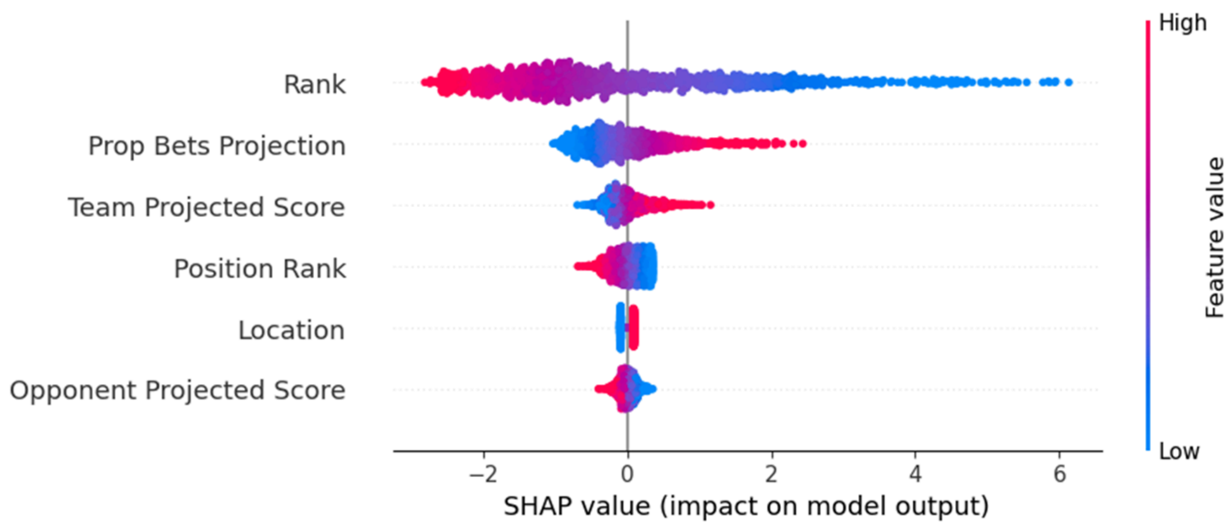

Figure 6: Shapley values of tight end projections

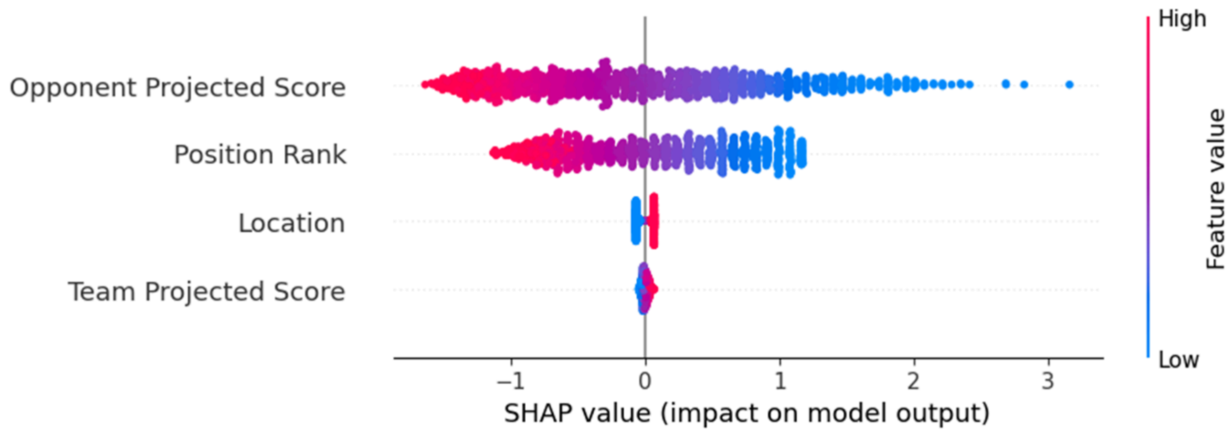

Figure 7: Shapley values of defense projections

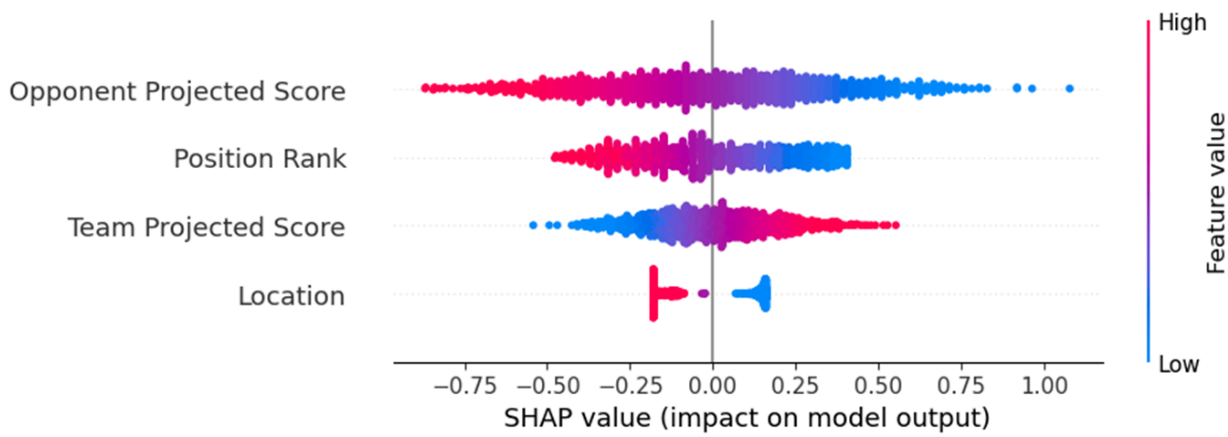

Figure 8: Shapley values of kicker projections