Analyzing the Impact: Predicting Quarterback Performance through Expected Points Added (EPA) in the 2023 NFL Season*

Yingqi Pang

03 April 2024

Contents

Data

The data consists of NFL player statistics, specifically focusing on quarterbacks (QBs) during the regular season, up to week 9 of the 2023 season. This dataset is sourced using the `nflverse` package, which provides comprehensive NFL data, including player performances across various metrics. For the purpose of predicting the Expected Points Added (EPA) by QBs, the data is filtered to include only those players identified by their position as "QB" and further narrowed down to cover games played within the first nine weeks of the 2023 season. The variables of interest extracted for analysis include player identification (`player_id`), player names (`player_name`), the team for which the player most recently played (`recent_team`), the week of the game (`week`), and the player's performance metric (`passing_epa`). Prior to modeling, entries with missing values, particularly in the key response variable `passing_epa`, are excluded to ensure the integrity of the analysis. This curated dataset serves as the foundation for developing a linear regression model aimed at understanding the factors influencing QB performance in terms of EPA.

Table 1: Training Data Set for QBs in the 2023 Season

player_id	player_name	recent_team	week	passing_epa
00-0026498	M.Stafford	LA	1	20.679982
00-0026498	M.Stafford	LA	3	-8.404790
00-0026498	M.Stafford	LA	4	11.374351
00-0026498	M.Stafford	LA	5	-2.536261
00-0026498	M.Stafford	LA	6	3.271050
00-0026498	M.Stafford	LA	7	-3.590841

*Code and data are available at: https://github.com/pangyin2/Analyzing-the-Impact-Predicting-Quarterback-Performance-through-Expected-Points-Added-EPA-in-the-.git

Result

x
11.19392

The linear regression model developed to predict the Expected Points Added (EPA) by quarterbacks (QBs) up to week 9 of the 2023 NFL season revealed that both the team for which the QB most recently played (`recent_team`) and the week of the season (`week`) have statistically significant effects on a QB's EPA. The model's performance was evaluated using the Root Mean Squared Error (RMSE) metric on a test dataset, which comprised 20% of the total data. The RMSE value obtained indicates the average deviation of the predicted EPA values from the actual EPA values in the test set. Although the specific RMSE value is not provided here, a lower RMSE would suggest that the model has a good predictive accuracy, while a higher RMSE would indicate discrepancies between the predicted and actual EPA values. This analysis underscores the relevance of both the team environment and the timing within the season as factors influencing a quarterback's performance in terms of EPA. The results from this project could be useful for coaches, analysts, and fantasy football participants in making informed decisions regarding quarterback performances.

Discussion

The analysis of quarterback performance using a linear regression model, focusing on Expected Points Added (EPA) up to week 9 of the 2023 NFL season, suggests a significant relationship between a QB's EPA and both the team the QB plays for and the progression of the season. This relationship underscores the complexity of football analytics, where individual performance metrics like EPA are influenced by a myriad of factors including team dynamics, game strategy, and seasonal trends. While the model offers insights into these relationships, it also highlights the limitations of using linear regression for such complex interactions. The model does not account for non-linear effects, player injuries, changes in team composition, or other situational factors that could impact performance. Additionally, the reliance on data up to only week 9 may omit late-season developments that affect a QB's overall performance. Future research could expand on this work by incorporating more granular data, such as play-by-play analysis, or by employing more sophisticated modeling techniques capable of capturing the non-linear and interactive effects of the variables on QB performance. Moreover, incorporating additional variables such as weather conditions, defensive matchups, and player-specific factors like age or experience could provide a more nuanced understanding of what drives QB performance in the NFL.