NFL Passing EPA Forecast for Weeks 10-18, 2023 Season

AUTHOR
Erping Gong

PUBLISHED
April 4, 2024

Introduction

In the fiercely competitive world of professional football, the margin between victory and defeat often hinges on the ability to predict and capitalize on in-game occurrences. One of the pivotal metrics for understanding and forecasting game dynamics is the Expected Points Added (EPA) on passing plays. As we reach the midpoint of the NFL's 2023 season, the quest to accurately project `passing_epa` for each team in the forthcoming weeks becomes not only a matter of curiosity but a strategic imperative.

This report is tailored for the management and technical staff who are gearing up for the crucial second half of the season, specifically focusing on Weeks 10 to 18. It aims to serve as a compass to navigate the complex waters of NFL analytics by offering robust predictions of `passing_epa`, using sophisticated statistical models underpinned by the data collected up to Week 9. Through the lens of predictive analytics, we will dissect the performance of quarterbacks across different teams, uncover patterns and trends, and engineer features that can lead to the most accurate forecasts possible.

In the pages that follow, we will navigate through the methodology employed to train and validate our forecasting model, dissect the performance of our predictions, and provide a thoughtful discussion on the implications of our findings. Our aim is not only to predict the future but to understand it, thereby enabling the teams to craft strategies that harness the full potential of their offensive arsenal.

Data

The data for this analysis was sourced from the `nflverse` database, which is a comprehensive collection of National Football League (NFL) statistics, updated regularly to ensure the most current information is utilized in our analysis. The `nflverse` project provides an accessible gateway to NFL data for analysts, providing a platform to analyze player statistics, team performance, and other critical metrics (`nflverse` package, Carpenter et al., 2021).

Using the `dplyr` package (Wickham et al., 2021), we filtered the dataset to include only regular season quarterback (QB) statistics (`season_type == "REG"`), for the 2023 season up to and including Week 9. Further data preparation, such as selecting relevant variables and handling missing values, was also performed using functions from the `dplyr` package.

Methodology

Feature Engineering

Initial feature selection was performed by identifying variables that are known to influence the expected points added (EPA) on passing plays. This included the player's ID, name, the team they recently played for, the week of the season, and the passing EPA itself. The `tidyr` package (Wickham & Henry, 2020) aided in reshaping the data as needed for analysis.

Model

In our pursuit to quantify the influence of team dynamics and temporal progression throughout the season on quarterbacks' performance, we constructed a linear regression model with `passing_epa` as the response variable. The model is specified as `passing_epa ~ recent_team + week`, where `passing_epa` denotes the Expected Points Added by a quarterback's passing plays. The `recent_team` predictor encapsulates the categorical effect of a quarterback's current team on his performance, reflecting systematic variations in `passing_epa` across different teams. The `week` variable serves as a numerical predictor, allowing us to explore the trajectory of quarterbacks' contributions to their teams' scoring potential over time. The mathematical representation of our model:

[= *0* + {} *{} I(=)* + {} +]

where (*0) denotes the intercept, (*{}) are the coefficients for each team (with one team necessarily serving as the reference category), (* {}) is the coefficient for the week of the season, and ( ) represents the error term. This model assumes linearity in the parameters and additive effects of each predictor, allowing us to extrapolate the `passing_epa` for any given combination of team and week within the scope of our data.

Model Validation

```
predictions <- predict(model, newdata = test_set)
validation_rmse <- sqrt(mean((predictions - test_set$passing_epa)^2))

print(validation_rmse)
```

```
[1] 11.19392
```

Upon the completion of model training, we proceeded to validate its predictive capability on the unseen data present in our `test_set`. To quantify the model's accuracy, we deployed the Root Mean Square Error (RMSE) metric, which encapsulates the average magnitude of the errors between our model's predictions and the actual `passing_epa` values. The RMSE is particularly instructive as it punishes larger errors more severely by squaring the residuals before averaging, thus providing a stringent assessment of prediction quality.

After computing the predictions using our linear model, the resulting RMSE was calculated to be 11.19392. This value serves as an average estimate of the error magnitude across our predictions. In the context of our dataset, where `passing_epa` can exhibit substantial variability, an RMSE of 11.19392 suggests that while the model captures the general trend, there is a noteworthy deviation between the predicted and actual values that should be accounted for in strategic considerations. It is essential to benchmark this RMSE against domain-specific standards or naive/baseline models to contextualize its significance fully. For instance, comparing it against the RMSE of a model that merely predicts the average `passing_epa` of the training set can offer insights into the relative sophistication and added value of our predictive model.

Result

```
Warning: Item 2 has 62 rows but longest item has 335; recycled with remainder.
```

Team	Average Predicted passing_epa	Number of Predictions
ARI	-5.91	10
ATL	1.69	10
BAL	-1.32	12
BUF	2.01	12
CAR	-1.90	9
CHI	-2.53	11
CIN	3.86	9
CLE	2.09	10
DAL	0.38	12
DEN	1.96	8
DET	-0.90	9
GB	-2.31	9
HOU	-0.12	8
IND	-3.59	12
JAX	-3.68	10
KC	-4.49	10
LA	-2.57	10
LAC	0.10	8
LV	-3.34	11
MIA	-2.75	11
MIN	0.68	11
NE	-3.76	12
NO	-1.10	21
NYG	-1.20	14
NYJ	-3.18	9
PHI	1.90	9
PIT	1.81	10
SEA	-2.02	9
SF	1.69	12
TB	1.13	8

Team	Average Predicted passing_epa	Number of Predictions
TEN	0.36	10
WAS	-0.24	9

The aggregated predictions highlight distinct variations in quarterbacks' anticipated contributions to their teams' scoring capabilities. Remarkably, the Cincinnati Bengals' quarterbacks stand out with an average predicted `passing_epa` of +3.86, indicating a robust potential to enhance their team's scoring through effective passing. In stark contrast, the Arizona Cardinals' quarterbacks are at the other spectrum with an average predicted `passing_epa` of -5.91, suggesting challenges in achieving similar scoring contributions.

The spectrum of predictions across teams—from as high as +3.86 for the Cincinnati Bengals to as low as approximately -5.91 for the Arizona Cardinals—paints a comprehensive picture of the variable quarterback performance league-wide. These predictions are not merely numerical outputs but a reflection of the interplay between team strategies, quarterback capabilities, and the evolving dynamics of the NFL season.

Moreover, the "Number of Predictions" column sheds light on the data's breadth, with teams like the New Orleans Saints and New York Giants having a higher count, likely due to a combination of multiple quarterbacks being analyzed or a richer dataset. Such granularity in our predictions enables a more detailed exploration of team-specific quarterback performance as the season progresses.

Discussion

This analysis serves as a cornerstone for strategic decision-making, offering insights that teams can leverage to optimize their game plans and maximize scoring opportunities. However, it's crucial to recognize the inherent limitations of predictive modeling. The dynamic nature of NFL games, influenced by myriad factors such as player injuries, team morale, and unforeseen tactical shifts, means actual future performances may diverge from our predictions. Hence, while our model provides a valuable forecasting tool, it should be considered alongside other strategic elements to fully harness the potential of quarterbacks in driving team success in the 2023 NFL season.