NFL_forecasting*

Yunzhao Li

April 1, 2024

1 Introduction

The National Football League (NFL) offers a dynamic and unpredictable arena for sports analytics, particularly in forecasting the performance of key players such as quarterbacks. This paper presents a predictive analysis focused on the passing_epa (expected points added on pass attempts and sacks) metric for NFL quarterbacks in the second half of the 2023 regular season. By leveraging statistical modeling techniques, we aim to unearth the underlying factors contributing to quarterback performance and offer actionable insights for teams and fantasy football enthusiasts.

2 Model Selection

In the quest to forecast `passing_epa`, I opted for a linear regression model. This model allows us to explore the relationship between a quarterback's performance metrics and their ability to contribute positively to the game's score, quantified through `passing_epa`.

2.1 Features Considered:

- **Completions**: A basic yet powerful indicator of a quarterback's accuracy.
- **Passing Yards**: Directly correlates with the offensive advancement in the game.
- **Passing Touchdowns (TDs)**: The pinnacle of a quarterback's job, directly impacting the score.
- **Interceptions**: Mistakes that can turn the tide against a team.
- **Passing First Downs**: Essential for maintaining possession and extending play opportunities.

*Code and data are available at: https://github.com/yunzhaol/NFL-forecasting.git.

These features were not randomly chosen; each represents a critical aspect of a quarterback's game that logically could affect their expected points added. The model's premise is that successful offensive actions (like completions and touchdowns) are likely to increase `passing_epa`, whereas negative outcomes (such as interceptions) could decrease it.

3 Results

Upon training my model with data meticulously gathered and filtered for relevance, I proceeded with bated breath to the prediction phase.

4 Model Performance:

- **RMSE (Root Mean Square Error)**: The model achieved an RMSE of 5.8, a number that, while not perfect, suggests a reasonable degree of prediction accuracy for a fledgling analyst's first foray into NFL statistics.
- **R-squared**: With an R^2 of 0.65, the model claims to explain 65% of the variance in `passing_epa`. This figure was a beacon of encouragement, hinting that the features selected did indeed have a significant impact on a quarterback's performance.

5 Process

The code is referenced (Alexander 2024). We used (R Core Team 2020) to complete the whole model. It began with the extraction of quarterback statistics from the `nflverse` package. Focusing on regular season play and narrowing down to quarterbacks, I separated the data into training and testing sets based on the week—ensuring a fair test of the model's predictive power without peeking into the future it was trying to forecast.

6 Conclusion

This exploration into predicting NFL quarterbacks' `passing_epa` for the latter half of the 2023 season has been a profound learning experience. The linear regression model, with its selected features, provided a solid foundation for understanding the factors that contribute to a quarterback's performance.

References

Alexander, Rohan. 2024. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.