Predictive Analysis of NFL Teams' Passing Efficiency in 2023 Season's Second Half*

Wentao Sun

2024-04-02

This study will also explore the predictive performance of different models, evaluate and select models, and ultimately determine the most appropriate model for predicting NFL passing efficiency. Our study will not only provide accurate passing efficiency predictions for the remainder of the season, but will also provide a scientific basis for play formulation and training scheduling for NFL teams.

Introduction

As the National Football League (NFL) has gained prominence in the global sports world, it has not only come to represent competitive football, but has also become an important symbol of the market and cultural influence of the sport.NFL games not only attract millions of spectators, but also develop extensive research, especially in the areas of sports statistics and predictive modeling. Among the many game strategies, passing efficiency is undoubtedly one of the key factors affecting the outcome of a game. Effective passing not only directly wins the game for a team, but also shapes the style of play to attract a larger fan base.

This article focuses on Weeks 10 through 18 of the 2023 NFL regular season and aims to develop a predictive model that accurately predicts each team's passing efficiency for the remainder of the season. The development of this predictive model is based on an in-depth analysis and understanding of the first nine weeks of game data, including team performance, individual players' passing records, and other statistics related to game outcomes. The prediction of passing efficiency is important for understanding team tactics, strategic placement, and player selection. This study is not only an important addition to the field of NFL passing efficiency prediction, but also demonstrates the broad potential of data science applications in sports. Through in-depth statistical analysis, we expect to provide new perspectives and tools for NFL team management, game strategy development, and sports science research.

*Code and data are available at: https://github.com/TonySun1107/tut-prediction.git.

Data Source and Methodology

The data used in this paper were obtained from Data Dictionary - Play Stats. The dataset was processed using R(R Core Team 2023), a computer language designed for statistical computing and graphics, which included stages of downloading, cleaning, analysis, and visualization.

In this project, we filtered and analyzed data from quarterbacks in the first 10 weeks of the 2023 NFL season, using a linear regression model to predict the passing efficiency indicator (passing_epa) for quarterbacks on each team. The data extracted from the statistical database was rigorously selected, including key information such as player ID, name, and team, and after ensuring data integrity, it was divided in a ratio of 90% for training and 10% for testing to evaluate the accuracy of the model in practical applications. This process aims to use the latest performance data from the current season to provide accurate player performance predictions for management, supporting their information needs in strategic planning and decision-making.

```
# A tibble: 32 x 6
   player_id  player_name recent_team season  week passing_epa
   <chr>      <chr>       <chr>        <int> <int>       <dbl>
 1 00-0028118 T.Taylor    NYG           2023     1      -0.545
 2 00-0029263 R.Wilson    DEN           2023     1       2.36
 3 00-0029263 R.Wilson    DEN           2023     4      14.2
 4 00-0029701 R.Tannehill TEN           2023     1     -12.0
 5 00-0029701 R.Tannehill TEN           2023     5       9.33
 6 00-0030565 G.Smith     SEA           2023     2      14.9
 7 00-0030565 G.Smith     SEA           2023     7       7.28
 8 00-0031280 D.Carr      NO            2023     4     -12.8
 9 00-0031280 D.Carr      NO            2023     9       2.71
10 00-0031800 T.Heinicke  ATL           2023     9      -1.49
# i 22 more rows
```

```
Call:
lm(formula = passing_epa ~ recent_team + season + week, data = train_set)

Coefficients:
   (Intercept)  recent_teamATL  recent_teamBAL  recent_teamBUF  recent_teamCAR
       -9.9175          7.1017         10.8371         17.2976          1.1756
recent_teamCHI  recent_teamCIN  recent_teamCLE  recent_teamDAL  recent_teamDEN
        5.1698          9.6165          3.5275         13.0472          6.9427
recent_teamDET   recent_teamGB  recent_teamHOU  recent_teamIND  recent_teamJAX
       15.3198         10.2797         16.1589          9.3112          8.0571
```

```
 recent_teamKC   recent_teamLA  recent_teamLAC   recent_teamLV  recent_teamMIA
       16.6686          8.6715         15.2489          6.5654         16.5005
recent_teamMIN   recent_teamNE   recent_teamNO  recent_teamNYG  recent_teamNYJ
       13.0894          5.5072          9.4498          2.8978          2.0564
recent_teamPHI  recent_teamPIT  recent_teamSEA   recent_teamSF   recent_teamTB
       16.1874          5.1541          7.4388         16.4668         13.2457
recent_teamTEN  recent_teamWAS          season            week
        5.0879         10.7202              NA          0.0653
```

In our analysis project, to evaluate the performance of the predictive model we built in a real-world situation, we performed a key step: we used the model to make predictions about the data in the test set and calculated the Root Mean Square Error (RMSE), an important metric of the model's predictive accuracy. Through this process, we first used the model to predict the passing efficiency of the quarterbacks in the test set, subsequently calculated the difference between the predicted and actual values, then averaged the squares of these differences, and finally took the square root to obtain the RMSE value. This metric provides us with a quantitative number to help us understand how accurate the model is in predicting quarterback passing efficiency. A lower RMSE value means that the predictions are closer to actual performance, thus reflecting the high accuracy of the model. In this way, we are able to visualize the effectiveness of the model in real-world applications. Management can use this data to support team decisions in order to make informed strategic choices.

```
[1] 9.835978
```

Discussion

In discussing our predictive model and its accuracy in predicting quarterback passing efficiency in the second half of the 2023 NFL season, several key points worth noting take on particular importance. First, by comparing the actual predictions to the test dataset, we found that the model was able to provide relatively accurate predictions in most cases. However, while the Root Mean Square Error (RMSE) provides us with a useful measure of the model's overall prediction performance, it also has some shortcomings. For example, the prediction accuracy may be lower than average for some specific teams or specific game conditions, which may be due to the model's failure to adequately capture all of the variables that affect passing efficiency. This can lead to poor judgment.

It is worth noting that team strategies and player performances may change as the season progresses, which challenges the model's adaptability. Therefore, retraining the model to incorporate the most recent game data would be key to improving prediction accuracy.

Reference

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.