Predicting Quarterback Success: Passing EPA*

Timothius Prajogi

March 31, 2024

Table of contents

1 Introduction

2 Data

The data was retrieved using the `nflverse` (Carl et al. 2023) package, containing the various stats of an NFL player.

This data was downloaded, cleaned, parsed, analyzed, and visualized using R (R Core Team 2023), a statistical programming language, with package support from `tidyverse` (Wickham et al. 2019), a collection of libraries which included the following packages that were utilized:

- `ggplot2` (Wickham 2016)
- `dplyr` (Wickham et al. 2023)
- `readr` (Wickham, Hester, and Bryan 2023)
- `tibble` (Müller and Wickham 2023)

*Code and data are available at: https://github.com/prajogt/predict_passing_epa .

For additional assistance with data storage, the `arrow` (Richardson et al. 2024) was used and for report generation, the `knitr` (Xie 2023) package was used.

The models were created using `tidymodels` (Kuhn and Wickham 2020) and summarized using `modelsummary` (Arel-Bundock 2022).

Table 1: 2023 NFL data (Weeks 1-9)

Passing EPA	Completions	Attempts	Passing Yards	Passing Touchdowns	Interceptions	Sacks	PACR
-2.031960	0	1	0	0	0	1	0.0000000
20.679982	24	38	334	0	0	0	0.8166259
-5.089193	34	55	307	1	2	1	0.8457300
-8.404790	18	33	269	1	2	6	0.8485804
11.374351	27	40	319	1	1	2	1.0740741
-2.536261	21	37	222	2	0	4	0.6065574

As explained by `nflverse`'s data dictionary (Carl et al. 2023), passing EPA is the total expected points added on pass attempts and sacks, completions were the amount of passes that were completed, attempts were the amount of pass attempts made, passing yards are the amount of yards gained on pass plays, passing touchdowns are the amount of passing touchdowns, interceptions were the amount of interceptions thrown, the sacks are the number of times the player was sacked, and PACR is the Passing (yards) Air (yards) Conversion Ratio which is the number of passing yards per air yards thrown per game

As the purpose of this report is to create a model to predict passing EPA, theses features were selected as they are believed to be strongly correlated with passing EPA, meaning using them would create a better and more accurate model.

The original dataset was filtered to only include statistics from the 2023 season, separated into the first half of the season (weeks 1-9) and the second half of the season. Since passing EPA concerns the quarterback's stats, we only considered the rows which pertained to quarterbacks.

3 Model

The model that was created is:

$$
\begin{aligned}
y_i|\mu_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 \times \text{completions}_i + \beta_2 \times \text{attempts}_i + \beta_3 \times \text{passing-yds}_i \\
&+\beta_4 \times \text{passing-tds}_i + \beta_5 \times \text{interceptions}_i + \beta_6 \times \text{sacks}_i + \beta_7 \times \text{pacr}_i
\end{aligned}
$$

Where y is the predicted passing EPA of a player.

This model aims to use the amount of completions, attempts, passing yards, passing touchdowns, interceptions, sacks and PACR of a player to predict their passing EPA. Since passing EPA is defined as the expected points added, the more completions, passing yards, and passing touchdowns that were made it is reasonable to expect that the amount of points added would also increase. Then the amount of interceptions and sacks that were made would most likely mean that more of the passes that the players made ended up not contributing to the amount of points that they gained. Attempts is included as the attempts the quarterback makes to pass correlates with the offensive effectiveness of the team. Finally the PACR provides insights as to the effectiveness of a pass, which would correlate to the overall passing EPA of a player.

To ensure that we are not over-fitting, the dataset of the was split into 80% for training and 20% for testing. For 254 observations for training and 64 observations for testing.

4 Results

Table 2: Predicted Passing EPA of Testing Data

Predicted	Actual	Error	Squared Error
-2.6260098	-2.5362610	0.0897488	0.0080548
-0.4431122	-4.9787140	-4.5356018	20.5716834
-15.2135779	-13.5330319	1.6805459	2.8242347
-1.0187119	-0.5446569	0.4740551	0.2247282
-1.4864176	-0.6515430	0.8348746	0.6970156
0.0163133	2.4409939	2.4246807	5.8790763

The model overall does well to predict players Passing EPA, with a mean squared error of 14.26, from the predictions shown in Table 2.

From Table 3 we can see that completions and passing touchdowns has a strong positive correlation between passing EPA and attempts, interceptions, and sacks have a strong negative correlation between passing EPA. Surprisingly passing yards only has a small correlation with passing EPA, which can most likely be explained by the fact that gaining yards doesn't necessary result in touchdowns as strongly as passing touchdowns does.

Therefore, for the next half of the season to determine the success of the team it would be beneficial to look into their quarterback's passing completions, passing attempts, passing touchdowns, passing interceptions, and sacks to get a sense of the expected points per pass. Noting these stats each week would be helpful in seeing how well a team would do against a future opponent based on this predicted passing EPA. Going up against an opponent whose

Table 3: Explaining the Passing EPA given their completions, attempts, passing yards, passing touchdowns, interceptions, sacks and PACR

	Passing EPA
(Intercept)	−0.792
	(0.817)
completions	0.284
	(0.113)
attempts	−0.694
	(0.074)
passing_yards	0.099
	(0.006)
passing_tds	1.770
	(0.319)
interceptions	−3.943
	(0.339)
sacks	−1.901
	(0.148)
pacr	−0.021
	(0.442)
Num.Obs.	254
R2	0.845
R2 Adj.	0.841
AIC	1450.9
BIC	1482.8
RMSE	4.06

quarterback has been consistently delivering completions and passing touchdowns with a low amount of interceptions and sacks, would be troublesome for a quarterback with a lower predicted passing EPA.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Carl, Sebastian, Ben Baldwin, Lee Sharpe, Tan Ho, and John Edwards. 2023. *Nflverse: Easily Install and Load the 'Nflverse'*. https://nflverse.nflverse.com/.

Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. https://www.tidymodels.org.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.