Predicting Quarterback Success in the NFL

Jacob Ieyoub
Springboard
Houston, TX
jieyoub@gmail.com

***Abstract* - The success of an NFL quarterback is hard to predict and it is evident when looking at the [poor historical performance] of teams selecting quarterbacks in the NFL Draft . There are many factors that play into the outcome of the career of an individual quarterback. If it was as easy as going after the biggest player with the strongest arm, then Jamarcus Russell would be in his thirteenth NFL season rather than fizzling out of the league after thirty-one games in three years. It is one of the most important positions in all of professional sports. To be able to predict if a college quarterback will be successful in the NFL would lead to better decision making when building rosters, provide struggling franchises with hope and ultimately lead to a better product on the field across the league. The goal of this project is to use college quarterback statistics to create a model for better predicting whether or not a quarterback will be successful in the NFL.**

1. Introduction - The inspiration for this project is The Ringer's Ryen Russillo asking "Why Is It So Hard to Draft NFL Quarterbacks?". A similar study from FiveThirtyEight's Josh Hermsmeyer gave some direction for this project, however only basic college passing statistics are used as features in this study. The NFL is king and to be able to find a franchise quarterback is the first step in the path to winning championships. Many owners and general managers in the league would find a model that has success in predicting which quarterback prospects will produce in the NFL quite useful in their decision making on draft day. It could give them confidence in who they are selecting and also could lead to situations where they are able to fill other needs earlier in the draft and draft an undervalued quarterback later. The data used for this project were collected from pro-football-reference.com and sports-reference.com/cfb. It consists of NFL and college passing data from 1997 to 2019. The approach was to determine our criterion of success in the NFL and then build a model that uses that specified criterion as our target variable and the college passing data as the features for modeling.

2. Data Wrangling - To begin, the NFL data was concatenated into a single pandas DataFrame. Once concatenated, the DataFrame was subsetted to only contain players with more than 100 passing attempts. formatted by season. Since these player codes did not correspond to the player codes used in the college data, the "Player" column was split to only contain the player names. This was done by splitting the players codes from the player names.

In the NFL dataset there is a "QBrec" column that has the starting record by season for each player. The entries in this column contained record data that was incorrectly put into a date format. This included "mm/dd/yyyy" and "mm-dd-yyyy." The wins and losses were determined from these values and then a "win_pct" column was created. The next step for this data was to use .groupby() aggregated by average to determine the career averages for each player. For columns such as ‘Int’, ‘TD’, ‘Yds’, ‘G’ and ‘Sk’, the career totals were used to replace the career seasonal average that was output by .groupby(). Once this was done, several columns that are not needed for the scope of this project were dropped. The resulting dataset was 132 rows x 20 columns.

For the college passing data, the data was concatenated into a single pandas DataFrame. As with the NFL players, the college players had to be split from their player code. The necessary career totals were determined and this replaced the career averages calculated from using .groupby() aggregated by average. The resulting DataFrame is 635 rows x 14 columns.

After creating these two DataFrames, they were exported into a .csv file. In a separate Jupyter Notebook, an inner merge was performed on the “Player” column to obtain a DataFrame that consists of all the quarterbacks that were drafted between 1998 and 2019. This DataFrame will be used for our model and is 151 rows x 35 columns. It consists of NFL career and college career statistics.

3. Exploratory Data Analysis - The data was first explored and analyzed to determine the target variable from the NFL passing dataset. This was done by using bivariate analysis to evaluate the correlations of several NFL passing metrics vs NFL win percentage. Exploratory Data Analysis was performed on the original dataset which only contained the year 2009 - 2019. A correlation matrix was made and then eight NFL passing metrics that had a Pearson coefficient > 0.5 were then related to win percentage by making a scatter plot for each. It was clear that they all had a positive correlation. These 8 metrics were Passer Rating, AY/A (Adjusted Yards per Attempt), ANY/A (Adjusted Net Yards per Attempt), TD% (Touchdowns / Attempts), QBR (Total Quarterback Rating), NY/A (Net Yards per Attempt), Y/A (Yards per Attempt) and Completion Percentage.
The next step in this analysis was to determine what the target variable for determining NFL success is. Each of the 8 metrics were put into a correlation matrix with the available college quarterback data between 2009-2019. After doing this, YPA (Yards per Attempt) was chosen as the target as it had the most consistent Pearson coefficients across all the college passing metrics as seen in the table below.

Y/A_NFL Correlations	r-value
AY/A_College	0.11
Att_College	0.14
Cmp_College	0.18
G_College	0.17
Int_College	0.14
Pct	0.2
Rate_College	0.17
TD_College	0.26
Y/A_College	0.11
Yds_College	0.17
Y/G_College	0.036

After choosing YPA, the NFL Average YPA between 2009-2019 was calculated to be 7.1. This will be the threshold of success for a quarterback in the NFL. It should be noted here that YPA will end up not being used as the target variable as there was not much success with modeling for it. More on that in section 4. To simplify the analysis, a categorical column was made in the dataset that had a value of either 1 for above the

NFL YPA average or 0 for below that average. Of the 63 quarterbacks that had more than 100 passing attempts in the NFL and drafted between 2009-2019, 40 had below a 7.1 average YPA and 23 had above a 7.1 average YPA. Next, the distributions were examined by looking at distribution plots and box plots. They were made by grouping the quarterbacks that were above and below the average and seeing the distributions of each metric with each subgroup. This provided a great visualization that showed most passing metrics having a higher mean value for the quarterbacks with a greater than NFL Average YPA of 7.1. This difference shows that these metrics translate to NFL YPA. Bootstrap Inference was then used to examine the features further by looking at the subgroups in each. The p-values of the difference of the means for the subgroups was calculated and the following was found.

Diff of Subgroup Mean	p-values
AY/A_College	0.0799
Att_College	0.0679
Cmp_College	0.0415
G_College	0.0055
Int_College	0.0578
Pct	0.0377
Rate_College	0.0328
TD_College	0.0049
Y/A_College	0.0555
Yds_College	0.0196
Y/G_College	0.6254

When looking at the p-values, it can be seen that most of our features are right around the conventional threshold of 0.05. We will keep all of these except for Y/G_College, which is not very close to the threshold.

Next, a Seaborn pairplot was made to examine the relationships amongst our features. It was evident that several had strong correlations with each other. This will be explored further when selecting the features of our machine learning model. It should also be noted that both the continuous target variable (Y/A_NFL) and categorical target variable (above_nfl_ypa) were retained for later to see which approach works best.

This leaves us with 10 features moving forward that include:

1. AYA_College
2. Att_College
3. Cmp_College
4. G_College
5. Int_College
6. Pct
7. Rate_College
8. TD_College
9. Y/A_College
10. Yds_College

The next step will be to explore different machine learning models to see which provides the best prediction for quarterback success in the NFL.

4. Machine Learning - For the modeling phase of the project, several machine learning techniques were applied to the project dataset. After analyzing the dataset, a hypothesis was developed that the success in the NFL of a college quarterback could be predicted accurately enough for use when drafting a quarterback by only using basic college passing statistics. Two different target variables were analyzed in the modeling to see if there was more success modeling to predict one vs. the other. The two target variables were YPA(Yards Per Attempt) and Passer Rating. YPA was the original target variable as discussed in section 3. Passer Rating was chosen as the target variable to try after YPA because it had the strongest correlation to win percentage for quarterbacks amongst all the NFL passing metrics that were considered. The threshold of success for both of these target variables was a quarterback having a YPA or Passer Rating that was above the 20 year NFL average.

The modeling dataset consisted of quarterbacks drafted between 1998-2019. It was originally 2009 to 2019 but this only provided 63 data points. More data points would surely provide better modeling so the additional years of data were collected. The features included Adjusted Yards Per Attempt, Attempts, Completions, Games, Interceptions, Completion Percentage, Passer Rating, Touchdowns, Yards Per Attempt and Passing Yards. The target variable was a binary classification where 1 denotes above the NFL average and 0 below the NFL average.

<u>Yards Per Attempt</u> - As discussed in the Statistical Analysis, YPA was selected as the target variable because it showed the highest consistent correlation across the features along with a relatively significant correlation to NFL QB Win Percentage. The scikit-learn machine learning library was used to begin modeling the dataset. Starting off, a baseline classifier was created using the DummyClassifier class. This is a classifier that uses simple rules to gauge how accurately a model can be predicted by guessing. For YPA and all 10 features, the DummyClassifier yielded an accuracy of 60.4 % using the “most-frequent” strategy. Therefore, the goal was to have a model that would give a higher accuracy than 60.4%. A Logistic Regression model was the first model applied to the dataset. With all ten features, there was decent success with this model. Initial results using 5-fold cross-validation saw a mean score of 69.7%. This, however, needed to be further studied as from our statistical analysis we know there is multicollinearity between several of our features. By calculating the Variance Inflation Factor of our features, the dataset was reduced to four features, Adjusted Yards Per Attempt, Interceptions, Games and Touchdowns. This was with a threshold of 50.0 for the VIF. Logistic Regression was used again and the accuracy was 67.4%, which is a solid score. Next, the score of the test data was determined to be 50%. This is obviously very poor. Therefore, a DecisionTree classifier was created for our dataset to see if there is more success modeling the dataset. This provided a 5-fold cross-validation mean score of 53%. The model score on the test data was then determined to be 59%. This is much better than before but still poorer than the baseline. Because a very good model with YPA as the target variable could not be built, Passer Rating was used as the target variable to see if there would be more success.

<u>Passer Rating</u> - Passer Rating became the new target variable because it had the highest correlation amongst the basic NFL passing statistics to NFL QB Win Percentage. After seeing that the consistent correlation in the features to YPA did not lead to a good model, these correlations were disregarded. As with the YPA modeling, a DummyClassifier was created and this had an accuracy of 66.6%. Logistic Regression was then used with all ten features, ignoring multicollinearity at first, and the model predicted success with a 62% accuracy. The features were reduced like before and with the four features a new Logistic Regression model was created. The 5-fold cross-validation mean score of the

Logistic Regression model with these four features was 68%. This is a decent result and since this was better than the results from the modeling with YPA as the target variable, it appeared that passer rating is a better target variable. This was explored further by building a DecisionTree model. With no parameter tuning, the DecisionTree model produced poor results with a 5-Fold cross-validation mean score of 48.8%. By setting the hyperparameters max_depth = 2 and criterion = 'entropy', the 5-fold score improved to 61%. The test data then scored 77.3 % in accuracy. This result was the best of any of the modeling so far. To verify this high accuracy, the model was run 100 times. The mean of the scores for each player was then calculated and if the mean was above 0.5 then the predicted result was rounded to 1, or above the NFL Average Passer Rating. If the mean was below 0.5 then the predicted results were rounded to 0, or below the NFL Average Passer Rating. The resulting DataFrame can be seen in Table 1 along with feature importance in Table 2.

5. Conclusion - The results provide some interesting takeaways. First of all, there were zero false positives. This means if it did predict a player to be above the NFL average passer rating, then that player was above every time. There were, however, four false negatives. This included Gardner Minshew, Daniel Jones, Jacoby Brissett and most notably Patrick Mahomes. Obviously, the model missing on Patrick Mahomes is unforgivable but this follows the old adage that all models are wrong and some are useful. It should also be reiterated that this was only using basic college passing statistics as the features. Many factors play into the success of a quarterback's success but the results using these basic features are promising. Now, being above the NFL Average Passer Rating does not necessarily mean a quarterback is successful. It does however mean that a quarterback has the ability to be productive. It is up to the front office and coaches to optimize that productivity with the right offensive system, productive offensive weapons, a good offensive line and good defense.

Player	p	a	Pass Rating	T / F	Player	p	a	Pass Rating	T / F
Patrick Mahomes	0	1	109.6	False	**Jacoby Brissett**	0	1	84.8	False
Deshaun Watson	1	1	101.4	True	**Cody Kessler**	1	1	84.8	True
Lamar Jackson	1	1	98.9	True	**Mason Rudolph**	1	1	82.0	True
Dak Prescott	1	1	97.0	True	**Sam Darnold**	0	0	80.9	True
Gardner Minshew	0	1	91.2	False	**Kyle Allen**	0	0	80.0	True
Nick Mullens	1	1	90.8	True	**Jeff Driskel**	0	0	78.8	True
Drew Lock	1	1	89.7	True	**Josh Allen**	0	0	76.6	True
Jared Goff	1	1	87.9	True	**Dwayne Haskins**	0	0	76.1	True
Daniel Jones	0	1	87.7	False	**C. J. Beathar d**	0	0	75.5	True
Kyler Murray	1	1	87.4	True	**David Blough**	0	0	64.0	True
Baker Mayfield	1	1	86.2	True	**Josh Rosen**	0	0	59.4	True

Table 1. Test Results. p = predicted, a = actual.

Completions	0.41
Completion Percentage	0.35
Yards Per Attempt	0.24

Table 2. Feature Importance

Finding a good quarterback is just the tip of the iceberg. There are a few things to consider when looking at the feature importances. It can be safely said that the number of completions of a college quarterback does tell us whether or not they will be good in the NFL. To me it shows that experience is important as I am sure most quarterbacks that play more games will have more completions. The importance of completion percentage and yards per attempt is somewhat expected as an NFL quarterback needs to be accurate and efficient.

6. Further Considerations - As mentioned above, several factors affect a quarterback's success. Some of the considerations that would be beneficial for improving the model would include metrics that are not freely available. There are two ways to look at this project. The perspective of a fan or from the perspective of a general manager. This could also be looked at as pre-draft or post-draft. Pre-draft considerations and potential features would include more advanced passing data such as Tight-Window Throws, Average Depth of Target, Completion Percentage on Deep Throws and Deep Throw Rate. Other factors could include the strength of the defenses these quarterbacks faced and also incorporating combine results to see if they are significant.

7. References

1. Russillo, R., 2020. Why Is It So Hard To Draft NFL Quarterbacks?. [online] The Ringer. Available at: <https://www.theringer.com/video/2020/1/6/21051537/why-is-it-so-hard-to-draft-nfl-quarterbacks> [Accessed 6 January 2020].
2. Hermsmeyer, J., 2019. *The NFL Is Drafting Quarterbacks All Wrong*. [online] FiveThirtyEight. Available at: <https://fivethirtyeight.com/features/the-nfl-is-drafting-quarterbacks-all-wrong/> [Accessed 1 February 2020].
3. Barnwell, B., 2018. *Barnwell: Why Drafting A QB Is Often A Disaster ... And What That Means For 2018*. [online] ESPN.com. Available at: <https://www.espn.com/nfl/story/_/id/23039883/history-tells-us-nfl-terrible-evaluating-quarterbacks-means-2018-draft-prospects> [Accessed 1 February 2020].
4. Isenberg, S., 2018. *Assigning Success Likelihoods To Recently Drafted Quarterbacks*. [online] saisenberg.com. Available at: <https://saisenberg.com/projects/quarterbacks.html> [Accessed 1 March 2020].
5. <Pro-football-reference.com> [Accessed 6 January 2020]
6. <sports-reference.com/cfb> [Accessed 6 January 2020]