NFL quarterback regular season performance analysis and forecast for the rest of the season*

Dongli Sun

April 4, 2024

#Introduction

As the NFL at the midpoint of the 2023 season, I needed to develop a model trained through the nflverse package for each team for the remainder of the season (Weeks 10 through 18) to predict passing efficiency (passing_epa). Pass-expected points Added (passing_epa) is a key metric that quantifies the impact of quarterback passes on a team's expected points.

I utilize the nflverse package to obtain and analyze comprehensive statistics of NFL quarterbacks during the regular season. These statistics for me can be used to predict future results. In this report, I will present my approach to data analysis and modeling. The focus of my purpose for future data analysis is to develop predictive models based on various quarterback performance metrics to provide insights into future performance.

#Data

The data used in this analysis includes statistics from the NFL's 2023 regular season quarterbacks. The dataset includes information such as number of passes, passing yards, completions, opponent teams, and weeks. It is into a training set containing data up to week 9 and a test set containing data from week 10. This partitioning allows us to train our models on historical data and evaluate their performance on unseen data.

1 Model

Two models were considered to predict passing_epa at weeks 10-18:

Model 1: A linear regression model with the number of passes and passing yards as predictors. The model assumes a linear relationship between the predictor and the target variable.

---

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

Model 2: Linear regression model with completion, pass code and sack code as predictors. The model includes additional predictors to capture more nuanced aspects of quarterback performance.

Both models were trained using the training set and evaluated using the test set. The performance of each model is evaluated based on its ability to accurately predict the passing_epa of the test set. A model summary is provided to compare the performance of the two models and determine the most effective way to predict passing_epa.

1.1 Model set-up

2 References