

Report2

1 Introduction

This dataset contains both individual and team-related details about NFL players. This information serves as the foundation for computing projected fantasy points in the PPR (Points Per Reception) format.

In our initial analysis, we concentrated on determining the essential data required for forecasting projected fantasy points in the PPR format. A significant challenge surfaced when attempting to predict a player's performance for an upcoming season, such as the 2024 season. Since the season hadn't commenced yet, we lacked data regarding a player's performance during that specific season. Consequently, we needed to rely on personal information such as BMI, age, experience, and other pertinent factors.

To delve further, we initially investigated whether any noticeable correlations existed between this personal data and their corresponding PPR fantasy points. However, we discovered limited or no significant correlations. Consequently, we decided to utilize data from the preceding season to formulate predictions. For instance, when forecasting a player's performance in the 2024 season, we incorporated information from the 2023 season along with the player's individual characteristics as features. The goal was to predict the player's point total for the 2024 season.

We applied a similar methodology to our training data. For example, we regarded the points accrued during the 2016 season as sample features, with the final score for the 2017 season serving as the label. This approach yielded a dataset comprising information spanning 5 to 6 seasons for each player.

Ultimately, we concluded that three primary actions (Rushing, Receiving, and Passing) in combination with a player's performance in each game and their general information served as valuable data for constructing the model. Subsequently, we delved into the formulas required to calculate Rushing, Receiving, and Passing points. With the dataset prepared, we proceeded to build the model. I will explain the details of how this was accomplished in the following section

2 Methodology

Step 1: Data Preparation

I loaded the dataset into my preferred data analysis tool, such as Python, using libraries like Pandas and NumPy.

I explored the dataset to understand the distribution of features, checked for missing values, and identified any outliers or anomalies. I performed data cleaning and preprocessing as necessary.

Step 2: Feature Engineering

Based on the provided columns and additional context, I created new features that could potentially be valuable for predicting fantasy points. I considered the following feature engineering tasks:

Age and exp and bmi of some players were blank, but i obtained this information according to the information of previous years

I calculated fantasy points for passing, rushing, and receiving touchdowns, passing yards, rushing yards, and receiving yards using the provided conversion rates.

I computed per-game statistics for relevant features.

I created player-specific metrics based on target share, air yards share, and other advanced metrics.

I calculated expected points added (EPA) per game for passing, rushing, and receiving plays.

I chose these features at last: { 'years_exp', 'age', 'bmi', 'passing_fantasy_points', 'rushing_fantasy_points', 'receiving_fantasy_points', 'completions_per_game', 'attempts_per_game', 'passing_yards_per_game', 'passing_tds_per_game', 'interceptions_per_game', 'passing_air_yards_per_game', 'passing_yards_after_catch_per_game', 'rushing_yards', 'rushing_tds_per_game', 'targets_per_game', 'receptions_per_game', 'receiving_yards_per_game', 'receiving_tds_per_game', 'receiving_air_yards_per_game', 'receiving_yards_after_catch_per_game' }

Step 3: Model Building

I split the dataset into a training set and a testing set for model evaluation.

I chose an appropriate machine learning regression model for predicting fantasy points. I considered options like Random Forest, Gradient Boosting, or even a neural network as good choices.

At last I choose RandomForestRegressor

I trained the model using the training dataset and evaluated its performance using appropriate regression evaluation metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).

At last I choose Mean Absolute Error (MAE)

Step 4: Model Deployment

Once my model performed well on the testing dataset, I was ready to deploy it to make predictions for the upcoming NFL season.

Step 5: Prediction and Submission

I utilized the trained model to make predictions for each NFL player for the upcoming season. I created a CSV file with two columns: Player name and Projected fantasy points PPR, based on my model's predictions.

Step 6: Documentation and Reporting

I wrote a detailed report summarizing my data analysis, feature engineering, and model building pipeline. The report included visualizations, insights, and any challenges I faced during the project. I explained the rationale behind my feature engineering choices.

Step 7: Model Evaluation

I evaluated the model's performance using my testing dataset to ensure that my model's predictions aligned well with the actual player performance when the NFL season concluded.

I kept in mind that iterative model refinement and fine-tuning might be necessary to improve prediction accuracy. Additionally, I understood the importance of keeping the dataset updated with the latest player statistics if I intended to provide ongoing projections for fantasy football enthusiasts.

3 conclusion

We get this at last

Mean Absolute Error: 4.317499831365936

In this case, a MAE of approximately 4.32 indicates that, on average, your model's predictions for fantasy points in PPR (Points Per Reception) for NFL players are off by roughly 4.32 points when compared to the actual values. In many cases, achieving a MAE of approximately 4.32 would be considered a good result, especially if it aligns with the level of precision required for your application. It's important to consider the domain and practical implications of the error when interpreting MAE. If this level of error meets your requirements, your model appears to be performing well in predicting fantasy points in PPR for NFL players.