Fantasy Football

Predicting Player Performance

By: Bryson Sicotte

Why?

- Available player score projections are unreliable.

- Many participants rely on an unscientific approach to player selection.

- A robust, high-performing tool would be in high demand on both an individual, and league level.

Note - The fantasy football scores used in this model match ESPN's standard scoring format.

The Training Data – Observations

- Each observation is a player's statistics from a specific game.

Total observations by position:

Wide Receivers	Running Backs	Tight Ends	Quarterbacks	Kickers

84,038	80,325	49,318	31,031	22,050

The Training Data – Features

Features include:

- Individual statistics for each player
- The opposing team's statistics (taken as an average of all weeks up to that point in the season).

Target Variable:

- Total Fantasy Points

Feature Engineering

- The features that were ultimately used were comprised of game statistics (in the manner mentioned under the the next bullet) and opponent statistics.
- For each row (a week in the dataset for a specific player), a mean of the previous three weeks was used. It may later be determined that a mean of a different number of previous weeks has more predictive power. Three weeks is simply a starting point.
- Certain features were determined by subject-matter experts not to have a strong bearing on the target (e.g. high school, age), and were thus removed from the final model.

Choosing a Model

Model Used: Random Forest Regressor

1st Iteration:

Hyperparameters:

- Max depth: 3, 4, 5, 6, 7
- N estimators: 10, 100

2nd Iteration:

- Max depth: 8, 9, 10, 11, 12
- N estimators: 10, 50, 100

Model Performance

Difference between actual fantasy score and predicted:

1st Iteration (Hyperparameters: max_depth: 3, 4, n_estimators: 10)

- Mean: 3.74
- Median: 2.47

2nd Iteration (Hyperparameters: max_depth: 8, 9, 10, 11, 12, n_estimators: 10, 50, 100)

- Mean: 3.64
- Median: 2.41

MAE scores from 5-fold cross-validation: -3.59, -3.59, -3.73, -3.69, -3.61

Future Optimizations

To improve predictions:

- Create a separate model for each position
- Try new/more hyperparameters (as computing power allows)
- Select different features