STAT 6341: Sports Analytics

# Predicting NFL Player Outcomes through the Lens of Fantasy Football

Delaney Helgeson

# Project Goals

- Prospective Modeling
  - Predict the **Fantasy Points** that a player will score in a **given future game** based on his **past performance** and **game characteristics** known ahead of time
- Performance Indicators
  - Understand the factors that contribute to a player's performance, measured in Fantasy Points

# Consumers

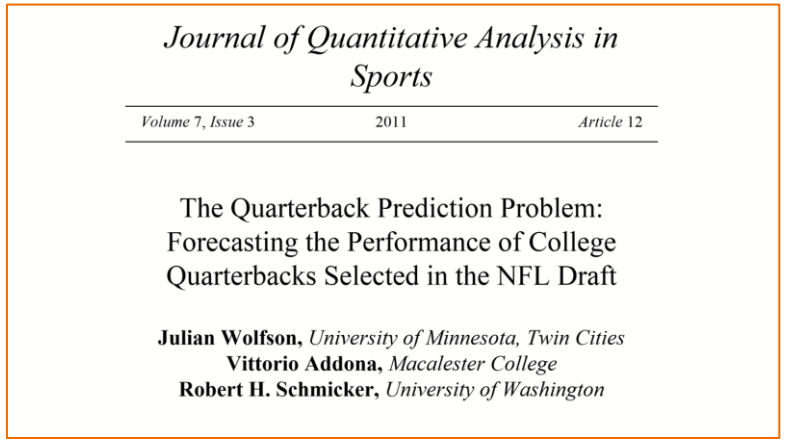
- People that participate in Fantasy Football and/or bet on professional football.
- Modeling Component
  - Sports bettors can **enter information** about their player of interest into the model and **receive a predicted fantasy score** in return. The score can help the person decide whether to bet on the player **each week**.
  - Used on a **weekly basis**
- Data Visualization Component
  - Sports bettors can explore the interactive visualizations to better understand **patterns in player performance**.
  - Used at the **beginning of the season** or on a **weekly basis**

# Academic Literature

**Estimating player value in American football using plus–minus models**

*R. Paul Sabin*

From the journal Journal of Quantitative Analysis in Sports
https://doi.org/10.1515/jqas-2020-0033



*Journal of Quantitative Analysis in Sports*

*Volume 7, Issue 3* 2011 *Article 12*

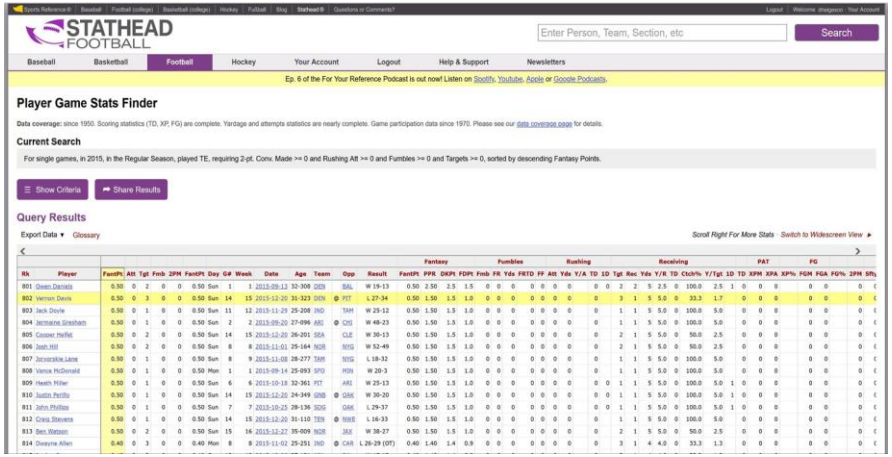
The Quarterback Prediction Problem: Forecasting the Performance of College Quarterbacks Selected in the NFL Draft

**Julian Wolfson,** *University of Minnesota, Twin Cities*
**Vittorio Addona,** *Macalester College*
**Robert H. Schmicker,** *University of Washington*

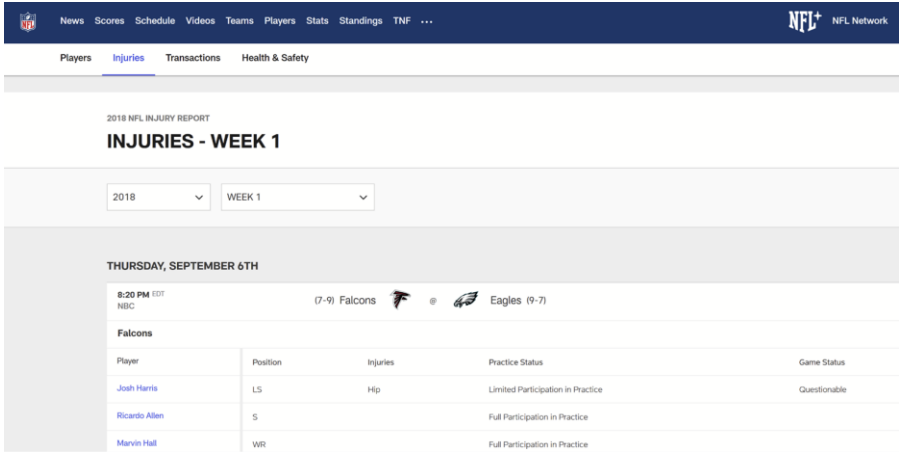
# Commercial Literature


# Data Source

StatHead.com

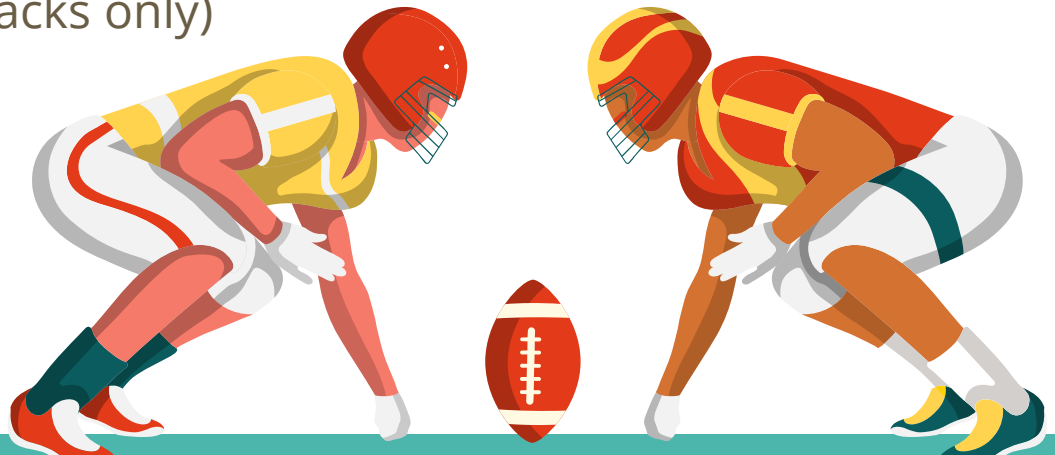



NFL.com



# Data Pre-Processing

- Standardizing Team Names
- Substituting 0’s for missing rates
- Grouping injury types
- Cleaning date formatting
- Exclude games with less than 5 pass attempts (Quarterbacks only)



# Feature Engineering

- Extract the month from date
- Extract the points scored by each team and game result
- Shift data so that past performance is in line with future game characteristics and fantasy points scored
- Calculate moving and cumulative averages of each player's past performance stats
    - Treat k as a tuning parameter
- Get averages of performance against opponent
- Join injury data to player-game data

# Recipe

- Remove near zero variance columns
- Remove correlated predictors
- Remove columns that were linear combinations of one another
- Center & scale continuous variables
- Fit natural cubic splines on numeric predictors with 3 degrees of freedom (regression only)
- Create dummy columns for categorical variables (regression only)
- One-hot encode categorical variables (ML only)

# Modeling Frameworks


Training-Validation-Test Split → Tune K on Linear Regression using training-validation split → Select optimal K and retrain models using training and validation data

Models & Tuning Parameters:

- Base Linear Regression
- Penalized Linear Regression
  - Penalty & mixture
- Random Forest for Regression
  - Minimum number of observations in terminal node
  - Number of predictors considered at each split
- XGBoost for Regression
  - Minimum number of observations in terminal node
  - Number of predictors considered at each split
  - Tree Depth
  - Learning Rate

# Modeling Frameworks

- Separate models for
  - Quarterbacks
  - Running Backs, Wide Receivers and Tight Ends
- 5-Fold cross validation repeated 5 times

# Model Performance: Quarterbacks

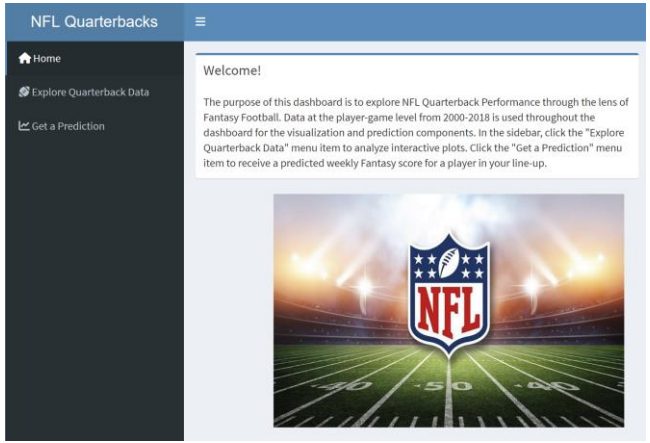
| Model | Resampled RMSE | Test RMSE |
|---|---|---|
| Base Linear Regression | 7.448 | 8.096 |
| Penalized Linear Regression | 7.458 | 8.130 |
| Random Forest | 7.451 | 8.196 |
| XGBoost | 7.446 | 8.151 |

# Model Performance: Running Backs, Wide Receivers, and Tight Ends

| Model | Resampled RMSE | Test RMSE |
|---|---|---|
| Base Linear Regression | 5.3348 | 5.5318 |
| Penalized Linear Regression | 5.3321 | 5.5316 |
| Random Forest | 5.3692 | 5.5763 |
| XGBoost | 5.3327 | 5.5447 |

# Shiny App

- Explore Tab
  - Consumers can use the interactive visualizations to explore the relationship between Fantasy Points and other factors, such as type of injury, performance history, and team.
  - Useful for high level overviews
- Predictions Tab
  - Consumers can enter information about a player of interest and receive a projected fantasy score in return.
  - Used on a weekly basis to make decisions about player line-up



# Limitations

- Training-Validation-Test splits were not proportional for Quarterbacks and Running Backs/Wide Receivers/Tight Ends
- Injury data was sparse prior to 2008

# Future Work

- Tune for a wider range of K
- Tune number of knots on natural cubic splines
- Gather more data for RB, WR, & TE

# Takeaways

- Simple > Complex
- Strong indicators of performance:
  - Player’s performance history
  - Track record of opponent
  - Home or Away

# Literature Citations


ESPN. (2022, August 31). *Mike Clay's Projections + Mid-Round Madness* [Video]. YouTube. https://www.youtube.com/watch?app=desktop&v=0aluQ44GsNM

Sabin, R. (2021). Estimating player value in American football using plus–minus models. *Journal of Quantitative Analysis in Sports*, *17*(4), 313-364. https://doi.org/10.1515/jqas-2020-0033

Sidell, T. (2022, October 24). *IBM Watson and ESPN use AI to Transform Fantasy Football Data*. Journey to AI Blog. Retrieved December 1, 2022, from https://www.ibm.com/blogs/journey-to-ai/2022/10/ibm-watson-and-espn-use-ai- to-transform-fantasy-football-data-into-insight/

Wolfson, J., Addona, V. & Schmicker, R. (2011). The Quarterback Prediction Problem: Forecasting the Performance of College Quarterbacks Selected in the NFL Draft. *Journal of Quantitative Analysis in Sports*, *7*(3). https://doi.org/10.2202/1559-0410.1302