

Which Fantasy Football Site has the Most Accurate Projections?

Amar Srinivas

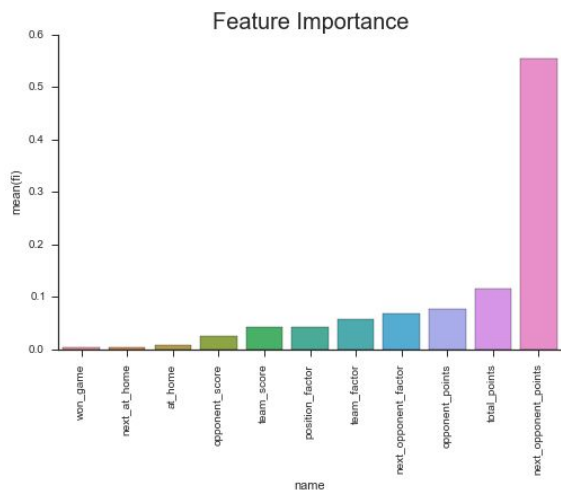


Project Overview:

Fantasy Football is a huge point of entry into American football in the United States. It can help casual fans better understand and become more involved in a sport in which they may have previously lacked investment. Many people playing Fantasy Football set their starting lineups based on the projections given to them by various fantasy football sites. However, on a week to week basis, there are projections that continue to fall short, resulting in a negative impact on people's chances of Fantasy Football success. This project is an attempt to determine whether or not ESPN, as well as other various Fantasy Football sites, provide accurate projections.

Process:

I began my project by scraping for relevant fantasy data from a variety of fantasy football prognosticators such as ESPN, numberFire, and CBS and examined the relative and absolute errors to quantify the accuracy of each site. I then moved onto feature engineering to see which features were more important relative to others.



Model:

For my model, I used a random forest regressor and then cross-validated with other models (Gradient Boost, for example) and used hyperparameter selections which worked best as opposed to other models. I used the RMSE score in this case to compare the outputs of my model versus the model predicted by ESPN. I then deployed my model on a WebApp using Flask to help recommend which player to start.

Tools Used:

- Flask, Psycopg2, Pandas, NumPy, Scikit-learn, Jupyter Notebook, AWS, EC2, HTML/CSS
- PCA, Seaborn, EDA, multicollinearity and linear regression, Random Forest, feature engineering

Challenges:

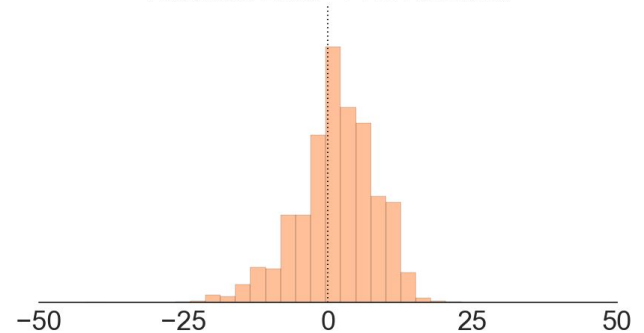
- Formatting the data that was scraped in a clean and methodical fashion that made sense relative to the various Fantasy Football sites.
- Finding which features to use based on signal and relevance was also particularly difficult.

Next Steps:

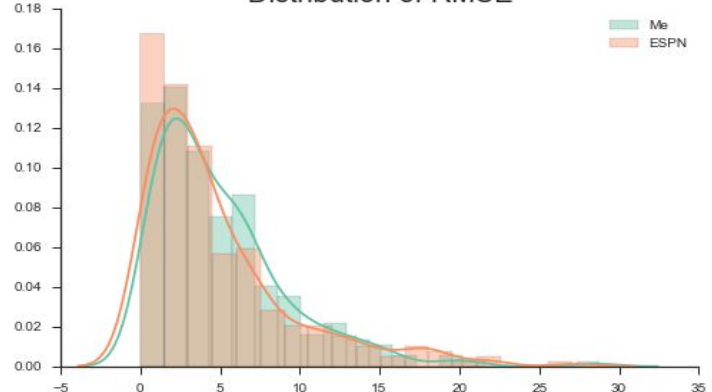
Since the data I am analyzing is only for the 2016 season, my model only predicts point totals from week to week for that season only. For my next step, I intend to scrape for more data from previous seasons so that I can predict point totals from season to season as opposed to week to week and provide more comprehensive predictions.

Project Link: <https://github.com/amarsrinivas/Capstone-Project>

Absolute Error - FFB Relevant



Distribution of RMSE



Fremont, CA
(510) 579-2519
asrinivas@utexas.edu

Amar Srinivas
US Citizen, Open to relocation

linkedin.com/in/amarsrinivas
github.com/amarsrinivas

SUMMARY

- Data Scientist with experience as a software engineer. Leveraging data science and machine learning to build data driven applications that translate data into actionable insights and drive company values.
- Applied statistics, data mining, and machine learning to discover useful insights in structured and unstructured data, and communicating those insights to both technical and non-technical audiences
- Fluent in Python, AWS; experienced with Java, SQL and Matlab; Map Reduce, Graph Theory, Spark ML, Recommendation Systems, EDA, probability and statistics, hypothesis testing, supervised and unsupervised ML algorithms, NLP, dimensionality reduction

SKILLS

- Webscraping/Bootstrapping
- Classification, regression, clustering
- Time series, dimensionality reduction
- Natural language processing
- Python
- SQL, PostgreSQL, MongoDB
- Scikit-learn, NumPy, Pandas
- JMP by SAS, Git, GitHub, Flask
- Statistics (Frequentist + Bayesian)
- Hypothesis testing, A/B Testing
- Visualization with matplotlib, seaborn
- Parallel Processing

EDUCATION

Galvanize Data Science Immersive Program, San Francisco 2018
BSA Mathematics, University of Texas at Austin

EXPERIENCE

Galvanize, San Francisco, CA 2018
Data Science

- Built a linear regression model to test and predict the sales value of tractors given various guidelines and parameters. Ran cross-validation checks to pull the most pertinent models needed.
- Built a statistical model to predict customer churn for a car rental company with above 0.96 recall score. Methodologies include logistic regression and decision trees.
- Currently working on a Fantasy Football Player recommendation system as my Capstone Project. Methodologies include Random Forest Regression and Gradient Boost.

Avi Networks, Santa Clara, CA 2015-2016
Intern - Engineering Team

- Re-configured Avi's BIG IP text files into the preferred JSON format using parsing tools such as Antlr, Bison, and YACC
Technologies used: Python, YAML, YACC, Antlr
- Created documentation for Avi's Protocol Buffers and created a standard interface for REST APIs using Swagger
Technologies used: Python, Swagger
- Managed a booth for Avi at the Open Stack Developer's Conference in Austin, TX. Spoke to over 150 leads about the benefits of Avi's product and other aspects such as load balancing and cloud apps.

The Qt Company, Santa Clara, CA 06/15-08/15
Engineering Intern

- Self learned the QML language and mastered the Qt IDE, and learned how to create cross-platform GUI applications using Qt.
- Researched and resolved bugs reported by various Qt licensees as well as Qt product questions from prospects and leads.
- Managed the booth at the Embedded Systems Conference explaining the capabilities of Qt Creator and how comprehensive Qt can really be to potential customers.

NeuroSky, San Jose, CA 06/15-08/15
Data Science Team

- Read the raw data from Amazon Web Servers (AWS) and created powerful visual charts using the d3.js JavaScript Library and HTML5
- Created a comprehensive User Interface for NeuroSky's users. Over the course of the summer, created every chart and graph to be implemented in the UI and successfully completed the project with the Data Science Team