# Optimizing Preseason Fantasy Football Rankings


**John O'Hollaren (jpo4@duke.edu)**
Department of Electrical and Computer Engineering
Duke University


## Motivation

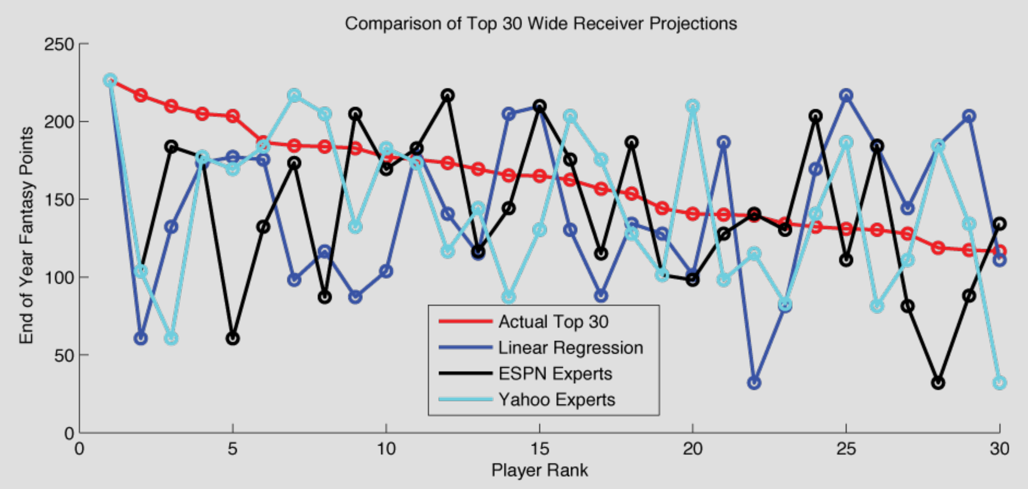


Figure 1: Total end-of-year fantasy points scored by ESPN / Yahoo's preseason top 30, compared to the actual top 30 and the featured reduced linear regression algorithm. This shows that ranking is very difficult.

**Why is predicting fantasy football stats important?**

- Profit: $15 billion per year is spent on fantasy football.
- Popularity: According to the Fantasy Sports Trade Association, 10% of the US population plays fantasy football.
- Difficulty: Ranking players is difficult, even the experts are rarely correct (see Figure 1). A small edge can therefore go a long ways.

**How does fantasy football work?**

Points are awarded when a player scores a touchdown or gains yards. Each owner picks their team at the beginning of the year. Correctly predicting which players will play well greatly improves your team's year-long performance. This paper focuses on an optimal strategy for predicting which NFL wide receivers will have the best fantasy football season using data available prior to the beginning of the season.

**Goal**

A top 30 list of NFL WRs ordered by predicted fantasy point output in the 2012 season, using only data available prior to the 2012 season.

## Data

The following key statistics were gathered for every single active wide receiver in the NFL from 2007 until 2012.

700+ Players Per Year ➔

13 Features

| |
|---|
| **Receiving Yards** |
| **Receiving Touchdowns** |
| **Receiving Targets** |
| **Receiving Catches** |
| **ESPN Preseason Ranking** |
| **Yahoo Preseason Ranking** |
| **Δ 1 year Receiving Yards**: (Yards in year Y-1) – (Yards in year Y-2) |
| **Δ 1 year Receiving TDs**: (TDs in year Y-1) – (TDs in year Y-2) |
| **Δ 2 year Receiving Yards**: (Yards in year Y-1) – (Yards in year Y-3) |
| **Δ 2 year Receiving TDs**: (TDs in year Y-1) – (TDs in year Y-3) |
| **Δ 3 year Receiving Yards**: (Yards in year Y-1) – (Yards in year Y-4) |
| **Δ 3 year Receiving TDs**: (TDs in year Y-1) – (TDs in year Y-4) |
| **Fantasy Points in Previous Year**: 6 • TDs + 0.1 • Yards |

**Performance Criteria**

Discounted cumulative gain (DCG) will be used to reward ranking good players highly and penalize ranking good players low. A higher DCG is better.

$$DCG = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2(i)} \qquad rel_i = \frac{1}{\text{actual end of year rank}}$$

Below we can visually view our data, as well as the inconsistency of professional rankings. These will serve as our benchmarks.

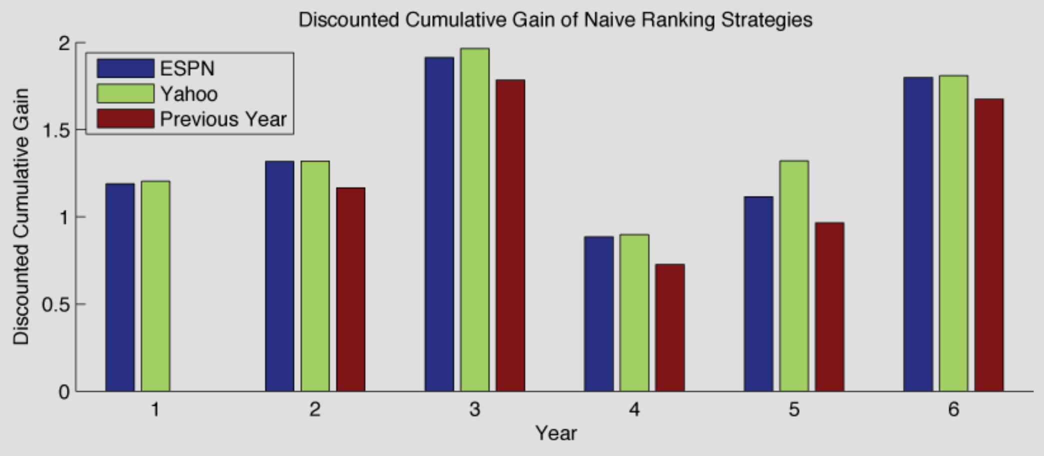


Figure 2: DCG of ESPN / Yahoo preseason rankings over the years, along with a naïve method of using the previous year's final rankings as the next year's preseason rankings.

| Expert | Discounted Cumulative Gain for 2012 Predictions |
|---|---|
| ESPN Expert (Matthew Berry) | 1.7991 |
| Yahoo Team of Experts | 1.8100 |

Figure 3: The numeric DCG of ESPN and Yahoo in 2012. This is the benchmark used for this research.

## Methods

**1) Linear Regression**

- $\beta = [\, w_{FantasyPts}\ w_{RecYds}\ w_{RecTDs}\ w_{Targets}\ w_{Catches}\ w_{\Delta 1yrRecYds}\ w_{\Delta 1yrRecTDs}\ w_{\Delta 2yrRecYds}\ w_{\Delta 2yrRecYds}\ w_{\Delta 3yrRecYds}\ w_{\Delta 3yrRecYds}\ w_{ESPN}\ w_{Yahoo}\,]$
- $\beta = (X^T \bullet X)^{-1} \bullet X^T \bullet Y$
- $Y = \beta \bullet X$

**2) Feature Reduction**

- Unimportant features, as determined by weight, will be removed to create an L-dimensional β for linear regression.
- $\beta = [\, w_1 \ldots w_L \,]$

**3) K-Means Mixture Model**

- 2 and 3 dimensional K-Means will be used to separate players, and linear regression will be run on each of the K mixtures individually.
- $\beta_1 = [\, w_1 \ldots w_N \,]$
  ↓
- $\beta_K = [\, w_1 \ldots w_N \,]$

**4) PCA Regression**

- PCA will be used to estimate the regression coefficients. Different numbers of principle components will be experimented with.
- $PCA(X) \rightarrow \Lambda_{p\times p}, \Omega_{n\times p}$
- $\Omega_{n\times p}$ = PC score. Representation of X in PC space
- $\Lambda_{p\times p}$ = PC loadings. Each column contains loadings for one PC
- $\beta = (\Omega^T \bullet \Omega)^{-1} \bullet \Omega^T \bullet (Y - \mu_Y)$
- Transform to regression coefficients for uncentered variables
- $\beta = [\, \mu_Y - \mu_X \bullet \Lambda \bullet \beta \mid \Lambda \bullet \beta \,]$
- $Y = [\, ones \mid X \,] \bullet \beta$

## Linear Regression

Using linear regression, predicted point totals are calculated for each player. Players are then ranked by predicted points for the final output.

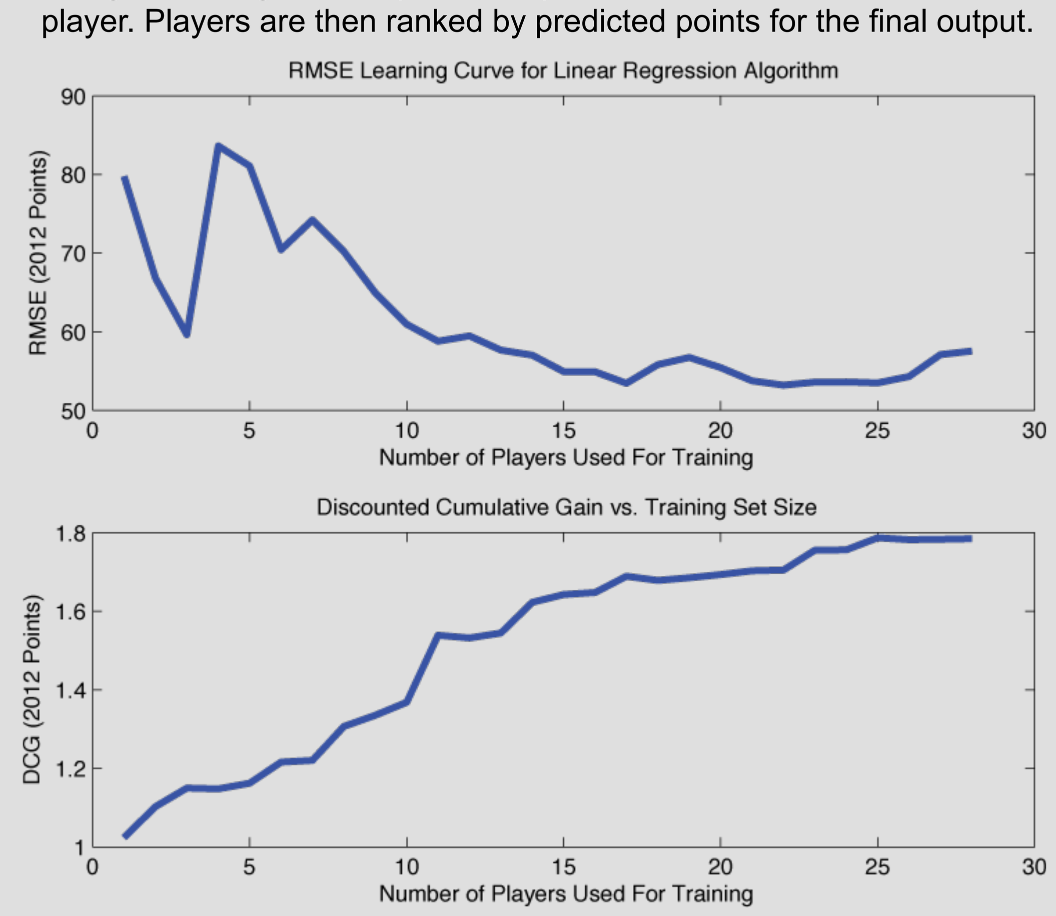


Figure 4: Linear regression versus accuracy as a function of training set size. As the RMSE between point predictions and actual end-of-year points decreases, the DCG of the resultant ranking increases.

## Feature Reduction

The best results are obtained by using a subset of 13 total features. The DCG for this 2-feature case is 1.7896, still short of what is achieved by ESPN and Yahoo, but very close, and better than the naïve ranking method proposed previously.

| Method | β | DCG |
|---|---|---|
| 13 Features | $\beta = [\, w_{FantasyPts}\ w_{RecYds}\ w_{RecTDs}\ w_{Targets}\ w_{Catches}\ w_{\Delta 1yrRecYds}\ w_{\Delta 1yrRecTDs}\ w_{\Delta 2yrRecYds}\ w_{\Delta 2yrRecYds}\ w_{\Delta 3yrRecYds}\ w_{\Delta 3yrRecYds}\ w_{ESPN}\ w_{Yahoo}\,]$<br>β = [0.17, 0.059, 12.91, -0.25, 0.18, -0.34, -4.82 -0.031, 0.61, 0.044, -3.36, -1.05, 1.03] | 1.7179 |
| 7 Features | $\beta = [\, w_{FantasyPts}\ w_{RecYds}\ w_{RecTDs}\ w_{\Delta 1yrRecYds}\ w_{\Delta 1yrRecTDs}\ w_{ESPN}\ w_{Yahoo}\,]$<br>β = [-0.18, 0.12, 11.10, -0.056, -3.35, -2.27, 1.67] | 1.7722 |
| 6 Features | $\beta = [\, w_{FantasyPts}\ w_{RecYds}\ w_{RecTDs}\ w_{\Delta 1yrRecYds}\ w_{\Delta 1yrRecTDs}\ w_{ESPN}\,]$<br>β = [0.10, 0.079, 10.23, -0.06, -3.09, -0.403] | 1.7621 |
| 3 Features | $\beta = [\, w_{FantasyPts}\ w_{RecYds}\ w_{RecTDs}\,]$<br>β = [1.72, -0.061, -5.08] | 1.7806 |
| 2 Features | $\beta = [\, w_{FantasyPts}\ w_{RecTDs}\,]$<br>β = [1.15, -2.031] | 1.7896 |
| 2 Features | $\beta = [\, w_{FantasyPts}\ w_{RecYds}\,]$<br>β = [0.92, 0.019] | 1.7863 |

Figure 5: Results of feature reduction on linear regression.

## K-Means Mixture Model

There are many types of receivers in the NFL, including those who play in the slot and catch lots of short passes, those who play out wide and catch very few longer passes, and all-stars who do it all. This insight prompts clustering receivers by type prior to prediction.

**K-Means Clustering**

K-Means will be used to cluster receivers based on various groups of 2 and 3 features.

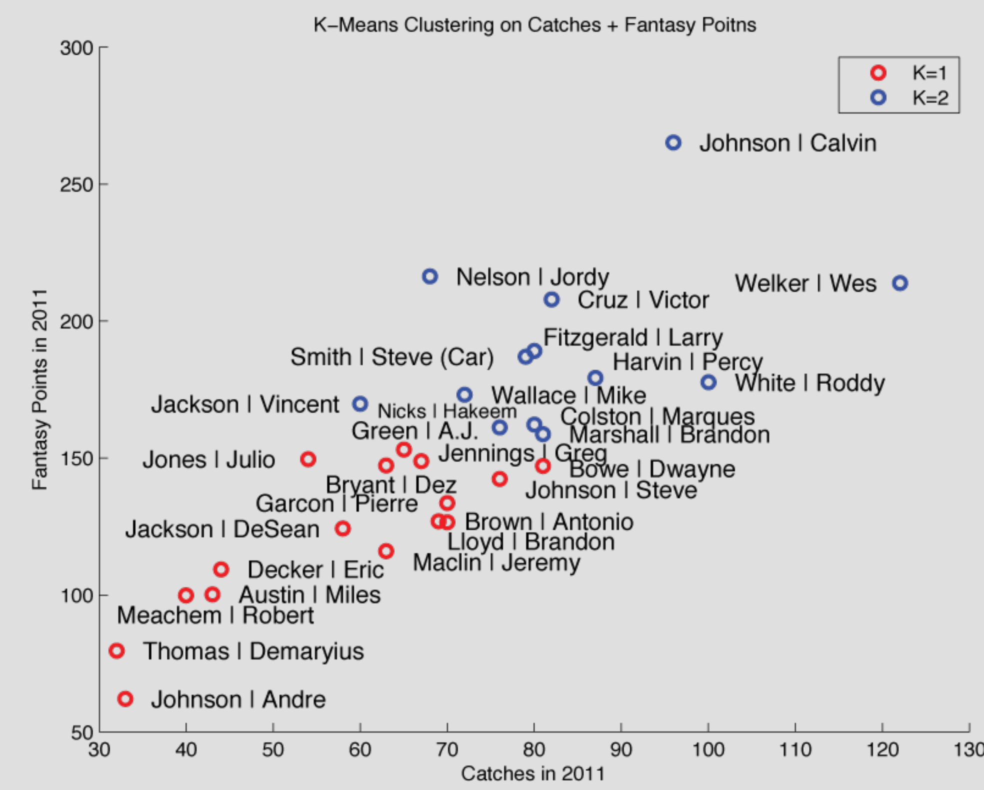


Figure 6: 2D K-Means clustering by previous year catches and fantasy points. Note how this tends to separate high production, elite WRs in blue from role players and secondary receivers in red.

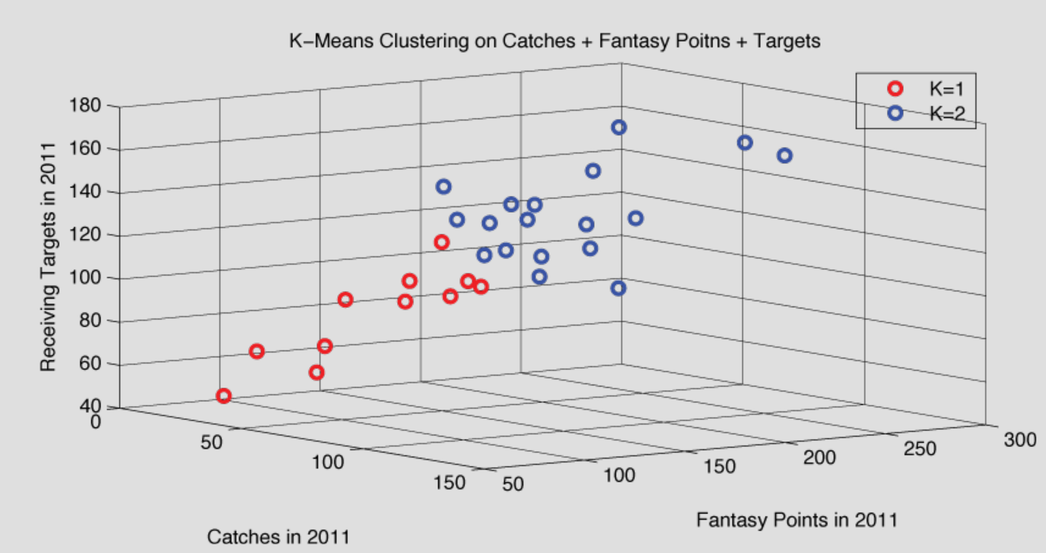


Figure 7: 3D K-Means clustering by previous year catches, fantasy points, and targets.

**Mixture Model**

Linear regression is performed on each K-Means grouping individually to produce each player's projected fantasy points for 2012.

**Feature Reduction**

Using the results from the feature reduction section, the experiment is repeated for 2 and 6 features.

**Combination**

The players, regardless of mixture, are then recombined to be ranked by the DCG algorithm.

| # Mixtures | K Means Features | DCG (6 Features) | DCG (2 Features) |
|---|---|---|---|
| 2 | Catches<br>Fantasy Points | 1.7601 | 1.8063 |
| 2 | Fantasy Points | 1.7601 | 1.7925 |
| 3 | Fantasy Points | 1.3621 | 1.7880 |
| 2 | Receiving Targets | 1.2053 | 1.7877 |
| 2 | Catches<br>Fantasy Points<br>Targets | 1.1887 | 1.7863 |
| 2 | Catches | 1.1811 | 1.7814 |
| 2 | Touchdowns | 1.1107 | 1.7747 |
| 2 | Receiving Yards | 1.0857 | 1.4428 |

Figure 8: K-Means mixture model results through several types of runs.

The best results are obtained by using 2D K-means on catches and previous year fantasy points to define our mixture model, and then using 2 features, points and TDs, for regression. The DCG is 1.8063, better than ESPN but not quite as good as Yahoo.

## PCA Regression

The results of PCA show that the majority of the variance in our data is explained by the first 4 principle components.

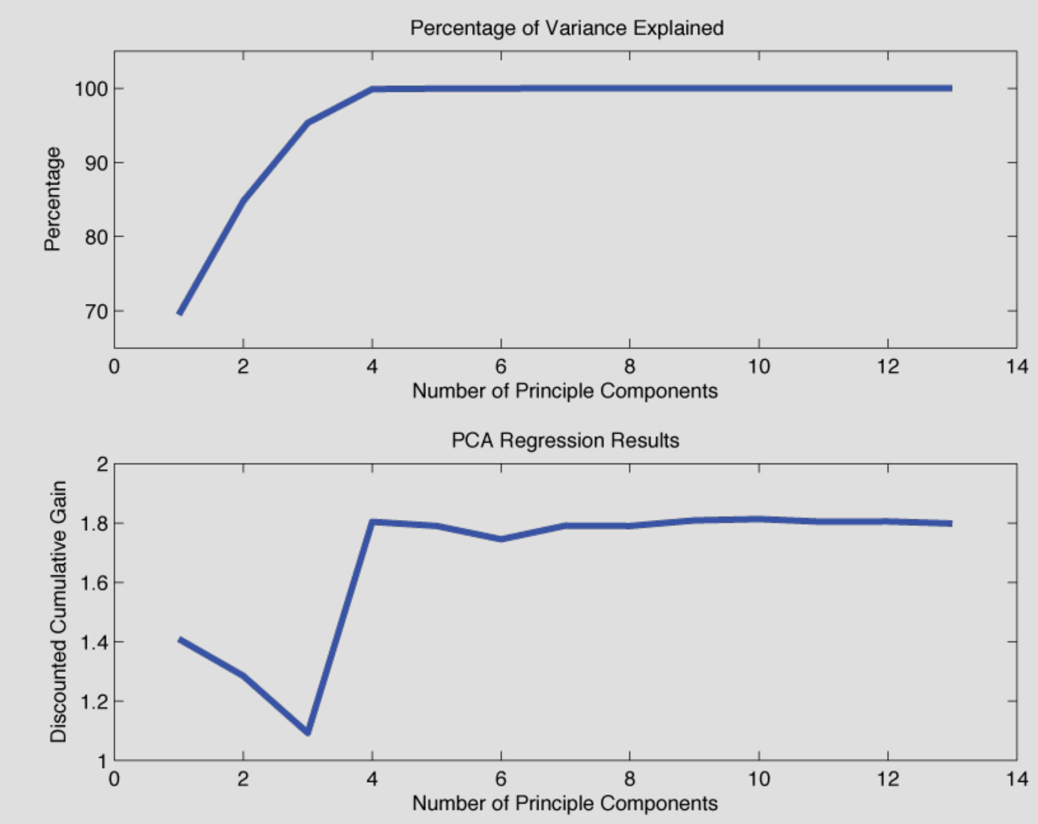


Figure 9: PCA and PCA regression run with varying numbers of principle components. Note how the DCG goes up as we choose enough principle components to explain the majority of the variance.

Using PCA regression with 10 principle components, we are able to out-perform the experts at both Yahoo and ESPN.

| Number of Principle Components | Discounted Cumulative Gain |
|---|---|
| 4 | 1.8040 |
| 10 | 1.8132 |

Figure 10: The PCA results are better than standard linear regression and the mixture model approach.

## Conclusion

Several machine learning approaches to ranking NFL receivers are shown. The best results come from PCA regression with 10 principle components. This method outperforms the experts at ESPN and Yahoo for the 2012 season.

| Actual | PCA Regression | ESPN | Yahoo |
|---|---|---|---|
| Johnson \| Calvin | Johnson \| Calvin | Johnson \| Calvin | Johnson \| Calvin |
| Marshall \| Brandon | Welker \| Wes | Fitzgerald \| Larry | Fitzgerald \| Larry |
| Bryant \| Dez | Johnson \| Andre | Johnson \| Andre | Jennings \| Greg |
| Green \| A.J. | Jennings \| Greg | White \| Roddy | White \| Roddy |
| Thomas \| Demaryius | Fitzgerald \| Larry | Jennings \| Greg | Cruz \| Victor |
| Jackson \| Vincent | White \| Roddy | Wallace \| Mike | Johnson \| Andre |
| Decker \| Eric | Wallace \| Mike | Welker \| Wes | Marshall \| Brandon |
| Johnson \| Andre | Nicks \| Hakeem | Nicks \| Hakeem | Green \| A.J. |
| Jones \| Julio | Cruz \| Victor | Green \| A.J. | Wallace \| Mike |
| White \| Roddy | Green \| A.J. | Cruz \| Victor | Jones \| Julio |
| Colston \| Marques | Jones \| Julio | Jones \| Julio | Welker \| Wes |
| Welker \| Wes | Colston \| Marques | Marshall \| Brandon | Nelson \| Jordy |
| Cruz \| Victor | Nelson \| Jordy | Nelson \| Jordy | Smith \| Steve (Car) |
| Crabtree \| Michael | Marshall \| Brandon | Smith \| Steve (Car) | Nicks \| Hakeem |
| Wayne \| Reggie | Smith \| Steve (Car) | Bryant \| Dez | Austin \| Miles |
| Jones \| James | Bryant \| Dez | Colston \| Marques | Thomas \| Demaryius |
| Cobb \| Randall | Harvin \| Percy | Lloyd \| Brandon | Colston \| Marques |
| Williams \| Mike (TB) | Bowe \| Dwayne | Jackson \| Vincent | Maclin \| Jeremy |
| Smith \| Steve (Car) | Lloyd \| Brandon | Harvin \| Percy | Harvin \| Percy |
| Johnson \| Steve | Jackson \| Vincent | Bowe \| Dwayne | Bryant \| Dez |
| Moore \| Lance | Maclin \| Jeremy | Maclin \| Jeremy | Bowe \| Dwayne |
| Shorts \| Cecil | Johnson \| Steve | Johnson \| Steve | Lloyd \| Brandon |
| Smith \| Torrey | Austin \| Miles | Austin \| Miles | Britt \| Kenny |
| Wallace \| Mike | Brown \| Antonio | Thomas \| Demaryius | Johnson \| Steve |
| Hilton \| T.Y. | Jackson \| DeSean | Brown \| Antonio | Jackson \| Vincent |
| Austin \| Miles | Garcon \| Pierre | Decker \| Eric | Jackson \| DeSean |
| Maclin \| Jeremy | Meachem \| Robert | Jackson \| DeSean | Brown \| Antonio |
| Blackmon \| Justin | Thomas \| Demaryius | Meachem \| Robert | Decker \| Eric |
| Rice \| Sidney | Decker \| Eric | Garcon \| Pierre | Smith \| Torrey |
| Nelson \| Jordy | Smith \| Torrey | Smith \| Torrey | Meachem \| Robert |

Figure 11: Final outputs compared. Using PCA Regression would have beat ESPN and Yahoo in 2012.

## References


1) "Fantasy Football Homepage." ESPN.com. 2007 – 2012.
2) "Fantasy Football Homepage." Yahoo.com. 2007 – 2012.
3) Kapina, Nitin. "Predicting Fantasy Football Performance With Machine Learning Techniques." Final Project for Andrew Ng's AI Course. December 14, 2012.
4) Mathews, Tim, et. al. "Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domain." Association for the Advancement of Artificial Intelligence, 2012.
5) Farquhar, Lee, et. al. "Types of Fantasy Sports Users and Their Motivations." Journal of Computer-Mediated Communication, 2007.
6) Wood, Frank. "Principle Component Analysis." Tutorial PDF. December 2009.
7) Torres-Reyna, Oscar. "Linear Regression, version 6.0" Data and Statistical Services, Princeton University.
8) Rasmussen, Carl Edward. "The Infinite Gaussian Mixture Model." Advances in Neural Information Processing Systems, 2012.
9) McLachlan, G. J., et. al. "A mixture model-based approach to the clustering of microarray expression data." Bioinformatics, 2001.
10) Steinbach, Michael, et. al. "A Comparison of Document Clustering Techniques." Technical Report, University of Minnesota. May 2000.