EECS/MSAI 349 Project Status Report

Submitted by: Weilin Ma

My task is to predict an active NBA player's most fitted position out of the 8 "new positions" based on the attributes of the player's physicals and basketball game's normal stats categories. This idea is based on the research project I found by Alex Cheng. My task is important because it's the job every team manager and sports analyzer will do prior to the rookies draft night and constantly during the first 1-2 years of their career. Teams are making choices how to pick the rookies, how to make trades to build their rosters, and how to manage the players' playing time based on this important fact.

The dataset was scraped (otherwise it takes forever to scrape the data) from basketball-reference.com of some active players since 2013-2014 season. Because it pins as the start of a new era of basketball, the "small ball". Teams are favoring speed, pace, 3pt shooting more than height and 2pts.
There are 10 attributes that I used:

1. AST% (assist percentage)
2. TRB% (total rebound percentage)
3. TS% (true shoorting perccentage)
4. 3PAr (3-point attempt rate)
5. FTr (free throw attempt rate)
6. STL% (steal percentage)
7. BLK% (block percentage)
8. USG% (usage percentage)
9. OWS (offensive win share)
10. DWS (defensive win share)

The players will be assigned into 8 categories:

1) Floor General
2) Combo Guard
3) Scoring Wing
4) Shooting Wing
5) 3-and-D Wing
6) Versatile Forward
7) Scoring Center
8) Supporting Center

I am using about 600 players' data who are active from 2012 to the end of 2017 season. 10-fold cross-validation was used to train the model. Then I'll need to use NCAA datasets of the 2017 rookie class to test the model. In the end, I can compare the test results to how the teams are actually using the Class 2017 Rookies.

Up until now I am about to finish the dataset scraping and started looking for packages that perform the best. According to my research on similar projects, I will be most likely using Scikit-learn KMeans Clustering. I'm also using Tableau for visualization. Decision Tree and Random Forest are gonna be tried first.

My plans for the rest of the project will be doing the training, testing the 2017 rookie class with NCAA dataset, and write up the project report and the website.