Project 1 proposal: Do fans know who the best nba players are?

By Noel Mathew

What is the problem I want to solve?

Every year from November to June, fans watch the best basketball players and teams compete for the NBA championship. The NBA is the a star-driven league where fans gravitate to certain high performing players. Those players usually generate a lot of buzz on social media platforms like Twitter, and people will buy tickets just to watch that player play. The problem I want to solve is how much does of the players' performance factor into if a player is considered a star or not. In other words, do fans know who are the best players are in the NBA.

Who is my client and why do they care about this problem?

One of the clients in this case will be the NBA teams. If a NBA team has a high performing player and that team is having low attendance numbers then that team officials could use this project to help them realize and promote their players. Another set of clients could be the agents of the players. The agent can use this project to find out that he has a star player/high performing player who is not generating any buss. Then the agent could use the knowledge and help get fans notice that player. This in turn will help the player get more endorsements as his marketing appeal will go up. Another client could be the NBA itself. SInce the NBA is a star driven league, it can use this model to help promote unnoticed high performing players and turn them into stars.

What data am I using? How will I acquire the data?

I will start with basketball statistics of the players during the 2017 season. There are the basic statistics: points, assists, rebounds, shooting percentage, steals, blocks, and minutes played (all per game). There are also what are called advanced statistics which require some more complicated calculations and are designed to show how a player contributes to team success beyond the basic stats, e.g. plus-minus, win shares, player efficiency rating, etc. Then I will look how much a player's tweets get retweeted or favorited. In addition, I will look at number of hits that the player's wikipedia page had during that season. In addition, I will look at attendance for each team to see fans are seeing the high performing players. All of these statistics are widely available at https://www.kaggle.com/noahgift/social-power-nba.

Briefly outline how I'll solve this problem

I believe that a regression model will be used to highlight relationship between a players stats and their social media buzz. I will come up with a value which determines if a player is a high performing player or not and a value for their social media buzz and model that relationship.

Here's my outline as I envision the project right now:

1. Gather, wrangle, and tidy up data for as many players that played in the 2016-2017 NBA season as I can get enough data for
2. Train various regression machine learning on the data
3. Use cross-validation techniques to tune hyper-parameters for each of the model
4. Use the results of the model to see if a high-performing player is going unnoticed or a low-performing player is getting too much noticed.

What are my deliverables?

At the end of the project, I will have a Jupyter Notebook with both the project code and my report comments. I will also have a slide deck for presenting the project. I will have a list of high performing players who are unnoticed and what could be possible reasons that they are unnoticed.