

# NBA Positions and Machine Learning

by Tejas Prakash

## Introduction

So recently as a side project I decided to venture into the realm of NBA machine learning. Last week I did a project that estimated value of a player. This week I'm doing an estimation of a player's position, purely based on stats. We now have to tackle the problem of: Which stats could we possibly use to quantify a playstyle? After some thought, I decide to use nine statistics: TS%, 3-point rate, free throw rate, rebounding percentage, assist percentage, steal percentage, block percentage, turnover percentage, and usage rate. My data source is basketball-reference. Instead of crawling bbref, I decided it was easier to just export the data as a CSV and iterate through that, storing results in a dictionary. The model I used was a LinearSVC model. This model was quite simple to use, but I quickly found it was far too inconsistent for my needs. So I decided to create a function that would run the model fifty times and average the results. This caused my results to be far more consistent. My results are stored [here \(link\)](#). Please remember when viewing the data, I did not consider actual performance in this (other than true shooting percentage). Most of it is just usage rates, so keep that in mind. You can view my code [here](#).

# The Ideal Players

According to my model, the ideal players at each position are:

Ideal Point Guard: Chris Paul/Ricky Rubio/John Wall

Ideal Shooting Guard: Klay Thompson/JJ Redick/Bradley Beal

Ideal Small Forward: Otto Porter/Jae Crowder/Trevor Ariza

Ideal Power Forward: Nikola Vucevic/Karl-Anthony Towns/Kevin Love

Ideal Center (many ties): Rudy Gobert/Joakim Noah/Embiid/DeAndre Jordan/Dwight Howard

## Point Guards

### What makes a PG a PG?

The top ideal PG was CP3, followed by Ricky Rubio, and then John Wall. These three PG's are all pass-first, traditional point guards. Chris Paul especially is known as one of the last true point guards. Rubio has always been a passer, and only recently has developed a decent jump shot. Wall can do everything, but is also a great passer. View the table below for their stats during the 2016-17 season (the ones that were considered).

	TS%	3PAr	FTr	ORB %	DRB%	TRB %	AST%	STL%	BLK%	TOV%	USG%
Chris Paul	.614	.385	.331	2.4	15.0	8.9	46.8	3.1	0.3	14.0	24.4

Rubio	.539	.302	.438	3.2	11.4	7.2	38.9	2.6	0.3	20.1	17.4
John Wall	.541	.190	.367	2.4	10.6	6.5	46.9	2.7	1.4	16.2	30.6
Drummond	.518	.008	.390	15.1	36.2	25.2	6.0	2.6	3.1	12.5	22.4

TABLE A-1

For contrast, I have included Andre Drummond, one of the lowest PG matches (0.0025). A PG seems to be characterized by decently high usage rates. Wall leads with 30.6% of these three PG's. Paul and Rubio have high three point attempt rates at .385 and .302, respectively. Wall is a little lower at .190, but still higher than Drummond's .008. The guards have very low ORB% rates compared to Drummond. Rubio and Wall have  $\frac{1}{3}$  of Drummond's defensive rebound rates. PG's, quite obviously, have a high AST rate. Point guards are known as the distributors. They run the ball down the court, execute plays, and gameplay for the most part runs through them. They have decent STL rates, and low BLK rates. So the ideal PG is a great passer and distributor, can shoot the three, and can steal the ball.

## PG's in Other Positions

The most interesting point guard in another position is Michael Carter-Williams. After winning ROTY in his rookie year, he has become nothing more than a role player. Carter-Williams is a 6'6 PG who cannot really shoot that well. However, my model seems to think that he's a center. See Table A-2.

	TS%	3PAr	FTr	ORB %	DRB%	TRB %	AST%	STL%	BLK%	TOV%	USG %
MCW	.473	.169	.317	3.6	15.7	9.5	30.3	2.5	1.6	18.2	24.4
Chris Paul	.614	.385	.331	2.4	15.0	8.9	46.8	3.1	0.3	14.0	24.4
Gobert	.681	.002	.762	13.6	29.5	21.8	5.7	0.9	6.4	15.0	16.7

Table A-2

In this table, we have MCW's stats, compared to the ideal PG and C. Immediately, we see MCW's low TS%. This might be the cause of the mess here. This value seems to have thrown the model off. We also see terrible rebounding rates, and a high AST rate. I cannot think of any other reason that the model would have predicted MCW as a C, as everything else points to him being a guard. [0.2137225815328117, 0.15601459330304984, 0.20995262024971836, 0.1986333699474779, 0.22167683496694215] are the prediction values. It seems there was a very small difference between him being a PG, SF, PF, and C. This is one of the worst and most disappointing results of the model.

Centers typically have low 3PAr, high rebounding rates, higher FT rates, lower AST rates, lower STL rates, higher BLK rates, and obviously lower USG rates. The only center that came up as a PG was Marc Gasol, and I would guess that this is a result of many things. First, he is starting to come into his own as a three point shooter. In the 15-16 season, he hit two threes. In the 16-17 season, he hit 104 threes. And he did it at a surprising efficiency. Another reason that he might be listed as a PG is his low offensive rebound rates. However, I believe that the main reason that he is listed as a PG is his 26.4% usage rate, which is actually higher than CP3's.

Other than Gasol, the second closest center to PG was Jokic. Again, Jokic has the three point shooting game at a pretty good efficiency, and he has a high AST rate and STL rate. I believe the reason he is a C and not a PG is his higher rebounding rates. The below table might help to show the true differences.

	TS%	3PAr	FTr	ORB %	DRB %	TRB %	AST %	STL%	BLK %	TOV %	USG %	
Gobert	.681	.002	.762	13.6	29.5	21.8	5.7	0.9	6.4	15.0	16.7	
CP3	.614	.385	.331	2.4	15.0	8.9	46.8	3.1	0.3	14.0	24.4	
Wall	.541	.190	.367	2.4	10.6	6.5	46.9	2.7	1.4	16.2	30.6	
Gasol	.554	.231	.286	2.6	18.8	10.5	24.8	1.4	3.7	11.3	26.4	
Jokic	.616	.153	.287	11.5	26.0	18.8	23.9	1.8	2.2	14.4	21.8	

I truly enjoyed doing this project. It gave me insights into basketball and furthered my knowledge of machine learning.