

Winning is the Name of the Game

Predicting Individual NBA Player Win Share

Joe Earnshaw

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

What do the players below have in common?

Nikoloz Tskitishvili



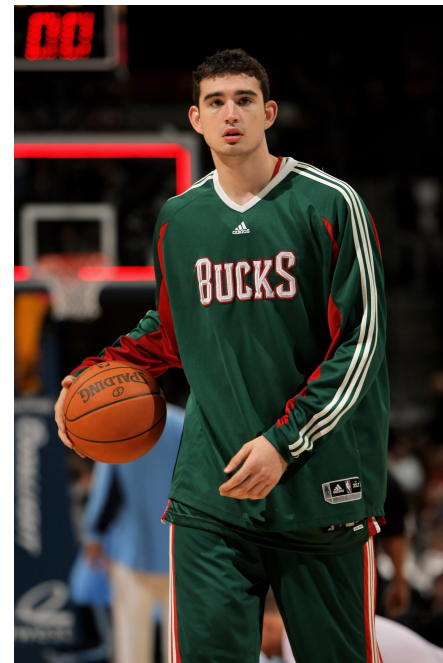
Source: [Mile High Sports](#)

Rafael Araujo



Source: [holycrapshots.wordpress.com](#)

Joe Alexander



Source: [Bleacher Report](#)

All three passed the
'Eye Test'



Source: [Sports Illustrated](#)

Basketball and...psychology?

“Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance.”

-- Daniel Kahneman, *Thinking, Fast and Slow*

Just a few examples...

Anchoring

“tendency to rely too heavily on the first piece of information offered (the ‘anchor’) when making decisions (1)

Confirmation Bias

“the tendency to search for, interpret, favor, and recall information in a way that confirms one’s preexisting beliefs or hypotheses” (2)

Present Bias

“tendency of people to give stronger weight to payoffs that are closer to the present time when considering trade-offs between two future moments” (3)

Hindsight bias

“the tendency to see past events as being predictable at the time of those events [happening]” (4)

The Old & New: Evolution of the NBA

Analytics now widely used throughout the league

Data analysts can be found in most teams front office

Massive amounts of data being generated, largely thanks to sophisticated tracking camera systems



Source: [UMKC](#)

‘Google Boy’ versus Old-School



Charles Barkley is not a fan either...



What do these teams have in common?



Source: [Rolling Stone](#)



Source: [NBA Futures](#)



Source: [Spectrum News](#)

They win
A LOT...
with the help
of analytics!



Winning in the NBA

How do you win in the NBA?

Simple, having better players!

Issue: how do you define 'better'?

Solution: WS/48

What is 'WS/48'?

[Basketball-Reference](#): “a player statistic which attempts to divvy up credit for team success to the individuals on the team [per 48 minutes]”

Or more simply, does this player help us win?

Takes both offensive and defensive statistics into account

Important to note: impressive performance but it is not 100% perfect

“All models are wrong, but some are useful”

Next question: can we predict this?

With vast amounts of data available, combined with machine learning, yes we can!

Or at least get very close...

With machine learning, important to remember “all models are wrong, but some are useful”

Can help us get a good picture of how good any given NBA player is

The Data

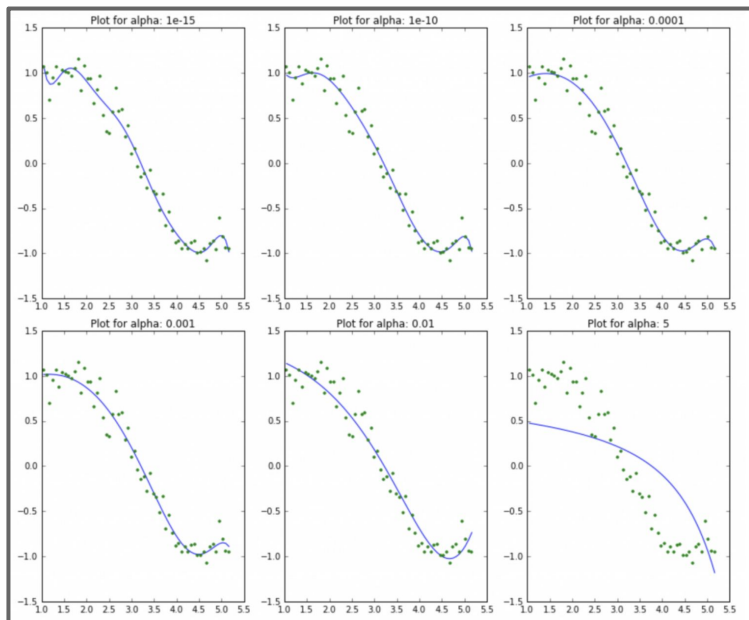
51 Variables (i.e. statistics)

Nearly 5000 records from the past 10
NBA seasons

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4759 entries, 0 to 4758
Data columns (total 51 columns):
Player      4759 non-null object
Pos         4759 non-null object
Age         4759 non-null int64
Tm          4759 non-null object
G           4759 non-null int64
GS          4759 non-null int64
MP          4759 non-null int64
FG          4759 non-null float64
FGA         4759 non-null float64
FG%         4759 non-null float64
3P          4759 non-null float64
3PA         4759 non-null float64
3P%         4759 non-null float64
2P          4759 non-null float64
2PA         4759 non-null float64
2P%         4759 non-null float64
FT          4759 non-null float64
FTA         4759 non-null float64
FT%         4759 non-null float64
ORB         4759 non-null float64
DRB         4759 non-null float64
TRB         4759 non-null float64
AST         4759 non-null float64
STL         4759 non-null float64
BLK         4759 non-null float64
TOV         4759 non-null float64
PF          4759 non-null float64
PTS         4759 non-null float64
ORtg        4759 non-null float64
DRtg        4759 non-null float64
PER         4759 non-null float64
TS%         4759 non-null float64
3PAr        4759 non-null float64
FTr         4759 non-null float64
ORB%        4759 non-null float64
DRB%        4759 non-null float64
TRB%        4759 non-null float64
AST%        4759 non-null float64
STL%        4759 non-null float64
BLK%        4759 non-null float64
TOV%        4759 non-null float64
USG%        4759 non-null float64
OWS         4759 non-null float64
DWS         4759 non-null float64
WS          4759 non-null float64
WS/48       4759 non-null float64
OBPM        4759 non-null float64
DBPM        4759 non-null float64
BPM         4759 non-null float64
VORP        4759 non-null float64
Year        4759 non-null int64
dtypes: float64(43), int64(5), object(3)
```

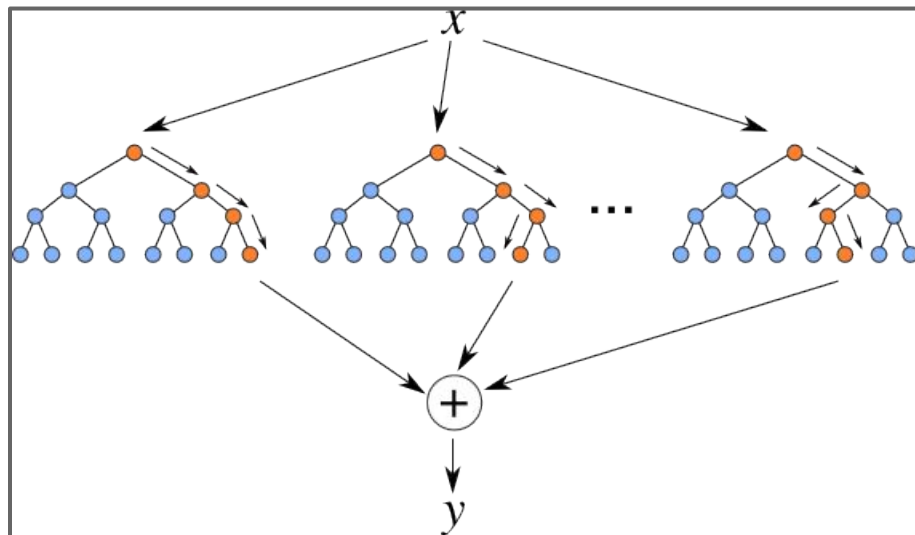
Machine Learning: The Models

Ridge Regression



Source: [Analytics Vidhya](#)

Random Forest



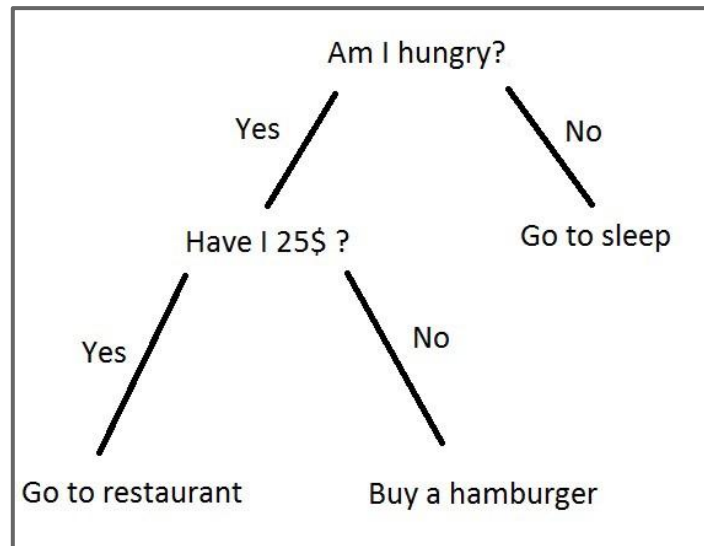
Source: [HARP](#)

The Models In Everyday Terms

Ridge Regression



Random Forest



Machine Learning: The Metrics

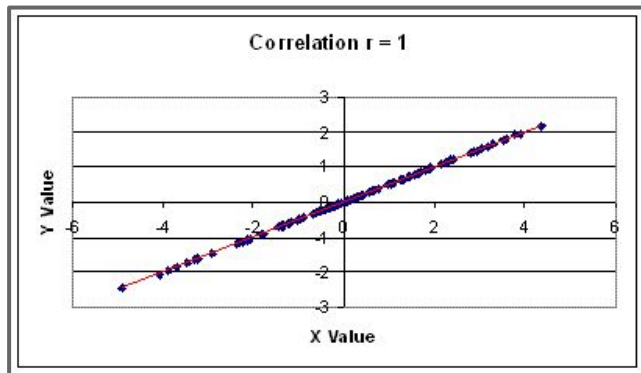
Evaluating our models

R-squared

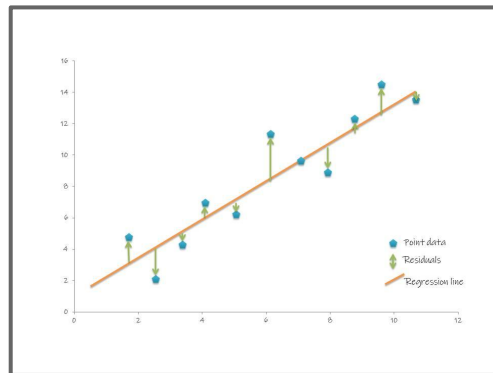
- 'Coefficient of determination'
- Provides goodness of fit

RMSE

- 'Root mean Squared Error'
- helps us see how far off predictions are from actual observations



Source: [DePaul](#)



Source: [Hatari Labs](#)

The Results are in....

Performance for each model

Ridge Regression

- R-squared = 0.975
- RMSE = ~0.01

Random Forest

- R-squared = 0.895
- RMSE = ~0.03

So what does this mean?

Can use machine learning to get a very approximate picture of how 'good' a player is

Help us overcome things like confirmation bias

In conjecture with scouts, can assemble teams that consistently win (think Golden State Warriors)

In summary...

Charles should probably just stick to working on his golf swing...

