

A.I. Scout: Improving NBA Player Evaluation with Machine Learning

By Chum Mapa

CONTEXT

- NBA G League creation
- Player development strategy

PURPOSE

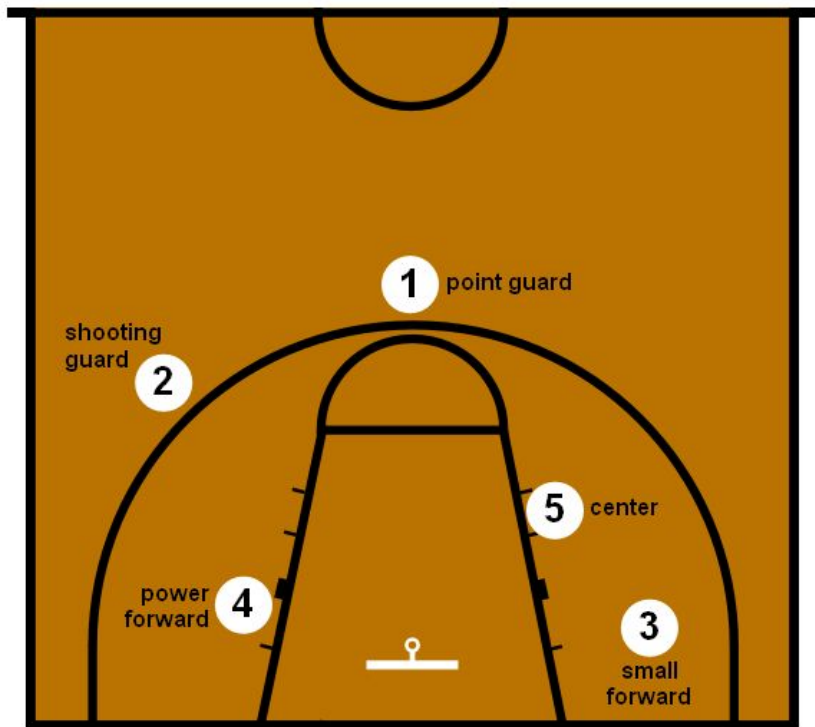
- Track player development
- Inform decisions

OBJECTIVES

- Have positions actually become more fluid in the modern NBA?
- Create/define new player molds
- Create prediction model



Traditional Basketball Positions



Point Guard: Primary ball handler and passer

Shooting Guard: 3-point range and long mid range

Small Forward: Versatile

Power Forward: score close to the rim and shoot mid-range

Center: Slower, taller "back to the basket" players

DATA

FiveThirtyEight  **Sports**

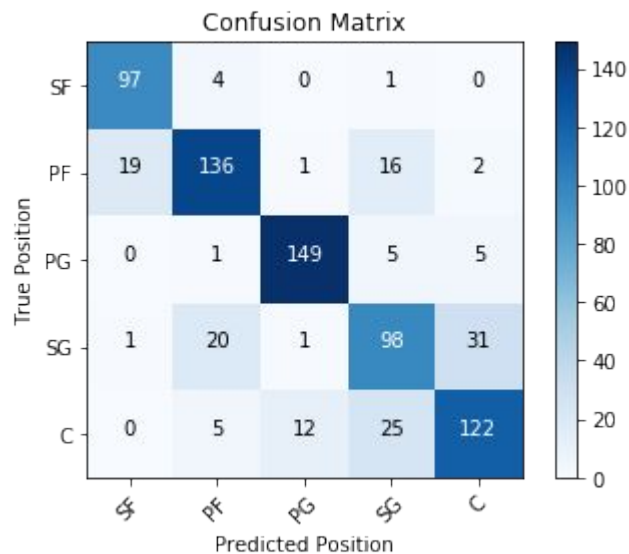
 **BASKETBALL**
REFERENCE
via Bart Torvik (barttorvik.com)

"I don't have the five positions anymore...
we've become more versatile as the years have gone on."

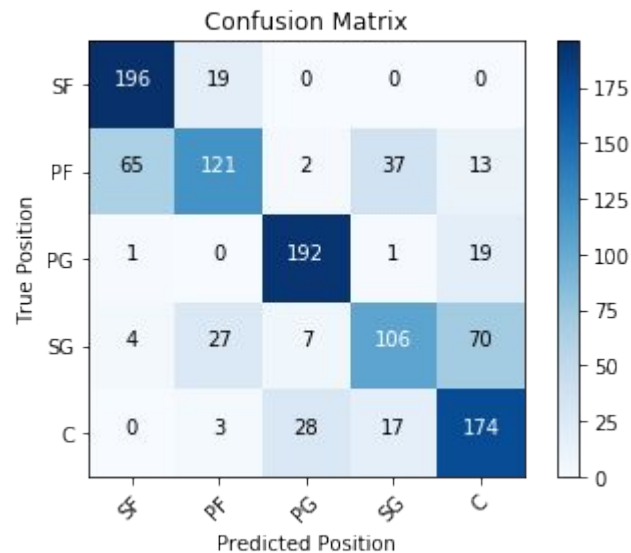
Brad Stevens,
Boston Celtics Head Coach

Source: Kareem Copeland of the Associated Press

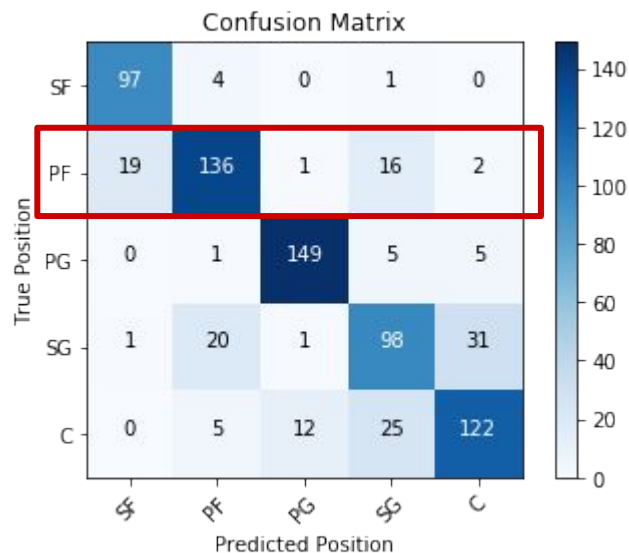
90's Era



Modern Era

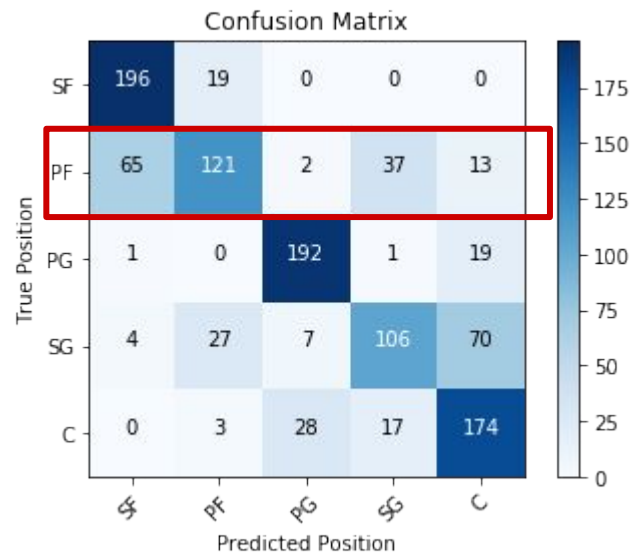


90's Era



Accuracy: 78.1%

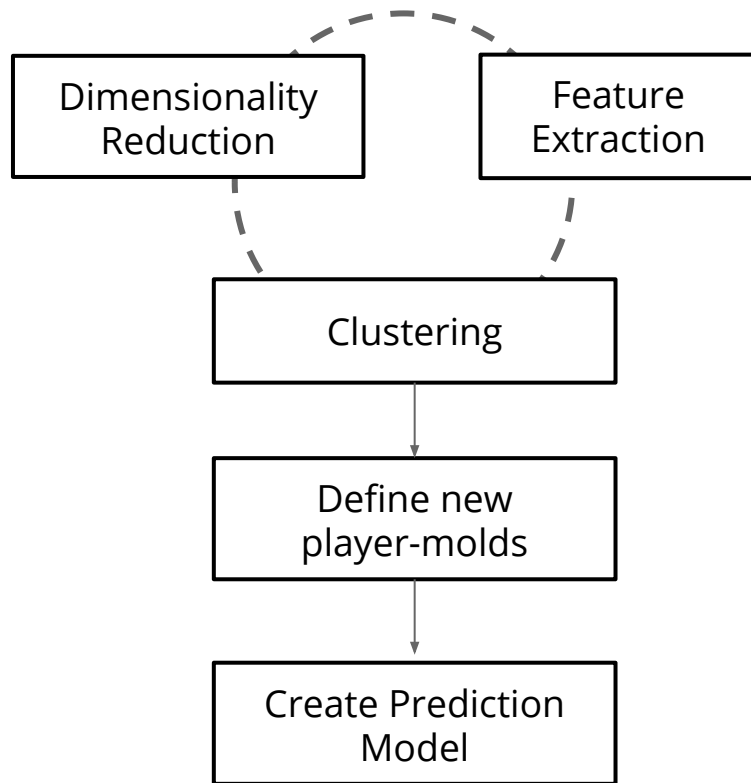
Modern Era



Accuracy: 50.1%

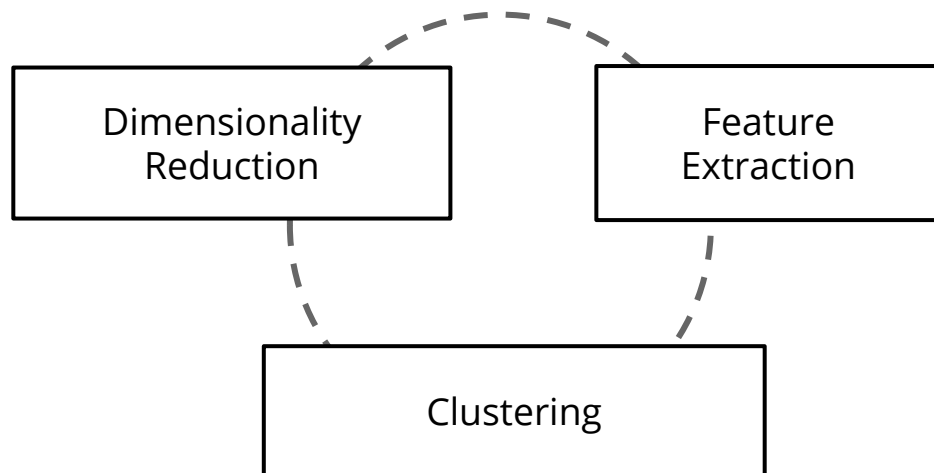
WORKFLOW

- Creating modern player-molds



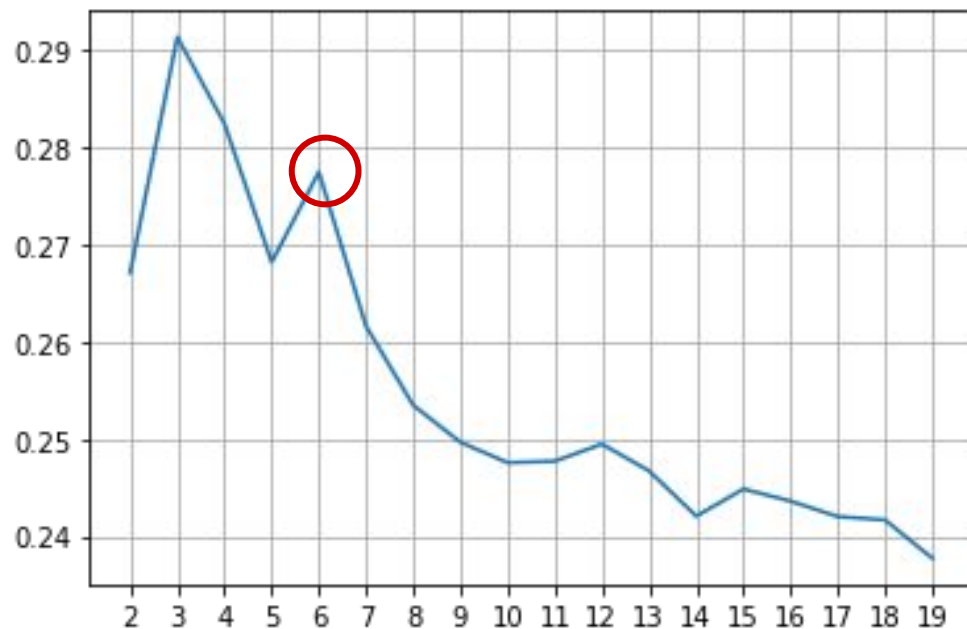
Step 1: Unsupervised Learning

Finding an optimal number of distinct groups using clustering methods



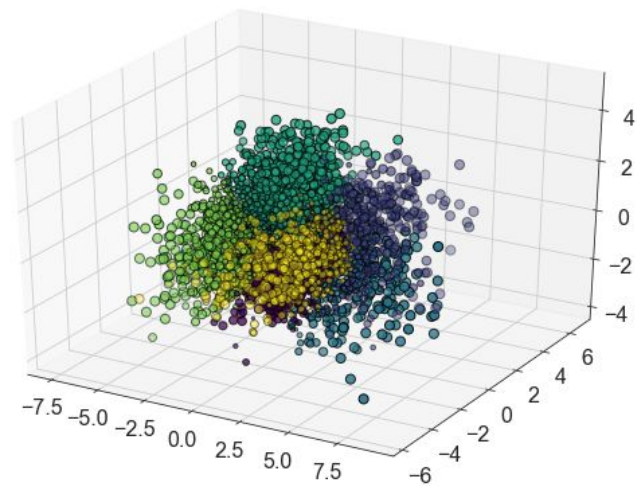
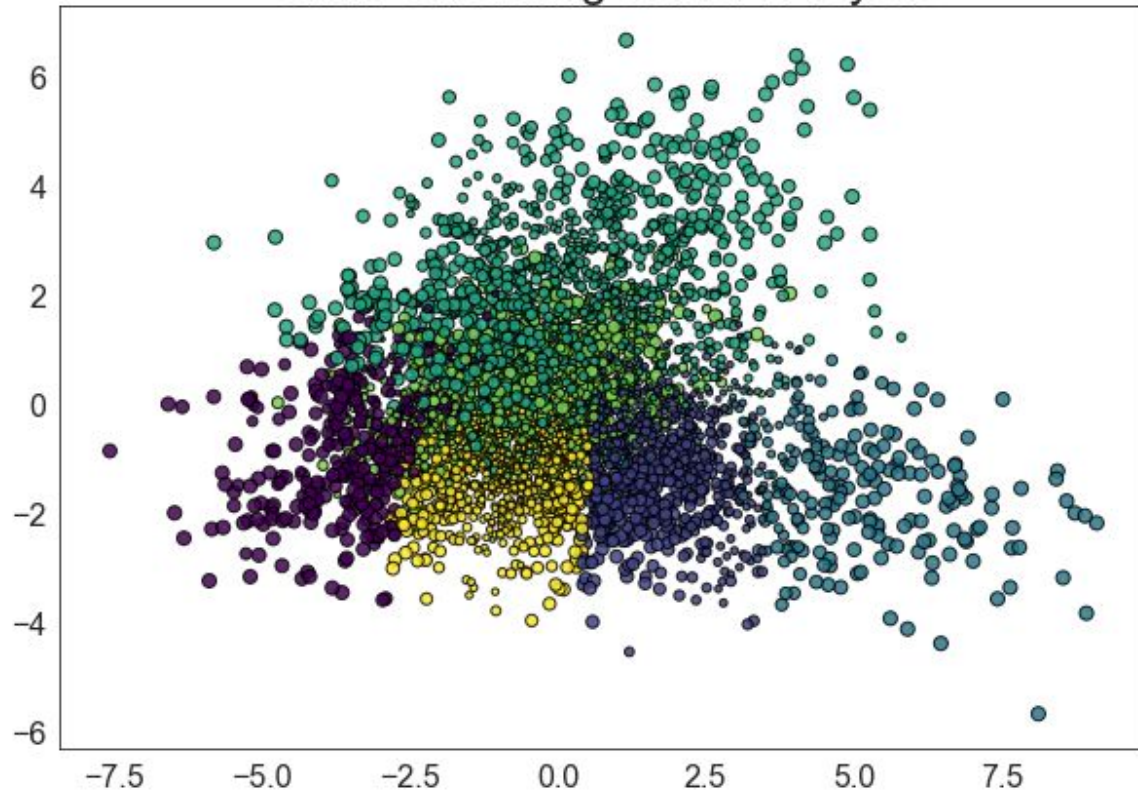
Step 1: Unsupervised Learning

Silhouette Score



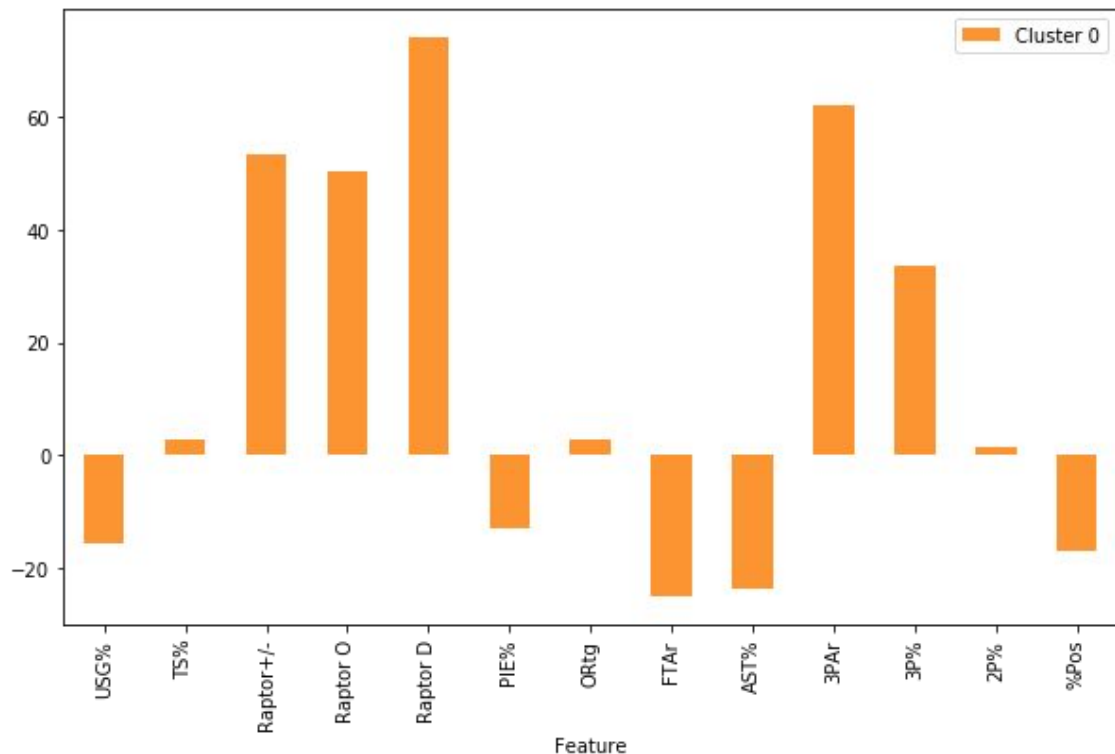
6 Clusters with 3
Principle Components

GMM Clustering of NBA Players



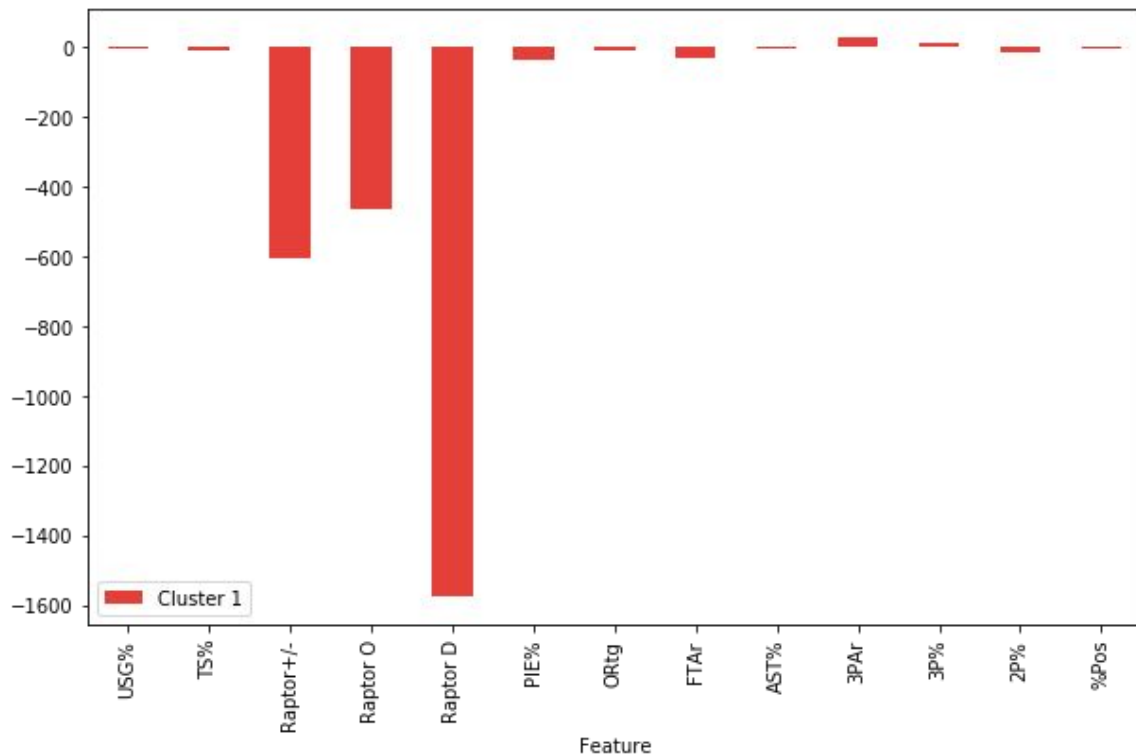
GAUSSIAN MIXTURE MODEL

Co: Two-Way Long Distance Shooter



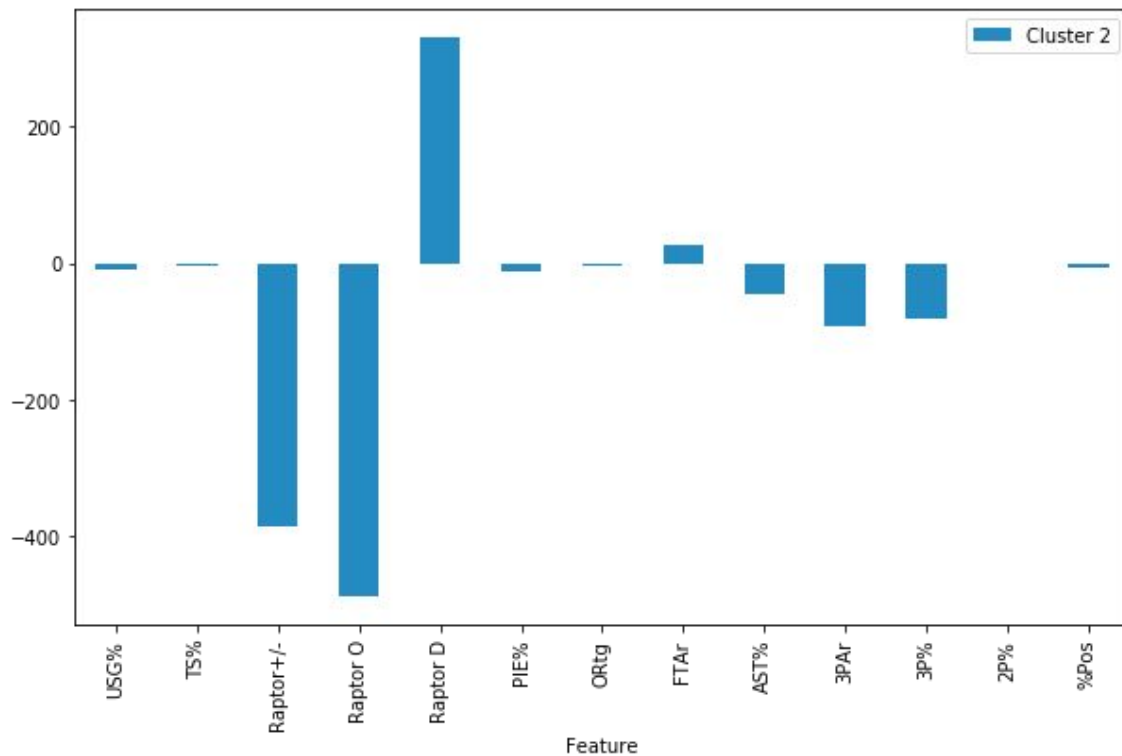
- Ray Allen
- Danny Green
- JJ Redick
- Kyle Korver

C1: Perimeter Role Player



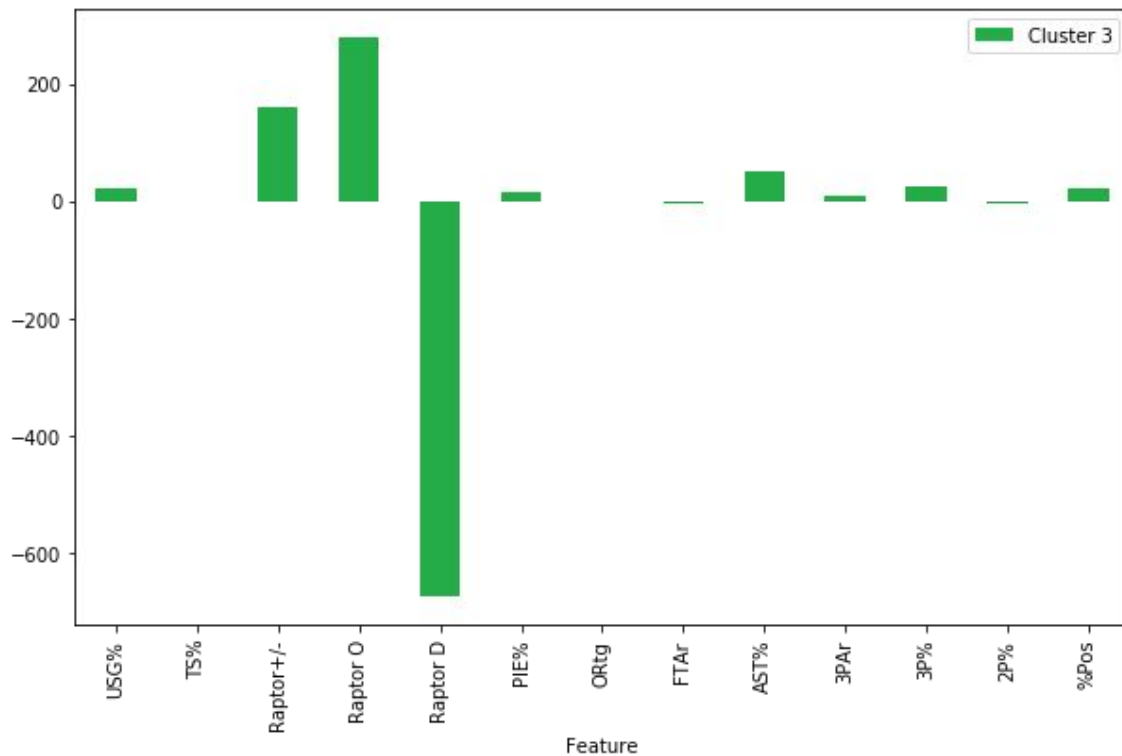
- Michael Carter-Williams
- Kent Bazemore
- Justin Holiday

C2: Limited Big Man



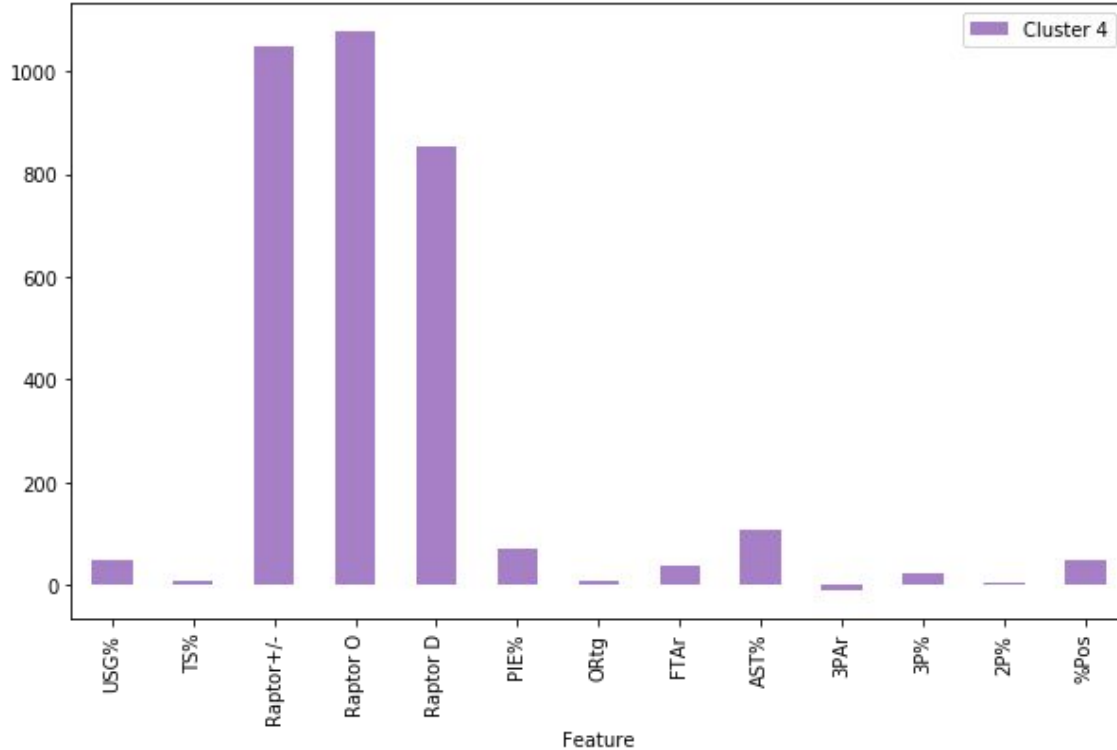
- Kendrick Perkins
- Ben Wallace

C3: Versatile Sharpshooters



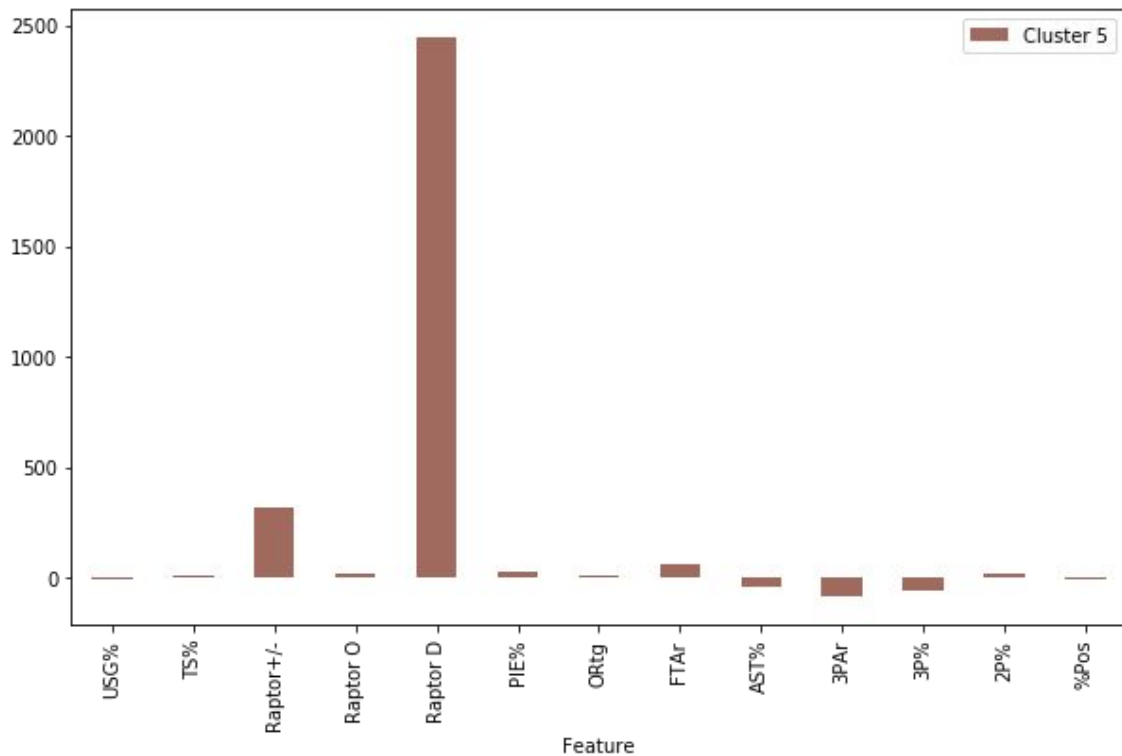
- Kyle Lowry
- Jamal Crawford
- Deron Williams
- Tony Parker

C4: Superstars



- Steph Curry
- LeBron James
- Nikola Jokic
- Kevin Durant

C5: Traditional 'Big Man' Mold

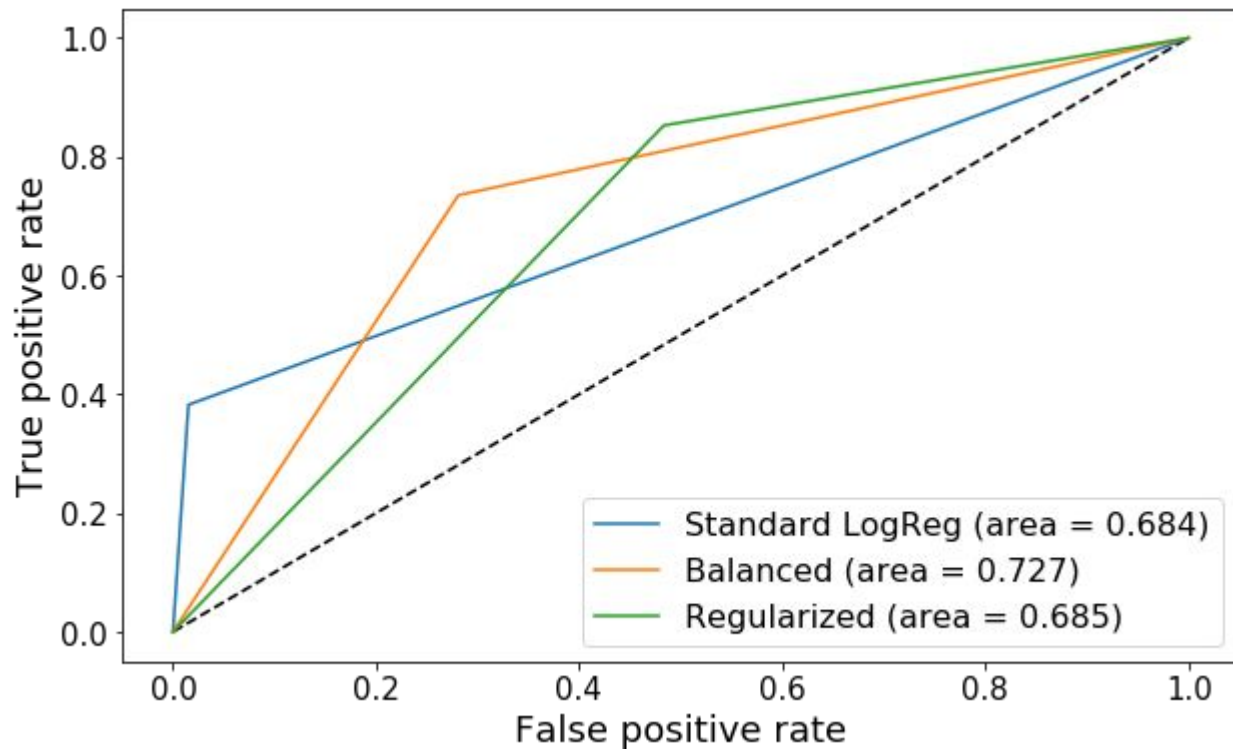


- Kevin Garnett
- Dwight Howard
- Rudy Gobert
- Tyson Chandler

Prediction Model

Logistic Regression:
Identifying Future 'C4'
Superstars

ROC curve



C4 Predictions:

- Recall: 76%

BEST PERFORMING: Model corrected for class imbalance

player_name	year	class	conf	prediction	probability1
Josh Sharkey	2020	Sr	SC	c4	100.0
Antoine Davis	2020	So	Horz	c4	99.9
Braden Norris	2019	Fr	Horz	c4	99.9
Matt Frierson	2019	Sr	SC	c4	99.8
Jagan Mosely	2020	Sr	BE	c4	99.7
Chase Freeman	2020	Jr	NEC	c4	99.7
Matt Halvorsen	2019	So	SC	c4	99.7
Josh Sharkey	2019	Jr	SC	c4	99.7
Connor McCaffery	2020	So	B10	c4	99.4
Kenny Williams	2019	Sr	ACC	c4	99.1
Jermaine Marrow	2020	Sr	BSth	c4	98.9

NEXT STEPS

- Feature Engineering
- Multiclass Predictions
- International Stats
- Flask/Dash Application

CONTACT

- CHUM MAPA
 - chaminda.mapa@gmail.com
 - github: chum46
 - linkedin.com/in/chaminda-chum-mapa