Defining Offensive and Defensive Positions in College Basketball to Build Optimal Rosters for Maximum Tournament Success

Colin Krantz and Kushal Shah, Syracuse University

Abstract

While some say the game of basketball is "positionless," we believe that providing more specific position definitions could help to better understand the value of each player. In college basketball, March Madness runs can put a team on the map and possibly make them a contender for years to come. As we said, position definitions need to be more specific to provide the true value of a player, so we create offensive and defensive position definitions for players based on their physical attributes and performance statistics in order to exhibit a better interpretation of their role on the court than just Guard, Forward, or Center. Certain positions are more closely related to winning tournament games, so recruiting high schoolers and transfers who more closely conform to those positions could become the key to making deep runs in March Madness.

Methodology

There were 3 total components to our newly defined positions:
1. Offensive Cluster
2. Defensive Cluster
3. Random Forest Predicted Position

Our positions were based on 3 major aspects of a player. Their physical build, style of play, and shot selection. We used a collection of metrics for players, however we avoided singular advanced performance metrics (ORTG, DRTF) as that would simply cluster players based on their skill which would not provide insight for a team or recruiter. Our final labelled position would be a combination of these, for example 4_5_F would be a player in offensive cluster 4, defensive cluster 5, and a predicted Forward

Since we used model-based clustering and we calibrated the random forest in a certain manner, each player received a probability to be a certain position or in each cluster.

Clusters and Prototypes

Below are our offensive and defensive cluster with player prototypes for each cluster

Offensive Cluster 1 – Elfrid Payton (Fr)

Offensive Cluster 2 – Montrezl Harrel (Fr)

Offensive Cluster 3 – Kyle Anderson (Fr)

Offensive Cluster 4 – Tereke Eckwood (Sr)

Offensive Cluster 5 – Alex Len (Fr)

Offensive Cluster 6 – Devon Reed (Fr)

Offensive Cluster 7 – Kemba Walker (Jr)

Offensive Cluster 8 – Jimmer Fredette (Jr)

Offensive Cluster 9 – Joe Harris (So)

Defensive Cluster 1 – Reggie Jackson (Jr)

Defensive Cluster 2 – Marcus Smart (Fr)

Defensive Cluster 3 – Colin Sexton (Fr)

Defensive Cluster 4 – Anthony Davis (Fr)

Defensive Cluster 5 – Grayson Allen (Sr)

Defensive Cluster 6 – C.J. McCollum (Sr)

Defensive Cluster 7 – Tony Snell (Fr)

Defensive Cluster 8 – OG Anunoby (So)

Defensive Cluster 9 – Matisse Thybulle (Sr)

Creating the Ideal Roster

Every college basketball team's goal is to win March Madness. Hence, we attempted to use our defined positions to understand which rosters we predicted to do well in the tournament. To do this, we created team ratings based on the probability of its players to be within each cluster. For example, if a team has 5 players with qualified minutes and each of them have a 0.25 probability of falling within Offensive Cluster 1, then the team Offensive Cluster 1 rating is 1.25. We repeated this process for every offensive and defensive cluster to create 18 total cluster ratings for each team. Two logistic regressions were made for each round based on these ratings to determine which clusters were most significant towards winning a given round. Using the significant coefficients from these regressions, we predict the probability that a team will win each round of the tournament based on their offensive and defensive roster buildup. We then created a third regression for each round that predicts tournament win percentage based on the number of players on each roster that fell within each of our three-part positions. Therefore, each team has 3 win probability values for each round, one based on their offensive cluster rating, one based on their defensive cluster ratings, and the third based on their three-part position counts. We add the results of all 3 of these regressions for our final value to compare teams matched up in the tournament.

Average Round by Round Win Percentage
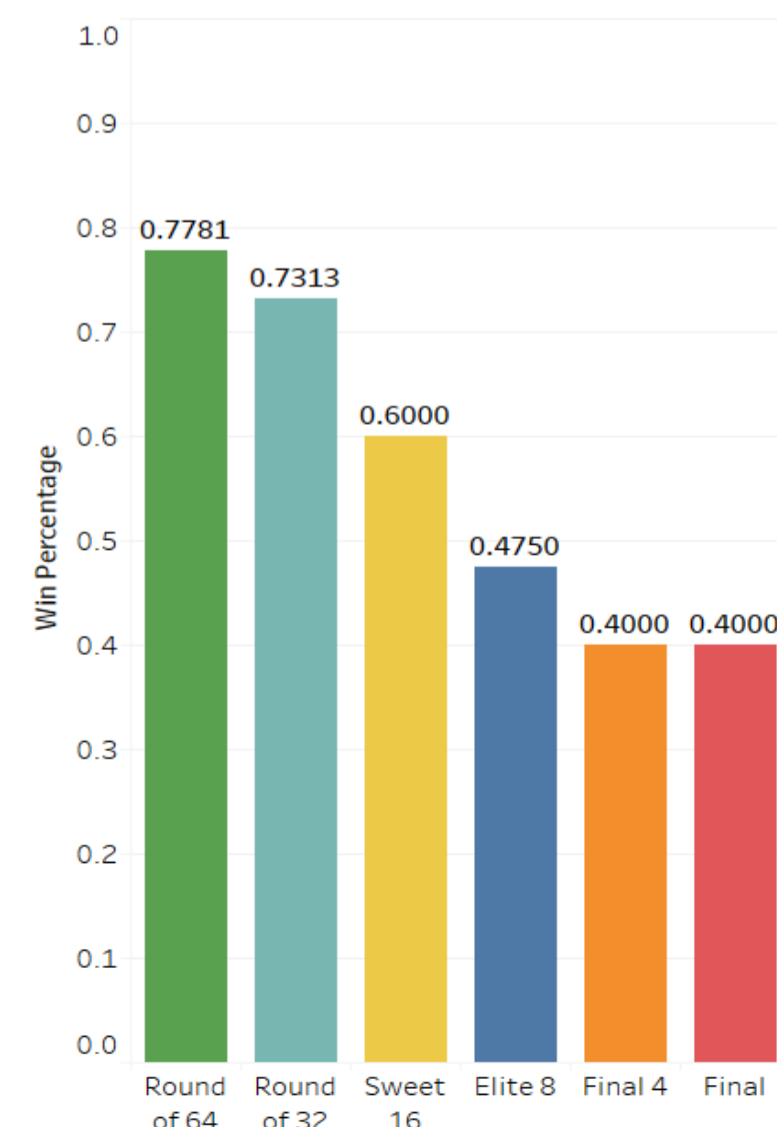
Notable Tournament Picks

(16) UMBC defeating (1) Virginia in 2018
- Only 16 seed we predicted to win

(15) Florida Gulf Coast defeating (2) Georgetown and (7) San Diego St in 2013

(7) UConn defeating (8) Kentucky in the Championship Game in 2014
- Highest sum of seeds in a Championship

Kushal Shah – kshah07@syr.edu
Colin Krantz – cmkrantz@syr.edu