Jackson Kennedy & Ian Weihe
2-25

An Analysis of 3-point Shooting Indicators in Collegiate Basketball

3-point shooting is undeniably essential to winning basketball games in 2024, but it's also scarce. Thus, professional teams, often through more of a scouting lens, are constantly searching for ways to identify players who may be better outside shooters than their college numbers suggest. Kawhi Leonard, Pascal Siakam, Karl-Anthony Towns… The list of players who didn't shoot well (or at all) in college yet developed into above-average NBA shooters goes on. The aim of this project is to develop a simple, straightforward process for identifying such players.

Once our data was collected, we began by applying the best subsets technique to find the most effective multiple linear regression model for predicting 3-point makes per game. Our goal here was to essentially find what non-3-point shooting stats are most correlated to (predictive of) 3-point shooting. The four-predictor model using free throw percentage, height (in.), turnovers, and two-point percentage emerged as the best model with an adjusted R-squared of 0.4818.

From there, we transitioned into principal component analysis. PCA seemed appropriate due to the fact that many of our predictors (features) are closely related. Free throws attempted, free throws made, and free throw percentage were all separate features during our multiple linear regression modeling process, for example. The hope was that we could express these groups of similar features as principal components, and thus account for the same amount of variability but with significantly less dimensions. Our PCA yielded three primary components — one which takes into account players' shooting profiles (i.e. how frequently and efficiently they shoot from different places on the floor), another which detects whether or not a player is a big man who shoots most of his shots at the rim, and a third which measures how often players turn the ball over. We then constructed a second regression model with these three components as its features. Somewhat surprisingly, the principal component model did *not* perform any better than the original MLR model, as evidenced by its higher RMSE (.568 > .528).

For the third and final part of our project, we decided to take a look at the studentized residuals of the four-predictor MLR model in order to 1) identify players who made significantly less threes than they should have according to our model, and then 2) find historical comparisons for these specific players. Using college data for 2021 and 2022 NBA Draft prospects, we examined the five largest negative studentized residuals for each draft class. Then, in the interest of time, we selected two observations to perform 'case studies' on — Josh Minott and Johnny Davis. These case studies did not consist of any actual *K*th Nearest Neighbor analysis, but the concept wasn't too different. In short, we made use of the Bart Torvik database, which consists of every division-1 college player since 2008, to filter for players with the most similar statistical profiles (i.e. nearest neighbors) who also had multi-year NBA careers. Then, we simply made note of how those players progressed as 3-point shooters in the NBA.

Our findings reveal free throw shooting to be the clear-cut best indicator of shooting ability, concurring with much of the literature we encountered beforehand. Additionally, we found that principal components are not necessarily 'better' than simple, unrelated additive predictors. If given more time, we'd like to explore the differences between partial least squares components and principal components; in retrospect, a more supervised technique (such as PLS) may have been better suited for our project, considering our goal was always to explain the most amount of variability in the response as possible.