Abstract

This paper investigates whether machine learning models can provide meaningful insights into the concept of "clutchness" for National Basketball Association (NBA) players. Shots late in close NBA games are widely considered higher pressure and more difficult. While fans and analysts frequently debate which players are the most "clutch" in these late game situations, the idea of "clutchness" has been notoriously difficult to define and measure, especially with traditional statistical methods. This paper aims to better determine "clutchness" by using machine learning models.

The main dataset used has raw shot-by-shot data for every shot attempt in the NBA over the last 19 seasons. For each shot the dataset specifies who was shooting, where and when the shot was taken, the type of shot (driving dunk, step back jump shot, etc.), the defense, and more. I also merged this data set with another that contained player statistics from previous seasons. We applied multiple machine learning models and techniques to train a predictive model with the ability to predict shot outcomes under these various conditions. This was the main focus of the project as it required extensive model tuning through trial and error.

Finally, to evaluate clutchness, this predictive model was applied while withholding information about the time and score of the game, effectively isolating player performance from contextual pressure. By comparing the model's expected outcomes to the actual results of the high-pressure shots, we can quantify how individual players perform relative to statistical expectations in late-game situations. This approach allows for a data-driven assessment of "clutchness," highlighting which players consistently exceed or fall short of predicted performance when the stakes are highest.

This research contributes to the current body of knowledge by applying machine learning to the often difficult-to-measure notion of "clutchness." There has been little publicly available research on this topic, especially research utilizing newer, evolving machine learning models.