



Social Basketball Analytics: Predicting Excitement

Team We The Hack: Raymond Chou, Alison Chow and Victor Denisov

Background

In basketball, a run is when a team scores consecutive unanswered points to surpass the opponent. This creates excitement for fans cheering on the team behind the run, while the opponents are at the edge of their seats. For instance, fans enjoy watching the Golden State Warriors go on “exciting” runs; however, it can devastate the opponent’s morale. The ability to predict an “exciting” run can benefit a team as it mitigates the potential onslaught before it occurs.

Our Hack

We classified a run as 8 unanswered points based on a play-by-play estimation, validated with historical box score data. Now, what’s an “exciting” run? Statistically, we can measure excitement with drastic fluctuations in win percentage since these deviate from average fan expectations.

In today’s age, social media is a barometer for fan sentiment. Since excitement is subjective, this can be a qualitative measure. As a result, we analyzed Twitter data during games.

Data and Process

As an initial starting point, we performed exploratory analysis of NBA play-by-play data and developed a linear regression model to predict win percentage at a 92.9% accuracy with Point Differential as the most significant feature.

We then used Breakout Detection, a machine learning algorithm that detects change in a given time series, to identify “exciting” highs and lows in win percentage on a play-by-play basis. Afterwards, we validated these results by calculating the sentiment, with a natural language processor, of a random sample of tweets at the apexes of the win percentages. Our model also interacts with the Twitter API to monitor spikes in tweet traffic during games.

Overall, we created an ensemble model, incorporating machine learning algorithms, such as linear regression and Breakout Detection to predict the occurrence of “exciting” runs.