

Springboard--Data Science Career Track

Capstone Project 1

What are the factors that contribute to a home team winning an NCAA basketball game?

By Adeyemi Adejuwon

February 2020

- **What is the problem you want to solve?**

Most people in sports have heard about the perceived advantage of teams playing at their home field. The goal of my project is to study the measurable factors that affect a home team winning an NCAA basketball game.

- **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**

The clients of my project are NCAA basketball fans, coaches, and teams. For instance, a coach of a basketball team can take into account factors that influence his team winning a game while playing at home, and use them to tune tactics for the game, and thereby taking advantage of these factors.

- **What data are you using? How will you acquire the data?**

I will be using the dataset, available from Kaggle¹, associated with an annual competition sponsored by Google.

The dataset is comprised of about 350 college basketball teams playing in the NCAA since 1985, and contains 82,000 data points consisting of 30 variables such as number of assists, three-point percentages per game, win and loss records per season, location of matches played, etc.

- **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

My current vision is to model the business problem as a classification data science problem with the goal of building machine learning models to estimate the probability that a home team will win a game. I will be using features such as assists, three-point percentages, location of matches (home vs. away), to create such classification models. I am planning on using parametric and non-parametric models, studying their relative performance.

¹ <https://www.kaggle.com/ncaa/ncaa-basketball>

- **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**

As required by the program, I will submit the following deliverables upon completion of this project: all Jupyter notebooks, a written final report, and a presentation slide deck.