

IDS 702 Final Project Proposal

How to Win NBA Betting

Leon Zhang

Due date: Oct 21st 2020

1 Overview

In 2019, the National Basketball Associations, NBA, announced to legalize sports betting across United States. This brings the field of data science with tremendous opportunities in the sports betting market. There are many advantages in using statistical modeling in sporting betting because it offers a quantitative insight of expected returns. For this project, we will find out what factors is worth taking into considerations when placing NBA bets. There are three common types on bet: moneyline (which team wins the game), spread (how many points a team will win or lose by), and over/under totals (the total points scored by both teams). To narrow down the scope, this analysis will be only focusing on the moneyline bets.

2 Research Questions

The overall goal of this project is to explore the best set of predictors a bettor should consider when placing moneyline bets. The predictors are the quantifiable information we have for both teams before they play against each other. In particular, I've outlined three questions that I believe is interesting to investigate on:

1. Can we quantify the effect of winning/losing streak on a team's chance of winning?
2. To what degree does "back to back" games influence a team's winning outcome?
3. What in-game statistics affects a team's odds of winning the most?

3 Data

It was difficult to find a data set that includes all the information I desire, so I created my own data set from multiple sources. The final columns of the dataset are listed with description below:

Game: The i th game of a particular season

Team: The team which played the game

Opponent: The opponent which the team played against

Date: Date of the game which was played

Home: Is the game is played on the team's home court? 1 for yes

is_b2b: Is this game a back to back game for the team? 1 for yes

is_b2b_opp: Is this game a back to back game for the team's opponent? 1 for yes

streak: Current winning/losing streak of the team. >0 for winning; <0 for losing.

opp_streak: Current winning/losing streak of the opponent. >0 for winning; <0 for losing.

record: Total wins - total loses of the team.

opp_record: Total wins - total loses of the opponent.

And all the cumulative average game stats for each team, such as total score, rebound, assist, etc..

4 Project plan

I plan on exploring multilevel logistic regression model for grouping each NBA team.

Milestones:

1. Prepare dataset: Done)
2. Complete EDA: Sunday, Oct 25th
3. Model building and check for assumptions: Sunday, Nov 1st
4. Interpret result: Sunday, Nov 8th
5. Complete presentation: Tuesday, November, 17th
6. Finish report: Tuesday, November 24th