Predict results of NBA games

Michał Kubica

1 Introduction

Nowadays, data science and machine learning is used to solve all kinds of problems. We try to apply it everywhere we can and wherever it does make sense or money. At universities, in buisness and in sports.

As we know, Americans love baseball, football and basketball. But also they love numbers. That's why they were collecting data for years about major sports leagues such as NBA, NFL and MLB. They have stored every possible statistic, which was possible to measure. Now it is time to use it.

2 Goal

The goal of this project is to build the best model for predicting NBA games results using methods from the course.

3 Data

The data about NBA games is available for free. There should no problem with collecting needed datasets.

4 Problems to face

For sure there will be a lot of problems, few examples:

- defining target value,

- small number of observations - in NBA there are 30 teams, every team plays about 80 games per season,

- in NBA rotation of players is huge. So building model, which is based on previous season results is questionable,

- huge number of predictors

- there will be many predictors correlated with each other: total points, attempts, assists, rebounds . . .

- injuries - injury od key player is a very important information to predict results - how to include such an information in a model?,

- in one season the teams meet with each other maximum 4 times - there will be need to somehow estimate power for each team