NBA Home Court Advantage

Edvard Notberg, Timo Kaasik, Villem Paabo

Introduction

Basketball is a very known game all around the world and it offers lots of enjoyment to anyone who follows it. The biggest organisation for it is the NBA (National Basketball Association). In NBA, there can only be one winner. This opens up a dialogue, to try and guess or reason the odds of one team winning or losing.

Data

The dataset, containing data of 2314 games played over 2 seasons, comes from Advanced Sports Analytics. Every game has 55 attributes for it, out of which we used 16.

Used methods

Preprocessing

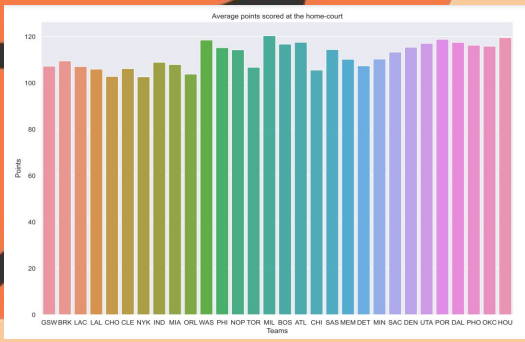
A lot of columns in the data were duplicated due to player stats which we removed. We also had to add a column with the winning teams abbreviation because of the weird styling of the original csv file. The nominal attributes were encoded using 1-hot vectors.

Training and evaluating models

To train and evaluate models, the dataset was split into training (80%) and test set (20%).
We trained our models using different sets of features and used four different supervised learning algorithms. The algorithms were KNN classifier & regression, GaussianNB and Linear regression models. The average accuracy of the models are below..

Conclusion

We were able to replicate and prove the significance of the home court advantage.
We were able to quite accurately predict the winner of a game, given the playing teams and the home court.
Also, we will use our models to predict the games that took place at the start of this week (after the deadline).



ML models

GaussianNB	14.0%
KNNR	7.0%
LRM	60.9%
KNNC	50.1%