Final Project Data Memo

Albert Li

April 22, 2018

Data Source

The dataset I obtained for the final project involves the NBA shot logs during the 2014-2015 regular season. This dataset was obtained from Kaggle.com. The data within this dataset includes every shot attempt made during the course of the 2014-2015 season and includes useful information regarding the shot attempt, such as the player taking the shot, the distance from the basket, the nearest defender, time left on the shot clock, and more. The dataset includes 128,056 rows with 21 fields. At this point of time, I do not anticipate joining this current dataset with any external datasets.

Research Questions

The main goal for this research project is to create a model that will accurately predict if a shot taken in an NBA regular season game in the 2014-2015 season will make or miss. I am also interested in which fields are most important in determining whether or not a shot attempt will make or miss. In this case, my problem is a classification problem. Although I am able to create a model to classify a shot as a make or miss, interpretation of the model is limited to only the 2014-2015 NBA season, as players tend to increase or decrease in skill through time.

Potential Data Issues

Looking at my dataset, some potential issues involve formatting the dataset. One field called “MATCHUP” is a character type field which contains the date of the matchup and the teams that are playing. I will need to extract the characters of the opposing team if I want to include the opposing team as a potential field. Another field I will need to adjust is the “GAME_CLOCK” field, which is formatted in minutes and seconds left in the quarter. I will need to figure out a method to quantify this field.

Citation

DanB. (2016). *NBA shot logs*. Retrieved from https://www.kaggle.com/dansbecker/nba-shot-logs (https://www.kaggle.com/dansbecker/nba-shot-logs)