

Introduction

The topic of our team final project is predicting which college basketball team would be the champion of NCAA 2015 tournament. We choose this topic because all the members in our team are energetic people and we are interested in sports. Two of them are very familiar with basketball. Besides, Duke Men's basketball team is very popular and has won the NCAA 2015 tournament championship. Everyone in Duke feels proud of them, so do we. Therefore, when we found the data set of NCAA basketball to predict 2015 tournament champion, we decided the topic immediately. That is what we want for our final project, applying what we have learned to our interested field. In addition, this data set once was provided for a official competition, thus the data would be reliable and rich enough for doing prediction. Actually, there are more than 5000 pieces of match data for each year and 34 columns of many kinds of information for each match including 26 technical skill indexes, which are useful for us to compare the performance of each team. Except for the detailed data of regular season for 2015, there are also detailed data of both regular season and tournament for 2003-2014, which can be helpful if we want to consider or compare the performance of past years. With this data set, we can use many methods that we have learned from the class, such as regression and LASSO. Therefore, it is an appropriate data set for the final project.