As early as professional sports had occurred, people started to bet money on the results. With the arrival of digital age and the increasingly expanded market, people try to reach the final outcome as close as they can, whether for gambling or simply having something to talk about with their pals. It is widely accepted that accurate prediction is much easier in professional leagues, such as National Basketball Association(NBA) and National Football League(NFL), than that in student sports league, such as the National College Athletics Association(NCAA), for there are fewer teams and more matches and the media pay more attention and even rate them in virtual games which value a sense of reality. Meanwhile in the student sports leagues, only those that traditionally stay competitive stay competitive. For other teams, people don’t know much about them, therefore have no idea which team is better. In this paper, I will provide a rather credible model to estimate the result of NCAA basketball teams in their journey in March Madness.

Here is a brief introduction about NCAA and March Madness. The March Madness is like the playoffs of a college basketball season. It is a tournament that involves 68 college teams that are champions of conferences in Division I and the remaining teams chosen by the tournament selection committee. In the first round, eight teams play in pairs to compete for four at-large seats called “First Four”. Then the 64 teams will be divided into four regions and in each region they will be seeded from #1 to #16. Within each region, the 16 teams play a 4-round single- elimination tournament with matchups determined by seed (1 vs. 16, 2 vs. 15, etc.); the winner of each region goes to the Final Four. And at last, a 2-round single elimination tournament is played to decide the championship. It is important to know that every game is played at a neutral site. Unlike in regular season, no team has so-called “Home Court Advantage”.

In this model, we will use the statistics of each team during its regular season to predict the result of that particular team in the March Madness. The data that is used includes box scores as well as players’ personal statistics like assists, rebounds, steals, blocks and field goal percent.