James Bloor

Boston University Metropolitan College

MET CS 677

Project Write Up

Fall 2 2021

BU ID: U45448564

For my project topic I am going to look at College Football statistics. More specifically, I want to look at the top teams each year and see how they compare to other teams that were not ranked as high in the College Football Playoff committee rankings. Each year there are huge debates on non-Power Five teams competing to be put into the College Football Playoffs. Year over year, these undefeated “outliers” aren’t given a chance in the playoffs. I am going to look at how these teams look on paper and take the human bias out of the decision. Looking at a team’s final season stats, who really stands out? I will use a KNN method that we learned in this class to help me decide who the top teams really are and if the playoff committee got their decisions right based purply on statistical background.

In the 2017 College Football season, UCF went undefeated through regular season play. At the end of the year rankings, they were slotted in at 12th in the nation. While UCF is not in a Power Five Conference, I believe going undefeated is an achievement that should have gotten UCF into the College Football Playoff or at least closer in the final CFP rankings. The College Football Playoff or CFP was first introduced in the 2014 season where the top four national ranked teams are placed in a small elimination-based tournament at the end of the season. With the introduction of these playoffs came a new method of ranking college teams, The College Football Playoff Committee. A group of 13 individuals who convene weekly and rank the top 25 college football teams during the season. While the committee members talk amongst themselves and are allowed use various methods to determine the ranking of teams, there is no requirement that they must use any certain mathematical of statistical methods when deciding their top 25. Here is where I believe human bias comes into play.

It does not take a genius to know that Alabama is a force in college football and that Georgia has the best defense in the nation; however, I want to look if ‘bigger’ name teams receive an unfair advance when being considered for the CFP. Using a KNN method to classify college football teams, human bias can be moved out of the equation, and we will be able to see what teams stand out amongst the rest.

2017 is the season I want to take a dive into first. With 2014-2016 being playoff years I will be using the data from these years to train my classifier. I was able to find a great database with college football data at this link: <https://www.kaggle.com/jeffgallini/college-football-team-stats-2019> . With there being dozens and dozens of metrics to choose from, I chose my features in my python code that you can view in this zipped folder. I also created my own metrics that was not included in the data, ‘Winning Percentage.’ Not every college football team plays the same number of games. With this, some teams can have more wins than another but a lesser winning percentage. To ensure that this data column does not give a team an unfair advantage, I took out the wins and loses columns and replaced it with a winning percentage column. This creates an even playing field for all the teams we will be looking at.

For the rankings of college football teams, I created four buckets, CFP (teams ranked 1-4), OLI (teams ranked 5-10), Top 25 (teams ranked 10-25), NR (teams that were unranked). I added these buckets to each dataset that I used in my data to train the team rankings and then compare my results when needed. Next, I wanted to look at what would be the most accuracy number of neighbors for my classifier. After running a quick for loop, I saw that seven neighbors were my highest accuracy.

After training my classifier on 2014-2016 data I predicted the 2017 CFP teams, my results are as follows. My prediction was that Alabama, Clemson, Georgia, Ohio St., Penn St., TCU, Washington, and Wisconsin have a resume to get into the playoffs. The teams that were ranked in the top four by the committee were Alabama, Clemson, Georgia, and Oklahoma. Oklahoma was the only team to be in the playoffs but not make it into my 2017 prediction. It should also be recognized that even though UCF had an undefeated season in 2017, they still did not make the playoffs in my classifier. These results can show that there isn’t much president for a 12-0, non-Power five team to make the playoffs.

While analyzing my results, I wondered what would happen if I did the same steps as above for the 2018 season. I can retrain my classifier from 2014-2017 data and see what 2018 was predicting. Nicely enough, UCF also went undefeated in the 2018 season and ended up being ranked 8th in the nation. By running the numbers again, I could see if UCF made it into the playoffs this go around and who stood out in the nation.

For my 2018 predictions I came up with one team predicted in the CFP bucket, Clemson. Even though these were four undefeated teams in the nation (three of them make the playoffs) my predictor saw Clemson as the one team to make the playoffs. Again, UCF does not have the statistics to make into in the playoffs without human bias.

Overall, I am pleased with the results of my KNN classifier. While the results are not exactly what I anticipated, I find it interesting who stands out on a given college football season by just looking at the numbers. Using methods to train and test data can help predict the future and find results that the human world is looking for. Take the human bias out of college football is a large and daunting task. With the rich history of this sport and the zealots that come with it, I do not know how they would take incorporating some mathematics or statistics into their rankings of teams. If there is one thing I know for sure, it is that Nick Saban and the Crimson Tide are always bound for a CFP appearance.