Predicting Teams Best Suited for Participation in the College Football Playoff

Richard E. Molina
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556
rmolina1@nd.edu

ABSTRACT

Over the past 2 decades, college football has been plagued by controversy over how it determines its national champion. College football recently attempted a method based on computer algorithms in determining their champion with the Bowl Championship Series (BCS) system, but there were many flaws with it. Now, with the onset of the College Football Playoff, the same questions and problems still remain. Given the number of teams that play division 1 football, the constraints of academic concerns and time spent away from the classroom, as well as player safety, it is virtually impossible to have a larger playoff similar to the NFL. As a result, the need to find a fair and impartial way in which 4 teams should be represented in the playoffs needs to be discovered.

Keywords

Data Mining, Sports Analytics, College Football.

1. INTRODUCTION

Over the past 2 decades, college football has been plagued by controversy over how it determines its national champion. Prior to the 1998 season, there was no system in place to determine a nation champion other than poll voting among the Associated Press and Coaches polls. However, quite often there was a difference in the final results of two polls, which would result in what was termed among the sporting community as a "split national champion." The idea of a split championship was inconceivable for fans, especially where, prior to 1996, a game could end in a tie. Fans had enough of ties, and demanded an outright champion be crowned.

Beginning with the 1998 season, the NCAA adopted a new system where the top two teams in their rankings would be matched up against one another in a game to determine a national champion. This system was known as the Bowl Championship Series [1], or more commonly referred to as the BCS. This system relied on a combination of the two polls mentioned earlier and a combination of computer algorithms to rank the top teams in the country. The top two teams in the final rankings would then be matched up in a championship game and the two would play for the national title. This was a good idea in principle, but it had seemingly become an annual ritual among college football fans to lament and voice their displeasure over the top two teams chosen for the championship game. Quite often over the course of this period of college football, there were more than two teams who had an argument to be included in the championship game. This would result in an argument among experts as to if the system got it correct or not. Clamoring for a football playoff system began to gain momentum around 2006 and every year gained more support as more and more controversy occurred.

The fans finally got what they asked for with the 2014 season, as the old BCS system was abandoned and the new College Football Playoff system [2] was adopted. A committee of 13 members would collaborate and vote on the top teams and rank them accordingly by only taking into consideration what is seen on the field. The top four teams would then participate in a seeded playoff to determine a national champion.

Though it has yet to be seen how this system will perform compared to its predecessor, the same fundamental issues and problems exist: how do we determine the teams that are most deserving of a spot in the 4 team playoff? Given the number of teams that play division 1 football, the constraints of academic concerns and time spent away from the classroom, as well as player safety, it is virtually impossible to have a larger playoff similar to the NFL. As a result, it becomes imperative that the committee accurately identifies the four most deserving teams in a fair and impartial way.

This work aims to identify the key attributes of the best teams of season's prior and attempt to solve one of the age-old questions of college football: "Who is the best team in the country?"

2. RELATED WORK

Given the unpredictable nature of the sport and, until recently, lack of a playoff system, college football has been a prime research focus in the area of sports analytics. Researchers have attempted to use data mining techniques to predominantly be able to predict game outcomes in the sport.

Work done in [5] utilized historical statistic data to develop classification and regression type models to assess the ability of these methodologies to accurately predict the outcomes of games. This work had promising results in that they showed that classification-type models predict game outcomes better than regression-type models. Their work also showed that decision trees produced better results (over 85% precision accuracy) over artificial neural networks and support vector machines. This study utilized 28 different features ranging from points and yards scored/allowed per game to the percentage of games played by upper- and lower- classmen on the team. Of the 28 features, this study found that the two most important in accurately predicting who would win a bowl game on a neutral field were non-

conference team winning percentage and average margin of victory.

Likewise, the work done in [6] also focuses on predicting winners of college football bowl games. However, their approach differs by basing their predictions on historical results of games rather than a direct comparison of statistics of the two teams that are scheduled to compete. Essentially, the teams being considered are compared to other teams in order to find comparable teams statistically. Once identified, game data and results are pulled from these teams and utilized in their prediction algorithm. There were four key features that they focused their algorithm on: RPI, Pythagorean wins, Offensive strategy, and turnover differential. The results from their approach show relatively high accuracy in correct result prediction from the bowl games (32 correct out of 35 games; 91.43%).

Finally, [7] essentially conduced a survey of various data mining techniques and how they perform in regards to game result prediction. The various techniques looked in to were ANN, decision trees, Bayesian method, logistic regression, SVM, and fuzzy methods. They detected two major challenges: 1) low prediction accuracy and 2) the lack of a general and comprehensive set of statistics forces researchers to collect data from a variety of different sports websites/databases. Not being able to utilize the same datasets across a variety of different methods make the comparison of different experiments and results inconclusive. They also suggest some solutions to these problems, such as improving prediction accuracy through the use of machine learning and data mining techniques and the utilization of hybrid algorithms. They also claim that including more features can help contribute to more accurate predictions.

3. DATA

Data was pulled and collected from [3] with the help of the import.io tool. Historical data from 2004 to present day statistics have been collected. ESPN has their own Stats & Information Department that has already verified and corrected any discrepancies in the statistics they have and those on record with the NCAA. As a result, data correctness is not an issue. There is no missing data and, as mentioned earlier, much of the discrepancies that may have existed have already been corrected by ESPNs Stats & Information Department. Some of the statistics collected from ESPN include total team offensive and defensive statistics, team rushing and passing statistics, overall record, and the final rankings of the USA Today and Associated Press polls. In addition to these seasonal statistics, I was also able to obtain individual game statistics from [4] which was used to calculate the strength of schedule feature in the algorithm. This data includes game scores, location of the game, number of turnovers, and total yards gained/allowed during the game. The data from the current season (2014) will be used as testing data while the data collected from 2004 to 2013 will serve as the training data.

4. FEATURE SELECTION

Before I could construct the model and ranking algorithm discussed in Section 5, key statistical categories first needed to be identified that corresponded to the top teams in the country. Several statistical categories were looked at ranging from yards gained/allowed, pass completion percentage, average rushing yards per attempt, as well as numerous others. Out of all the statistical categories considered, the following were determined to be the most statistically significant and correlated well to the top teams:

Winning Percentage

Of all the statistical categories, this represents the most obvious selection. A team's winning percentage is the numerical representation of the number of games the team has won.

$$Win\% = \frac{Wins}{Games\ Played}$$

For example, if team A has won 9 games and lost 3, then the teams winning percentage would be $9/(9+3)$, which is equivalent to a winning percentage of 0.750.

Points Scored/Allowed

Points scored are the average number of points scored by the team per game. Points allowed are the number of points scored against the team by their opponents per game.

Margin of Victory

Margin of Victory is the average difference between the number of points the team scored and the number of points their opponent scored. If a team lost a game, their margin of victory for that particular game would be negative. As a result, this would negatively impact the teams overall margin of victory.

$$MOV = \frac{\sum(Pts\ Scored - Pts.\ Allowed)}{Games\ Played}$$

For example, let's assume that Team B played 3 games and won game 1 by 10 and game 2 by 21 points. Let's also say that Team B lost game 3 by 10 points as well. This would mean that their average margin of victory would be equal to $(10 + 21 + (-10))/3 = 21/3$. This results in an average margin of victory for team B of 7.0.

Strength of Schedule

Strength of Schedule is perhaps the best measurement of the quality of a team's schedule. This feature not only measures the records of the teams it has played, but also takes into account the opponents of the opponents they've faced. A team that plays opponents who have a high winning percentage would have a higher strength of schedule than a team that plays against weaker teams with a lower winning percentage.

$$SOS = \frac{2(owp) + oowp}{3}$$

In the above equation, owp is the team's opponent's winning percentage and oowp is the opponents' opponents winning percentage. The range of acceptable strength of schedules is between 0.0 and 1.0, inclusive.

Strength of Victory/Defeat

Strength of Victory is similar to strength of schedule except that it only takes into consideration the opponent's winning percentage and the opponents' opponents winning percentage of teams whom they have beaten. Teams that have not won a game will have Strength of Victory of 0.0. Strength of Defeat keeps track of these same features for teams that the team has lost to. Teams that have not lost a game will have Strength of Defeat of 1.0. Overall, the range of this statistical feature will be in the range of 0-1.

Quality Wins/Bad Losses

For the purposes of this work, a quality win is defined as a win on the road over a team who's winning percentage is over 0.7. In contrast to this, A bad loss is one in which a team loses a game at

home against a team with a winning percentage less than 0.5. Quality wins and bad losses are factored into strength and victory and strength of defeat features.

Strength of Conference

Strength of Conference is the numerical representation of how teams in a conference performed in games played outside of the conference. Essentially, it is the average winning percentage of all teams in the conference in out-of-conference games. For this feature, we only consider out-of-conference games since games played in conference would basically be adding one win and one loss every game. Similar to strength of schedule, the range of acceptable strength of conference is between 0.0 and 1.0, inclusive.

Conference Championship

This feature is simply a nominal value that describes if a team won it's conference championship outright or split the championship with another team. A team is considered an outright conference champion if it wins a conference championship game hosted by the conference or is in sole possession of the best record in games played among conference members. A team is considered a co-champion if it shares the best conference record with one or more other team(s).

Most of these features were identified because they are commonly associated with top teams. Statistical categories like winning percentage, strength of schedule, points scored/allowed, and conference championships have historically been associated with good to great college football teams. Also, prior to the start of the season, the playoff selection committee stated that conference championships would be rewarded in their rankings.

5. METHOD

To accomplish the proposed goals of this project, a two-part approach was utilized in determining the best teams in the country. The first approach was a decision tree model that is utilized as a filter to remove teams from consideration to improve algorithm performance. The second approach utilizes an algorithm that calculates seasonal performance and ranks the teams accordingly.

5.1 Decision Tree Filter

As of 2014, there are currently 128 schools eligible for participation in the NCAA Division 1 FBS level and that number increases every year. Given the large number of schools to evaluate, we need a way to reduce the number of teams from consideration. To do this, we will utilize a decision tree model to conduct this reduction of the dataset.

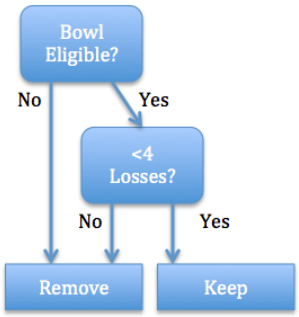

Figure 5.1 – Decision Tree Filtering Model

Since the goal of this project is to identify the top 4 teams in the country, we can eliminate any teams right away that are not eligible for postseason play under NCAA rules. This would eliminate any teams that have a losing record (winning percentage < .5) as well as any that have been sanctioned by the NCAA and forbidden from postseason competition. Furthermore, utilizing a binning technique, I was able to see that in every year in the testing set, mostly all of the teams considered to be the top 25 teams in the country did not have more than 4 losses (Figure 5.2). If we limit that to only the top 10 teams, then they have no more than 3 losses on the season. As a result, for our purposes, we can eliminate any teams from consideration that have a winning percentage less than 50% and more than 4 losses.

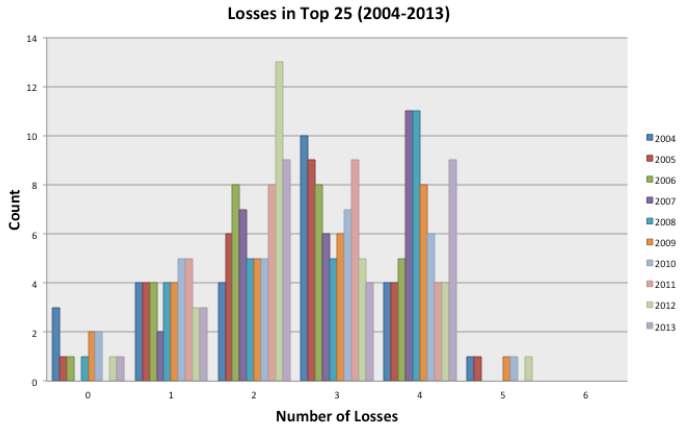

Figure 5.2 – Losses by teams in the top 25 final rankings (2004-2013)

Using this approach reduces the amount of teams in consideration substantially. On average, we see a reduction from approximately 125 teams to about 30, representing a reduction of approximately 75% of the dataset.

5.2 Ranking Algorithm

With these remaining teams in consideration, we want to be able to rank them according to an algorithm based off of the historical data we collected from 2004-2013. The trick was being able to identify which features mentioned in section 4 or combination of features can correctly identify the top teams in the country. Utilizing the training data and experimentation with several different algorithms, I found an algorithm that performs very well when compared with the final rankings of the playoff selection committee.

The chosen algorithm for this application is broken up into 4 different factors: Points ratio, Strength of Margin of Victory, Strength of Record, and Conference Factor.

Points Ratio

The Points Ratio factor of the ranking algorithm utilizes the points scored/allowed features. This metric is simple to calculate as it is simply a ratio of points scored and points allowed by the team:

$$PR = \alpha \cdot \frac{Pts.Scored}{Pts.Allowed}$$

In the equation, α is a multiplier determined by analysis done on the historical data. This analysis revealed that α provided the most precise rankings when it was set to 1.75.

Strength of Margin of Victory

Strength of Margin of Victory rivals the simplicity of the Points Ratio factor since it is merely the product of the team's strength of schedule and margin of victory.

$$SMV = SOS \times MOV$$

Utilizing these two features in this way helps to reduce the bias of "running up the score" on weaker opponents. Consider the following example: Team A has a margin of victory of only 10 points while Team B has a much larger margin of victory of 32 points. However, Team A has a high strength of schedule at 0.8 while Team B had a really weak schedule of 0.25. Utilizing the strength of margin of victory factor, both teams will be perceived as equal in the eyes of this metric:

$$Team\ A = 0.8 * 10 = 8$$

$$Team\ B = 0.25 * 32 = 8$$

This factor does not reward teams for winning by a large margin if they are playing a weak schedule.

Strength of Record

The Strength of Record factor takes into account strength of victory and strength of defeat as well as the record of the team (number of wins and losses). The strength of record factor is calculated as follows:

$$SOR = 2(SOV * Wins) - \left(\frac{1}{SOD} * Losses\right)$$

This factor multiplies the number of wins a team has by the strength of their victories and also by a factor of 2. Similarly, the number of losses by a team is multiplied by the inverse of its strength of defeat. The difference is then taken between the two components. The inverse of the strength of defeat is taken because we do not wish to penalize teams that have lost to strong teams. The higher the team's strength of defeat, the lower the portion taken out of the strength of victory component will be and vice versa.

The idea behind the strength of record factor was to dive deeper into a teams record and evaluate the quality of their wins and losses. Unfortunately, in college football, not all teams with 9-3 records are equal. If they were, it would make evaluation of them very simple. A team can be a weak 9-3 as a result of them playing a weak schedule, an average 9-3 by beating weak teams while losing to above average teams, or a strong 9-3 by beating some quality teams and narrowly losing close games against excellent teams. As a result, the strength of record factor is needed to determine the true value of a team's record.

Conference Factor

The conference factor is taken by utilizing the strength of conference feature as well as the conference championship feature. However, because the conference championship is a nominal value, I needed a way to quantify this feature numerically in a way that rewarded the champion in a fair manner. To do this, the following formula was utilized:

$$Conf\ factor = \alpha * Strength\ of\ Conf$$

Where α is equal to 5 if a team is the outright conference champion, 3 if the team split the conference championship with 1 or more other team(s), or 1 otherwise. This formula rewards teams that win their conference in a manner that is equivalent to how strong their conference is. This way a team that wins an outright conference championship in which their conference had an out-of-conference winning percentage around 0.80 is rewarded more than a co-conference champion who's conference managed to only have an out-of-conference winning percentage of 0.55. It is worth noting that for teams that are independent of any conference affiliation, their strength of schedule is substituted in place of strength of conference for this factor.

Utilizing these four portions of the ranking algorithm, we can then calculate a performance score for the team by summing these four factors together:

$$Score = PR + SMV + SOR + Conf\ factor$$

Once these scores are calculated for every team in the reduced dataset, they are then ranked in order from greatest to least. The resulting list represents the final performance rankings for the teams. As a result, the top four teams with the highest scores can be perceived to be the best four teams for the season and advance to the college football playoff.

6. RESULTS & EVALUATION

To evaluate the performance of this method, a comparison was made between the rankings my algorithm produced and the final rankings released by the Playoff Committee. The top 10 for both can be seen in Figure 6.1 below.

	Playoff Committee	My Rankings (score)
1	Alabama	Oregon (19.661)
2	Oregon	Alabama (19.181)
3	Florida State	Florida State (16.942)
4	Ohio State	Ohio State (16.488)
5	Baylor	TCU (16.085)
6	TCU	Baylor (12.774)
7	Mississippi State	Michigan State (11.316)
8	Michigan State	Boise State (10.982)
9	Mississippi	Marshall (9.894)
10	Arizona	Mississippi State (9.514)

Figure 6.1 – Final top 10 rankings from Playoff Committee and the Ranking Algorithm described in this work

My algorithm was able to correctly pick the same top 6 teams as the committee did within one spot as well as 8 of the top 10 teams (Mississippi and Arizona were ranked 11th and 14th, respectively, in my rankings). The main differences between my top 10 rankings and those of the playoff committee were at the top between Oregon and Alabama, the 5 and 6 teams of TCU and Baylor, and the inclusion of Boise State and Marshall in my top 10 at the expense of Mississippi and Arizona.

Oregon and Alabama each had outstanding seasons as any college football analyst can attest to. Both are equally deserving of the top spot in the rankings and the close proximity in scores between the two in my rankings conveys that.

The 4,5,6 teams in the playoff committee were a hot topic of discussion leading up to the release of the selection committee's final rankings. Each team of Ohio State, Baylor, and TCU were

well deserving of the fourth and final playoff spot. Ultimately, the deciding factor came down to the conference factor of the algorithm. Due to its outright Big Ten Conference Championship, Ohio State received the boost it needed to make it into the top 4 over the Big 12 Conference co-champions of TCU and Baylor. It is also worth noting that, even though Baylor beat TCU head-to-head during the season, TCU ultimately had a stronger strength of record (2.629 vs. 0.662) as well as marginally better points ratio (4.034 vs. 3.528) and strength of margin of victory (6.489 vs. 5.649). This is why TCU is ahead of Baylor in my rankings.

Boise State and Marshall, two teams from outside of the Power Five[1] conferences, we able to crack the top 10 in my rankings at 8 and 9, respectively. While ultimately well outside of the top 4, their inclusion in the top 10 shows how undervalued teams from outside the Power Five Conferences are. The committee had Boise State (11-2 Mountain West Conference Champion) ranked 20th in their rankings behind a 4 loss Auburn team. Marshall (12-1 Conference USA Champion) went unranked in the committee's rankings even though their 1 loss was only by 1 point. Overall, I believe the selection committee has a bias towards the Power Five conference members which results in unfair exclusion towards teams from the other conferences.

A full detailed comparison of the top 25 teams in both rankings as well as a linear correlation of the ranking algorithm described in this work to the playoff selection committee can be seen below.

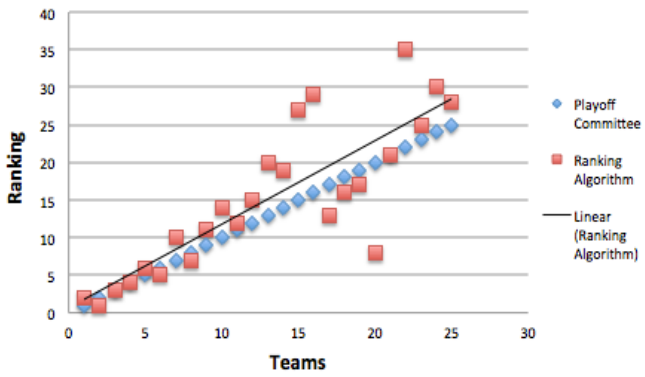

Figure 6.2 – Comparison of playoff committee rankings to ranking algorithm with a linear fit

7. CONCLUSION

Overall, this work successfully completed the goals that were set out at the beginning of the work. This work was able to successfully identify key features and attributes that are most indicative of the top teams in college football. The rankings produced by the algorithm in this work were in direct agreement with those in the committee with 100% accuracy of the top 6 +/- 1 spot and 80% agreement of the top 10 teams. Overall, even though conference championships were supposed to be rewarded in the playoff committee's rankings, they seem to only reward teams from inside the 'Power Five' Conferences. While analysts and apparently committee members often perceive teams from outside of the group of conferences as inferior, their performance on the field is what should matter. The algorithm described in this work removes that bias and stigma from non-Power Five conference teams and evaluates them on level footing with those perceived stronger teams.

[1] Power Five Conferences is a term used when referring to the group of Big Ten, Pacific 12, Big 12, Southeastern (SEC), and Atlantic Coast (ACC) conferences

8. FUTURE WORK

I would like to expand this work to eventually factor in head-to-head game results into the algorithm if two teams are very close in the rankings. Also, other potential statistical factors like Turnover Differential and Offensive strategy (pass/rush attempts ratios) and how efficient they are in their game plan could play a roll in determining the top teams.

9. ACKNOWLEDGMENTS

Special thanks goes out to Dr. Nitesh Chawla for providing the knowledge necessary to complete this work as well as for the freedom to pursue a topic in a field I find very interesting. Special Thanks also goes out to ESPN.com and Warren Repole with Sunshine Forecast Enterprises for allowing the use of their data in this project.

REFERENCES

[1] Bowl Championship Series Homepage. http://www.bcsfootball.org

[2] College Football Playoff Homepage. http://www.collegefootballplayoff.com

[3] ESPN: The Worldwide Leader in Sports. http://espn.go.com/college-football/statistics

[4] Warren Repole. *Sunshine Forecast Enterprises*. http://www.repole.com/sun4cast/data.html

[5] Delen, D., Cogdell, D., Kasap, N. *A Comparative Analysis of Data Mining Methods in Predicting NCAA Bowl Outcomes.* International Journal of Forecasting, Volume 28, Issue 2, Pages 543-552.

[6] Leung, C.K., Joseph, K.W. *Sports Data Mining: Predicting Results for the College Football Games.* Procedia Computer Science, Volume 35, 2014, Pages 710-719.

[7] Haghighat, M., Rastegari, H., Nourafza, N. *A Review of Data Mining Techniques for Result Prediction in Sports.* Advances in Computer Science: an International Journal, Volume 2, Issue 5, No.6, November 2013.

[8] Schumaker, R., Solieman, O., Chen, H. *Sports Data Mining*. New York, NY: Springer; 2010.

[9] McCabe, A., Travathan, J. *Artificial Intelligence in Sports Prediction.* IEEE Computer Society. Washington, DC, USA, 2008, pp. 1194-1197

[10] Seif El-Nasr, M., Drachen, A., Canossa, A., eds. *Game analytics: maximizing the value of player data.* London, UK: Springer; 2013

[11] Fainmesser, I., Fershtman, C., & Gandal, N. *A consistent weighted ranking scheme with an application to NCAA college football rankings*. 2009. CEPR discussion papers 5239

[12] Farlow, S. J. *A computer program for ranking athletic teams*. International Journal of Mathematical Education in Science and Technology. 1984. 15(6), 697–702.