Michael Loos

We're Number 1!

The champion of college football has often been a source of controversy. The first Associated Press (AP) Poll was published in 1936 (Brown). One year, the AP declared a three-way tie for national champion, leading to much dissatisfaction and protest from fans of each team. For much of college football history, multiple polls were used to rank teams throughout the season and select a champion at the end of each season. On several occasions, these polls had different number one teams, and a champion was crowned by each poll. The process eventually evolved to a computer ranking system, the Bowl Championship Series (BCS) rankings, for several years until a new system was put in place, the College Football Playoff (CFP). This system uses a committee to rank the top four teams to participate in a playoff. Many claim the rankings could be better and there could be bias in these rankings. Can we create a better ranking system? In fact, could there be a better system that not only ranked teams, but used a rating to help communicate the gap between each team?

Data Explanation

Data was collected for the 2019 FBS season through the collegefootballdata.com API. This API contains access to many different data sets from games scores to summarized game statistics to play by play data. For this project, team game stats were used. Variables within this data set include the following for both home and away teams within the requested game:
- School
- Home/Away
- Points
- Total Yards
- Rushing Yards
- Passing Yards
- Tackles
- Sacks
- Turnovers

To obtain this data, a JSON file was created that includes all FBS schools. The JSON file was opened and iterated through each team. For each team, the API was called and returned a JSON file for the specified week's game. This data was then written to a new JSON file under that team and week. This was done team by team, week by week, until all game data was acquired. This process was largely built on code created by Byron Slabach (4.) to pull game data.

Methods

Once all game data was collected, the data could be used to create new weekly rankings. For each week, data such as win/loss, points for and points against were used to create a rating for each team. The final output of the model is the M-Score, named after Michael Loos. The M-

Score is comprised of two main components. The first component is a simple rating derived from the number of wins and games played. This formula is based off the Colley Rankings model which uses Laplace's rule of succession to modify a basic winning percentage calculation (Lagnville). The formula becomes:

$$\text{Rating} = (1 + \text{Wins}) / (2 + \text{Games Played})$$

The second component of the rating is a Pythagorean expected win percentage. We use Football Outsider's formula (Schatz):

$$\text{Expected Win Percentage} = \text{Points For}^{2.37} / (\text{Points For}^{2.37} + \text{Points Against}^{2.37})$$

While Point Differential is not the same calculation, here we can see the comparison of overall Points For-Points Against compared to the Expected Win Percentage.

Exp Win Pct vs Point Differential
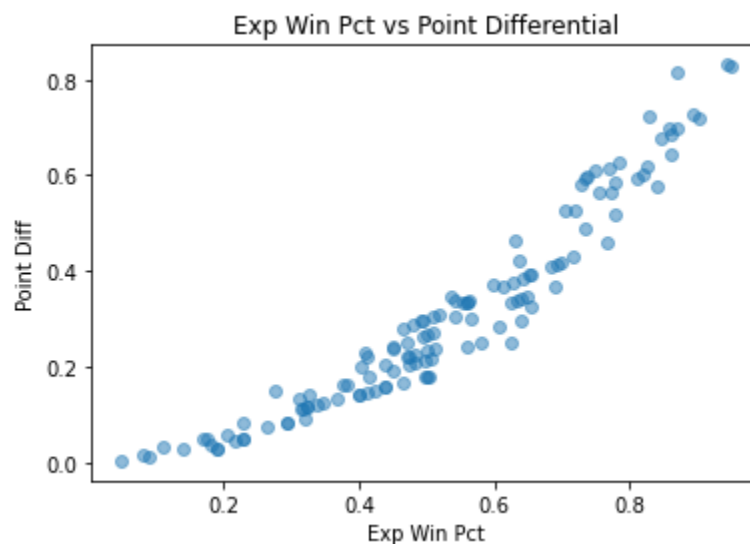
These two components are multiplied to create the M-Score:

$$\text{M-Score} = (1 + \text{Wins}) / (2 + \text{Games Played}) * \text{Expected Win Percentage}$$

This rating system combines the actual win-loss results on the field with the expected outcome for a team's season. If two teams have the same record, the team that wins games by larger margins will be rewarded. The team with closer margins of victory, or worse defeats would be punished. This can help take some of the "luck" out of game results.

The M-Score was calculated for each team each week. The values were then written to a new JSON file. Each week, the new weekly data was added to the previous week's JSON file, until we have our final rankings.

Model Analysis

The results of the model at the end of the season:

| Ranking	School	Wins	Losses	Point Diff	Exp Win Pct	M-Score
1	Clemson	14.0	1.0	456.0	0.942172	0.831328
2	Ohio State	13.0	1.0	464.0	0.948433	0.829879
3	LSU	15.0	0.0	398.0	0.867961	0.816905
4	Georgia	12.0	2.0	255.0	0.893083	0.725630
5	Appalachian State	13.0	1.0	263.0	0.827741	0.724273
6	Alabama	11.0	2.0	372.0	0.900841	0.720673
7	Oregon	12.0	2.0	264.0	0.858911	0.697865
8	Penn State	11.0	2.0	257.0	0.870644	0.696515
9	Florida	11.0	2.0	231.0	0.859765	0.687812
10	Notre Dame	11.0	2.0	245.0	0.845928	0.676743
11	Utah	11.0	3.0	242.0	0.860180	0.645135
12	Air Force	11.0	2.0	185.0	0.782662	0.626129
13	Louisiana	11.0	3.0	255.0	0.825034	0.618776
14	Navy	11.0	2.0	193.0	0.770126	0.616101
15	Boise State	12.0	2.0	180.0	0.749588	0.609040
16	UCF	10.0	3.0	265.0	0.818174	0.599995
17	Oklahoma	12.0	2.0	208.0	0.736963	0.598783
18	Memphis	12.0	2.0	196.0	0.732524	0.595176
19	Iowa	10.0	3.0	153.0	0.809381	0.593546
20	Baylor	11.0	3.0	194.0	0.778698	0.584023
21	Minnesota	11.0	2.0	150.0	0.727059	0.581647
22	Wisconsin	10.0	4.0	240.0	0.839931	0.577452
23	San Diego State	10.0	3.0	111.0	0.771933	0.566084
24	Florida Atlantic	11.0	3.0	192.0	0.753894	0.565421
25	Cincinnati	11.0	3.0	127.0	0.703872	0.527904

Right off the bat, we can see that the winner of the CFP, LSU, finished third. This could be considered an unacceptable result. Especially since 2019 LSU went undefeated and is considered potentially one of the greatest teams of all time. We also see Appalachian State at 5, compared to their average rating of 18.5 in the AP and Coaches polls. This might point to an issue with the

model. However, the overall model was not too far off. Compared to the AP Poll and Coaches Poll, the model was off by an average of 5 places for each team in the top 25.

Some observations from the ratings show that the model is strongly correlated with wining. This is expected as win percentage is a main component of the model.

Win Pct vs M-Score

Ranking vs Wins
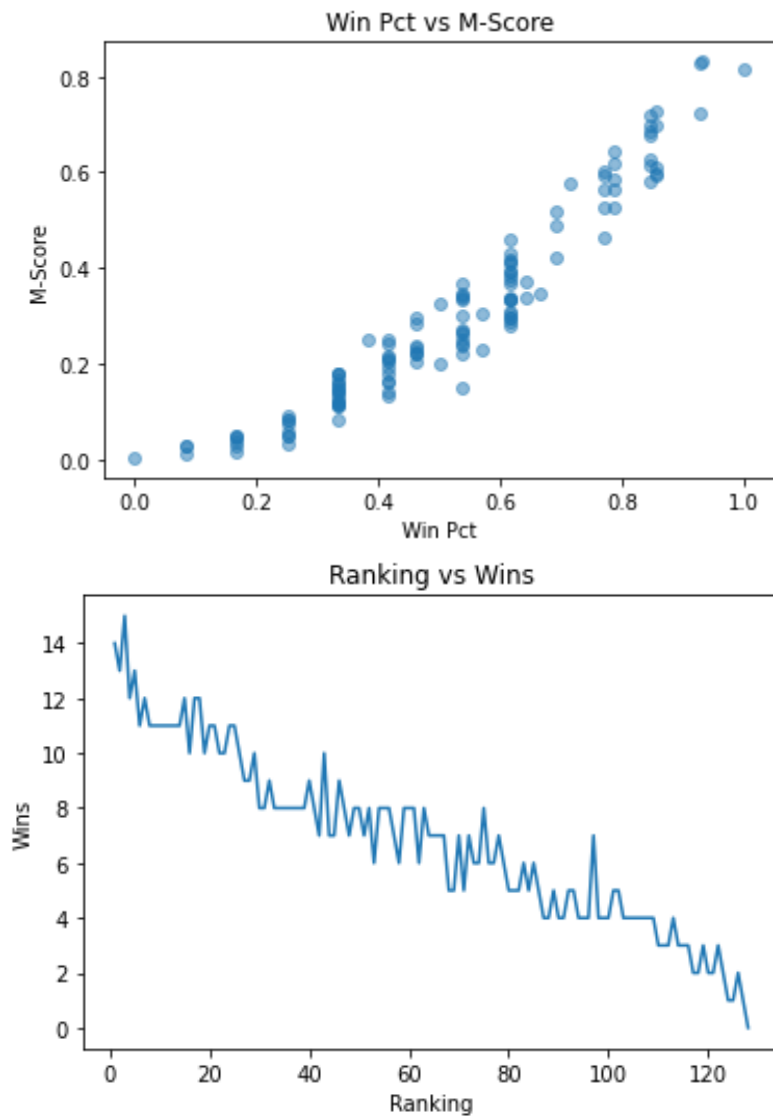
We see a similar correlation with point differential and M-Score. Although, it appears that point differential at the lower and upper bounds appear to be less influential.

Point Diff vs M-Score
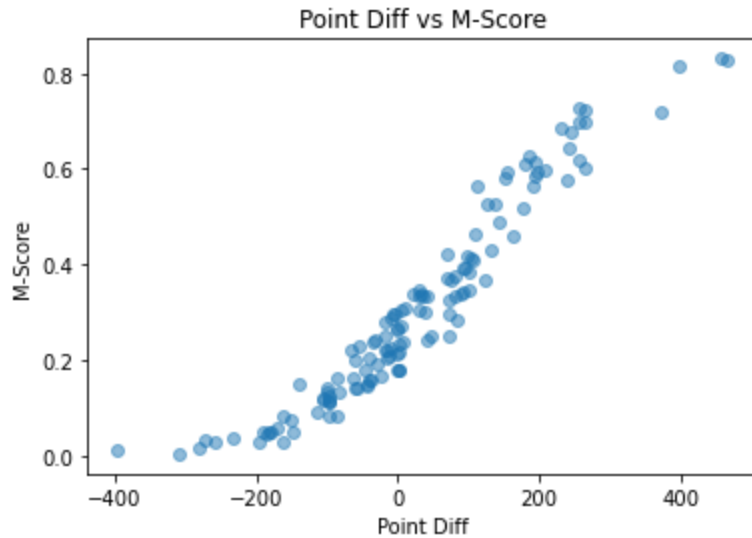
Conclusion

While this model is not perfect, it has the starting framework for a good ratings system. This model can be compared against the major polls and not look too out of place. The model could certainly use some additional metrics, such as a strength of schedule metric. This would help put some additional context around winning percentage and expected winning percentage.

Assumptions

One assumption that was made, due to the volume of the data, was that the scores and data for each game are correct and in the correct week. Any incorrect scores would certainly have an influence on the model. While any scores in the incorrect week would not have an impact on the final ratings, data in the wrong week would impact the weekly rankings.

Limitations and Recommendations

The biggest limitation within this project is not being able to obtain calculations such as opponents win percentage, to incorporate a strength of schedule metric, due to time limitations. This would certainly have an impact on the ratings as a team could produce better results while facing an easier schedule, which leads to a higher win percentage and Pythagorean win expectation, than a team with a more difficult schedule. This metric will be added to the model in the future.

Future Use and Implementation

There are several potential uses for this ratings model. The ultimate use case would be the sole ranking system to crown the college football national champion. As the current college football playoff model stands, this model could be used as a resource by the playoff selection committee.

Another potential use in the future could be to incorporate this model, along with other computer rankings and polls, into a ranking system, much like the old BCS rankings.

Ideally, implementation of the model would be extremely simple if able to use the exact same data source, api.collegefootball.com. Once this data source is updated with the newest week of data, simply run the script to pull the new data and update the ratings.

Ethical Assessment

In recent years, the college football world seems to value money just as much as, if not more than, winning. While there should be no ethical concerns about the current state of the model, any intentional introduction of bias towards any specific schools or conferences could have significant financial consequences.

References

1. Brown, M. (2018, October 19). What we can learn from college football's very first AP poll, where Minnesota was no. 1. SBNation.com. Retrieved March 5, 2022, from https://www.sbnation.com/college-football/2018/10/19/18001148/ap-top-25-poll-history

2. Langville, A. N. (2013). Chapter 3 Colley's Method. In C. D. Meyer (Ed.), Who's #1?: The Science of Rating and Ranking (pp. 21–22). essay, Princeton University Press.

3. Schatz, A. (2022, July 14). Pythagoras on the Gridiron. Football Outsiders. Retrieved March 5, 2022, from https://www.footballoutsiders.com/stat-analysis/2003/pythagoras-gridiron

4. Slabach, B. CFBRating, (2020), GitHub repository, https:// https://github.com/slabach/CFBRating