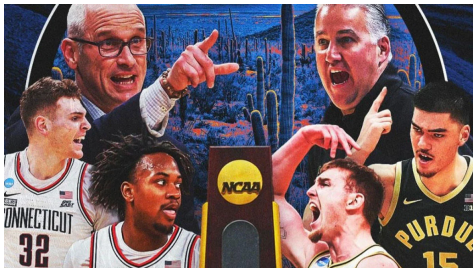# Predictability in College Sports: Comparing the Accuracies of Prediction Models for College Football and Basketball Games

Presenters: Yihan Gao, Sam Lam, David Xia
Research Mentors: Prof. AJ Hildebrand, Bingyan Liu

University of Illinois Urbana-Champaign

Rose-Hulman Undergraduate Math Conference,
Rose-Hulman Institute of Technology

April 27, 2024

College football and basketball are a huge deal. The recent mens' and womens' NCAA basketball championship games both drew well over 10 million viewers, broadcasting rights deals for college sports approach $1 billion, and billions of dollars are bet each year on college football and basketball.

Introduction and Motivating Questions

We investigate the predictability of college football and basketball games. The predictability of a game tells us how closely matched the teams are, as well as how much the outcome depends on skill as opposed to luck.

We focus on questions such as the following:

- How effective is the betting market in predicting the outcomes of games compared to more naive prediction methods?
- How does the predictability of college football compare to that of college basketball?
- How do we measure and compare the accuracies of probabilistic predictions?

Prediction Methods We Analyzed

- **Poll Based Predictions:** We make predictions based on who is better ranked in the AP Top 25 Poll, which ranks the top teams in college football and basketball.

- **Betting Market:** The team favored to win by the betting market is predicted to win. We use moneyline scores as a proxy for betting market predictions.

- **Elo Ratings:** Elo rating is a point system updated after every game to depending on the strength of its opponent (1500 is an average rating, 2000 represents a very good team). It was originally developed to rate chess players. The team with the higher Elo rating is predicted to win.
  Elo ratings obtained from `collegefootballdata.com` *(college football only).*

- The AP Top 25 Poll is a weekly poll ranking college sports teams, which aggregates the individual rankings by approximately 60 sports journalists.
- Ballots are aggregated into the official ranking using a point system: A team receives 25 points for being ranked first, 24 points for being ranked second, etc., and the teams are ranked by how many points they receive.
- When two ranked teams play against each other, the higher ranked team is predicted to win.
- Poll-based predictions are somewhat subjective, because the poll is based on analysts' individual opinions.

AP Top 25 Poll: Example

1	UCONN	**UCONN** (58)	**37-3**	Previous: 1 High: 1 - 58 Voters Low: 1 - 58 Voters	**1450** Points
2	P	**PURDUE**	**34-5**	Previous: 3 ▲ High: 2 - 57 Voters Low: 8 - Nick Alvarez	**1386** Points
3	A	**ALABAMA**	**25-12**	Previous: 19 ▲ High: 3 - 21 Voters Low: 11 - Rick Bozich	**1216** Points
	UH	**HOUSTON**	**32-5**	Previous: 2 ▼ High: 3 - 34 Voters Low: NR - Tony Garcia	**1216** Points
5	T	**TENNESSEE**	**27-9**	Previous: 6 ▲ High: 4 - 24 Voters Low: 18 - Nick Alvarez	**1203** Points

Final AP Top 25 Poll of the 2023-24 Basketball Season (Alabama and Houston are tied for third).
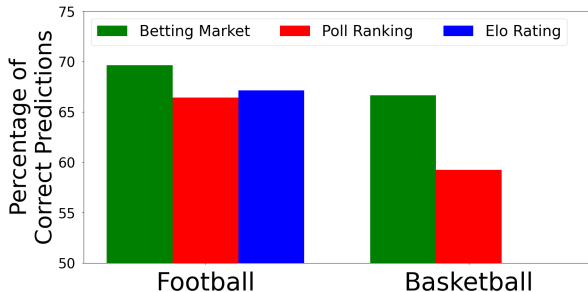
Poll Based Prediction Accuracy
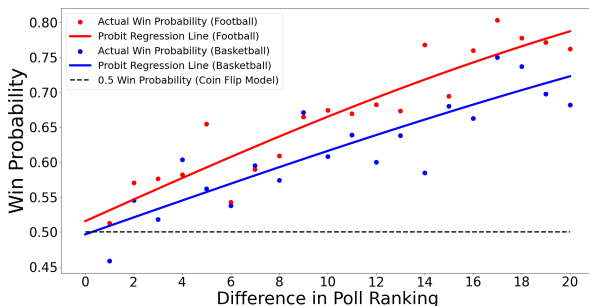
Proportion of games won by higher ranked team in AP Top 25 matchups in college football (1936) and college basketball (1950 to present).

- Overall, the higher ranked team is more likely to win the game, but there is significant variation in the win probabilities for individual rank differences. The variations may be explained in part by home advantage, which can offset the advantage provided by a higher ranking.
- The regression lines show that the accuracy of the predictions grows approximately linearly as a function of the rank difference, and is higher for college football than college basketball.

Prediction Accuracies by Method and Sport

Overall prediction accuracy of AP Top 25 matchups between 2007 and 2018 in College Football (551 games) and College Basketball (1026 games).

- Of the three methods, the betting market performs better than the other methods.
- Overall, College Football is more predictable than College Basketball.

Yearly Variations in Accuracies (College Football)
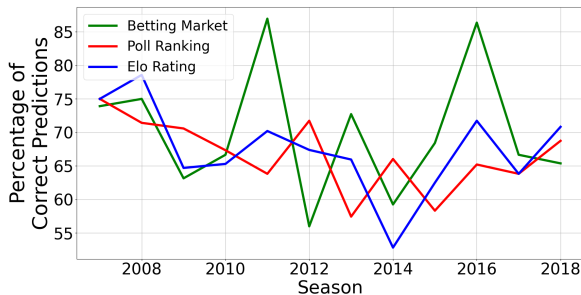
Prediction accuracy for AP Top 25 matchups in College Football for each
season between 2007 and 2018 (approximately 40 games per season).

- There are large variations from season to season for each of the three
  prediction methods, but most of the time, the betting market
  outperforms the other methods.
- The variations may be explained in part by the relatively small sample
  sizes per season.

How Moneyline Scores Work

Team	Spread	Moneyline
Los Angeles Lakers	+7.5 (-115)	+240
Denver Nuggets	-7.5 (-105)	-303

- Moneyline scores are a pair of numbers, one positive, the other one negative.
- The team with the negative moneyline score is the betting market's favorite and thus is predicted to win. The team with the positive moneyline score is the underdog.
- The positive money line score represents the amount of money won on a bet of $100.
- In the above example, the Nuggets are the favorite and the Lakers are the underdog. If you bet $100 on the Lakers and the Lakers win, then you get your money back and win $240.

- **Implied win probabilities (unnormalized):** If Teams *A* and *B* have moneyline scores of $M_A > 0$ and $M_B < 0$, then

$$P_A^* = \frac{100}{M_A + 100}$$
$$P_B^* = \frac{-M_B}{-M_B + 100}$$

- **Normalized win probabilities:**

$$P_A = P(\text{A wins over B}) = \frac{P_A^*}{P_A^* + P_B^*}$$
$$P_B = P(\text{B wins over A}) = \frac{P_B^*}{P_A^* + P_B^*}$$

Converting Moneyline Scores to Win Probabilities: Example

In the previous example (Lakers vs Nuggets) $M_{Lakers} = +240$ and $M_{Nuggets} = -303$.

- **Implied win probabilities (unnormalized):**

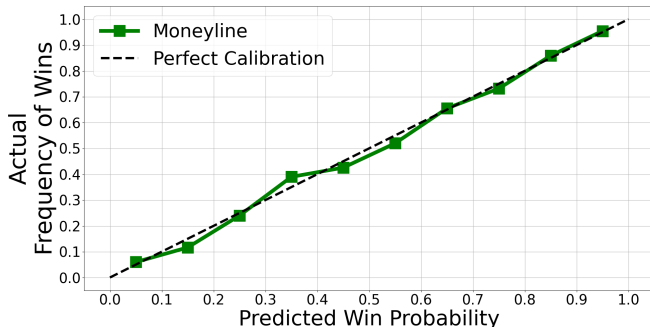$$P^*_{Lakers} = \frac{100}{240 + 100} = \frac{100}{340} = 0.294$$

$$P^*_{Nuggets} = = \frac{-(-303)}{-(-303) + 100} = \frac{303}{403} = 0.752$$

- **Normalized win probabilities:**

$$P_{Lakers} = \frac{P^*_{Lakers}}{P^*_{Lakers} + P^*_{Nuggets}} = \frac{0.294}{0.294 + 0.752} = 0.281$$

$$P_{Nuggets} = \frac{P^*_{Nuggets}}{P^*_{Lakers} + P^*_{Nuggets}} = \frac{0.752}{0.294 + 0.752} = 0.719$$

Accuracy of Moneyline Based Probabilistic Predictions

The above graph is a calibration plot, showing the actual proportion of games won versus the implied win probabilities based on moneyline scores.

- Moneyline scores are converted to win probabilities, which are compared with actual win frequencies using bins of width 0.1.
- The graph shows that the moneyline based probabilistic predictions are very close to the actual win frequencies.

- Among the methods we investigated, the betting market is the most accurate predictor of game outcomes.
- College football is more predictable than college basketball.
- There is dramatic variation in predictability from season to season.
- Probabilistic win predictions based on moneyline scores are remarkably close to the actual win frequency.

References

- College Football Data, `collegefootballdata.com`
- College Poll Archive, `collegepollarchive.com`
- Langville, A. N., Meyer, C. D. (2012). "Who's # 1? The science of rating and ranking", Princeton University Press.
- Sportsbook Reviews Online, `sportsbookreviewsonline.com`
- Sports Reference, `sports-reference.com`
- Stern, H. S. (1997). "How accurately can sports outcomes be predicted?". Chance 10, 19-23.

Thank you!

Questions?