COPA AMÉRICA 2024
PREDICTIONS

Julio Cesar Alonso Cifuentes
Diego Antonio Bohórquez Ordoñez
Juan Camilo Osorio Colonia

Design by:
Laura Barrera Valdés

June 2024

COPA AMÉRICA
2024 PREDICTIONS

Cienfi centro de investigación en economía y finanzas | ICESI

CONMEBOL
COPA AMERICA
USA 2024

One of the most anticipated competitions globally will begin on Thursday, June 20, at Mercedes-Benz Stadium of Atlanta (USA): the Copa América. The American powers gather their stars to fight for a new continental title. For every football fan, a period of maximum expectation begins to determine who will be America new king: will it be Messi's Argentina, Vinícius's Brazil, James´s Colombia, Suárez's Uruguay, or Sánchez's Chile?

At the Universidad Icesi's Center for Research in Economics and Finance (Cienfi), we are dedicated to answering business questions with data and statistical and machine learning tools. As has been our pattern for the past six years, we are no strangers to this football world question, and we have used analytics to predict the tournament's winner. In this document, we present our methodology and prediction of the new champion of the old continent.

THE MODEL

The matches' results and the information of the participating teams in the last five editions of the Copa América (2007, 2011, 2015, 2016, 2019 y 2021) were used to construct the model.

The variables included in this study were the following:

- **Manager's age**: a measure of the manager's experience.

- **Manager tenure**: a measure of how well adapted or developed the coach's idea of the game is with the team. A newly selected coach has needed more time to build the player base and establish the foundations of their play system.

- **Host country**: In sports, especially football, the positive effect of home advantage on a team is widely documented.

- **Number of World Cup appearances**: a measure of each team's reputation and success on the continent. Qualifying for the World Cup requires being better than most teams on the continent.

- **ELO rating**: The rating system was developed by the mathematician Arpad Elo to calculate the relative ability of chess players. Applied to football, it is a measure to calculate a ranking considering the performance of each team (in fact, since 2018, FIFA adopted a new methodology based on this scoring system to generate its ranking).

- **Performance in points:** a measure of each team's performance by assigning 3 points to a win, 1 to a draw, and 0 to a loss and calculating the proportion of points scored out of the maximum possible points that could be achieved. This performance was calculated in different time windows up to the day before the start of the tournament: i) from January 1st, 2000, ii) the last 4 years, and iii) the last 1.5 years.

- **Average goals scored and conceded**: it is also a measure of the team's performance. This performance was also calculated in time different windows up to the day before the start of the tournament: i) from January 1st, 2000, ii) the last 4 years, and iii) the last 1.5 years.

- **Players' age:** Age can be a relevant variable in a team's performance. Senior players can bring experience at crucial moments of a match, but their physical or technical performance may be in decline. Meanwhile,

young players can bring physical deployment and talent but are often inexperienced. It was decided to calculate the average age by position groups on the field (goalkeeper, defender, midfielder, and striker).

- Player performance: We aimed to have player performance data from public portals such as Transfermarkt[1] and sofifa.com[2]. However, the need for more data before 2012 precludes its usage. It was decided to design an annual scoring system for clubs, and each player in each year was assigned the performance score of the club he belonged to. Subsequently, the average score per position group is calculated for each team (goalkeeper, defender, midfielder and striker). This assumes that if a player plays for a club and the club is a high performer, the player must also be a high performer[3].

We employed a combination of statistical and machine learning models, specifically Poisson regression with Lasso regularization, Random Forest, and Extreme Gradient Boosting. These models were used to estimate the number of goals a team would score in each match. Our assumption was that the number of goals scored by a team follows a Poisson distribution and is statistically independent of the number of goals scored by the opponent.

Our results conclude that the best model is a *XGboost*, which had a hit rate of 59% over the evaluation sample[4], being equal to the hit rate of the bookmakers.

Prediction

After running our model's 50,000 simulations for Copa América 2024, we found that there is a great favorite to win the tournament (see Table 1), Argentina, with a 29.34% probability of being champion.

A second block of favorites is formed by Uruguay, Brazil and Colombia.

Table 1 shows the probability of each of the 16 teams reaching the quarterfinals, semifinals, and final and the probability of being champion.

---

[1] This portal stores statistics on players and their market value.

[2] This portal has player performance scores for the EA Sports FIFA video game.

[3] We recognize that this assumption could be controversial, but it was the best approximation to address the data limitations discussed above.

[4] For more details, please get in touch with the authors.

**COPA AMÉRICA
2024 PREDICTIONS**

Cienfi — centro de investigación en economía y finanzas · ICESI

CONMEBOL
COPA AMERICA
USA 2024

Table 1. Simulated probabilities of reaching the different stages of Copa América 2024 and being champion (in %).

National Team	QUARTERFINALS	SEMIFINAL	FINAL	CHAMPION
Argentina	89,50	66,94	49,38	29,34
Uruguay	85,58	50,29	30,52	17,49
Brazil	75,28	46,78	26,18	14,83
Colombia	71,40	43,44	23,83	12,70
México	73,26	34,90	15,97	6,44
Ecuador	64,37	27,60	11,54	4,51
Chile	50,79	26,55	9,67	3,56
USA	58,35	22,98	8,04	3,05
Perú	41,06	20,00	6,73	2,41
Paraguay	34,17	14,68	5,47	2,03
Venezuela	34,77	10,41	3,13	0,90
Panamá	30,50	8,69	2,33	0,64
Costa Rica	19,14	6,51	1,91	0,60
Jamaica	27,58	6,87	1,91	0,51
Canadá	18,63	6,69	1,63	0,50
Bolivia	25,55	6,58	1,65	0,41

Source: Cienfi

It is important to remember that even a team with a low probability of winning the tournament can surprise and win it. Just as there is a low probability of a person being struck by lightning, we still observe people being killed by lightning. Probabilities are only an indicator of how likely a situation is to occur.

The odds of winning a game only materialize on the playing field. The excitement and uncertainty that make football the most followed sport in the world are always present. This uncertainty motivates the Cienfi team to use data to generate forecasts for this football tournament. As we all know, "**Football is the most important thing among the least important things.**