# Euro 2024 predictions

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

Design by:
Laura Barrera Valdés

June 2024

Cienfi
centro de
investigación
en economía
y finanzas

UNIVERSIDAD
ICESI
TU FUTURO A OTRO NIVEL

One of the most anticipated competitions globally will begin on Friday, June 14, at Munich, Germany's legendary Allianz Arena: the European Championship. The European 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 Europe's new king: will it be Mbappé's France, Kane's England, Cristiano's Portugal, Kroos' Germany, Modric's Croatia, Morata's Spain, Jorginho's Italy, or De Bruyne's Belgium?

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 European Championship (2004, 2008, 2012, 2016, and 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 Random Forest, which had a hit rate of 58% over the evaluation sample[4], being above the hit rate of the bookmakers (50%).

## Prediction

After running our model's 50,000 simulations for Euro 2024, we find that no great favorite to win the tournament exists (see Table 1). Spain and England have the highest probability (although not very high) of winning the cup.

---

[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.

France is between the prior group and a group of five teams (Italy, Portugal, Belgium, Germany, and the Netherlands) with an equal probability of being champions.

Table 1 shows the probability of each of the 24 European teams reaching the round of 16, quarterfinals, semifinals, and final and the probability of being champion.

## Table 1. Simulated probabilities of reaching the different stages of Euro 2024 and being champion (in %) (1 of 2).

| National Team | ROUND OF 16 | QUARTERFINALS | SEMIFINAL | FINAL | CHAMPION |
|---|---|---|---|---|---|
| Spain | 70,92 | 54,88 | 33, 04 | 19,93 | 11,73 |
| England | 79,13 | 56,19 | 33,27 | 19,97 | 11,56 |
| France | 65,07 | 50,55 | 30,19 | 16,71 | 9,30 |
| Italy | 62,71 | 47,88 | 26,80 | 14,81 | 7,86 |
| Portugal | 73,57 | 48,50 | 27,50 | 14,75 | 7,83 |
| Belgium | 76,53 | 49,83 | 27,37 | 14,73 | 7,83 |
| Germany | 72,30 | 48,50 | 26,00 | 14,21 | 7,36 |
| Netherlands | 58,53 | 44,98 | 25,41 | 13,53 | 7,06 |
| Switzerland | 67,46 | 44,81 | 23,26 | 12,16 | 5,98 |
| Austria | 44,90 | 33,73 | 16,78 | 7,77 | 3,60 |
| Denmark | 51,75 | 32,05 | 14,14 | 6,29 | 2,65 |
| Croatia | 39,02 | 28,79 | 13,16 | 5,70 | 2,42 |
| Turkey | 51,64 | 29,99 | 13,54 | 5,69 | 2,38 |
| Ukraine | 51,41 | 29,03 | 12,29 | 4,88 | 1,94 |
| Poland | 31,49 | 23,16 | 10,22 | 4,13 | 1,73 |
| Czech Republic | 40,42 | 23,19 | 9,52 | 3,76 | 1,44 |
| Serbia | 39,16 | 24,77 | 9,78 | 3,77 | 1,36 |
| Hungary | 36,84 | 23,26 | 9,01 | 3,44 | 1,27 |
| Slovakia | 40,21 | 21,51 | 8,25 | 3,03 | 1,04 |

## Table 1. Simulated probabilities of reaching the different stages of Euro 2024 and being champion (in %) (2 of 2).

| National Team | ROUND OF 16 | QUARTERFINALS | SEMIFINAL | FINAL | CHAMPION |
| --- | --- | --- | --- | --- | --- |
| Albania | 27,35 | 19,36 | 7,63 | 2,90 | 1,03 |
| Georgia | 34,37 | 18,38 | 7,00 | 2,51 | 0,90 |
| Slovenia | 29,96 | 17,44 | 6,14 | 2,21 | 0,74 |
| Romania | 31,85 | 16,45 | 5,78 | 1,99 | 0,64 |
| Scotland | 23,39 | 12,77 | 3,93 | 1,13 | 0,35 |

**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."**