Predicting the FIFA World Cup 2026

A Monte Carlo Simulation of Modern Tournament Football

Javier Ruano Hernández

javierruanohdez

February 3, 2026

Abstract

Predicting the outcome of a football World Cup is inherently uncertain. Traditional odds and statistical models often fail to capture the complex interplay of modern football dynamics, including tactical evolution, squad rotation, momentum, and tournament-specific randomness.

This work presents a Monte Carlo simulation framework for the FIFA World Cup 2026, integrating historical team performance, FIFA rankings, recent form, head-to-head statistics, and a modern football strength metric.

By running thousands of tournament simulations, the model generates probabilistic estimates for championship outcomes, highlighting both favorites and potential dark horses. Our approach emphasizes structural and tactical consistency over mere historical prestige, providing a complementary perspective to conventional betting markets. Notably, the final showdown between Spain and Argentina emerges as a particularly revealing scenario, pitting two of the most robust and tactically adaptive teams against each other under conditions of extreme variance.

1. Introduction

Football resists precise prediction. Unlike league competitions, World Cups are short, intense, and unforgiving. A single draw, red card, or penalty shootout can overturn years of planning.

Traditional approaches rely heavily on historical prestige or betting odds. However, recent tournaments have shown that:

- Tactical cohesion often outweighs individual talent
- Teams peak at different moments
- Variance increases dramatically in knockout stages

The 2026 World Cup introduces an expanded format and three host countries, further increasing uncertainty. This project aims to model that uncertainty explicitly.

2. World Cup 2026: Structural Challenges

The 2026 FIFA World Cup differs fundamentally from previous editions:

- Hosted across the United States, Canada, and Mexico
- Long travel distances between venues
- Climate and time-zone variability
- Expanded format with 48 teams

The tournament structure consists of:

- 12 groups of 4 teams
- 12 group winners qualify automatically
- Best 8 runners-up advance
- Remaining runners-up play a preliminary knockout round
- Standard knockout stages from Round of 16 onward

This structure increases exposure to randomness, particularly for traditional powerhouses that historically rely on slow tournament starts.

3. Data Sources

The model integrates multiple data sources:

3.1. Match Results

Historical international match results from 2000 onward were derived from a publicly available dataset hosted on Kaggle, originally compiled by Mart Jürisoo. The dataset aggregates official international matches dating back to 1872 and is widely used in football analytics research.

3.2. FIFA Rankings

FIFA ranking points were collected across multiple years. While FIFA rankings are imperfect, they provide a standardized proxy for baseline team strength and seeding expectations.

3.3. Tournament Achievements

A custom dataset of tournament achievements was constructed with the assistance of AI-based data structuring. This dataset encodes:

- World Cup titles, finals, and semifinal appearances
- Continental titles, finals, and semifinal appearances
- Recency of achievements with exponential decay

The goal is to reward recent competitive relevance rather than distant history.

4. Methodology

4.1. Match-Level Model

Individual matches are evaluated using a Gradient Boosting classifier trained on:

- FIFA point differentials
- Recent form (smoothed last five matches)
- Head-to-head statistics
- Interaction features between form and ranking

Predicted probabilities are adjusted to reflect tournament round variance.

4.2. Tournament Simulation

Each World Cup simulation represents one plausible tournament universe. For each simulation:

1. Group stages are played match by match
2. Qualification follows official 2026 rules
3. Knockout rounds introduce increasing randomness

A total of 5,000 simulations were executed to obtain stable probability estimates.

4.3. Variance by Tournament Round

Randomness increases as the tournament progresses:

```
Group stage      → Low variance
Round of 16      → Medium variance
Quarterfinals    → High variance
Final            → Maximum variance
```

This reflects empirical tournament behavior.

5. Modern Football Strength

A key modeling decision is the inclusion of a "modern football strength" component.

Historical success alone does not guarantee present competitiveness. This component rewards:

- Tactical continuity
- Pressing systems and compact blocks
- Squad automation and positional discipline
- Post-2018 tournament performance

Teams such as Argentina, Spain, France, England, and Morocco score highly under this framework.

6. Results

After 5,000 simulations, the estimated probabilities of winning the tournament are summarized below:

Team	Probability (%)
Argentina	16.1
Spain	13.1
France	12.8
Portugal	8.5
England	8.4
Croatia	5.2
Germany	4.9
Morocco	4.7
Belgium	4.1
Brazil	2.9

While betting markets often favor Spain marginally, the simulation consistently places Argentina at the top due to stability, adaptability, and recent tournament performance.

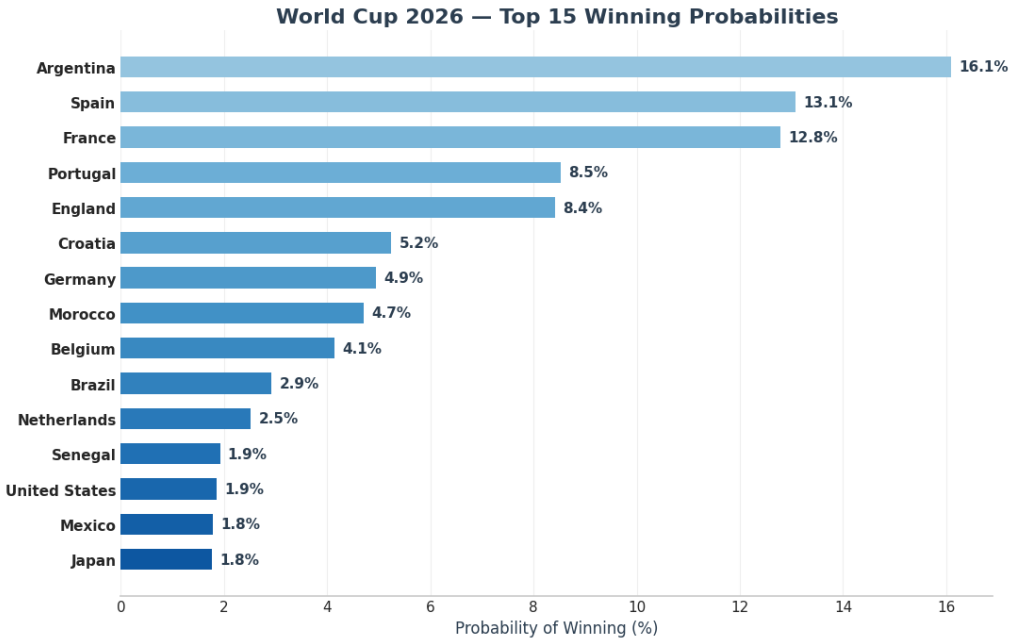

Figure 1: Top 15 team probabilities of winning the World Cup 2026.

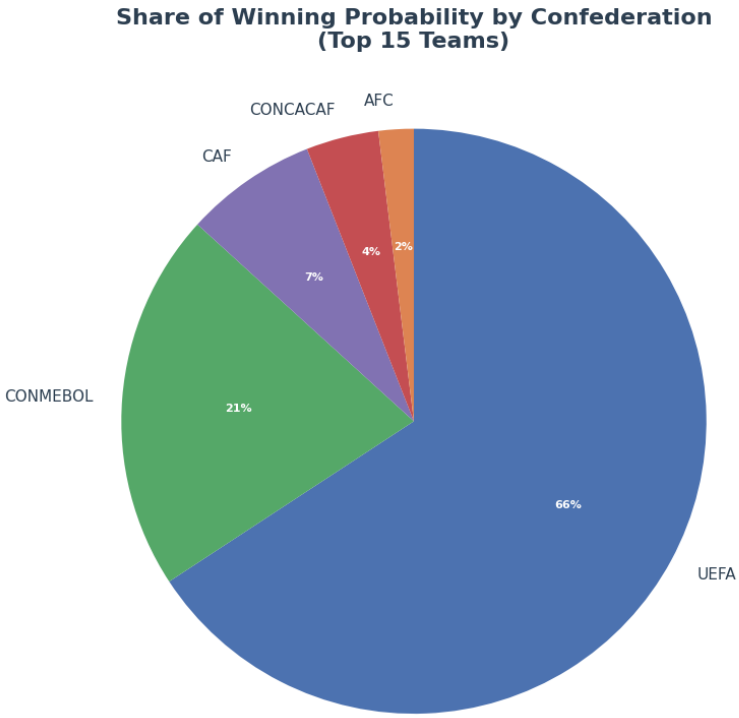

Figure 2: Distribution of winning probabilities by confederation (Top 15 teams).

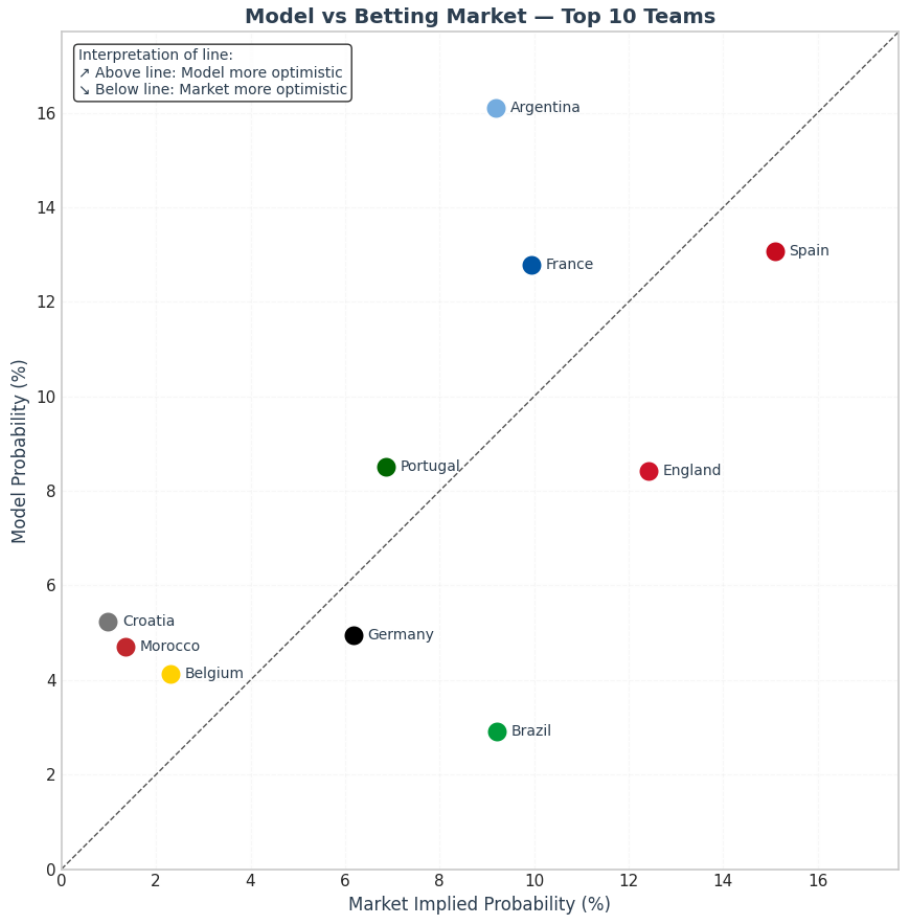

Figure 3: Comparison between model predictions and market implied probabilities (Top 10 teams).

7. Interpretation

The model answers a different question than bookmakers:

> Which teams are best adapted to modern tournament football under extreme variance?

Under this lens:

- Argentina exhibits the highest consistency
- Spain and France form a closely matched second tier
- Portugal and England benefit from tactical maturity
- Croatia and Morocco emerges as dark horses due to structure and resilience

The upcoming Finalissima between Spain and Argentina is particularly revealing, as it pits the two most consistently favored teams in the simulation against each other in a competitive setting.

8. Limitations

This model has important limitations:

- No explicit modeling of injuries or squad selection
- Coaching changes are not dynamically updated
- Player-level form is aggregated at team level
- FIFA rankings introduce structural bias

Most importantly, football remains inherently unpredictable.

9. Future Work

Potential extensions include:

- Player-level embeddings based on club performance
- Travel and climate fatigue modeling
- Tactical matchup vectors
- Real-time odds comparison

10. Conclusion

This project does not aim to predict the World Cup with certainty. Instead, it embraces uncertainty and uses it as a modeling feature.

The Monte Carlo approach allows us to see not just a winner, but the distribution of possible outcomes, highlighting which teams are structurally robust under variance.

Graphical results demonstrate that:

- Argentina consistently ranks highest across simulations
- Spain and France remain highly competitive, often neck-and-neck
- England, Portugal, Croatia, Germany and Morocco show variability that could make knockout surprises likely
- Confederation distribution highlights the relative strength of UEFA and CONMEBOL teams in the top tier
- Market comparison plots reveal where betting odds diverge from model insights

In summary, modern football prediction benefits from combining historical data, tactical evaluation, and simulation under uncertainty. While no model can fully capture the chaos of a tournament, this approach provides both insightful probabilities and a framework for strategic understanding.