

PREDICCIONES DE LA COPA AMÉRICA 2024



- CONMEBOL -
**COPA AMERICA
USA 2024**

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

Diseño por:
Laura Barrera Valdés

Junio de 2024

Cienfi

centro de
investigación
en economía
y finanzas





El jueves 20 de junio, en el Mercedes-Benz Stadium de Atlanta (Estados Unidos), se dará inicio a una de las competiciones que más genera expectativa alrededor del mundo: la Copa América. Los equipos del continente reúnen a sus estrellas para luchar por un nuevo título continental. Para todo aficionado del fútbol comienza una etapa de máxima expectativa por saber quién será el nuevo rey de América: ¿Será la Argentina de Messi, Brasil de Vinícius, Colombia de James, Uruguay Suárez o Chile de Sánchez?

En el Centro de Investigación en Economía y Finanzas -Cienfi- de la Universidad Icesi, nos dedicamos a responder preguntas de negocio con datos y herramientas estadísticas y de *machine learning*. Como ya es costumbre, desde hace 6 años, no somos ajenos a esta pregunta que se hace el mundo del fútbol y nos aventuramos a emplear datos para pronosticar el ganador del torneo. A continuación, presentamos nuestra metodología y pronóstico del nuevo campeón del continente americano.

EL MODELO

Para la construcción del modelo, se tomó como base los resultados de los partidos y la información de las selecciones participantes en las últimas 6 ediciones de la Copa América (2007, 2011, 2015, 2016, 2019 y 2021).

Las variables incluidas en el estudio fueron las siguientes:

- **Edad del entrenador:** medida de la experiencia del entrenador.
- **Tiempo en el cargo del entrenador:** medida de qué tan adaptada o desarrollada está la idea de juego del entrenador con el equipo. Un entrenador recién seleccionado en el

cargo ha tenido poco tiempo para construir la base de jugadores y establecer las bases de su sistema de juego.

- **País anfitrión:** en el mundo de los deportes, especialmente el fútbol, está ampliamente documentado el efecto positivo de la localía para un equipo.
- **Número de participaciones en la Copa del Mundo:** medida de la reputación y éxito de cada equipo en el continente. Clasificarse a la Copa del Mundo requiere haber sido mejor que la mayoría de las selecciones del continente.
- **Puntaje ELO:** El sistema de puntuación fue desarrollado por el matemático Arpad Elo para calcular la habilidad relativa de los jugadores de ajedrez. Aplicado al fútbol, constituye una medida para calcular un ranking de acuerdo con el rendimiento de cada equipo (de hecho, desde el 2018, FIFA adoptó una nueva metodología basada en este sistema de puntuación para generar su ranking).
- **Rendimiento en puntos** medida del rendimiento de cada equipo al asignar 3 puntos a una victoria, 1 a un empate y 0 a una derrota (sin diferenciar qué tipo de partido es) y calcular la proporción de puntos obtenidos sobre el máximo posible de puntos que se podría alcanzar. Este rendimiento se calculó en diferentes ventanas hasta el día antes del inicio del torneo: i) desde el 1 de enero del 2000, ii) desde hace 4 años y iii) desde hace 1.5 años.
- **Gol promedio a favor y en contra:** medida de rendimiento de cada equipo. Este rendimiento se calculó para las mismas tres



ventanas empleadas en el rendimiento en puntos.

- **Edad de los jugadores:** La edad puede ser una variable relevante en el rendimiento de un equipo. Jugadores veteranos pueden aportar experiencia en momentos cruciales de un partido, pero su rendimiento físico o técnico puede estar en declive. Mientras tanto, jugadores jóvenes pueden aportar despliegue físico y talento, pero suelen ser inexpertos. Se decidió calcular la edad promedio por grupos de posición en el campo (arquero, defensa, mediocampista y delantero).
- **Rendimiento de los jugadores:** Nuestro objetivo era contar con datos del rendimiento de los jugadores provenientes de portales públicos como Transfermarkt¹ y sofascore.com². Sin embargo, la escasez de datos para versiones antes del año 2012 impidió su uso. Se decidió diseñar un sistema de puntuación anual para los clubes y a cada jugador en cada año se le asignó la puntuación de rendimiento del club al que pertenece. Posteriormente, se calcula el puntaje promedio por grupos de posición para cada equipo (arquero, defensa, mediocampista y delantero). Esto supone que si un futbolista juega en un club y el club tiene alto rendimiento, entonces el jugador también debe tener alto rendimiento³.

Se consideraron modelos estadísticos y de *machine learning* (regresión de *Poisson* con regularización *Lasso*, *Random Forest* y *Extreme Gradient Boosting*) para estimar la cantidad de goles que un equipo anotaría en cada partido, bajo la aceptación del supuesto de que la cantidad de goles anotados por un equipo sigue una distribución de *Poisson* y que es estadísticamente independiente de la cantidad de goles que anota el rival.

Nuestros resultados concluyen que el mejor modelo es un *XGboost*, que tuvo una tasa de acierto del 59% sobre la muestra de evaluación⁴, estando igual que el acierto de las casas de apuestas.

Pronóstico

Al realizar las 50.000 simulaciones de nuestro modelo para la Copa América 2024, encontramos que **existe un gran favorito para ganar el torneo (Ver Tabla 1) como lo es Argentina, con una probabilidad del 29,34% de ser campeón.**

Un segundo bloque de oponentes lo conforma Uruguay, Brasil y Colombia.

En la Tabla 1 presentamos la probabilidad de cada una de las 16 selecciones de llegar a la etapa de octavos, cuartos, semifinales y final. Asimismo, la probabilidad de ser campeón.

¹ Este portal almacena estadísticas de los jugadores y su valor de mercado.

² Este portal tiene los puntajes de rendimiento de los jugadores para el videojuego EA Sports FIFA.

³ Reconocemos que es un supuesto que podría discutirse, pero fue la mejor aproximación para hacer frente a las limitaciones comentadas sobre los datos.

⁴ Para conocer mayores detalles consultar con autores.

Tabla 1. Probabilidades simuladas de alcanzar las diferentes etapas de la Copa América 2024 y de ser campeón (en %).

| Selección | CUARTOS | SEMIFINAL | FINAL | CAMPEÓN |
|------------|---------|-----------|-------|---------|
| Argentina | 89,50 | 66,94 | 49,38 | 29,34 |
| Uruguay | 85,58 | 50,29 | 30,52 | 17,49 |
| Brasil | 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 |
| EEUU | 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 |

Fuente: Cienfi

Es importante recordar que incluso una selección con una baja probabilidad de ganar el torneo puede sorprender y ganarlo. Así como existe una baja probabilidad de que una persona sea alcanzada por un rayo, pero aún así observamos que personas mueren por un rayo. Las probabilidades solo son un indicador de que tan posible es que ocurra una situación.

Las probabilidades de ganar un partido solo se materializan en el campo de juego. La emoción y la incertidumbre que hacen del fútbol el deporte más seguido en el mundo siempre están presentes. Y es precisamente esa incertidumbre la que motiva al equipo de Cienfi a emplear datos para generar pronósticos de este torneo de



fútbol. Pues como todos sabemos “El fútbol es lo más importante entre lo menos importante”.