Revista Ciencia, Tecnología e Innovación en Ingeniería y Arquitectura
2(1) (2026), pp. 5-12
Facultad de Ingeniería y Arquitectura
Universidad de El Salvador
https://revistas.ues.edu.sv/index.php/cti-fia

CTI-FIA

Artículo | Article

Predicción Probabilística del Mundial de Fútbol FIFA 2026 mediante Gradient Boosting y Simulación Monte Carlo

Probabilistic Prediction of the 2026 FIFA World Cup Using Gradient Boosting and Monte Carlo Simulation

David Deras[1] , Bladimir Diaz[1,2]
1 Facultad de Ingeniería y Arquitectura, Universidad de El Salvador
2 https://orcid.org/0009-0007-3328-5179

Correspondencia: ***bladimir.diaz@ues.edu.sv***
Enviado: ***4 de junio de 2026*** Aceptado: ***1 de julio de 2026***

DOI: https://doi.org/10.66778/CTI.v02n01.01

PALABRAS CLAVE

Fútbol, predicción probabilística, XGBoost, Monte Carlo, ranking ELO, Mundial 2026, calibración de probabilidades

RESUMEN

Se presenta un sistema end-to-end para estimar la probabilidad de campeonato de cada una de las 48 selecciones que disputarán la Copa Mundial de la FIFA 2026 bajo el nuevo formato de 12 grupos. El pipeline integra fuentes heterogéneas (resultados internacionales 1872–2026, datos de eventos, goleadores y tandas de penales, Expected Goals (xG) de StatsBomb, valor de mercado de Transfermarket y ranking FIFA histórico extendido con un snapshot de 2026) para construir nueve características incluyendo features derivadas calculadas estrictamente as of date que alimentan cuatro modelos multiclase (Regresión Logística, XGBoost, LightGBM y XGBoost calibrado con Platt) entrenados con ponderación por decaimiento temporal exponencial y rebalanceo de clases. El mejor modelo (XGBoost calibrado con método Platt) obtiene un Log-Loss de ,855 y un Brier de ,166 sobre un conjunto de prueba temporal y se acopla a un motor Monte Carlo de 10,000 iteraciones que respeta el reglamento 2026 (2 primeros de grupo + 8 mejores terceros, desempates puntos - DG - GF, definición por penales). Los intervalos de confianza de Clopper–Pearson sitúan a un grupo reducido de selecciones históricamente fuertes entre los favoritos. Una sección dedicada discute las limitaciones del modelo: equipos debutantes sin historial, cobertura parcial de xG y ausencia de modelado estructural de lesiones.

KEYWORDS

Soccer, probabilistic prediction, XGBoost, Monte Carlo simulation, ELO ranking, 2026 FIFA World Cup, probability calibration

ABSTRACT

A comprehensive end-to-end system is presented to estimate the championship probability of each of the 48 national teams participating in the 2026 FIFA World Cup under the new 12-group format. The pipeline integrates heterogeneous data sources, including international match results from 1872–2026, event data (goalscorers and penalty shootouts), Expected Goals (xG) from StatsBomb, Transfermarkt market values, and historical FIFA rankings extended with a 2026 snapshot, to construct nine features—including derived features calculated strictly as-of-date—which are used to train four multiclass models: Logistic Regression, XGBoost, LightGBM, and Platt-calibrated XGBoost. The models are trained using exponential temporal decay weighting and class rebalancing. The best-performing model, Platt-calibrated XGBoost, achieves a Log-Loss of 0.855 and a Brier score of 0.166 on a temporal test set and is coupled with a 10,000-iteration Monte Carlo engine that complies with the 2026 competition regulations (top two teams from each group + eight best third-placed teams, tiebreakers based on points–goal difference–goals scored, and penalty shootouts). Clopper–Pearson confidence intervals place a small group of historically strong national teams among the favorites. A dedicated section discusses the model's limitations, including debutant teams with no historical record, partial xG coverage, and the absence of structural injury modeling.

Publicado por la Facultad de Ingeniería y Arquitectura de la Universidad de El Salvador.
Este es un artículo *Open Access* bajo la licenica *Creative Commons* Atribución 4.0 Internacional (CC BY 4.0)

INTRODUCCIÓN

La Copa Mundial de la FIFA 2026 estrena el mayor cambio estructural de su historia: 48 selecciones repartidas en 12 grupos de 4, de los cuales avanzan los dos primeros lugares más los ocho mejores terceros, totalizando 32 equipos en la fase eliminatoria. El torneo agrega una ronda adicional (Round of 32) y eleva el total de partidos a 104 (FIFA, 2023). Este cambio reglamentario invalida directamente los modelos calibrados sobre formatos de 32 equipos y motiva un rediseño del motor de simulación. Este trabajo introduce un sistema reproducible para estimar la probabilidad de título de cada selección. Las contribuciones son: (i) un pipeline de datos automatizado que combina cuatro fuentes históricas y modernas; (ii) un benchmark de tres clasificadores multiclase entrenados con time decay exponencial y calibración de Platt (sigmoid) sobre un hold-out temporal; (iii) un motor Monte Carlo que implementa fielmente el reglamento 2026, incluyendo el desempate de los ocho mejores terceros y la simetría local/visitante con excepción para los anfitriones; (iv) una validación histórica sobre el Mundial 2022 sin data leakage y un análisis de sensibilidad a lesiones (−30 % del valor de plantilla).

METODOLOGÍA

Trabajo relacionado

El árbol genético de la predicción futbolística parte del modelo Poisson de Maher (M. J. Maher, 1982) y de su refinamiento Dixon–Coles (M. J. Dixon and S. G. Coles, 1997), que captura la correlación de marcadores bajos. El sistema de puntuación Elo (A. E. Elo,1978) proporciona una métrica de fuerza relativa actualizable partido a partido, hoy estandarizada para selecciones nacionales. Trabajos más recientes combinan *Random Forest* con ratings de equipo para Mundiales y Eurocopas (A. Groll, C. Ley, G. Schauberger, and H. Van Eetvelde, 2019). La popularización de *XGBoost* (T. Chen and C. Guestrin, 2016) y *LightGBM* (G. Ke et al., 2017) ha desplazado la frontera hacia modelos de *gradient boosting*, mientras que SHAP (S. M. Lundberg and S.-I. Lee, 2017) permite interpretar sus predicciones. Para evaluar la calibración de probabilidades el estándar son el *Log-Loss* y el *Brier score* (G. W. Brier, 1950), complementados con calibración isotónica o de Platt (J. C. Platt,1999), (A. Niculescu-Mizil and R. Caruana, 2005).

Pipeline de datos

Los datos crudos se descargan o cargan en src/data/ y se materializan en data/raw/; la limpieza y la unión producen data/processed/features.csv. El pipeline integra cinco fuentes, resumidas en la Tabla 1.

Fuente principal de resultados

El dataset de Mart Jüri-soo (M. Jürisoo, 2026) disponible tanto en GitHub como en Kaggle (M. Jürisoo, 2026) constituye el núcleo del entrenamiento con ≈49,000 partidos internacionales (1872–2026); según su documentación, los datos se recopilan de Wikipedia, RSSSF y los sitios web de las federaciones nacionales. De la misma colección provienen los archivos de eventos goalscorers (goleadores con minuto y penal), shootouts (tandas) y former_names (continuidad de federaciones), usados para las features derivadas y la unificación de nombres. El ranking FIFA histórico (Cashncarry, 2023) serie 1993–2024 extendida con un snapshot manual del 2026-05-30–alimenta el feature ranking_diff.

Tabla 1

Fuente	Contenido	Cobertura
martj42 [12], [13]	Resultados internacionales	1872–2026, ≈49 000 partidos
martj42 (eventos) [12]	Goleadores (min, penal), tandas y nombres históricos	47.6k goles; 677 tandas
FIFA Ranking [14]	Posición y puntos (+ *snapshot* 2026)	1993–2026, ≈210 selecc.
StatsBomb [15]	xG por equipo	109 selecc.
Transfermarkt [16]	Valor de mercado plantilla	62 selecc., snapshot

Estadísticas avanzadas y valor de plantilla

Los *xG* históricos provienen del open data de StatsBomb (*StatsBomb, 2024)*, consumido vía statsbombpy; las selecciones sin cobertura StatsBomb reciben la media global (≈1,2) como imputación. El *ELO* se calcula in-house con la fórmula de Elo (A. E. Elo, 1978) sobre el historial completo de resultados, sin depender de un proveedor externo. El valor de mercado de las plantillas se obtiene con un snapshot manual de Transfermarkt (Transfermarkt, 2026).

Limpieza y unión

El proceso incluye: estandarización de nombres de países a un canonical name (USA - United States) y unificación de federaciones sucesoras reconocidas por FIFA (Yugoslavia - Serbia, Checoslovaquia - Chequia) para dar continuidad al ELO; deduplicación por (date, home, away); geocodificación de sedes con *geopy* (distancia 0 cuando

falla); *merge-asof backward del ranking FIFA* extendido con el snapshot del 2026-05-30, libre de leakage pues todo partido solo ve ranks anteriores a su fecha– e imputación del *xG* con la media global (≈1,2) para selecc. sin cobertura StatsBomb. El ELO se acumula sobre el universo completo de partidos (todos los torneos, incl. amistosos, desde 1872), no sobre el subconjunto filtrado de filas a emitir.

Licencias de datos

Los *xG* provienen del open data de StatsBomb (StatsBomb,2024), empleado bajo su licencia de uso no comercial con atribución. El valor de mercado de Transfermarkt (Transfermarkt, 2026) es un snapshot manual utilizado exclusivamente con fines académicos, sin redistribución del dato original. Las demás fuentes (martj42 y el ranking FIFA) son datasets públicos de uso para investigación.

RESULTADOS

Ingeniería de características

Se construyen nueve características justificadas tanto teóricamente como mediante EDA (Tabla 2). Tres de ellas se derivan de datos de eventos (goleadores y tandas) y se calculan estrictamente as-of-date mediante merge_asof(backward, allow_exact_matches=False), de modo que cada partido solo observa eventos anteriores a su fecha (sin leakage). La Figura 2 ilustra la dinámica del ELO en cinco selecciones representativas; la Figura 4 muestra colinealidad alta entre elo_diff y ranking_diff (r ≈ 0,88), mientras que el resto de pares se mantiene en niveles moderados.

Tabla 2

Caracteristicas del modelo y su justificacion

Feature	Justificación
elo_diff	ELO local − ELO visitante (ELO sobre historia completa); métrica dinámica superior al ranking FIFA estático [4].
squad_value_diff	$\log(val_L) - \log(val_V)$; proxy de calidad de plantilla.
xg_avg_for	xG prom. a favor: local − visitante; eficiencia ofensiva.
xg_avg_against	xG prom. en contra: local − visitante; solidez defensiva.
travel_distance_diff	dist. visitante−local a la sede (km); fatiga de viaje (no se anula en sede neutral).
ranking_diff	$rank_V - rank_L$; ranking FIFA (refrescado al 2026)
penalty_share_diff	% goles de penal, local − visitante (as-of)
striker_concentration_diff	Herfindahl de goleadores, local − visitante (as-of); dependencia de una figura.

Figura 1

Distribución de elo_diff por clase de resultado (H/E/V). La separación monotónica entre las tres distribuciones respalda el poder predictivo del feature

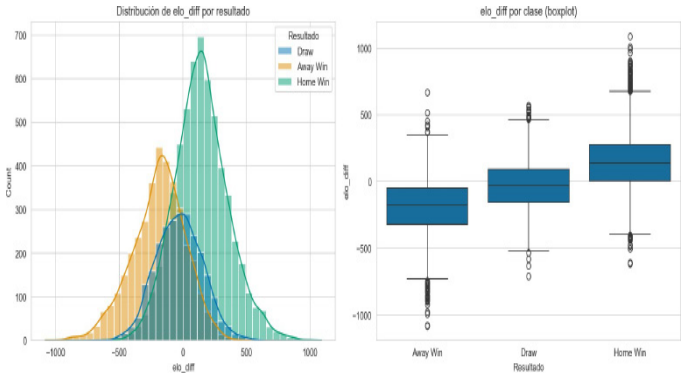
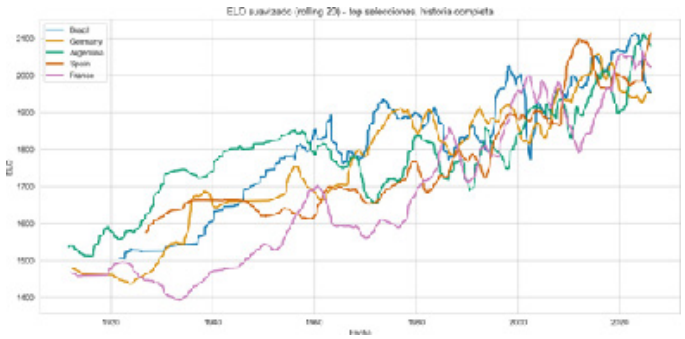


Figura 2

Evolución del rating ELO para cinco selecciones representativas a lo largo de cuatro décadas



Decaimiento temporal y ponderación de clases.

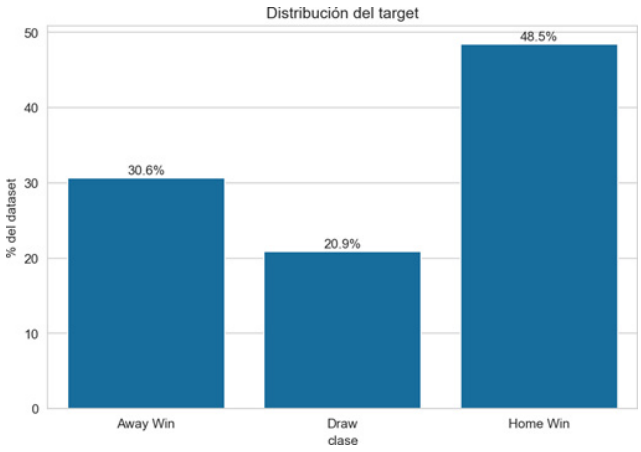
Cada partido contribuye al entrenamiento con un peso

$$W(t) = e^{-\lambda \Delta t} \cdot cy, \quad \lambda = 1 \times 10^{-3} \text{ días}^{-1}, \qquad (1).$$

donde Δt son los días entre la fecha del partido y la más reciente del dataset, y *cyi* es el peso balanced de sklearn sobre la etiqueta *yi* $\{H, E, V\}$. Esto corrige el desbalance natural H ≈ 49 %, E ≈ 21 %, V ≈ 30 % Figura 3 y prioriza la señal reciente. Los pesos se renormalizan por su media para que el aprendizaje permanezca a escala unitaria.

Figura 3

Distribución de la variable objetivo. El desbalance hacia la clase Local motiva la ponderación class_weight= balanced



Modelado

El problema se formula como clasificación multiclase a nivel de partido *{H, E, V }*. Se entrenan tres modelos: Regresión Logística multinomial *(baseline, con StandardScaler), XGBoost (multi:softprob) y LightGBM.* La búsqueda de hiperparámetros emplea Optuna (Transfermarkt, 2026) con 100 trials y *TimeSeriesSplit(n = 5)* interno sobre el conjunto de entrenamiento, optimizando *neg_log_loss*; admite paralelización opcional de *trials (-n-jobs)*, con valor por defecto en serie para garantizar reproducibilidad bit a bit del TPE. La Figura 5 resume la importancia *SHAP* agregada *(mean |SHAP|)* y muestra que *elo_diff* domina (≈1,48), seguido de forma notable por *striker_concentration_diff (≈0,36)* y *ranking_diff (≈0,25)*. La concentración de goleadores, derivada de los datos de eventos, captura un eje (estructura del ataque) ortogonal al nivel del equipo. Las *features de xG y penalty_share_diff forman* un tercer nivel (≈0,10–,12), mientras que *shootout_winrate_diff, squad_value_diff y travel_distance_diff* aportan marginalmente (≈0,09, ,04 y ,03, respectivamente).

División temporal del dataset

Se evita rigurosamente el data leakage mediante un split es-trictamente temporal (Tabla 3). La calibración Platt (sigmoid) del XGBoost se ajusta sobre la franja 2021 (validación) para no contaminar el conjunto de prueba. El método sigmoid es más estable que el isotónico cuando el conjunto de validación es pequeño (un año de datos, tres clases), evitando sobreajuste en la curva de calibración.

Recorte mínimo

El entrenamiento usa partidos con date ≥ 2010, aunque *features.csv* contiene datos desde 1993. Los *features estáticos* (xG, valor de plantilla, ranking moderno) no son representativos del fútbol pre-2010, y el time *decay* con *λ = ,001* ya reduce el peso de esos partidos a menos del 2 %. El ELO sigue calculándose sobre el histórico completo para mayor precisión.

Tabla 3

Slipt temporal y uso de cada conjunto

Conjunto	Filtro	Uso
Train	date < 2021-01-01	LogReg, XGBoost, LightGBM
Validación	2021-01-01 ≤ date < 2022-01-01	Calibración Platt (sigmoid)
Test	date ≥ 2022-01-01	Evaluación final

Doble pipeline

Para validar el Mundial 2022 sin *leakage* se reejecuta el entrenamiento con *cutoff = 2022-01-01*, produciendo *xgboost_pre2022.joblib*. Cuando este cutoff vacía el conjunto de validación temporal, se activa un *fallback* 85/15 que reserva el último 15 % temporal del train para la calibración.

Motor de simulación del torneo

El motor (*src/simulation/*) ejecuta N =10,000 realizaciones del torneo completo (104 partidos × N = 1,04×106 partidos simulados). Cada partido i, j pasa por las siguientes etapas:

1. Probabilidades simétricas. Salvo cuando uno de los equipos es anfitrión jugando en su país, se promedia P (i vs j) con la inversión de P (j vs i) para eliminar el sesgo *home/away* heredado del *dataset*.

2. Marcador y *outcome*. El *outcome {H, E, V }* se sortea con las probabilidades del modelo. En fase de grupos se generan goles *gi Poisson(μi), gj Poisson(μj) con μ basado* en *xG* a favor/en contra, y el marcador se ajusta para ser coherente con el *outcome* (desempates por DG/GF).

3. Fase de grupos. 12 grupos de 4; clasifican los dos primeros más los 8 mejores terceros, con desempates, puntos, diferencia de goles, goles a favor según el reglamento FIFA 2026 (FIFA, 2026).

4. Eliminatorias. Round of 32 Round of 16 Round, Cuartos, Semifinales, Final. Si hay empate tras los 90, la definición por penales se modela como Bernoulli(0,5).

Se contabilizan, porequipo, las veces que alcanza cada ronda y los intervalos de confianza al 95 % se calculan con el método binomial exacto de Clopper–Pearson (C. J. Clopper and E. S. Pearson, 1934) vía *scipy.stats.beta*. Las *N* iteraciones, independientes entre sí, se reparten en bloques fijos ejecutados en paralelo (*joblib*); cada bloque usa un *stream RNG* propio derivado de *SeedSequence(seed).spawn*, de modo que el resultado es idéntico con cualquier número de workers (reproducibili-dad independiente del paralelismo).

Figura 4

Evolución del rating ELO para cinco selecciones representativas a lo largo de cuatro décadas.

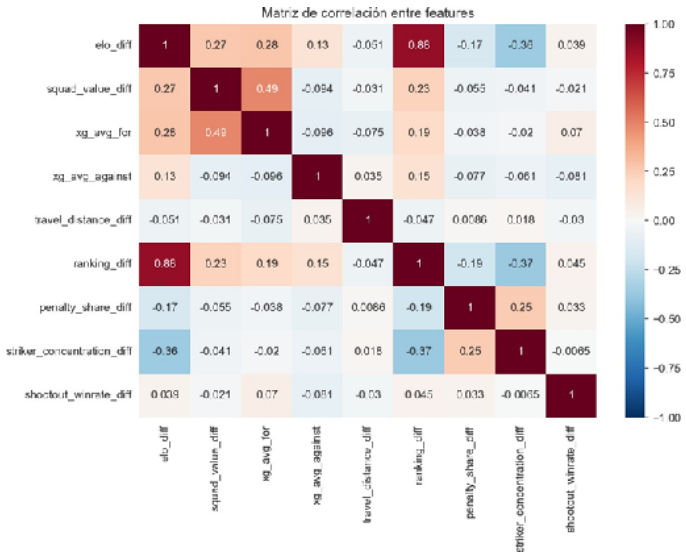
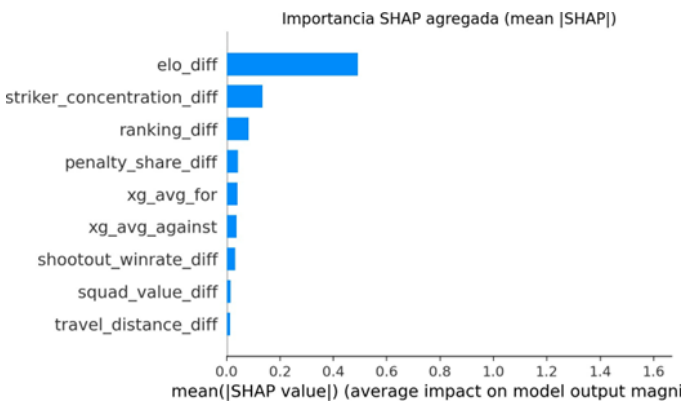


Figura 5

Resumen SHAP agregado (mean |SHAP|) del XGBoost. elo_diff domina la contribución, con striker_concentration_diff en segundo lugar y ranking_diff en tercero



DISCUSIÓN

Evaluación

Se reportan *Log-Loss y Brier score* sobre el conjunto de prueba (date ≥ 2022). Se evita accuracy por ser inadecuada para predicciones probabilísticas desbalanceadas (G. W. Brier, 1950). Nota metodológica sobre el Brier. El valor reportado es el promedio de los tres scores binarios one-vs-rest (cada clase *{H, E, V }* contra el resto), de modo que se sitúa en el rango [0, 1] y no en la formulación multiclase de suma de diferencias cuadráticas (rango [0, 2]). Ambas medidas son equivalentes salvo un factor: el Brier de suma multiclase es K veces el promediado, con K = 3 clases; así, el ,166 reportado corresponde a ,499 en la escala [0, 2]. La Tabla 4 muestra que *XGBoost* calibrado es el mejor modelo en ambas métricas el único que supera a todos los demás simultáneamente en *Log-Loss y Brier*. La calibración Platt sobre datos de validación 2021 corrige la sobreconfianza del *XGBoost* base sin degradar su discriminación (*Log-Loss ,866 ,855*). El *XGBoost* sin calibrar supera a *LightGBM*, y *LogReg* queda en último lugar pese al rebalanceo de clases y el *time decay.*

Tabla 4

Métricas de evaluación sobre test temporal (≥2022)

Modelo	Log-Loss ↓	Brier ↓
XGBoost-Cal	**,8547**	**,1662**
XGBoost	,8657	,1701
LightGBM	,8713	,1717
LogReg	,8832	,1728

Estudio de ablandación

Para cuantificar la contribución marginal de cada grupo de features se reentrena el XGBoost calibrado manteniendo fijos el split temporal, los pesos (time decay + balanced) y los hiperparámetros óptimos; lo único que varía entre filas es el conjunto de features. La Tabla 5 muestra que el grupo de mayor aporte son las features estáticas (xG + valor de plantilla, ΔLL = +0,0041 al retirarlas) seguido de las features derivadas de eventos (+0,0035, impulsado sobre todo por striker_concentration_diff). Las features de aporte marginal (travel_distance_diff, penalty_share_diff y shootout_winrate_diff) se documentan con filas propias: sus Δ caen dentro del ruido de una sola corrida (|ΔLL| ≤ 0,0008). En particular, penalty_share_diff muestra un Δ ligeramente negativo,

indistinguible de cero. Una cuarta feature derivada evaluada (late_goal_ratio, proporción de goles tardíos) se descartó por completo del modelo: su ablación leave-one-out mostró aporte negativo (ruido neto, el modelo mejora sin ella).

Tabla 5

Ablación de Features (XGBOOST CALIBRADO, TEST ≥2022)

Configuración	Log-Loss	Brier	ΔLL	ΔBrier
Completo (9 feat.)	,8547	,1662	—	—
Sin xG	,8587	,1670	+0,0040	+0,0008
Sin squad_value	,8553	,1664	+0,0006	+0,0002
Sin derivadas	,8582	,1668	+0,0035	+0,0006
Sin travel_distance	,8548	,1663	+0,0001	+0,0001
Sin penalty_share	,8542	,1661	−0,0005	−0,0001
Sin shootout	,8555	,1664	+0,0008	+0,0002
Sin estáticas (xG+squad)	,8588	,1670	+0,0041	+0,0008

Validación histórica

El modelo xgboost_pre2022 (entrenado exclusivamente con date<2022-01-01) se evaluá sobre los partidos del Mundial 2022, sirviendo como prueba out-of-time sin contaminación. Los resultados numéricos se imprimen en consola al ejecutar make evaluate.

RESULTADOS DE LA SIMULACIÓN

La Tabla 6 lista el Top-10 de candidatos al título y la Tabla 7 resume la progresión por fase de los tres anfitriones, ambos con intervalos Clopper–Pearson al 95 %.

Análisis de sensibilidad a lesiones

Aunque las lesiones no se modelan estructuralmente, el script src/analysis/sensitivity.py aplica una perturba-ción de −30 % sobre squad_value para cada equipo del top-5 y re-ejecuta la simulación. La Tabla 8 muestra que el efecto es modesto (|Δ| ≤ ,40 pp en todos los casos), siempre en la dirección esperada (reducción de la probabilidad de título).

Esto es coherente con la importancia SHAP: squad_value_diff ocupa una de las últimas posiciones (Figura 5), muy por debajo de elo_diff y de las features de estructura del ataque.

Limitaciones

El modelo presenta cinco limitaciones reconocidas:

1. Debutantes. Equipos como Uzbekistán, Curazao o Cabo Verde tienen historial internacional escaso; su ELO se inicializa con el valor neutro 1500 y converge lentamente, lo que probablemente infla su probabilidad relativa frente a rivales subestimados de la misma confederación.
2. Cobertura xG. El open data de StatsBomb [15] privilegia torneos UEFA/CONMEBOL; las selecciones de AFC, CAF y CONCACAF reciben un xG por defecto de 1,2, que es la media global. Esto subestima la varianza inter-equipos en estas confederaciones.

Tabla 6

Top-10 candidatos a campeón (10 000 iteraciones)

Equipo	P(Campeón)	IC 95 %
España	23,72 %	[22,89, 24,57]
Argentina	11,27 %	[10,66, 11,91]
Francia	10,88 %	[10,28, 11,51]
Brasil	6,35 %	[5,88, 6,85]
Inglaterra	5,55 %	[5,11, 6,02]
Marruecos	4,51 %	[4,11, 4,94]
Alemania	4,18 %	[3,80, 4,59]
México	4,06 %	[3,68, 4,47]
Colombia	3,61 %	[3,25, 3,99]
Países Bajos	3,48 %	[3,13, 3,86]

Tabla 7

Progresion por fase de los anfitriones (%, IC 95 %).

Fase	México	EE.UU.	Canadá
Round of 32	94,5	66,3	96,5
Round of 16	63,3	30,3	53,6
Cuartos	33,5	16,9	25,8
Semifinales	18,4	7,7	11,7
Final	9,4	2,9	5,0
Campeón	4,1	1,1	1,9

Tabla 8

Top 10 candidatos a campeon (10000 iteraciones).

Equipo	Base %	Lesión %	Δ pp
España	23,72	23,38	−0,34
Argentina	11,27	10,87	−0,40
Francia	10,88	10,57	−0,31
Brasil	6,35	6,17	−0,18
Inglaterra	5,55	5,34	−0,21

3. Features estáticas anacrónicas. El xG y el valor de plantilla son snapshots (un valor por equipo, sin fecha) aplicados a todos los partidos históricos, lo que introduce un anacronismo/leakage. No existe serie temporal histórica de estos datos, así que versionarlos sería fabricar información; el efecto se acota con el time decay (< 2 % pre-2010), el recorte a date≥2010 y el uso de diffs relativos, y se cuantifica en la fila "Sin estáticas" de la Tabla V. El valor de Transfermarkt es además una fotografía manual (no live scraping).

4. Cobertura de travel_distance_diff. Las coor-denadas solo están precargadas para los 48 equipos del Mundial; los partidos históricos entre otros equipos quedan en 0 (82,5 % del dataset). La feature ya no se anula en sedes neutrales, corrige el caso de los Mundiales y en la simulación (host_distance) está 100 % poblada.

5. Lesiones de última hora. No hay modelado estructural de bajas individuales; el sensitivity.py es única-mente un stress test agregado sobre squad_value. Una baja clave en una posición específica (e.g., portero titular) tiene un impacto que el modelo no puede capturar.

Reproducibilidad

El repositorio mantiene una arquitectura modular bajo src/ (data/, features/, models/, simulation/, analysis/, visualization/). Los CSVs versionables se producen en data/ processed/ y las figuras en reports/figures/. Toda la cadena se reproduce con:

pip install -e .

make all # features -> train -> evaluate # -> simulate -> sensitivity

streamlit run \ src/visualization/dashboard.py

La cobertura de pruebas (pytest tests/ -v) valida la mecánica del ELO, la lógica de desempates de grupos y el muestreo Poisson de goles.

CONCLUSIONES

Se ha presentado un sistema reproducible para predecir la Copa Mundial 2026 que cumple con los requisitos planteados: múltiples fuentes limpias, nueve características justificadas por EDA y SHAP –tres de ellas derivadas de datos de eventos y calculadas as-of-date–, benchmark de cuatro modelos con calibración Platt sobre validación temporal y un motor Monte Carlo de 10,000 iteraciones fiel al nuevo reglamento. El XGBoost calibrado logra Log-Loss de ,855 y Brier de ,166 — mejor en ambas métricas simultáneamente. Más allá del núcleo elo_diff/ranking_diff, el hallazgo más relevante es que striker_concentration_ diff (estruc-tura del ataque) es la segunda feature más informativa. Las simulaciones ubican a España (23,7 %), Argentina (11,3 %) y Francia (10,9 %) como principales favoritas; los tres anfitriones alcanzan probabilidades de título entre 1,1 % y 4,1 %, con ventaja de local coherente en rondas iniciales pero convergencia en etapas avanzadas.

Trabajo futuro

1. Incorporar un modelo Dixon–Coles (M. J. Dixon and S. G. Coles, 1997) para los marcadores en lugar de Poisson independiente.

2. Reemplazar el snapshot Transfermarkt por un scraper programado y versionar xG/valor de plantilla por temporada para eliminar el anacronismo.

3. Escalar a 105−106 iteraciones mediante vectorización completa del bucle Monte Carlo (ya paralelizado con joblib); y (iv) modelar lesiones a nivel de jugador integrando features de minutos disputados y posición.

Disponibilidad de código y recursos

El código fuente, los scripts de entrenamiento, los datos procesados y los experimentos necesarios para reproducir los resultados de este estudio se encuentran disponibles públicamente en (D. Deras, 2026). Asimismo, se proporciona un dashboard interactivo para la exploración de resultados y simulaciones mediante (D. Deras, 2026).

REFERENCIAS

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2623–2631. https://doi.org/10.1145/3292500.3330701

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078%3C0001:VOFEIT%3E2.0.CO;2

Cashncarry. (2023). FIFA world ranking 1993–2023 [Data set].

Kaggle. https://www.kaggle.com/datasets/cashncarry/fifaworldranking

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26(4), 404–413. https://doi.org/10.1093/biomet/26.4.404

Deras, D. (2026a). Predictions—World Cup 2026 [Computer software]. GitHub. https://github.com/daiv05/fifa-world-cup-model

Deras, D. (2026b). Predictions—World Cup 2026 [Computer application]. Streamlit. https://fifa-world-cup-model.streamlit.app/

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265–280. https://doi.org/10.1111/1467-9876.00065

Elo, A. E. (1978). The rating of chessplayers, past and present. Arco Publishing.

FIFA. (2023). FIFA World Cup 2026 competition format and regulations. https://www.fifa.com/fifaplus/en/tournaments/mens/worldcup/canadamexicousa2026

Groll, A., Ley, C., Schauberger, G., & Van Eetvelde, H. (2019). A hybrid random forest to predict soccer matches in international tournaments. Journal of Quantitative Analysis in Sports, 15(4), 271–287. https://doi.org/10.1515/jqas-2018-0060

Jürisoo, M. (2026a). International football results from 1872 to 2026 (results, goalscorers, shootouts, former_names) [Data set]. GitHub. https://github.com/martj42/international_results

Jürisoo, M. (2026b). International football results from 1872 to 2017 [Data set]. Kaggle. https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Maher, M. J. (1982). Modelling association football scores. Statistica Neerlandica, 36(3), 109–118. https://doi.org/10.1111/j.1467-9574.1982.tb00782.x

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning, 625–632. https://doi.org/10.1145/1102351.1102430

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), Advances in large margin classifiers (pp. 61–74). MIT Press.

StatsBomb. (2024). StatsBomb open data [Data set]. GitHub. https://github.com/statsbomb/open-data

Transfermarkt. (2026). National teams market values [Data set]. https://www.transfermarkt.com