

Football Match Prediction: A Monte Carlo Simulation Approach and Perspectives with Neural Networks

Warner Pereira Cardoso Junior¹, Criscilla Maia Costa Rezende¹,
Dióscoros Brito Aguiar Junior¹

¹Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Jataí (UFJ)
Jataí – GO – Brasil

warner.junior@discente.ufj.edu.br, criscillarezende@ufj.edu.br
dioscoros.junior@ufj.edu.br

Abstract. Introduction: *The unpredictability and low scoring in football make predictive modeling a high-variance challenge. The use of statistical databases, such as Football Manager, opens new avenues for sports analysis. Objective:* *This paper presents preliminary results of a heuristic Monte Carlo simulation applied to two knockout scenarios: the Campeonato Goiano (two-legged) and the 2026 World Cup playoffs (single-elimination), treated as independent case studies. Methodology:* *The Poisson distribution was used within 10,000 Monte Carlo iterations, with a dynamically modeled λ parameter based on each team's relative strength from Football Manager 2026. Two pipelines were developed: an additive linear approach for the Goiano and a proportional ratio approach for the World Cup playoffs. Preliminary Results:* *The model achieved 83.3% accuracy (5/6) for the World Cup playoffs and 71.4% (5/7) for the Goiano, with Brier Scores of 0.2269 and 0.1229, respectively — both below the 0.25 random-guess threshold. These results support the progression towards a comparative study with Artificial Neural Networks and multiple linear regression in the full 2026 World Cup simulation.*

Keywords: *Football, Monte Carlo Simulation, Artificial Neural Networks, Predictive Modeling.*

1. Introduction

Football is the most popular sport in the world, and its unpredictability is one of its main factors of engagement. With a low goal average [Anderson e Sally 2013, Bandara et al. 2024], random events such as red cards and penalties have a disproportionate weight on the outcome [Sumpter 2016], making football a challenging environment for predictive models.

Since the proposition that goals could follow a Poisson distribution [Moroney 1956], the model was consolidated [Maher 1982] and later refined [Dixon e Coles 1997]. In recent years, Data Science and Big Data have transformed high-performance football, guiding from signings to match predictions [Klaiber e Rössle 2026, Memmert e Raabe 2018].

On one side of the methodological spectrum, Artificial Neural Networks (ANN) stand out for their ability to map non-linear relationships in large volumes of data, although they are subject to overfitting [Goodfellow et al. 2016]. On the other hand, stochastic models such as Monte Carlo (MC) Simulation inherently address uncertainty

by simulating thousands of possible scenarios [Metropolis e Ulam 1949]. While Bayesian Networks [Constantinou et al. 2012] and Logistic Regression [Pál e Bíró 2025] are also employed in sports analytics for their interpretability and ability to handle probabilistic dependencies, Monte Carlo Simulation was selected for this initial study because it natively models the stochastic nature of low-scoring matches through repeated sampling, without requiring strong assumptions about feature independence or linear decision boundaries. Both approaches have been applied in football: ANNs are widely used [Elstak et al. 2024], with reported accuracies of 52.8% [Luiz et al. 2024], while MC proves competitive in direct comparisons with other methods [Pál e Bíró 2025].

This paper presents preliminary results of a heuristic MC-based model to simulate the Campeonato Goiano and the 2026 World Cup playoffs, using data from Football Manager 2026 [Sports Interactive 2025]. The model uses Poisson with dynamic λ , parameterized according to data availability for each tournament. Validation was performed using the Brier Score (BS) [Steyerberg et al. 2010], a metric where 0 indicates perfect prediction and 0.25 the randomness threshold. Despite the reduced sample (14 matches in the Goiano, 6 in the playoffs), consistent with the exploratory scope of an undergraduate research project, the results indicate predictive capability and establish a methodological foundation for future work: a comparative study between MC, Artificial Neural Networks (ANN), and multiple linear regression in the complete 2026 World Cup simulation.

2. Materials and Methods

The methodology applied in this work was based on computational statistical modeling, with algorithms developed in Python. The Pandas library was used for data manipulation, and NumPy was used for pseudo-random number generation and iteration processing.

2.1. Data Collection and Construction

To mitigate the scarcity of unified technical-tactical performance data, variables were extracted from the *Football Manager 2026* simulator [Sports Interactive 2025] using FM Scout tools [FM Scout 2025] and the fminside.net portal [FMInside 2025]. The use of this database as a reliable representation of the football ecosystem has robust support in the professional and academic literature [Hocquet 2016], supporting everything from national performance analyses [Costa e Rocco Jr. 2015] and predictive recruitment models [Nunes 2025] to large-scale international tournament simulations [da Silva et al. 2026].

The data were structured at two levels of granularity adapted to the case studies. For the Campeonato Goiano, macrostructural team variables were collected: Global Reputation, Training Level, and Average Attendance. For the World Cup Playoffs, a microattribute-based approach was used, extracting Current Ability (CA) and Potential Ability (PA) metrics from the 16 main athletes of each national team (1 goalkeeper, 5 defenders, 5 midfielders, and 5 forwards), in addition to the FIFA ranking and coaches' technical rating. The formulation that consolidates these variables into the λ parameter is detailed in Subsection 2.3.

2.2. The Poisson Distribution in the Sporting Context

The statistical modeling of results assumed that scoring goals in a time interval follows a stochastic counting process. The model uses the Poisson distribution [Maher 1982, Dixon e Coles 1997] to calculate the probability of a team scoring a discrete number of goals k , where λ represents the expected average goal rate for that team:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

2.3. Dynamic Formulation of the Lambda Parameter (λ)

The parametrization of λ was developed in two pipelines adapted to the data granularity: (i) a linear additive approach for the Campeonato Goiano (macrostructural data and historical performance) and (ii) a proportional ratio approach for the World Cup Playoffs (CA/PA microattributes).

In the additive approach, λ_i is computed as a convex combination of six normalized scores:

$$\lambda_i = w_d S_{i,d} + w_o S_{i,o} + w_f S_{i,f} + w_q S_{i,q} + w_t S_{i,t} + w_m S_{i,m},$$

with weights $w_d = 0.15$ (recent performance), $w_o = 0.05$ (attack), $w_f = 0.05$ (defense), $w_q = 0.55$ (reputation), $w_t = 0.05$ (infrastructure), and $w_m = 0.15$ (home advantage). The scores are: $S_{i,d}$ (average points, wins, and goal difference per game); $S_{i,o}$ (goals scored per game); $S_{i,f} = [\max(\text{GoalsConceded}_i / \text{Games}_i, 0.1)]^{-1}$ (inverse of goals conceded); $S_{i,q}$ (FM reputation); $S_{i,t}$ (training facilities); and $S_{i,m} = 1 + \text{AvgAttendance}_i / 20000$ for the home team and 0 for the away team. The resulting λ_i is clamped to $[0.2, 6.0]$ to preserve unpredictability and prevent implausible scorelines.

In the proportional ratio approach, Score is computed as a convex combination of team strength components:

$$\text{Score} = 0.45 \cdot V_1 + 0.15 \cdot V_2 + 0.15 \cdot V_3 + 0.10 \cdot V_5 + 0.05 \cdot V_6,$$

where V_1 is the positional-weighted average of the 16 main athletes (1 GK, 5 defenders, 5 midfielders, 5 forwards) from CA/PA; V_2 is the coach rating; V_3 is the FIFA ranking score; V_5 is the average of the top 3 players; and V_6 is the average potential. The final score incorporates home advantage as a discrete scalar: $\text{Score}_{\text{final}} = \text{Score} + 0.10 \cdot V_4$, with $V_4 = 100$ (home), 50 (neutral), or 0 (away). The Poisson parameters are then coupled by the ratio between final scores:

$$\lambda_A = 1.35 \cdot \frac{\text{Score}_A}{\text{Score}_B}, \quad \lambda_B = 1.35 \cdot \frac{\text{Score}_B}{\text{Score}_A}$$

where $\mu = 1.35$ is the average goals per team (2.7 goals/game), consistent with the literature [Maher 1982, Dixon e Coles 1997, Bandara et al. 2024]. The fractional ratio eliminates the need for truncation.

For penalties, the probability of victory is:

$$P_{\text{pen},A} = 0.20 + 0.60 \cdot \frac{v_{1,A}}{v_{1,A} + v_{1,B}} + 0.20 \cdot \frac{\text{GKRate}_A}{\text{GKRate}_A + \text{GKRate}_B} \quad (1)$$

with 20% randomness, 60% squad strength (v_1), and 20% goalkeeper rating. For the Goiano, $P = 0.50$ due to the lack of individual data. Additional contextual factors (weather, injuries, pressure, refereeing and recent match form) were not modeled due to unavailability of structured data; their variability is partially captured by the stochastic nature of the simulation.

2.4. Monte Carlo Simulation Dynamics

The stochastic engine executed $N = 10,000$ replicates per scenario. For each matchup, Poisson vectors $\mathbf{X}_A, \mathbf{X}_B \in \mathbb{N}^N$ are generated with $X_{A,m} \sim \text{Poisson}(\lambda_A)$ and $X_{B,m} \sim \text{Poisson}(\lambda_B)$, where λ_A and λ_B are computed as described in Section 2.3. The probability of team A advancing is:

$$P_{\text{adv},A} = \frac{1}{N} \sum_{m=1}^N \mathbb{I}(X_{A,m} > X_{B,m}) + \left[\frac{1}{N} \sum_{m=1}^N \mathbb{I}(X_{A,m} = X_{B,m}) \right] \cdot P_{\text{pen},A}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $P_{\text{pen},A}$ comes from Eq. 1. For two-legged ties, goals from both legs are summed per iteration before applying the same rule. The simulation runs sequentially through all knockout stages, with N iterations per stage. At each stage, the probability of team T advancing is:

$$P_{\text{advance},T}^{(f)} = \frac{1}{N} \sum_{m=1}^N \mathbb{I}(\text{winner}_f(T, m) = T)$$

At the end of each stage, probabilities were compared to actual results for Brier Score calculation. Simulation stability was assessed via the standard error $\sqrt{p(1-p)/N}$, which with $N = 10,000$ yields a maximum theoretical standard deviation of 0.5%.

2.5. Declaration of Artificial Intelligence Use

For the preparation of this manuscript, the Gemini 3.1 Pro tool was used. The use of the artificial intelligence tool occurred exclusively for the following purposes: language review (improvement of grammar and spelling) and translation (conversion of parts of the manuscript into another language). All supervision, validation, and intellectual authorship of the manuscript are the sole responsibility of the human authors.

3. Results and Discussion

The simulation of the 2026 World Cup playoff matches and the final stages of the 2026 Campeonato Goiano allowed us to evaluate the model's predictive capability. The generated probabilities were validated by nominal accuracy and Brier Score (BS).

3.1. Validation in Single-Elimination Matches (World Cup Playoffs)

Table 1 compares the model's probabilities with the actual results.

Table 1. Comparison between Monte Carlo Predictions and Actual Results (Playoff Finals).

Slot (Bracket)	Model Probability	Actual Match Result	Prediction
Slot 1 (UEFA D)	Czech Rep. (52.3%) vs Denmark (47.7%)	Czech Rep. 2 x 2 Denmark (3-1 Pen.)	Correct
Slot 2 (UEFA A)	Bosnia (43.9%) vs Italy (56.1%)	Bosnia 1 x 1 Italy (4-1 Pen.)	Incorrect
Slot 3 (UEFA B)	Sweden (60.4%) vs Poland (39.6%)	Sweden 3 x 2 Poland	Correct
Slot 4 (UEFA C)	Kosovo (45.7%) vs Turkey (54.3%)	Kosovo 0 x 1 Turkey	Correct
Slot 5 (World)	DR Congo (54.2%) vs Jamaica (45.8%)	DR Congo 1 x 0 Jamaica	Correct
Slot 6 (World)	Iraq (50.6%) vs Bolivia (49.4%)	Iraq 2 x 1 Bolivia	Correct

The model achieved 83.3% accuracy (5/6), with $BS = 0.2269$ (a 9.2% improvement over the 0.25 threshold). The error in Slot 2 (Bosnia-Italy) reflects the high variance of the format: a balanced matchup (56.1%) decided on penalties, whose 80% ceiling was respected (Eq. 1).

3.2. Validation in Two-Legged Matches (Campeonato Goiano)

Table 2 shows the results of the knockout stage.

Table 2. Comparison between Predictions and Actual Standings (Goianão Knockout Stage).

Matchup (Round)	Model Favorite (Adv. Probability)	Qualified Team	Prediction
Quarter-finals	Goiás (92.2%) vs CRAC (7.8%)	Goiás	Correct
Quarter-finals	Atlético-GO (86.6%) vs ABECAT (13.4%)	Atlético-GO	Correct
Quarter-finals	Vila Nova (86.1%) vs Anápolis (13.9%)	Vila Nova	Correct
Quarter-finals	Jataiense (52.7%) vs Anapolina (47.3%)	Anapolina	Incorrect
Semi-finals	Goiás (67.8%) vs Anapolina (32.2%)	Goiás	Correct
Semi-finals	Vila Nova (54.4%) vs Atlético-GO (45.6%)	Atlético-GO	Incorrect
Final	Goiás (62.6%) vs Atlético-GO (37.4%)	Goiás (Champion)	Correct

The model achieved 71.4% accuracy (5/7), with $BS = 0.1229$, considerably lower than for the World Cup, validating the hypothesis that the two-legged format (180 min) reduces Poisson variance, allowing long-term strength to prevail. Errors occurred in matchups with probabilities close to 50% (Jataiense and Vila Nova), indicating that the model captured the inherent uncertainty of these games. The model correctly predicted the champion (Goiás) from the quarter-finals (54.8%) and in the final (62.6%).

The two scenarios are not directly comparable due to their distinct formats and data granularity, serving as independent case studies.

4. Final Considerations

The preliminary results indicate that Monte Carlo (MC) simulation, combined with the Poisson distribution and powered by Football Manager data, is a promising tool for predictive modeling in football. The dynamic formulation of the λ parameter proved effective in balancing statistical rigor and sporting realism across different data granularities.

The Brier Scores obtained — 0.2269 (World Cup) and 0.1229 (Goiano) — surpass the 0.25 randomness threshold, confirming that the model captures relevant information. The difference between scenarios validates the hypothesis that the two-legged format acts as a statistical filter, reducing Poisson variance. We acknowledge that the limited number of matches (6 and 7) constrains inference, an aspect inherent to the exploratory scope of this undergraduate research.

The heuristic approach establishes a methodological foundation for a comparative study in the 2026 World Cup context, testing three frameworks on the same dataset: the current stochastic MC model with fixed weights, MC with parameters estimated by multiple linear regression, and supervised ANNs for direct outcome prediction. The model is primarily designed for knockout tournaments; adaptation for league applications, where draws are standalone outcomes, remains as future work.

References

- Anderson, C. e Sally, D. (2013). *The Numbers Game: Why Everything You Know About Soccer Is Wrong*. Penguin Books.
- Bandara, I., Shelyag, S., Rajasegarar, S., Dwyer, D., Kim, E.-J., e Angelova, M. (2024). Winning with chaos in association football: Spatiotemporal event distribution randomness metric for team performance evaluation. *IEEE Access*, 12:83363–83376.
- Constantinou, A. C., Fenton, N. E., e Neil, M. (2012). pi-football: A bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 36:322–339.
- Costa, A. Q. d. e Rocco Jr., A. J. (2015). Jogos digitais e a gestão do esporte: o caso football manager. In *Anais do XIX Congresso Brasileiro de Ciências do Esporte e VI Congresso Internacional de Ciências do Esporte*, Vitória. Colégio Brasileiro de Ciências do Esporte.
- da Silva, L. P., Schincariol, T., e Wack, M. (2026). Football empires: Simulating the effects of colonialism on the 2026 world cup. Technical report, Center for Open Science.
- Dixon, M. J. e 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.
- Elstak, I., Salmon, P., e McLean, S. (2024). Artificial intelligence applications in the football codes: A systematic review. *Journal of sports sciences*, 42(13):1184–1199.
- FM Scout (2025). Fmscout editor 2026. FMScout.com, Software Tool.
- FMInside (2025). Football manager database and player ratings.
- Goodfellow, I., Bengio, Y., e Courville, A. (2016). *Deep Learning*. MIT Press.
- Hocquet, A. (2016). Football manager: Mutual shaping between game, sport, and community. *Journal of Media Studies and Popular Culture = Revue d'études des médias et de culture populaire*, 6. Special Issue.
- Klaiber, M. e Rössle, M. (2026). Uncovering data-driven football: Topic modeling from a data analytics perspective. *Journal of Sports Sciences*, pages 1–35.
- Luiz, L. E., Fialho, G., e Teixeira, J. P. (2024). Is football unpredictable? predicting matches using neural networks. *Forecasting*, 6(4):1152–1168.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.
- Memmert, D. e Raabe, D. (2018). *Data Analytics in Football: Positional Data Collection, Modelling and Analysis*. Routledge.
- Metropolis, N. e Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341.
- Moroney, M. J. (1956). *Facts from figures*, volume 236. Penguin books Baltimore, MD.

- Nunes, F. A. d. S. F. (2025). *Modelos preditores de contratação de profissionais do futebol baseado em Inteligência Artificial por mecanismos Machine Learning*. Tese (doutorado em administração), Universidade Federal de Minas Gerais, Belo Horizonte.
- Pál, J. G. e Bíró, C. (2025). Evaluating profitability in sports betting using probabilistic models and betting strategies. In *Annales Mathematicae et Informaticae*, volume 61, pages 202–214.
- Sports Interactive (2025). Football manager 2026. PC/Mac/Mobile/Console.
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., e Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, 21(1):128–138.
- Sumpter, D. (2016). *Soccermaths: Mathematical Adventures in the Beautiful Game*. Bloomsbury Publishing.