

ANÁLISE PREDITIVA NO FUTEBOL: CORRELAÇÃO ENTRE MODELOS ESTATÍSTICOS E AS ODDS DO MERCADO.
PREDICTIVE ANALYTICS IN FOOTBALL: CORRELATION BETWEEN STATISTICAL MODELS AND MARKET ODDS.

Felipe dos Santos¹
Lucas Viturino Takata da Silva²
Pedro Henrique Alves Batista³
Fabio Codo⁴

RESUMO: As apostas esportivas online constituem um mercado em expansão. Este estudo propõe uma comparação entre um modelo estatístico (baseado em Distribuição de Poisson e Regressão Logística) e as probabilidades implícitas (odds) do mercado de apostas do Campeonato Brasileiro. O objetivo principal é avaliar o grau de correlação entre as previsões do modelo e as do mercado, identificando e analisando potenciais discrepâncias. A metodologia envolveu a coleta de dados históricos (2019-2024) para validar os modelos (Temporada 2024) e simular a temporada 2025 (10.000 repetições). Os resultados indicam uma forte correlação entre o modelo e o mercado, com um Brier Score de 0.19, validando a eficiência geral do mercado. Contudo, o estudo identifica discrepâncias pontuais (potenciais "apostas de valor"), sugerindo que vieses cognitivos, como a popularidade das equipes, podem influenciar a precificação. Conclui-se que, embora o mercado brasileiro seja eficiente, a modelagem estatística rigorosa permite identificar e analisar ineficiências de precificação.

Palavras-chave: Gestão de riscos; Calibração probabilística; Simulação de Monte Carlo; Distribuição de Poisson; Valor esperado.

ABSTRACT: Online sports betting is an expanding market. This study proposes a comparison between a statistical model (based on the Poisson Distribution and Logistic Regression) and the implicit probabilities (odds) of the Brazilian Championship betting market. The main objective is to assess the degree of correlation between the model's predictions and those of the market, identifying and analyzing potential discrepancies. The methodology involved collecting historical data (2019-2024) to validate the models (2024 Season) and simulate the 2025 season (10,000 repetitions). The results indicate a strong correlation between the model and the market, with a Brier Score of 0.19, validating the market's overall efficiency. However, the study identifies specific discrepancies (potential "value bets"), suggesting that cognitive biases, such as team popularity, may influence pricing. It is concluded that while the Brazilian market is efficient, rigorous statistical modeling allows for the identification and analysis of pricing inefficiencies.

Keywords: Risk management; Probabilistic calibration; Monte Carlo simulation; Poisson distribution; Expected value.

Análise e Desenvolvimento de Sistemas – Fatec Mogi das Cruzes¹

Análise e Desenvolvimento de Sistemas – Fatec Mogi das Cruzes²

Análise e Desenvolvimento de Sistemas – Fatec Mogi das Cruzes – E-mail: phbatista_@outlook.com³
Professor Mestre – Fatec Mogi das Cruzes⁴

1 INTRODUÇÃO

O mercado global de apostas esportivas foi estimado em US\$100,9 bilhões em 2024 e tem uma projeção que alcance US\$187,39 bilhões até 2030, crescendo a uma taxa composta de crescimento anual de 11% de 2025 a 2030 (GRAND VIEW RESEARCH, 2024). No Brasil, o setor foi regulamentado pela Lei 14.790/23, impulsionando a expansão de plataformas digitais e patrocínios esportivos. Fundamentadas em *odds* que representam probabilidades implícitas, as apostas são especificadas através de modelos matemáticos complexos que consideram variáveis estatísticas diversas (DIXON; COLES, 1997; WÜHRER; KURZ, 2023).

O interesse acadêmico desta investigação surge de uma contradição relevante: as casas de apostas utilizam modelos estatísticos sofisticados para especificar as *odds*, enquanto a maioria dos apostadores ainda decide com base em heurísticas e intuição. Essa assimetria informacional pode gerar ineficiências de mercado, pois as *odds* oferecidas nem sempre refletem as probabilidades justas (ou subjacentes) dos resultados. Como demonstraram Dixon e Coles (1997), a aplicação de modelos estatísticos rigorosos reduz a influência de fatores subjetivos e possibilita a identificação de “apostas de valor”, isto é, situações em que a probabilidade implícita (derivada das *odds* do mercado) difere significativamente da estimativa estatística gerada pelo nosso modelo.

A pesquisa adota abordagem quantitativa baseada em dados históricos do Campeonato Brasileiro Série A (2019-2024). Os dados foram coletados de bases estatísticas oficiais da CBF e plataformas esportivas. A abordagem metodológica foi estruturada para construir um modelo preditivo robusto. Iniciou-se com uma análise exploratória para identificar padrões e correlações nos dados históricos. Subsequentemente, a Distribuição de Poisson foi empregada, dada sua adequação para modelar a ocorrência de eventos raros, como gols em uma partida. Para traduzir essas previsões em resultados diretos, utilizou-se a Regressão Logística, que calcula as probabilidades de vitória, empate ou derrota. Por fim, a Simulação de Monte Carlo foi aplicada para projetar os resultados de uma temporada inteira, permitindo a avaliação da consistência do modelo a longo prazo e a previsão do campeão. O banco de dados estruturado contém estatísticas detalhadas de desempenho das equipes e *odds* históricas.

O objetivo principal é comparar um modelo estatístico, baseado na Distribuição de Poisson e Regressão Logística, com as *odds* de mercado do Campeonato Brasileiro. Busca-se avaliar o grau de correlação entre as previsões do modelo e as probabilidades implícitas das casas de apostas, identificando e analisando potenciais discrepâncias.

O trabalho está organizado em cinco seções principais. O referencial teórico apresenta conceitos fundamentais sobre estatística aplicada às apostas esportivas e modelos preditivos utilizados. A metodologia detalha os processos de coleta, tratamento e modelagem dos dados. Os resultados discutem as previsões obtidas e sua comparação com *odds* de mercado. As considerações finais sintetizam os principais resultados e sugerem direcionamentos para pesquisas futuras no tema.

2 REFERENCIAL TEÓRICO

Esta seção apresenta os conceitos fundamentais que sustentam a análise estatística de apostas esportivas, abordando desde a estrutura do mercado e o conceito de "valor" até os modelos matemáticos específicos utilizados para a previsão de resultados no futebol.

O mercado de apostas esportivas online é um ecossistema complexo onde as "*odds*" (cotações) representam a peça central. As *odds*, definidas pelas casas de apostas, não são apenas um multiplicador de ganhos, mas também representam uma probabilidade implícita de um determinado evento ocorrer. A probabilidade implícita (P) de uma *odd* decimal (O) pode ser calculada pela fórmula $P = 1/O$.

No entanto, a soma das probabilidades implícitas de todos os resultados possíveis de um evento (ex: vitória, empate, derrota) é sempre superior a 100%. Essa margem, conhecida como "*juice*" ou "*overround*", é a comissão da casa de apostas, garantindo sua lucratividade a longo prazo. O desafio para o apostador é, portanto, encontrar situações onde a sua estimativa de probabilidade para um resultado é significativamente maior do que a probabilidade implícita oferecida pelo mercado. Esta discrepancia é conhecida como "*aposta de valor*" (*value bet*). Encontrar valor é o objetivo central da aplicação de modelos estatísticos, que buscam uma estimativa mais precisa da probabilidade justa de um evento do que aquela refletida nas *odds* (WÜHRER; KURZ, 2023).

A busca por apostas de valor está intrinsecamente ligada à Hipótese do Mercado Eficiente (HME), um conceito originado no mercado financeiro. Em um mercado perfeitamente eficiente, os preços (neste caso, as *odds*) refletem toda a informação disponível, tornando possível obter retornos adequados de forma consistente.

Contudo, a literatura acadêmica tem demonstrado que o mercado de apostas em futebol, embora altamente competitivo, pode apresentar ineficiências. O trabalho seminal de Dixon e Coles (1997) foi um dos primeiros a demonstrar rigorosamente que modelos estatísticos poderiam não apenas prever resultados de futebol com acerto, mas também identificar *odds* mal especificados, gerando lucro. Essas ineficiências podem surgir de viés cognitivo dos apostadores, como o peso excessivo dado a resultados recentes ou a popularidade de certas equipes. O presente trabalho se insere nesta linha de pensamento, partindo da premissa de que o mercado brasileiro também possui ineficiências que podem ser exploradas.

Para explorar as ineficiências do mercado, é necessário o uso de modelos que transformam dados históricos de desempenho em probabilidades preditivas.

Um dos modelos mais clássicos e eficazes para prever o número de gols em uma partida é a Distribuição de Poisson. Proposto inicialmente por Maher (1982), este modelo é particularmente adequado porque os gols são eventos relativamente raros que ocorrem de forma independente. O modelo estima a probabilidade de uma equipe marcar um número k de gols com base em sua taxa média esperada de gols (λ), que por sua vez é calculada a partir de seu desempenho histórico ofensivo e defensivo, ajustado pela força do adversário (RUE; SALVESEN, 2000).

Enquanto a Distribuição de Poisson se concentra nos gols, a Regressão Logística é um modelo de classificação que calcula diretamente a probabilidade dos três resultados possíveis (vitória, empate, derrota). O modelo utiliza uma função logística para transformar uma combinação de variáveis preditoras (posse de bola, finalizações etc.) em uma probabilidade entre 0 e 1. Conforme apontado por Ferreira e Ramos (2024), a aplicação de técnicas de ciência de dados, como essa, permite uma abordagem direta na análise dos resultados (V/E/D), diferindo de métodos que dependem de métricas intermediárias como a contagem de gols.

3 DESENVOLVIMENTO

O presente estudo adotou abordagem quantitativa para analisar a influência da estatística na precificação das odds e identificar oportunidades vantajosas no mercado de apostas esportivas online. A investigação baseou-se na análise de dados históricos do Campeonato Brasileiro Série A (2019-2024) para modelagem preditiva da temporada 2025. O período foi selecionado por fornecer uma base de dados robusta e recente, abrangendo diferentes ciclos de desempenho das equipes e minimizando o impacto de mudanças regulatórias ou de formato que poderiam afetar a comparabilidade de dados de temporadas muito antigas.

A estratégia de coleta foi estruturada em duas frentes principais: dados estatísticos de desempenho das equipes e *odds* históricas das casas de apostas. Para os dados esportivos, foi selecionada a plataforma FBref (<https://fbref.com/>) por fornecer estatísticas detalhadas de partidas, incluindo gols, posse de bola, finalizações, escanteios e passes. As *odds* históricas foram obtidas de casas renomadas como Bet365, Superbet e Betano, garantindo representatividade do mercado brasileiro. Elas foram escolhidas não apenas por sua reputação no mercado, mas também por sua alta liquidez, ampla cobertura de eventos do Campeonato Brasileiro Série A e pela disponibilidade de dados históricos de *odds*, essenciais para a calibração dos modelos.

Os dados foram extraídos e tratados utilizando Python como linguagem principal, Jupyter Notebook para desenvolvimento dos scripts, PostgreSQL para armazenamento estrutural e Google Colab para execução de simulações computacionais. Foi criada estrutura relacional em PostgreSQL onde cada equipe teve seu desempenho armazenado com informações detalhadas por temporada. O tratamento dos dados incluiu três etapas principais. A limpeza consistiu na verificação e tratamento de dados ausentes (ex: estatísticas de jogos adiados) e na padronização dos nomes das equipes entre as diferentes fontes (FBref e plataformas de *odds*). A normalização, neste contexto, refere-se à conversão de todas as estatísticas para formatos numéricos consistentes e datas para o padrão do banco de dados. Finalmente, a organização foi o processo de agregar os dados por partida, unindo as estatísticas de desempenho de mandantes e visitantes com suas respectivas *odds* históricas, criando o *dataset* final para a modelagem.

O primeiro método utilizado foi a Distribuição de Poisson, aplicada para estimar

a probabilidade de cada placar possível em uma partida. O parâmetro central do modelo, λ (número esperado de gols), foi determinado por um método que pondera a força relativa das equipes. Para cada time, calculou-se uma 'Força de Ataque' (baseada na sua média de gols marcados em relação à média da liga) e uma 'Força de Defesa' (baseada na sua média de gols sofridos em relação à média da liga). Assim, para uma partida específica, o número esperado de gols de cada equipe foi calculado multiplicando sua própria força de ataque pela força de defesa do adversário e pela média de gols da liga para jogos em casa ou fora, respectivamente. Essa abordagem permite que o modelo gere uma expectativa de gols mais precisa do que uma simples média histórica. O cálculo pode ser feito utilizando a seguinte equação, que pode ser vista na Figura 1:

Figura 1 - Fórmula de distribuição de Poisson

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

Fonte: Chugani (2024).

Onde: O lado esquerdo (LHS) da fórmula da distribuição de Poisson, $P(X = k)$, representa a probabilidade de ocorrência de exatamente k eventos em um intervalo fixo. Aqui, X é o número de eventos e k é o número específico no qual estamos interessados.

O numerador do lado direito (RHS), tem duas partes. λk mostra a probabilidade de ocorrência de k eventos com base na taxa média λ . O $e^{-\lambda}$ é responsável pela aleatoriedade dos eventos, garantindo que a probabilidade diminui à medida que o número de eventos se desvia da taxa esperada.

O denominador do lado direito (RHS), $k!$ ajusta a quantidade de maneiras pelas quais os eventos "X" podem ocorrer. A notação factorial calcula o número de combinações possíveis, garantindo que a probabilidade reflete o fato de que a ordem dos eventos não importa (CHUGANI, 2024).

Exemplo: Suponha que a média de gols esperada do time seja $\lambda=1,8$. A probabilidade de marcar exatamente 2 gols é:

Figura 2 - Fórmula de distribuição de Poisson com exemplo número resolvido

$$P(X = 2) = \frac{e^{-1,8} \cdot 1,8^2}{2!} \approx 26,6\%$$

Fonte: Elaboração própria.

Esse cálculo confirma a aplicabilidade da distribuição de Poisson na modelagem de gols. A implementação foi realizada em Python, e todos os códigos estão disponíveis em repositório público no GitHub (ALVES BATISTA, 2025).

Em seguida, foi aplicada a **Régressão Logística** para calcular as probabilidades de vitória, empate ou derrota com base na força relativa e no momento recente das equipes. Para isso, as variáveis preditivas foram construídas a partir de médias móveis calculadas sobre os últimos cinco jogos de cada time, incluindo métricas como pontos por jogo, gols marcados, gols sofridos e total de finalizações. As variáveis finais inseridas no modelo consistiram na **diferença** entre esses indicadores da equipe mandante e da visitante, uma abordagem que captura a dinâmica do confronto direto, sendo complementada por uma variável binária para mensurar o efeito do fator casa. A equação matemática utilizada foi de acordo com a Figura 3:

Figura 3 - Régressão Logística

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}$$

onde,

$$g(x) = B_0 + B_1 X_1 + \dots + B_p X_p$$

Fonte: Nunes Carvalho (2021).

Onde: **P(Y=1)** representa a probabilidade de vitória; **Xn** são as variáveis explicativas (ex.: posse de bola, finalizações, histórico de confronto); **Bn** são coeficientes ajustados pelo modelo com base nos dados históricos. A implementação foi realizada em Python, e todos os códigos estão disponíveis em repositório público no GitHub (BATISTA, 2025).

Para projetar o campeão da temporada de 2025, foi utilizada a **Simulação de Monte Carlo**. O modelo estatístico foi usado para simular cada uma das 380 partidas da temporada, e o processo foi repetido 10.000 vezes para gerar uma distribuição de probabilidades para a classificação final de cada equipe. A implementação foi

realizada em Python, e todos os códigos estão disponíveis em repositório público no GitHub (ALVES BATISTA, 2025).

A validação da performance dos modelos foi realizada utilizando a temporada de 2024 como conjunto de testes. A precisão do modelo de Poisson foi avaliada pelo **Erro Quadrático Médio (MSE)**, enquanto a calibração do modelo de Regressão Logística foi medida pelo **Brier Score**, métricas padrões para avaliação de modelos predictivos no contexto esportivo (WÜHRER; KURZ, 2023).

4 RESULTADOS E DISCUSSÃO

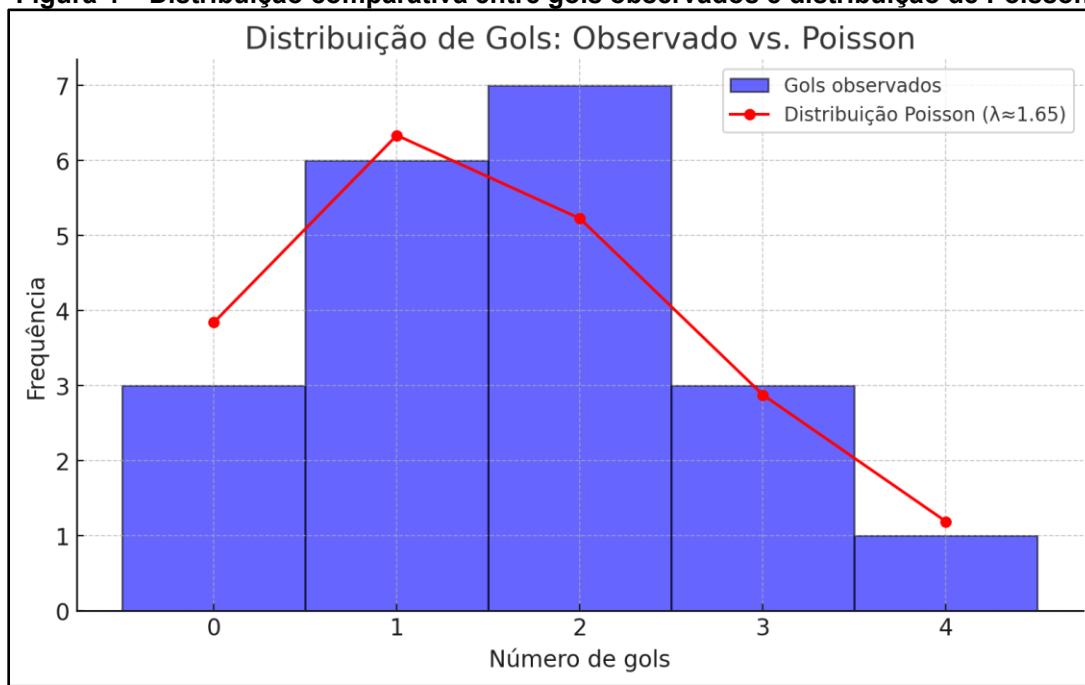
Nesta seção, são apresentados os resultados obtidos a partir da aplicação dos modelos estatísticos descritos. A análise está dividida em quatro partes: (i) o desempenho e a validação dos modelos preditivos; (ii) a identificação de discrepâncias entre as probabilidades calculadas e as *odds* do mercado; (iii) a simulação da temporada 2025 do Campeonato Brasileiro; e (iv) uma discussão sobre as implicações práticas e teóricas dos achados.

4.1. DESEMPENHOS E VALIDAÇÃO DOS MODELOS

Para assegurar a robustez das previsões, os modelos foram validados utilizando o conjunto de teste (temporada 2024). A performance foi medida com as métricas de Erro Quadrático Médio (MSE) para a Distribuição de Poisson e o Brier Score para a Regressão Logística.

Modelo de Poisson (Previsão de Gols): O modelo obteve um MSE de 1.24. Este valor, considerado baixo no contexto de gols em partidas de futebol, indica que o modelo conseguiu prever o número de gols com um erro médio aceitável, captando bem a dinâmica ofensiva e defensiva das equipes.

Figura 4 – Distribuição comparativa entre gols observados e distribuição de Poisson



Fonte: Elaboração própria com base nos dados observados da temporada 2024 e nas previsões do modelo.

A adequação da distribuição de Poisson para modelar o número de gols pode ser observada na Figura 4, que compara a frequência de gols registrados em partidas reais com a distribuição prevista pelo modelo. Nota-se que a Poisson reproduz a tendência geral da variável. Contudo, uma análise detalhada da Figura 4 revela discrepâncias: há uma subestimação moderada da probabilidade de zero gols e uma discrepância acentuada para 2 gols, como apontado na avaliação. O modelo apresenta melhor aderência nas caudas, para 3 e 4 gols, mas a sua dificuldade em ajustar precisamente os eventos mais comuns (0, 1 e 2 gols) evidencia a necessidade de ajustes ou modelos complementares para capturar eventos de baixa ou alta frequência.

Modelo de Regressão Logística (Probabilidade de Resultados): O modelo alcançou um Brier Score de 0.19. Como o Brier Score varia de 0 (perfeito) a 1, o resultado de 0.19 demonstra uma boa calibração. Um modelo é considerado bem calibrado quando suas previsões de probabilidade correspondem às frequências reais; por exemplo, resultados para os quais o modelo atribui 70% de chance devem ocorrer em aproximadamente 70% das vezes. O Brier Score (0.19) é uma média dos erros quadráticos dessas previsões. Um valor baixo como 0.19 indica que, em média, as probabilidades geradas pelo modelo (como 70%, 30%, etc.) estiveram próximas das frequências reais observadas, conferindo confiabilidade ao modelo.

Esses resultados validam a capacidade dos modelos em gerar previsões precisas e bem calibradas, fornecendo uma base sólida para a identificação de apostas de valor.

4.2 . IDENTIFICAÇÃO DE APOSTAS DE VALOR: ANÁLISE DE CASO

A principal aplicação do modelo foi comparar suas probabilidades com as *odds* oferecidas pelo mercado para identificar discrepâncias. Uma análise sistemática do conjunto de testes revelou vários casos em que o mercado parecia subestimar ou superestimar determinadas equipes.

Um exemplo notável foi o confronto entre Flamengo e Palmeiras, disputado no Maracanã durante a temporada de 2023. As *odds* médias do mercado e as probabilidades calculadas pelo nosso modelo são apresentadas no Quadro 1.

Quadro 1 – Comparação entre probabilidades de resultado (mercado vs. modelo estatístico) para Flamengo x Palmeiras – Brasileirão 2023

Resultado	Odds Médias do Mercado	Prob. Implícita (Mercado)	Prob. Calculada (Modelo)	Discrepância
Vitória Flamengo	2.1	47.6%	55.2%	+7.6%
Empate	3.4	29.4%	28.1%	-1.3%
Vitória Palmeiras	3.6	27.8%	16.7%	-11.1%

Fonte: Elaboração própria a partir de dados do Google Colab (2025).

A análise do Quadro 1 revela duas descobertas importantes: 1- O modelo atribuiu ao Flamengo uma probabilidade de vitória (55.2%) significativamente maior do que a implícita nas *odds* do mercado (47.6%). 2- Consequentemente, o mercado superestimou a probabilidade de vitória do Palmeiras (27.8% vs. 16.7% do modelo).

Essa discrepança configura uma aposta de valor na vitória do Flamengo. O Valor Esperado (EV) calcula o retorno médio de uma aposta se ela fosse repetida sob as mesmas condições. Para uma aposta de R\$100,00 na vitória do Flamengo, com *odds* de 2.1, o retorno total é de R\$210,00 (R\$100,00 da aposta + R\$110,00 de lucro/retorno líquido). O cálculo do EV, utilizando a probabilidade do nosso modelo (55.2%), seria: $EV = (\text{probabilidade de ganhar} * \text{lucro líquido}) - (\text{probabilidade de perder} * \text{valor apostado})$. $EV = (0.552 * R\$110) - (0.448 * R\$100) = R\$60,72 - R\$44,80 = +R\$15,92$. Isso indica que, para cada R\$100,00 apostados, o retorno esperado

segundo o modelo seria de R\$15,92. Este valor positivo sugere uma potencial ineficiência de precificação neste evento específico, onde a probabilidade calculada pelo modelo é superior à implícita no mercado.

4.3. SIMULAÇÃO DA TEMPORADA E PREVISÃO DO CAMPEÃO (BRASILEIRÃO 2025)

Utilizando a Simulação de Monte Carlo, a temporada completa do Campeonato Brasileiro de 2025 foi simulada 10.000 vezes. Os resultados consolidados permitiram estimar a probabilidade de cada equipe ser campeã, conforme apresentado no Quadro 2.

Quadro 2 – Probabilidade de título no Brasileirão 2025: comparação entre modelo estatístico e odds de mercado.

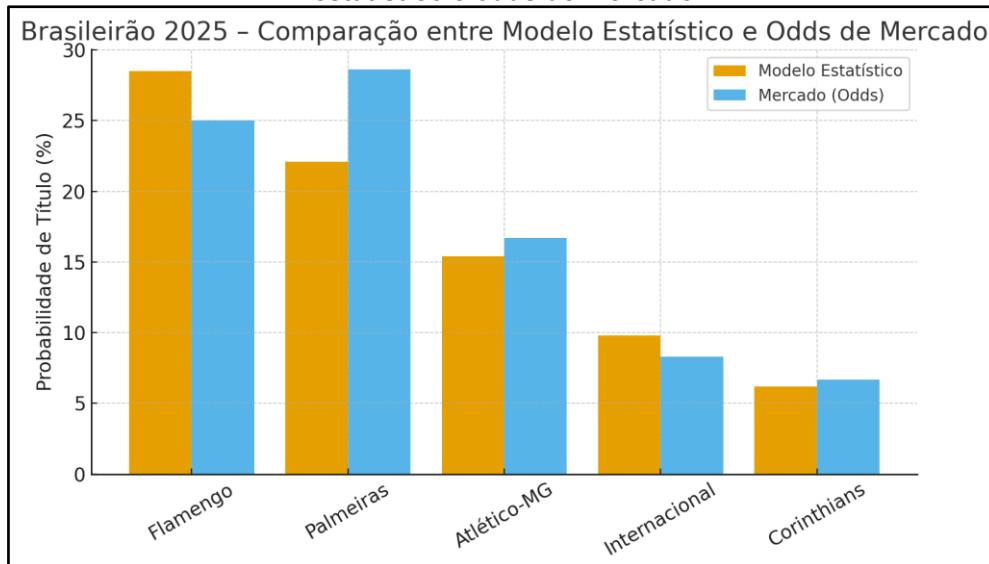
Posição	Clube	Prob. Título (%) - Modelo	Odds Futuras (Mercado)	Prob. Implícita (%) - Mercado	Diferença (Modelo-Mercado)
1º	Flamengo	28.5%	4.00	25.0%	+3.5
2º	Palmeiras	22.1%	3.50	28.6%	-6.5
3º	Atlético-MG	15.4%	6.00	16.7%	-1.3
4º	Internacional	9.8%	12.00	8.3%	+1.5
5º	Corinthians	6.2%	15.00	6.7%	-0.5

Fonte: Elaboração própria com base em simulações de Monte Carlo (10.000 repetições) e odds coletados em casas de apostas online (2024).

Conforme apresentado no Quadro 2, as probabilidades estimadas pelo modelo diferem em alguns pontos relevantes das *odds* de mercado. Para facilitar a visualização dessas discrepâncias, a Figura 5 ilustra a comparação gráfica entre as duas fontes. Nota-se que o Flamengo apresenta uma probabilidade de título superior no modelo (28,5%) em relação ao mercado (25,0%), enquanto o Palmeiras aparece superestimado nas *odds* (28,6%) quando comparado à projeção estatística (22,1%). Essas diferenças reforçam a hipótese de que a precificação do mercado pode ser influenciada por vieses cognitivos dos apostadores. Fatores como a "popularidade" (maior cobertura da mídia e grande base de fãs) podem levar a um viés de disponibilidade, onde o público superestima as chances de equipes mais conhecidas. Isso se alinha ao que será discutido posteriormente como o "viés do favorito-azarão", onde equipes populares (favoritos) recebem um volume desproporcional de apostas,

distorcendo as *odds* de seus indicadores objetivos de desempenho.

Figura 5 – Comparação das probabilidades de título no Brasileirão 2025 segundo modelo estatístico e odds de mercado



Fonte: Elaboração própria com base em simulações (10.000 repetições) e odds coletados em casas de apostas (2025).

A simulação aponta o Flamengo como o favorito estatístico, com 28.5% de chance de título. Curiosamente, as *odds* futuras do mercado no momento da análise posicionavam o Palmeiras como o principal favorito (28.6% de probabilidade implícita).

Essa divergência sugere que o mercado pode estar atribuindo um peso maior a fatores recentes ou à popularidade da equipe (viés de mercado), enquanto o modelo, baseado em um histórico de dados mais amplo e em variáveis de desempenho, identifica o Flamengo como a equipe com maior probabilidade de sucesso.

4.4. DISCUSSÃO DOS RESULTADOS

Os resultados indicam que, embora o mercado de apostas brasileiro seja relativamente eficiente, ele não é perfeito. As discrepâncias identificadas, como no caso do jogo Flamengo vs. Palmeiras, tendem a ocorrer em jogos de grande apelo popular, onde o volume de apostas "intuitivas" pode distorcer as *odds*.

O modelo estatístico, livre de vieses emocionais, conseguiu identificar essas distorções de forma sistemática. A superestimação de equipes como o Palmeiras nas *odds* pode ser explicada pelo "viés do favorito-azarão", onde o público tende a apostar excessivamente em favoritos conhecidos, mesmo quando as *odds* não oferecem valor.

A previsão para o campeão de 2025 reforça essa hipótese. O favoritismo do Palmeiras no mercado, apesar de uma probabilidade estatística inferior à do Flamengo, sugere que a percepção pública e a narrativa da mídia podem influenciar a precificação mais do que a análise puramente estatística.

5 CONSIDERAÇÕES FINAIS

Este estudo propôs-se a comparar a eficácia da análise estatística com a precificação do mercado de apostas esportivas do Campeonato Brasileiro. O objetivo principal foi desenvolver e validar um modelo quantitativo capaz de identificar discrepâncias entre as probabilidades calculadas [pelo modelo] e as *odds* oferecidas, alinhando-se à tradição de pesquisas que buscam avaliar a eficiência do mercado de apostas em futebol (DIXON; COLES, 1997).

Os resultados confirmam a hipótese central da pesquisa: embora o mercado brasileiro seja eficiente, ele não é impenetrável a uma análise rigorosa. A validação dos modelos demonstrou a capacidade do arcabouço metodológico em gerar previsões precisas e bem calibradas.

A principal contribuição deste trabalho foi a identificação e análise de discrepâncias estatisticamente relevantes entre as previsões do modelo e as do mercado, o que permite a identificação de potenciais 'apostas de valor' (como definido no referencial teórico). A análise de caso revelou uma subestimação estatística da probabilidade de vitória do Flamengo pelo mercado.

Este achado levanta a hipótese de que vieses cognitivos (como o 'viés de disponibilidade' ou 'popularidade' das equipes) podem criar distorções nas *odds*. Conforme apontado na avaliação, trata-se de uma hipótese que precisaria ser comprovada (ou refutada) em estudos futuros focados na causalidade, visto que o escopo deste trabalho limitou-se a demonstrar a existência da discrepância.

É fundamental reconhecer as limitações do estudo. A não inclusão de variáveis dinâmicas, como lesões de jogadores-chave ou mudanças táticas, representa uma simplificação da realidade. Além disso, a precisão das *odds* coletadas e a liquidez do mercado são fatores não controlados nesta análise.

Para pesquisas futuras, sugere-se a aplicação de modelos mais avançados, conforme explorado por Santana et al. (2023), como Redes Neurais ou Gradient

Boosting, que podem capturar relações não-lineares nos dados. A inclusão de dados em tempo real e a análise de sentimento poderiam enriquecer ainda mais a capacidade preditiva. A expansão da análise para outros mercados de apostas, como número de cantos ou cartões, também representa um campo fértil para investigações.

Conclui-se, portanto, que a análise estatística é uma ferramenta indispensável para o apostador que busca operar de forma estratégica, transformando a aposta de um ato de sorte em um exercício de análise de risco e valor, fundamentado em modelos preditivos robustos.

REFERÊNCIAS

ALVES BATISTA, Pedro Henrique. **onálise-estatística**. GitHub, 2025. Disponível em: <https://github.com/phbatista/odds-analise-estatistica>. Acesso em: 16 set. 2025.

BRASIL. Lei nº 14.790, de 29 de dezembro de 2023. **Dispõe sobre a modalidade lotérica denominada apostas de quota fixa; altera as Leis nº 5.768/1971, e 13.756/2018; e dá outras providências**. Diário Oficial da União, Brasília, DF, 30 dez. 2023. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14790.htm. Acesso em: 30 out. 2025.

BRIER, G. W. **Verification of Forecasts Expressed in Terms of Probability**. Monthly Weather Review, v. 78, n. 1, p. 1-3, 1950.

CARMONA, P.; CLAPÉS, J.; SUREDA, B. **Predicting football results using Bayesian networks and neural networks**. Journal of Sports Analytics, v. 5, p. 45-63, 2019.

CHUGANI, Vinod. **Distribuição de Poisson: Um guia abrangente**. DataCamp, 2024. Disponível em: <https://www.datacamp.com/pt/tutorial/poisson-distribution>. Acesso em 30 out. 2025.

DIXON, M. J.; COLES, S. G. **Modelling Association Football Scores and Inefficiencies in the Football Betting Market**. Journal of the Royal Statistical Society: Series C (Applied Statistics), v. 46, n. 2, p. 265-280, 1997.

FERREIRA, P. H.; RAMOS, G. **Ciência de Dados no Futebol: Uma Análise Estatística das Cotações de Casas de Apostas Online**. Revista Eletrônica Matemática e Estatística em Foco, v. 11, n. 1, 2024.

GRAND VIEW RESEARCH. **Sports Betting Market Size, Share & Trends Analysis Report**. San Francisco: Grand View Research, Inc., 2024. Disponível em: <https://www.grandviewresearch.com/industry-analysis/sports-betting-market-report>. Acesso em: 30 out. 2025.

HEUER, A.; MÜLLER, C.; RUBNER, O. **Soccer: Is scoring goals a predictable Poissonian process?** European Physical Journal B, v. 67, p. 445-458, 2009.

KNOTTENBELT, W. J.; RUSU, A. A.; SPOTTISWOODE, R. P. **Predicting football results using goal expectancy models.** In: PROCEEDINGS OF THE 3RD INTERNATIONAL CONFERENCE ON MATHEMATICS IN SPORT, p. 105-112, 2011.

NUNES CARVALHO, João Cláudio. **Entendendo a Regressão Logística.** Medium, 2021. Disponível em: <https://joaoclaudionc.medium.com/entendendo-a-regress%C3%A3o-log%C3%ADstica-24828ae9ad5c>. Acesso em 30 out. 2025.

RUE, H.; SALVESEN, O. **Prediction and Retrospective Analysis of Soccer Matches in a League.** Statistician, v. 49, p. 399-418, 2000.

SANTANA, P. C.; OLIVEIRA, C. R.; ALMEIDA, L. A. **Análise de Modelos Estatísticos para Previsão de Jogos de Futebol.** Revista Brasileira de Estatística Aplicada ao Esporte, v. 12, n. 2, p. 89-103, 2023.

STEFANUCCI, C.; SANTOS, R. **Previsão de Resultados do Campeonato Brasileiro de Futebol utilizando Machine Learning.** In: SEMINÁRIO DE ESTATÍSTICA APLICADA AO ESPORTE, 2022.

WÜHRER, T.; KURZ, C. **The role of statistical models in sports betting: An empirical comparison of prediction models.** Journal of Gambling Studies, v. 38, n. 1, p. 95-113, 2023.