UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS – CCSA
CURSO DE BACHARELADO EM CIÊNCIA DE DADOS PARA NEGÓCIOS

Previsão de Rebaixamento no Campeonato Brasileiro Série A: Uma Comparação de Modelos de *Machine Learning* com Validação *Walk-Forward*

Leonardo Feitosa Barroso
leonardo.feitosa@academico.ufpb.br

Orientador: Prof. Dr. Hilton Martins de Brito Ramalho
31 de julho de 2026

LEONARDO FEITOSA BARROSO
Matrícula: 20220097193

PREVISÃO DE REBAIXAMENTO NO CAMPEONATO BRASILEIRO SÉRIE A: UMA COMPARAÇÃO DE MODELOS DE MACHINE LEARNING COM VALIDAÇÃO WALK-FORWARD

Artigo tecnológico apresentado ao Curso de Bacharelado em Ciência de Dados para Negócios do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba, como requisito parcial para a obtenção do título de Bacharel em Ciência de Dados para Negócios.

Orientador: Prof. Dr. Hilton Martins de Brito Ramalho

João Pessoa
2026

Catalogação na publicação
Seção de Catalogação e Classificação

B277p Barroso, Leonardo Feitosa.
Previsão de rebaixamento no campeonato brasileiro Série A: uma comparação de modelos de machine learning com validação walk-forward / Leonardo Feitosa Barroso. - João Pessoa, 2026.
40 f. : il.

Orientação: Hilton Martins de Brito Ramalho.
TCC (Graduação) - UFPB/CCSA.

1. Aprendizado de Máquina. 2. Previsão Esportiva. 3. Rebaixamento. 4. Futebol. 5. Validação Walk-Forward. I. Ramalho, Hilton Martins de Brito. II. Título.

UFPB/CCSA CDU 330:004.6(043)

Elaborado por ANDRE DOMINGOS DA SILVA FERNANDES - CRB-15/00730

PREVISÃO DE REBAIXAMENTO NO CAMPEONATO BRASILEIRO SÉRIE A: UMA COMPARAÇÃO DE MODELOS DE *MACHINE LEARNING* COM VALIDAÇÃO *WALK-FORWARD*

Artigo tecnológico apresentado ao Curso de Bacharelado em Ciência de Dados para Negócios do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba, como requisito parcial para a obtenção do título de Bacharel em Ciência de Dados para Negócios.

Aprovado em: 10 / 08 / 2026

BANCA EXAMINADORA

Prof. Dr. Hilton Martins de Brito Ramalho (Orientador)
Universidade Federal da Paraíba

Ignácio Tavares
Universidade Federal da Paraíba

Adriana Damasceno
Universidade Federal da Paraíba

RESUMO

O rebaixamento à Série B representa uma das maiores perdas esportivas e financeiras que um clube de futebol brasileiro pode sofrer, com impactos diretos sobre receitas de direitos de transmissão, patrocínios e valor de mercado do elenco. O objetivo deste trabalho é desenvolver e comparar modelos de *Machine Learning* capazes de prever, antes do início da temporada, quais clubes apresentam maior risco de rebaixamento no Campeonato Brasileiro Série A. Foram avaliados quatro algoritmos de classificação binária — Regressão Logística, *Random Forest*, XGBoost e LightGBM — sobre um conjunto de 15 variáveis preditoras (*features*) que combina dados financeiros do elenco, coletados do portal Transfermarkt, com janelas deslizantes de desempenho das três e cinco temporadas anteriores. A validação utilizou o esquema *walk-forward*, que respeita a ordem cronológica dos dados, e a otimização de hiperparâmetros foi realizada com RandomizedSearchCV combinado a TimeSeriesSplit. No conjunto de teste independente (2023–2024), o LightGBM obteve o melhor AUC-ROC (0,877), enquanto a Regressão Logística apresentou a maior estabilidade temporal na validação *walk-forward* (média de 0,794 ± 0,058). A previsão para a temporada 2025 apontou Juventude, Sport Recife, Vitória e Mirassol como os clubes de maior risco; a validação retroativa com os resultados oficiais do campeonato confirmou dois dos quatro rebaixamentos previstos e AUC-ROC de 0,922 na temporada, com os quatro clubes efetivamente rebaixados entre os sete maiores riscos apontados pelo modelo. O estudo contribui com um fluxo de processamento (*pipeline*) reprodutível para previsão esportiva com dados temporais, evidencia empiricamente a inadequação da acurácia como métrica única em problemas com classes desbalanceadas e entrega o modelo final em uma aplicação web de acesso aberto.

Palavras-chave: Aprendizado de Máquina. Previsão Esportiva. Rebaixamento. Futebol. Validação *Walk-Forward*.

ABSTRACT

Relegation to Série B is one of the greatest sporting and financial losses a Brazilian football club can suffer, directly affecting broadcasting rights revenues, sponsorships, and squad market value. The objective of this study is to develop and compare Machine Learning models capable of predicting, before the start of the season, which clubs face the highest relegation risk in the Brazilian Série A championship. Four binary classification algorithms were evaluated — Logistic Regression, Random Forest, XGBoost, and LightGBM — over a set of 15 features combining squad financial data, collected from the Transfermarkt portal, with sliding windows of performance from the previous three and five seasons. Validation used the walk-forward scheme, which respects the chronological order of the data, and hyperparameter optimization was performed with RandomizedSearchCV combined with TimeSeriesSplit. On the independent test set (2023–2024), LightGBM achieved the best AUC-ROC (0.877), while Logistic Regression showed the highest temporal stability in walk-forward validation (mean 0.794 ± 0.058). The forecast for the 2025 season pointed to Juventude, Sport Recife, Vitória, and Mirassol as the clubs at highest risk; retrospective validation against the official league results confirmed two of the four predicted relegations and an AUC-ROC of 0.922 for the season, with all four actually relegated clubs ranked among the seven highest predicted risks. The study contributes a reproducible pipeline for sports prediction with temporal data, empirically demonstrates the inadequacy of accuracy as a sole metric in imbalanced classification problems, and delivers the final model as an openly accessible web application.

Keywords: Machine Learning. Sports Prediction. Relegation. Football. Walk-Forward Validation.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

1 INTRODUÇÃO

O futebol profissional brasileiro opera sob um regime de promoção e rebaixamento em que, ao final de cada edição do Campeonato Brasileiro Série A, os quatro últimos colocados são rebaixados à Série B. Para os clubes, as consequências ultrapassam o campo esportivo: a queda de divisão implica redução expressiva das receitas de transmissão televisiva, perda de patrocínios, desvalorização do elenco e dificuldade de retenção de atletas, efeitos que frequentemente se prolongam por várias temporadas. Antecipar o risco de rebaixamento é, portanto, uma informação estratégica para a gestão esportiva e financeira dos clubes, permitindo ações preventivas como reforços pontuais, renegociação de contratos e planejamento orçamentário.

Nesse contexto, técnicas de *Machine Learning* vêm sendo aplicadas com sucesso à previsão de resultados esportivos, tanto em nível de partida quanto de temporada (BUNKER; THABTAH, 2019; BABOOTA; KAUR, 2019). A literatura, contudo, concentra-se majoritariamente na previsão de resultados de jogos individuais em ligas europeias (CONSTANTINOU; FENTON; NEIL, 2012), havendo lacuna de estudos que abordem a previsão de rebaixamento no campeonato brasileiro utilizando exclusivamente informações disponíveis antes do início da temporada — cenário em que a previsão tem maior valor prático para a gestão. Além disso, dados de valor de mercado de elencos, como os disponibilizados pelo portal Transfermarkt, mostraram-se preditores relevantes de desempenho esportivo (MÜLLER; SIMONS; WEINMANN, 2017), mas ainda são pouco explorados em combinação com o histórico recente de desempenho dos clubes.

O objetivo geral deste trabalho é desenvolver e comparar modelos de *Machine Learning* capazes de estimar, antes do início da temporada, a probabilidade de rebaixamento de cada clube participante do Campeonato Brasileiro Série A.

Como objetivos específicos, busca-se: (i) construir uma base de dados histórica (2014–2025) combinando informações financeiras de elenco e desempenho esportivo por temporada; (ii) realizar engenharia de variáveis preditoras (doravante *features*) com janelas deslizantes que resumam o histórico recente de cada clube sem vazamento de dados (*data leakage*); (iii) treinar e otimizar quatro algoritmos de classificação binária (Regressão Logística, *Random Forest*, XGBoost e LightGBM) com validação que respeite a ordem temporal dos dados; (iv) comparar os modelos em um conjunto de teste independente utilizando métricas adequadas a classes desbalanceadas; e (v) aplicar o modelo final à previsão da temporada 2025.

O restante do texto está organizado da seguinte forma: a Seção 2 apresenta a fundamentação teórica e os trabalhos correlatos; a Seção 3 descreve a metodologia, incluindo coleta de dados, engenharia de *features*, validação e otimização; a Seção 4 apresenta e discute os resultados; e a Seção 5 traz as considerações finais, com limitações e propostas de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Classificação Supervisionada

Segundo Hastie, Tibshirani e Friedman (2009), os métodos de classificação supervisionada buscam aprender, a partir de um conjunto de treinamento, uma função que mapeie as variáveis preditoras X ∈ R^p à variável-alvo Y, de modo a generalizar para observações não vistas. No problema aqui tratado, a variável-alvo é binária: Y = 1 indica que o clube foi rebaixado ao final da temporada e Y = 0 indica que permaneceu na Série A. A escolha do algoritmo deve considerar o compromisso entre viés e variância, sendo recomendável avaliar múltiplos modelos antes de selecionar o mais adequado ao problema (JAMES et al., 2023).

2.2 Regressão Logística

A Regressão Logística (COX, 1958) modela a probabilidade de pertencimento à classe de interesse por meio da função sigmoide, conforme a Equação 1:

$$P(Y = 1 \mid X) = 1 / (1 + e^{-(\beta_0 + \beta^T X)}) \quad (1)$$

Sua principal vantagem é a interpretabilidade: cada coeficiente βj indica a direção e a magnitude do efeito da respectiva variável sobre o risco previsto, o que é particularmente valioso quando os resultados precisam subsidiar decisões de gestão (JAMES et al., 2023).

2.3 Random Forest

O *Random Forest* (BREIMAN, 2001) constrói um conjunto (*ensemble*) de árvores de decisão treinadas sobre amostras *bootstrap* dos dados, selecionando em cada divisão um subconjunto aleatório de preditoras. A predição final é obtida por agregação das árvores, o que reduz a variância do modelo e o torna robusto a *outliers* e a relações não lineares.

2.4 Gradient Boosting: XGBoost e LightGBM

Os métodos de potencialização sequencial (*boosting*) constroem árvores sequencialmente, com cada nova árvore corrigindo os erros das anteriores. O XGBoost (CHEN; GUESTRIN, 2016) introduziu regularização explícita e otimizações computacionais que o tornaram referência em problemas com dados tabulares. O LightGBM (KE et al., 2017) aprimora essa família de algoritmos com amostragem baseada em gradiente (GOSS) e agrupamento exclusivo de *features* (EFB), além do crescimento de árvore por folha, resultando em treinamento mais eficiente com desempenho competitivo ou superior.

2.5 Métricas de Avaliação para Classes Desbalanceadas

Em problemas nos quais uma das classes é minoritária — como o rebaixamento, que atinge apenas 4 dos 20 clubes por temporada (~20% das observações) — a acurácia isolada é uma métrica enganosa, pois um classificador trivial que sempre prevê a classe majoritária já alcança acurácia igual à prevalência dessa classe (HE; GARCIA, 2009; GÉRON, 2022). Por essa razão, este trabalho adota como métrica principal a área sob a curva ROC (AUC-ROC), que mede a capacidade do modelo de ordenar corretamente as observações por risco, independentemente do limiar de decisão (FAWCETT, 2006). Um AUC-ROC de 0,5 equivale a um classificador aleatório, e 1,0 a um classificador perfeito. Complementarmente, são reportados matriz de confusão, precisão, *recall* e F1-score por classe.

2.6 Validação para Dados Temporais

A validação cruzada tradicional (*k-fold*) pressupõe observações independentes e identicamente distribuídas, hipótese violada em dados com estrutura temporal: embaralhar as temporadas permitiria ao modelo aprender com o futuro para prever o passado (BERGMEIR; BENÍTEZ, 2012). A alternativa adequada é a validação *walk-forward* (ou de origem móvel), na qual o modelo é sempre treinado com dados anteriores ao período de validação, simulando seu uso real em produção (TASHMAN, 2000). Essa abordagem corresponde ao TimeSeriesSplit da biblioteca scikit-learn (PEDREGOSA et al., 2011).

2.7 Trabalhos Correlatos

Bunker e Thabtah (2019) propõem um arcabouço metodológico para previsão de resultados esportivos com *Machine Learning*, destacando a importância de separar *features* conhecidas antes da partida daquelas obtidas após sua realização — princípio análogo ao adotado neste trabalho ao restringir os preditores a informações pré-temporada. Baboota e Kaur (2019) aplicam engenharia de *features* baseada na forma recente das equipes para prever resultados da Premier League inglesa, obtendo ganhos expressivos com variáveis de desempenho acumulado — estratégia que fundamenta as janelas deslizantes aqui utilizadas. Constantinou, Fenton e Neil (2012) desenvolvem o modelo pi-football, baseado em redes bayesianas, para prognósticos de partidas, evidenciando o valor de combinar dados objetivos e conhecimento de domínio. Por fim, Müller, Simons e Weinmann (2017) demonstram que valores de mercado estimados coletivamente, como os do Transfermarkt, são preditores robustos de desempenho no futebol, o que justifica seu uso como *proxy* da capacidade financeira dos clubes.

Em contextos metodologicamente mais próximos ao deste estudo, Arabzad et al. (2014) empregam redes neurais artificiais para prever resultados da liga profissional iraniana a partir do histórico das sete edições anteriores, ilustrando a aplicação de *Machine Learning* a ligas fora do eixo europeu. Carpita, Ciavolino e Pasca (2019) exploram a base aberta European Soccer Database (Kaggle), construindo indicadores de desempenho por função e estimando a

probabilidade de vitória do mandante com regressão logística binomial — evidência da competitividade de modelos lineares interpretáveis também em dados esportivos de grande escala. Hubáček, Šourek e Železný (2019), vencedores do 2017 Soccer Prediction Challenge, demonstram a superioridade de árvores potencializadas por gradiente (*gradient boosted trees*) sobre abordagens relacionais na previsão de resultados de futebol, o que fundamenta a inclusão de XGBoost e LightGBM na comparação aqui realizada. No cenário nacional, Maciel (2024) compara sete algoritmos supervisionados na predição de resultados de partidas do Campeonato Brasileiro, com foco em simulações de apostas esportivas. Registre-se, por fim, que buscas realizadas nas bases Google Scholar e Scopus com as expressões (relegation OR "rebaixamento") AND (machine learning OR prediction) retornaram predominantemente estudos de previsão de resultados de partidas; não foram localizados trabalhos dedicados à previsão de rebaixamento no futebol brasileiro com informações exclusivamente pré-temporada, o que corrobora a lacuna apontada na Introdução.

Este trabalho diferencia-se dos correlatos em três aspectos: (i) o objeto de previsão é o rebaixamento ao final da temporada, e não o resultado de partidas individuais; (ii) o contexto é o campeonato brasileiro, menos explorado na literatura que as ligas europeias; e (iii) todas as *features* estão disponíveis antes do início da temporada, o que confere valor prático à previsão para o planejamento dos clubes.

Diante das considerações teóricas apresentadas, este trabalho adota quatro algoritmos que cobrem diferentes pontos do espectro viés-variância (Regressão Logística, *Random Forest*, XGBoost e LightGBM), o AUC-ROC como métrica principal de avaliação — em razão do desbalanceamento inerente ao problema — e a validação *walk-forward* como esquema compatível com a natureza temporal dos dados. A seção seguinte detalha como essas escolhas se materializam no desenho metodológico.

3 METODOLOGIA

Esta pesquisa classifica-se como aplicada, de abordagem quantitativa e caráter descritivo-preditivo (GIL, 2019). O fluxo metodológico segue o fluxo de processamento (*pipeline*) padrão de Ciência de Dados: coleta de dados, engenharia de *features* com janelas deslizantes, análise exploratória, pré-processamento sem vazamento de dados, validação *walk-forward*, otimização de hiperparâmetros, avaliação em teste independente e aplicação à temporada 2025. Todo o código, os dados e os experimentos estão disponíveis em repositório público (https://github.com/leonardofeitos4/previsao-tcc), garantindo a reprodutibilidade integral dos resultados. O modelo final foi ainda disponibilizado em uma aplicação web de acesso aberto (https://previsao-tcc-tfdykpjzhid7qj7cqquwbr.streamlit.app), que permite consultar o ranking de risco da temporada, inspecionar as métricas de avaliação e simular cenários de elenco — caráter aplicado condizente com a natureza tecnológica deste trabalho. Os resultados aqui reportados são gerados pelo *pipeline* em Python descrito nesta seção; a

aplicação é uma camada de visualização sobre esse mesmo modelo (semente aleatória fixa igual a 42 em todos os experimentos; o valor 42 é uma convenção amplamente adotada na comunidade de Ciência de Dados, e a fixação da semente serve unicamente para garantir a reprodutibilidade exata dos resultados).

3.1 Coleta de Dados

Os dados foram coletados do portal Transfermarkt (Transfermarkt, 2025) por meio de raspagem automatizada (*web scraping*) com as bibliotecas requests e BeautifulSoup, em duas frentes: (i) dados de elenco por clube e temporada — número de jogadores, quantidade de estrangeiros e valor de mercado total; e (ii) estatísticas de desempenho por temporada — posição final, vitórias, empates, derrotas, gols marcados e sofridos, saldo de gols, pontos e aproveitamento. A base cobre as temporadas de 2014 a 2024 para treinamento e teste (11 temporadas × 20 clubes = 220 observações rotuladas) e a temporada 2025 para previsão (20 clubes). A temporada 2025 foi deliberadamente excluída da coleta de desempenho, pois seus resultados constituem exatamente o que se deseja prever — incluí-los configuraria vazamento de dados.

Quanto aos aspectos éticos e legais da coleta, os dados raspados são públicos e de livre acesso, e foram utilizados exclusivamente para fins acadêmicos e não comerciais, com atribuição expressa da fonte (Transfermarkt, 2025). A raspagem limitou-se a páginas abertas, sem contornar mecanismos de autenticação ou restrição de acesso, e foi executada uma única vez por temporada, em volume reduzido de requisições, de modo a não onerar os servidores do portal. Ressalta-se, ainda, uma limitação da fonte: os valores de mercado do Transfermarkt são estimativas produzidas por julgamento coletivo (*crowdsourcing*), e não cifras oficiais de transações; a literatura, contudo, indica que tais estimativas constituem preditores robustos de desempenho esportivo, com qualidade comparável à de avaliações profissionais (MÜLLER; SIMONS; WEINMANN, 2017).

3.2 Engenharia de Features

Cada observação corresponde a um clube em uma temporada e é descrita por 15 *features*, detalhadas na Tabela 1: 3 variáveis de elenco (estáticas, conhecidas no início da temporada) e 12 variáveis de janela deslizante, que resumem o desempenho médio do clube nas 3 e nas 5 temporadas anteriores para seis métricas (pontos, saldo de gols, gols marcados, gols sofridos, vitórias e aproveitamento).

Tabela 1 – Descrição das 15 features utilizadas — cada observação corresponde a um clube em uma temporada.

Grupo	Feature	Descrição
Elenco (3 features)	Plantel	Número total de jogadores no elenco registrado
	Estrangeiros	Número de jogadores estrangeiros no elenco

Grupo	Feature	Descrição
	Valor de Mercado Total (M€)	Valor financeiro agregado do elenco
Janelas deslizantes (12 features: médias das 3 e 5 temporadas anteriores)	Pts_media_{3,5}	Média de pontos conquistados
	SG_media_{3,5}	Saldo de gols médio
	Gols_Pro_media_{3,5}	Média de gols marcados
	Gols_Contra_media_{3,5}	Média de gols sofridos
	V_media_{3,5}	Média de vitórias
	Aproveitamento_media_{3,5}	Percentual médio de pontos aproveitados

Fonte: Elaborado pelo autor (2026).

As janelas deslizantes adicionam ao modelo a "memória" do desempenho recente em campo, complementando os dados estáticos de elenco, que descrevem o que o clube tem, mas não o que ele fez. Para garantir a ausência de vazamento temporal, o cálculo das médias da temporada T utiliza exclusivamente as temporadas T−1 a T−w (implementação com shift(1).rolling()), de forma que nenhuma informação da própria temporada prevista participa de suas *features*.

3.3 Análise Exploratória dos Dados

A Tabela 2 apresenta as estatísticas descritivas das 15 *features* na base rotulada (2014–2024, n = 220), antes da imputação de valores ausentes. Três aspectos merecem destaque. Primeiro, a prevalência de rebaixamento é de exatos 20% (44 rebaixamentos em 220 observações), confirmando o desbalanceamento discutido na Seção 2.5. Segundo, as 12 *features* de janela deslizante apresentam 15,5% de valores ausentes — observações de clubes recém-promovidos, sem histórico suficiente na Série A, cujo tratamento é descrito na subseção seguinte. Terceiro, o valor de mercado dos elencos exibe forte assimetria (média de €55,3 milhões, desvio padrão de €38,5 milhões e máximo de €214,2 milhões), refletindo a acentuada heterogeneidade financeira entre os clubes da Série A — justamente a característica que motiva o uso dessa variável como preditor.

Tabela 2 – Estatísticas descritivas das 15 features na base rotulada (2014–2024, n = 220), antes da imputação de valores ausentes.

Feature	Média	DP	Mín.	Máx.	% ausente
Plantel	55,57	8,22	40,0	80,0	0,0
Estrangeiros	5,37	3,08	0,0	14,0	0,0
Valor de Mercado Total (M€)	55,32	38,48	5,4	214,2	0,0
Pts_media_3	53,48	10,02	28,0	80,0	15,5
Pts_media_5	53,53	9,57	28,0	80,0	15,5

Feature	Média	DP	Mín.	Máx.	% ausente
SG_media_3	2,00	13,40	-35,0	34,0	15,5
SG_media_5	2,05	12,71	-35,0	30,6	15,5
Gols_Pro_media_3	45,34	8,95	23,0	74,3	15,5
Gols_Pro_media_5	45,22	8,49	23,0	68,4	15,5
Gols_Contra_media_3	43,35	6,80	24,0	60,0	15,5
Gols_Contra_media_5	43,17	6,30	24,0	60,0	15,5
V_media_3	14,47	3,48	7,0	24,0	15,5
V_media_5	14,52	3,33	7,0	24,0	15,5
Aproveitamento_media_3	46,93	8,79	24,6	70,2	15,5
Aproveitamento_media_5	46,97	8,40	24,6	70,2	15,5

Fonte: Elaborado pelo autor (2026).

A Figura 1 exibe a matriz de correlação de Pearson entre as *features*. Observa-se forte correlação positiva entre as janelas de 3 e de 5 temporadas de uma mesma métrica e entre as métricas de desempenho entre si — pontos, vitórias e aproveitamento são quase colineares, por construção. As *features* de elenco correlacionam-se moderadamente com as de desempenho (o valor de mercado, por exemplo, acompanha o histórico recente de pontos), indicando que os dois grupos carregam informação parcialmente distinta. Essa estrutura de multicolinearidade não compromete a capacidade preditiva dos modelos, mas recomenda cautela na interpretação isolada de coeficientes individuais da Regressão Logística, ponto retomado na Seção 4.4.

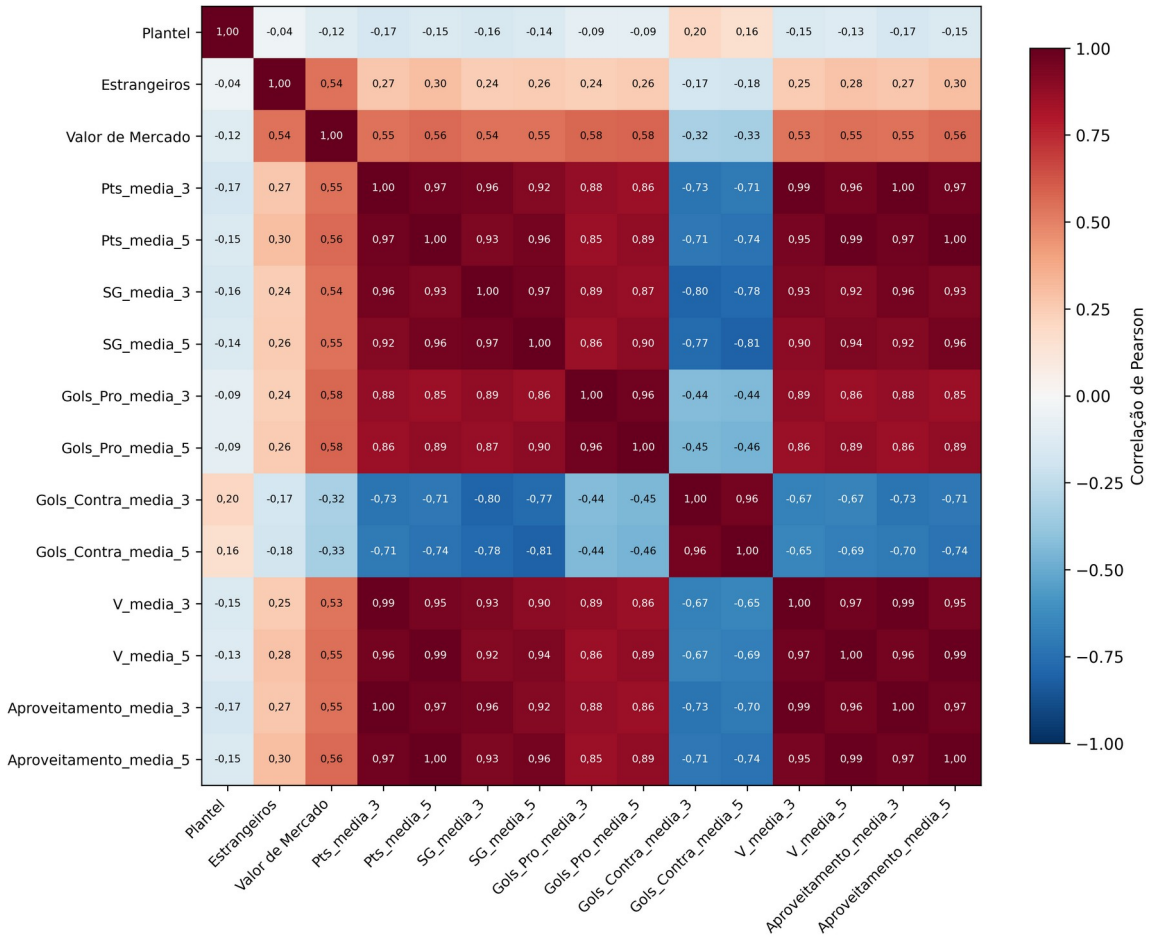

Figura 1 – Matriz de correlação de Pearson entre as 15 features (base rotulada 2014–2024).
Fonte: Elaborado pelo autor (2026).

3.4 Pré-processamento e Tratamento de Dados Faltantes

Clubes recém-promovidos da Série B não possuem histórico completo na Série A, o que gera valores ausentes nas *features* de janela deslizante. Esses valores foram imputados com a mediana calculada exclusivamente no conjunto de treinamento, decisão que se justifica por três razões: (i) a mediana é robusta a valores extremos, adequada a distribuições assimétricas como as de desempenho esportivo; (ii) calculá-la apenas no treino evita que estatísticas do conjunto de teste contaminem o ajuste do modelo; e (iii) a imputação por um valor central é conservadora — atribui ao clube sem histórico um comportamento "típico", sem induzir viés otimista ou pessimista. Em seguida, todas as *features* foram padronizadas com StandardScaler, com parâmetros ajustados somente no conjunto de treinamento e aplicados aos demais conjuntos, novamente evitando vazamento de dados.

O desbalanceamento de classes (~20% de rebaixados) foi tratado no nível dos algoritmos, com ponderação inversa à frequência das classes (class_weight='balanced' na Regressão Logística, no *Random Forest* e no LightGBM; scale_pos_weight=4 no XGBoost).

3.5 Separação Temporal e Validação Walk-Forward

Os dados rotulados foram divididos temporalmente em treinamento (2014–2022, 180 observações) e teste (2023–2024, 40 observações, sendo 8 rebaixamentos). Sobre o conjunto de treinamento aplicou-se a validação *walk-forward* com cinco partições (*folds*) de janela expansiva, conforme a Tabela 3: em cada *fold*, o modelo é treinado com todas as temporadas até o ano anterior e validado na temporada seguinte, com imputação e padronização recalculadas dentro de cada *fold* para preservar a integridade temporal.

Tabela 3 – Configuração dos cinco folds da validação walk-forward.

Fold	Conjunto de Treino	Validação	Obs. de Treino
1	2014–2017	2018	~80
2	2014–2018	2019	~100
3	2014–2019	2020	~120
4	2014–2020	2021	~140
5	2014–2021	2022	~160

Fonte: Elaborado pelo autor (2026).

3.6 Seleção de Algoritmos e Otimização de Hiperparâmetros

Foram selecionados quatro algoritmos que cobrem o espectro de complexidade usual em dados tabulares: a Regressão Logística, como modelo linear interpretável de referência; o *Random Forest*, como *ensemble* de *bagging*; e XGBoost e LightGBM, como representantes do estado da arte em *gradient boosting* para dados tabulares (CHEN; GUESTRIN, 2016; KE et al., 2017). Redes neurais profundas foram deliberadamente excluídas: com apenas ~220 observações, modelos com grande número de parâmetros tendem ao sobreajuste e não dispõem de volume de dados suficiente para exibir suas vantagens (GÉRON, 2022).

A otimização de hiperparâmetros utilizou RandomizedSearchCV com 30 candidatos por modelo, validação interna TimeSeriesSplit de 5 *folds* e AUC-ROC como critério de seleção. A busca aleatória foi preferida à busca em grade (*GridSearchCV*) porque, para um mesmo orçamento computacional, explora o espaço de hiperparâmetros de forma mais eficiente, especialmente quando apenas algumas dimensões são realmente relevantes (BERGSTRA; BENGIO, 2012); métodos de otimização bayesiana foram considerados, mas o ganho esperado não justificaria a complexidade adicional para um espaço de busca deste porte.

4 RESULTADOS E DISCUSSÃO

4.1 Validação Walk-Forward

A Tabela 4 apresenta o AUC-ROC de cada modelo em cada *fold* da validação *walk-forward*. A Regressão Logística obteve a maior média (0,794) com o menor desvio padrão (±0,058), evidenciando comportamento estável ao longo das temporadas. O LightGBM, por sua

vez, apresentou a maior variância entre *folds* (±0,151), com desempenho que oscila de 0,438 (temporada 2020, marcada pelas condições atípicas da pandemia) a 0,906 (2022). Essa instabilidade sugere que modelos de *boosting*, mais flexíveis, são também mais sensíveis a mudanças de regime nos dados — observação que corrobora a recomendação de Bergmeir e Benítez (2012) de avaliar preditores temporais em múltiplas janelas de validação, e não em um único corte.

O *fold* 3, correspondente à temporada 2020, merece exame específico: foi o pior para todos os modelos, com o LightGBM atingindo AUC-ROC de 0,438 — abaixo, inclusive, do classificador aleatório — e o XGBoost, 0,531. A temporada 2020 foi disputada sob os efeitos da pandemia de COVID-19: iniciada apenas em agosto de 2020 e encerrada em fevereiro de 2021, transcorreu com estádios vazios, calendário comprimido, surtos de contágio nos elencos e desgaste físico atípico. Nessas condições, a relação histórica entre as *features* pré-temporada e o desempenho em campo foi parcialmente rompida — um choque exógeno que modelos alimentados exclusivamente com informações estruturais não têm como antecipar. O episódio evidencia uma limitação inerente à abordagem pré-temporada e, ao mesmo tempo, fortalece a proposta de trabalho futuro de retreinamento parcial ao longo da temporada, capaz de incorporar sinais correntes e reagir a mudanças de regime. Nota-se, por fim, que a Regressão Logística foi o modelo menos afetado no *fold* pandêmico (0,703), comportamento coerente com a menor variância esperada de modelos mais simples.

Tabela 4 – AUC-ROC por fold da validação walk-forward.

Fold (ano val.)	Reg. Logística	Random Forest	XGBoost	LightGBM
Fold 1 — 2018	0,828	0,789	0,766	0,719
Fold 2 — 2019	0,859	0,734	0,719	0,750
Fold 3 — 2020	0,703	0,758	0,531	0,438
Fold 4 — 2021	0,750	0,656	0,594	0,688
Fold 5 — 2022	0,828	0,875	0,688	0,906
Média ± DP	0,794 ± 0,058	0,762 ± 0,071	0,659 ± 0,085	0,700 ± 0,151

Fonte: Elaborado pelo autor (2026).

4.2 Hiperparâmetros Selecionados

A Tabela 5 resume as melhores configurações encontradas pela busca aleatória. Nota-se que os *ensembles* convergiram para configurações conservadoras (profundidades reduzidas, taxas de aprendizado baixas), coerentes com o volume modesto de dados — árvores rasas limitam a capacidade do modelo e mitigam o sobreajuste.

Tabela 5 – Hiperparâmetros selecionados por RandomizedSearchCV (30 candidatos) com TimeSeriesSplit (5 folds) e AUC-ROC como critério de seleção.

Modelo	Melhores Hiperparâmetros (CV)	AUC-ROC CV
Regressão Logística	C = 0,985; solver = lbfgs	0,754
Random Forest	n_estimators = 100; max_depth = 3; min_samples_split = 10; max_features = sqrt	0,760
XGBoost	n_estimators = 50; max_depth = 5; learning_rate = 0,05; subsample = 0,7	0,742
LightGBM	n_estimators = 50; max_depth = 7; learning_rate = 0,01; num_leaves = 31	0,721

Fonte: Elaborado pelo autor (2026).

4.3 Avaliação no Conjunto de Teste Independente

Após a otimização, cada modelo foi reavaliado no conjunto de teste (2023–2024), não utilizado em nenhuma etapa anterior. A Tabela 6 apresenta o ranqueamento por AUC-ROC, e a Figura 2 exibe as curvas ROC correspondentes. O LightGBM alcançou o melhor AUC-ROC (0,877), seguido do *Random Forest* (0,844) e da Regressão Logística (0,828); o XGBoost, embora tenha registrado acurácia idêntica à do LightGBM (82,5%), obteve o menor AUC-ROC (0,652), caso examinado em detalhe na Seção 4.5. A Figura 3 compara visualmente as métricas, e a Figura 4 apresenta as matrizes de confusão dos quatro modelos.

Tabela 6 – Ranqueamento dos modelos por AUC-ROC no conjunto de teste independente (2023–2024).

Rank	Modelo	AUC-ROC	Acurácia
1.º	LightGBM	0,877	82,5%
2.º	Random Forest	0,844	80,0%
3.º	Regressão Logística	0,828	80,0%
4.º	XGBoost	0,652	82,5%

Fonte: Elaborado pelo autor (2026).

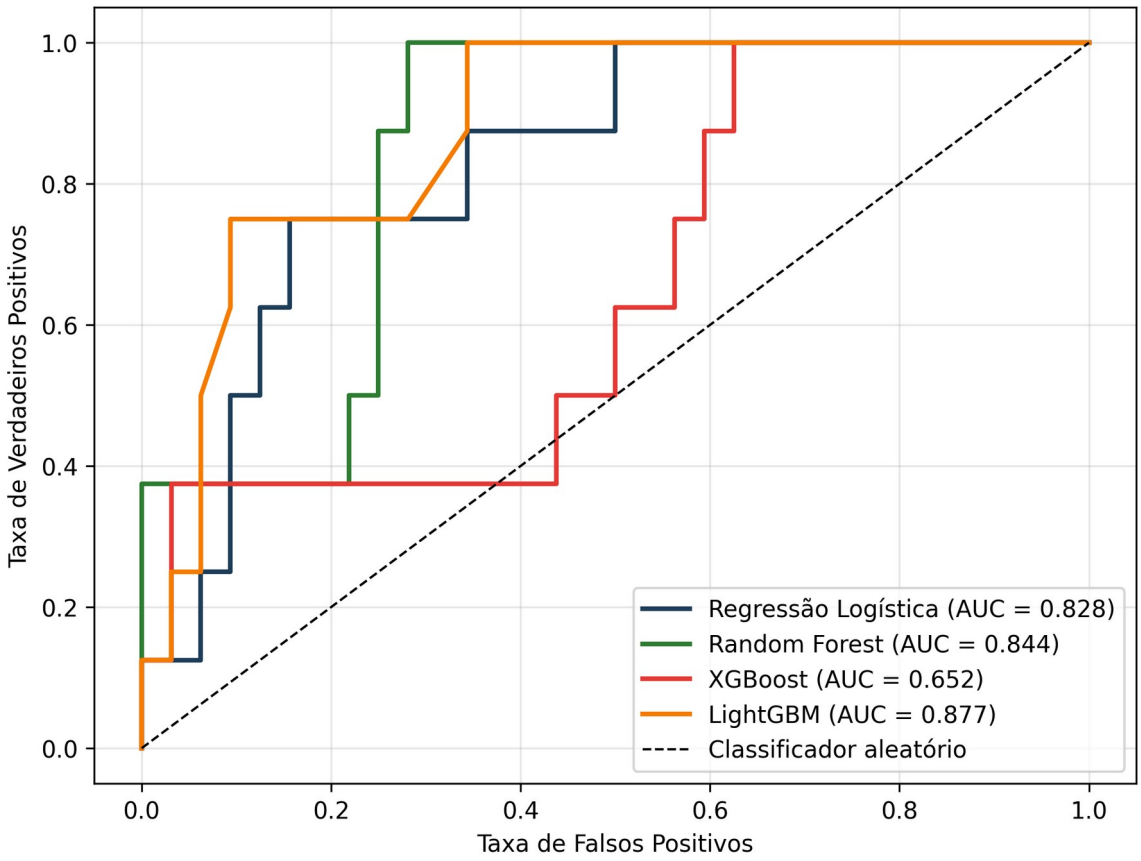

Figura 2 – Curvas ROC dos quatro modelos no conjunto de teste (2023–2024). A diagonal tracejada representa um classificador aleatório (AUC = 0,50).

Fonte: Elaborado pelo autor (2026).

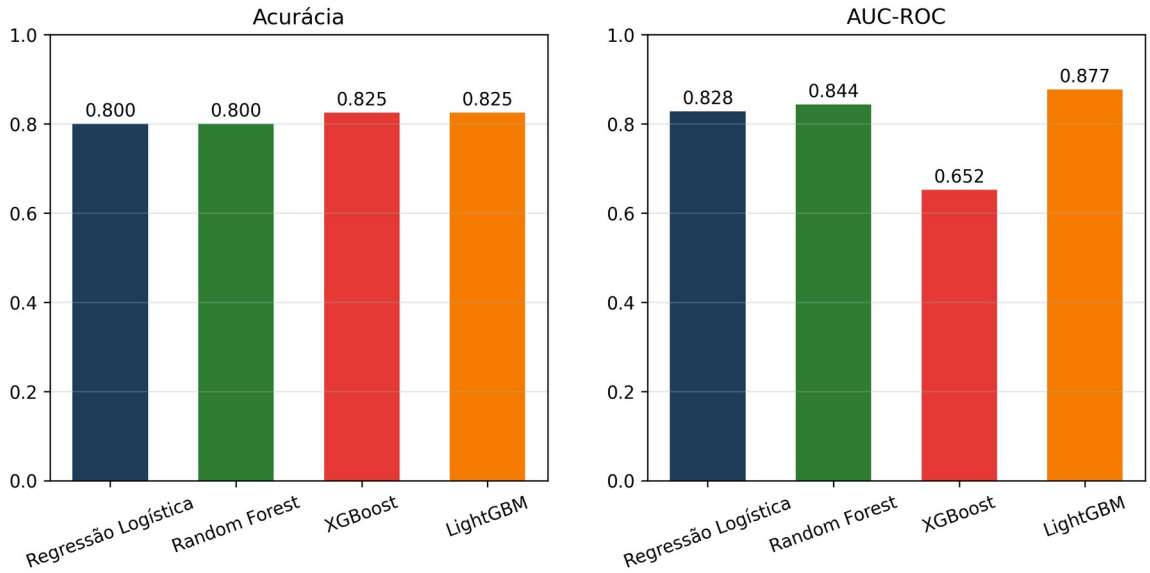

Figura 3 – Comparação das métricas de desempenho (Acurácia e AUC-ROC) entre os modelos no conjunto de teste.

Fonte: Elaborado pelo autor (2026).

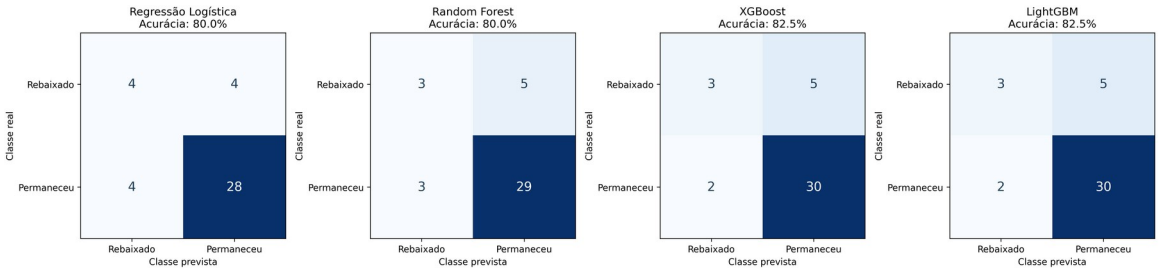

Figura 4 – Matrizes de confusão dos quatro modelos no conjunto de teste (2023–2024).
Fonte: Elaborado pelo autor (2026).

A Tabela 7 detalha o relatório de classificação da Regressão Logística. Observa-se o padrão típico de classes desbalanceadas: desempenho elevado na classe majoritária (F1 = 0,88) e moderado na classe minoritária (F1 = 0,50), reforçando que o valor do modelo está na *ordenação* dos clubes por risco — capturada pelo AUC-ROC — mais do que na classificação binária com limiar fixo.

Tabela 7 – Relatório de classificação da Regressão Logística no conjunto de teste (2023–2024).

Classe	Precisão	Recall	F1-Score	Suporte
Permanece (0)	0,88	0,88	0,88	32
Rebaixado (1)	0,50	0,50	0,50	8
Acurácia geral	—	—	0,80	40
Média macro	0,69	0,69	0,69	40
Média ponderada	0,80	0,80	0,80	40

Fonte: Elaborado pelo autor (2026).

Complementarmente, avaliou-se a calibração das probabilidades da Regressão Logística no conjunto de teste (Figura 5). Agrupando as previsões em quartis, a fração observada de rebaixamentos acompanha de perto a probabilidade média prevista em cada grupo (no quartil de maior risco, por exemplo, previsão média de 63,6% contra 50,0% de rebaixamentos observados), indicando probabilidades razoavelmente bem calibradas — sem o excesso de confiança que modelos mais flexíveis costumam exibir em amostras pequenas. Essa propriedade é relevante para o uso gerencial do modelo, pois permite interpretar as probabilidades previstas como estimativas honestas de risco, e não apenas como escores de ordenação.

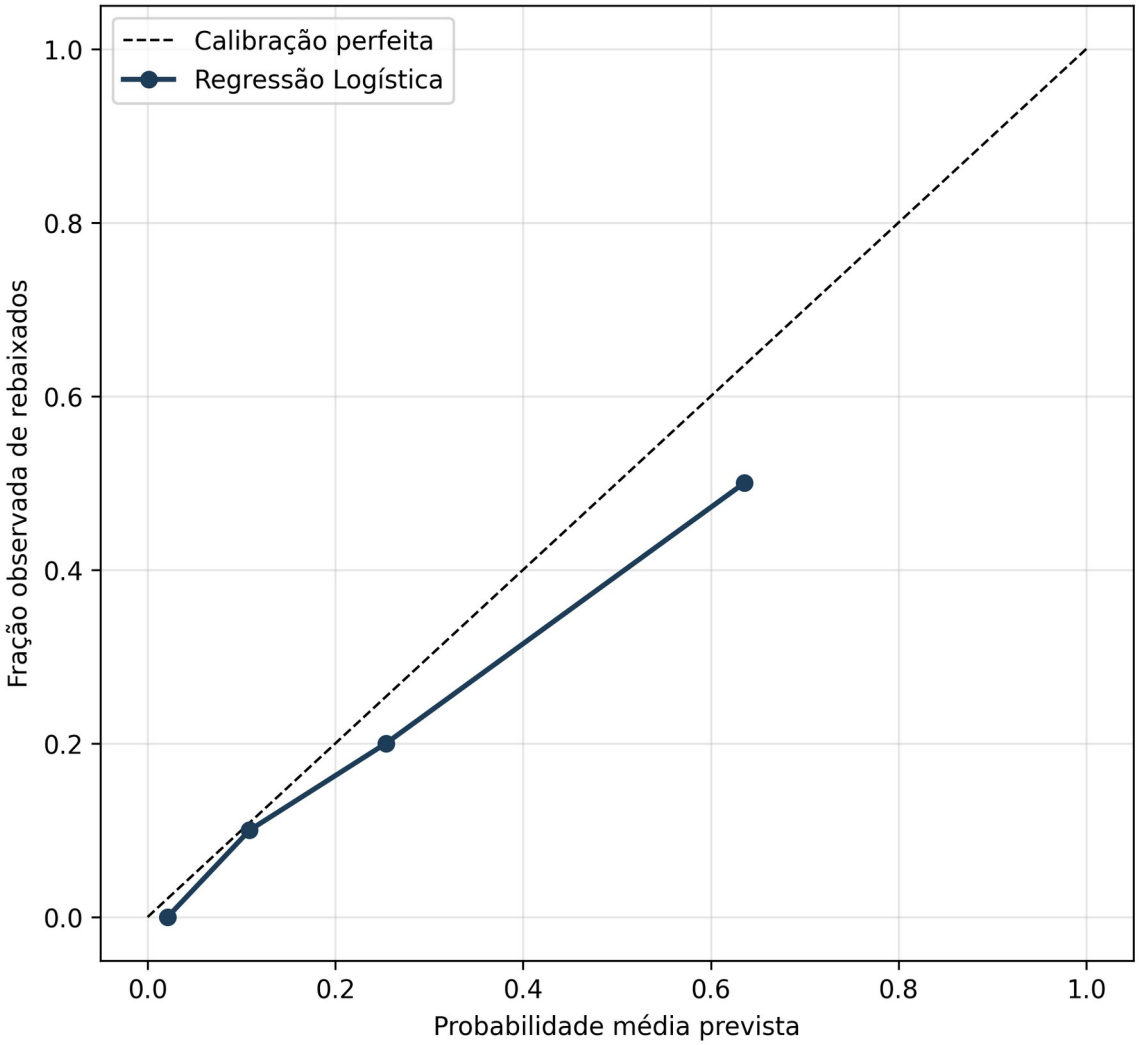

Figura 5 – Diagrama de confiabilidade (curva de calibração) da Regressão Logística no conjunto de teste (2023–2024), com agrupamento em quartis.

Fonte: Elaborado pelo autor (2026).

4.4 Importância das Features

A Figura 6 apresenta a importância relativa das *features* sob duas óticas complementares: os coeficientes padronizados da Regressão Logística, cujo sinal indica a direção do efeito sobre o risco, e a importância por ganho do LightGBM, que mede a contribuição de cada variável para as divisões das árvores. Os dois modelos convergem no essencial: o valor de mercado do elenco é a variável mais influente em ambos — coeficiente de −0,931 na Regressão Logística (quanto maior o valor, menor o risco previsto) e ganho cerca de quatro vezes superior ao da segunda colocada no LightGBM —, corroborando Müller, Simons e Weinmann (2017) quanto ao poder preditivo dessa medida. Entre as *features* de desempenho, a média de vitórias nas três temporadas anteriores destaca-se como fator de proteção (−0,703), confirmando que o histórico recente em campo agrega informação além da dimensão financeira. Chama atenção, ainda, o coeficiente positivo do tamanho do plantel (+0,549): controlados os

demais fatores, elencos mais numerosos associam-se a maior risco — resultado compatível com a hipótese de que plantéis inchados refletem menor qualidade média por atleta ou instabilidade de planejamento.

Dois cuidados de leitura se impõem, à luz da multicolinearidade diagnosticada na análise exploratória (Figura 1). Primeiro, coeficientes de *features* altamente correlacionadas devem ser interpretados em conjunto: os sinais aparentemente contraditórios entre janelas de uma mesma métrica (por exemplo, médias de pontos com coeficiente positivo e médias de vitórias com coeficiente negativo) decorrem da partição de efeito entre variáveis quase colineares, e não de relações causais opostas. Segundo, a concentração do ganho do LightGBM em poucas variáveis sugere redundância entre as janelas de 3 e de 5 temporadas — uma versão futura do modelo poderia ser simplificada com pouca perda de desempenho.

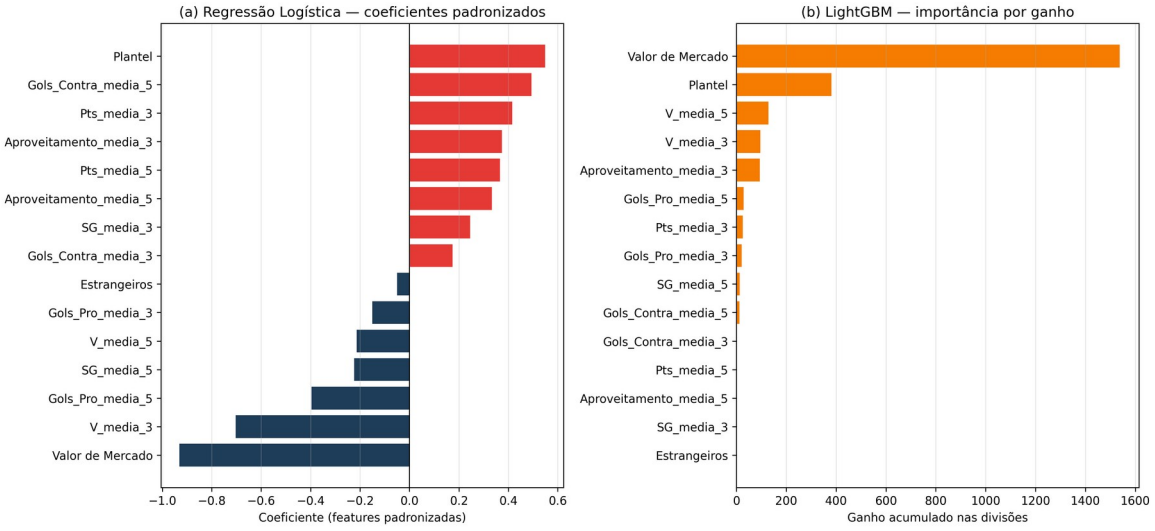

Figura 6 – Importância das features: (a) coeficientes padronizados da Regressão Logística final; (b) importância por ganho do LightGBM otimizado.

Fonte: Elaborado pelo autor (2026).

4.5 A Armadilha da Acurácia: o Caso do XGBoost

O resultado mais instrutivo da comparação é o contraste do XGBoost: 82,5% de acurácia — igual à do LightGBM — mas AUC-ROC de apenas 0,652. A comparação das matrizes de confusão (Figura 4) torna o caso ainda mais eloquente do que a mera igualdade de acurácia sugere: as matrizes dos dois modelos são idênticas. Ambos identificam 3 dos 8 rebaixamentos, incorrem em 5 falsos negativos e 2 falsos positivos e acertam 30 permanências. No limiar de 0,5, portanto, os dois modelos tomam exatamente as mesmas decisões — e ambos exibem o viés para a classe majoritária descrito por He e Garcia (2009), prevendo permanência para 35 dos 40 clubes.

Se as decisões binárias são as mesmas, a diferença de 0,225 no AUC-ROC não pode estar nelas: ela reside inteiramente na ordenação das probabilidades. Ordenando as 40 observações do teste por risco previsto, o LightGBM posiciona os oito clubes efetivamente

rebaixados nos lugares 1, 3, 5, 6, 7, 9, 17 e 19; o XGBoost, nos lugares 1, 3, 4, 18, 21, 24, 26 e 28 — isto é, atribui a cinco clubes que caíram um risco menor que o de mais de vinte clubes que permaneceram. A acurácia enxerga apenas o acerto no limiar e não distingue as duas situações; o AUC-ROC avalia a ordenação completa e as separa com nitidez.

O caso demonstra empiricamente por que a acurácia não deve ser utilizada como métrica única neste tipo de problema e reforça a escolha do AUC-ROC como critério principal deste trabalho. Reforça, ainda, o argumento gerencial desenvolvido na Seção 4.6: o que interessa ao gestor não é um rótulo binário, mas a lista ordenada dos clubes que merecem atenção — exatamente aquilo que a acurácia é incapaz de aferir.

4.6 Estabilidade versus Desempenho: Escolha do Modelo Final

Os resultados colocam em evidência um compromisso (*trade-off*) clássico. O LightGBM maximiza o AUC-ROC no teste (0,877), mas sua alta variância no *walk-forward* (±0,151) indica menor consistência entre temporadas. A Regressão Logística, embora terceira no teste (0,828), foi a mais estável na validação temporal (±0,058) e oferece coeficientes diretamente interpretáveis — *features* com coeficiente positivo aumentam o risco previsto de rebaixamento, informação diretamente utilizável pela gestão esportiva. Assim, para aplicação prática, a Regressão Logística foi adotada como modelo final de previsão, pela combinação de estabilidade e interpretabilidade; o LightGBM permanece como alternativa quando a maximização do AUC-ROC for o critério prioritário — escolha que a validação retroativa da temporada 2025 (Seção 4.8) viria a corroborar.

4.7 Previsão para a Temporada 2025

Aplicando o modelo final — a Regressão Logística, treinada com todo o período 2014–2024 —, aos 20 clubes participantes da Série A de 2025, obtém-se o ranking de risco da Tabela 8, ilustrado na Figura 7. Registre-se, para eliminar qualquer ambiguidade, que todas as probabilidades da Tabela 8 foram geradas exclusivamente pela Regressão Logística; as previsões correspondentes do LightGBM são apresentadas na Seção 4.8, no contexto da validação retroativa. Seguindo o formato do campeonato, os quatro clubes de maior probabilidade são classificados na zona de rebaixamento; destaca-se que apenas Juventude, Sport Recife e Vitória superaram o limiar de 50%, tendo o Mirassol sido incluído pela posição no ranking (4.º maior risco).

Tabela 8 – Ranking de risco de rebaixamento para a temporada 2025 (Regressão Logística treinada em 2014–2024). Os quatro primeiros são classificados na zona de rebaixamento.

#	Clube	Prob. Rebaixamento	Prob. Permanência	Previsão
1	Juventude	79,4%	20,6%	Rebaixado
2	Sport Recife	76,9%	23,1%	Rebaixado

#	Clube	Prob. Rebaixamento	Prob. Permanência	Previsão
3	Vitória	76,7%	23,3%	Rebaixado
4	Mirassol	43,6%	56,4%	Rebaixado
5	Ceará	35,3%	64,7%	Permanece
6	Bahia	33,1%	66,9%	Permanece
7	Fortaleza	27,9%	72,1%	Permanece
8	Fluminense	25,8%	74,2%	Permanece
9	Vasco da Gama	24,1%	75,9%	Permanece
10	Bragantino	13,1%	86,9%	Permanece
11	Atlético Mineiro	9,3%	90,7%	Permanece
12	Internacional	9,1%	90,9%	Permanece
13	São Paulo	9,0%	91,0%	Permanece
14	Corinthians	6,4%	93,6%	Permanece
15	Grêmio	5,0%	95,0%	Permanece
16	Cruzeiro	4,1%	95,9%	Permanece
17	Santos	2,9%	97,1%	Permanece
18	Botafogo	2,5%	97,5%	Permanece
19	Flamengo	0,2%	99,8%	Permanece
20	Palmeiras	0,0%	100,0%	Permanece

Fonte: Elaborado pelo autor (2026).

As previsões são coerentes com as características que o modelo penaliza: o Juventude acumula baixo aproveitamento nas temporadas recentes; Sport Recife e Vitória são recém-promovidos com elencos de menor valor de mercado; e o Mirassol faz sua estreia histórica na Série A com o menor valor de elenco entre os 20 participantes (€28,4 milhões), dependendo da imputação por mediana nas *features* de histórico — o que ilustra tanto o funcionamento quanto os limites do modelo para clubes sem passado na elite.

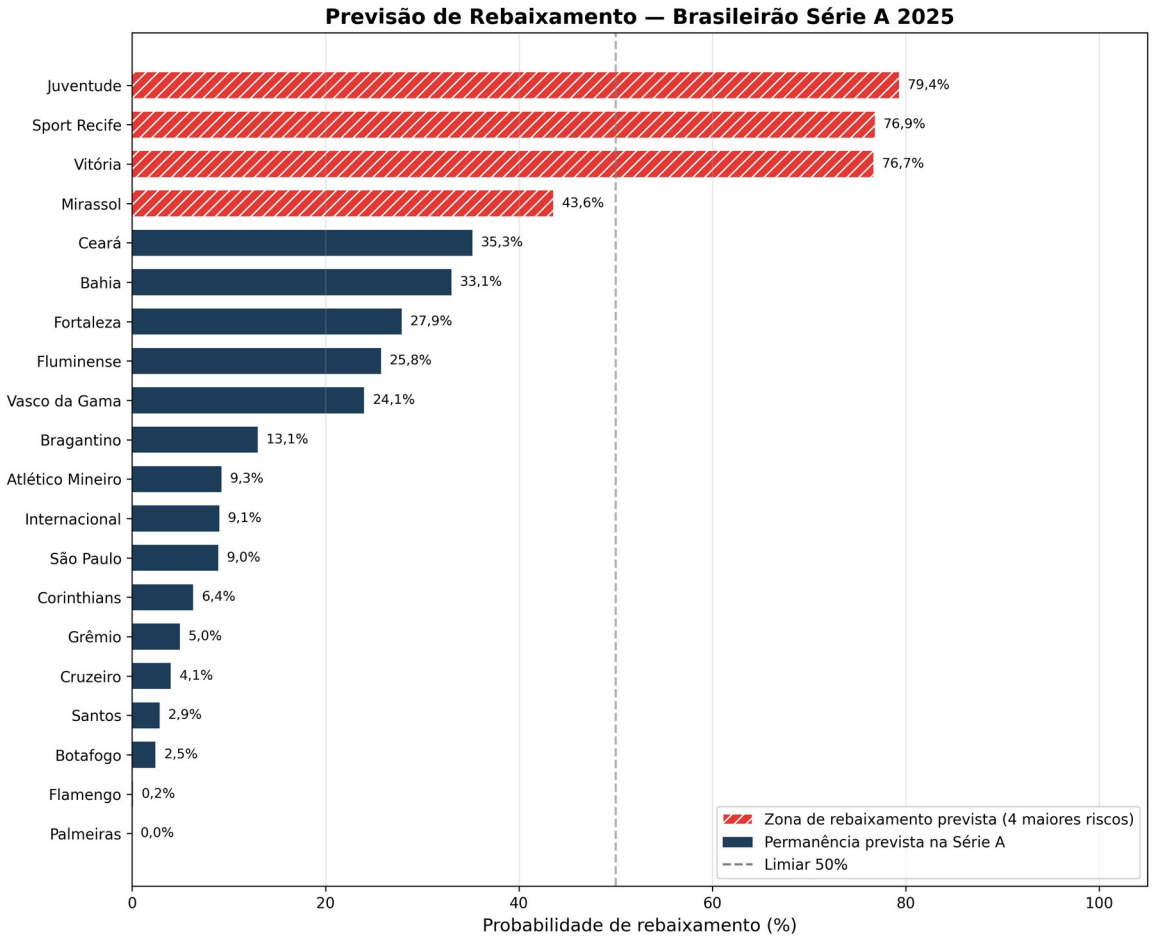

Figura 7 – Ranking de risco de rebaixamento — Brasileirão Série A 2025. Barras vermelhas hachuradas indicam os quatro clubes previstos para a zona de rebaixamento.

Fonte: Elaborado pelo autor (2026).

4.8 Validação Retroativa com os Resultados Oficiais de 2025

O Campeonato Brasileiro de 2025 foi concluído em dezembro de 2025, o que permite confrontar a previsão apresentada na seção anterior com o desfecho real — transformando-a de exercício prospectivo em resultado verificável. Os quatro clubes efetivamente rebaixados foram Sport Recife (20.º colocado), Juventude (19.º), Fortaleza (18.º) e Ceará (17.º). A Tabela 9 reapresenta o ranking de risco da Regressão Logística, agora acompanhado de intervalos de confiança de 95% (obtidos por *bootstrap* com B = 1000 reamostragens de todo o *pipeline*, da imputação ao ajuste), das probabilidades previstas pelo LightGBM treinado no mesmo período e da posição final real de cada clube.

Tabela 9 – Validação retroativa da previsão 2025: probabilidades da Regressão Logística (com intervalos de confiança de 95% via bootstrap, B = 1000) e do LightGBM versus o resultado real do campeonato. Clubes em negrito foram efetivamente rebaixados.

#	Clube	Prob. RL	IC 95%	Prob. LGBM	Pos. real
1	Juventude	79,4%	[65,6; 92,4]	62,4%	19.º
2	Sport Recife	76,9%	[42,0; 94,1]	52,8%	20.º

#	Clube	Prob. RL	IC 95%	Prob. LGBM	Pos. real
3	Vitória	76,7%	[53,1; 89,5]	47,2%	15.º
4	Mirassol	43,6%	[19,8; 56,9]	47,9%	4.º
5	Ceará	35,3%	[22,0; 77,1]	49,2%	17.º
6	Bahia	33,1%	[2,8; 66,7]	47,2%	7.º
7	Fortaleza	27,9%	[11,3; 69,0]	29,5%	18.º
8	Fluminense	25,8%	[7,0; 59,5]	31,5%	5.º
9	Vasco da Gama	24,1%	[5,5; 75,7]	52,8%	14.º
10	Bragantino	13,1%	[3,6; 42,9]	52,8%	10.º
11	Atlético Mineiro	9,3%	[0,2; 42,4]	35,0%	11.º
12	Internacional	9,1%	[2,3; 28,8]	32,7%	16.º
13	São Paulo	9,0%	[1,9; 38,0]	29,4%	8.º
14	Corinthians	6,4%	[1,3; 27,7]	29,4%	13.º
15	Grêmio	5,0%	[0,7; 31,1]	43,5%	9.º
16	Cruzeiro	4,1%	[0,6; 38,3]	45,6%	3.º
17	Santos	2,9%	[0,2; 34,5]	52,8%	12.º
18	Botafogo	2,5%	[0,0; 66,5]	37,8%	6.º
19	Flamengo	0,2%	[0,0; 6,9]	37,8%	1.º
20	Palmeiras	0,0%	[0,0; 2,4]	36,3%	2.º

Fonte: Elaborado pelo autor (2026).

Dos quatro clubes apontados na zona de rebaixamento, o modelo acertou dois — Juventude e Sport Recife, justamente o primeiro e o segundo maiores riscos previstos. Mais revelador, porém, é o desempenho de ordenação: o AUC-ROC realizado na temporada 2025 foi de 0,922, superior tanto ao obtido no conjunto de teste 2023–2024 (0,828) quanto à média da validação *walk-forward* (0,794). Os quatro rebaixados reais figuravam, todos, entre os sete maiores riscos do ranking: Ceará ocupava a 5.ª posição (35,3%) e Fortaleza a 7.ª (27,9%), imediatamente fora do corte. Dos 64 pares possíveis entre um clube rebaixado e um não rebaixado, o modelo ordenou corretamente 59.

Os dois erros do *top*-4 são instrutivos. O Vitória (3.º maior risco, 76,7%) salvou-se com 45 pontos, em 15.º lugar — um erro de magnitude moderada, já que o clube de fato flertou com a zona de rebaixamento. Já o Mirassol (4.º maior risco, 43,6%) constitui o maior erro do modelo, e o mais informativo: estreante na Série A, o clube realizou campanha extraordinária e terminou em 4.º lugar, com 67 pontos, classificando-se à Copa Libertadores. O caso expõe com nitidez a limitação da imputação por mediana discutida na Seção 3: sem qualquer histórico do clube na elite, as 12 *features* de janela receberam valores “típicos”, e nenhuma informação pré-temporada disponível ao modelo permitia antecipar desempenho tão atípico. Os intervalos de confiança reforçam a leitura prudente: com apenas ~220 observações de treinamento, os

intervalos são largos (o do Sport Recife, por exemplo, vai de 42,0% a 94,1%), recomendando que se interprete o *ranking* de risco, mais robusto, e não os valores pontuais isolados.

A comparação com o LightGBM corrobora, a posteriori, a escolha do modelo final feita na Seção 4.6 — estabilidade e interpretabilidade acima do pico de desempenho. O LightGBM também acertaria dois rebaixamentos (Juventude e Sport Recife), mas seu AUC-ROC realizado foi de apenas 0,711, com probabilidades comprimidas em uma faixa estreita (de ~29% a ~62%) e falsos alarmes relevantes — Vasco da Gama, Bragantino e Santos apareciam entre seus maiores riscos. O modelo que vencera no conjunto de teste foi visivelmente inferior na temporada seguinte, materializando exatamente a instabilidade entre períodos que a validação *walk-forward* havia diagnosticado (±0,151).

Por fim, para dimensionar o valor agregado pelo aprendizado de máquina, compararam-se duas heurísticas ingênuas de referência (*baselines*). A regra "rebaixar os quatro clubes de menor valor de elenco" acertaria dois rebaixamentos (Juventude e Ceará), com AUC-ROC de 0,906; a regra "rebaixar os quatro de pior média de pontos nas três temporadas anteriores" também acertaria dois (Juventude e Sport Recife), com AUC-ROC de 0,766. A Regressão Logística supera ambas (0,922), mas a margem estreita sobre a heurística de valor de elenco é coerente com a dominância dessa variável na análise de importância (Seção 4.4) e deve ser reconhecida com honestidade: em 2025, parte substancial do poder discriminativo já estaria ao alcance de um gestor que ordenasse os clubes apenas pelo valor de mercado. A contribuição do modelo está em integrar elenco e desempenho recente em um único escore probabilístico, calibrado e reprodutível — e em quantificar a incerteza associada.

4.9 Visão Longitudinal do Risco (2016–2025)

Para encerrar a análise, a Figura 8 consolida o comportamento do modelo ao longo de dez temporadas em um mapa de calor longitudinal: para cada temporada T entre 2016 e 2025, a Regressão Logística foi retreinada apenas com as temporadas anteriores a T (janela expansiva, com imputação e padronização recalculadas a cada retreinamento), e a probabilidade prevista de rebaixamento é exibida para cada clube participante — células contornadas em preto indicam rebaixamentos que de fato ocorreram. Três padrões emergem com nitidez. Primeiro, o "efeito elevador": clubes como Avaí, Goiás, Atlético-GO e Juventude alternam ausências (fora da Série A) com temporadas de risco previsto elevado, quase sempre confirmado — o modelo captura bem o perfil estrutural do clube recém-promovido de menor investimento. Segundo, a assimetria do risco: os clubes de maior valor de elenco (Flamengo, Palmeiras, Corinthians, São Paulo) permanecem em faixa de risco consistentemente baixa durante toda a década, enquanto a zona de perigo é disputada por um conjunto rotativo de clubes médios e recém-promovidos. Terceiro, os limites do modelo ficam igualmente visíveis: rebaixamentos de clubes tradicionais em crise aguda — Internacional em 2016 (risco previsto de 14%), Cruzeiro em 2019 (7%), Grêmio em 2021 (0,4%) e Santos em 2023 (11%) — são sistematicamente subestimados, pois

decorrem de colapsos dentro da temporada (crises institucionais, trocas sucessivas de treinador) que nenhuma *feature* pré-temporada antecipa. Essa leitura longitudinal complementa a validação retroativa de 2025 e reforça, em uma única visualização, tanto o valor prático quanto o alcance da abordagem proposta.

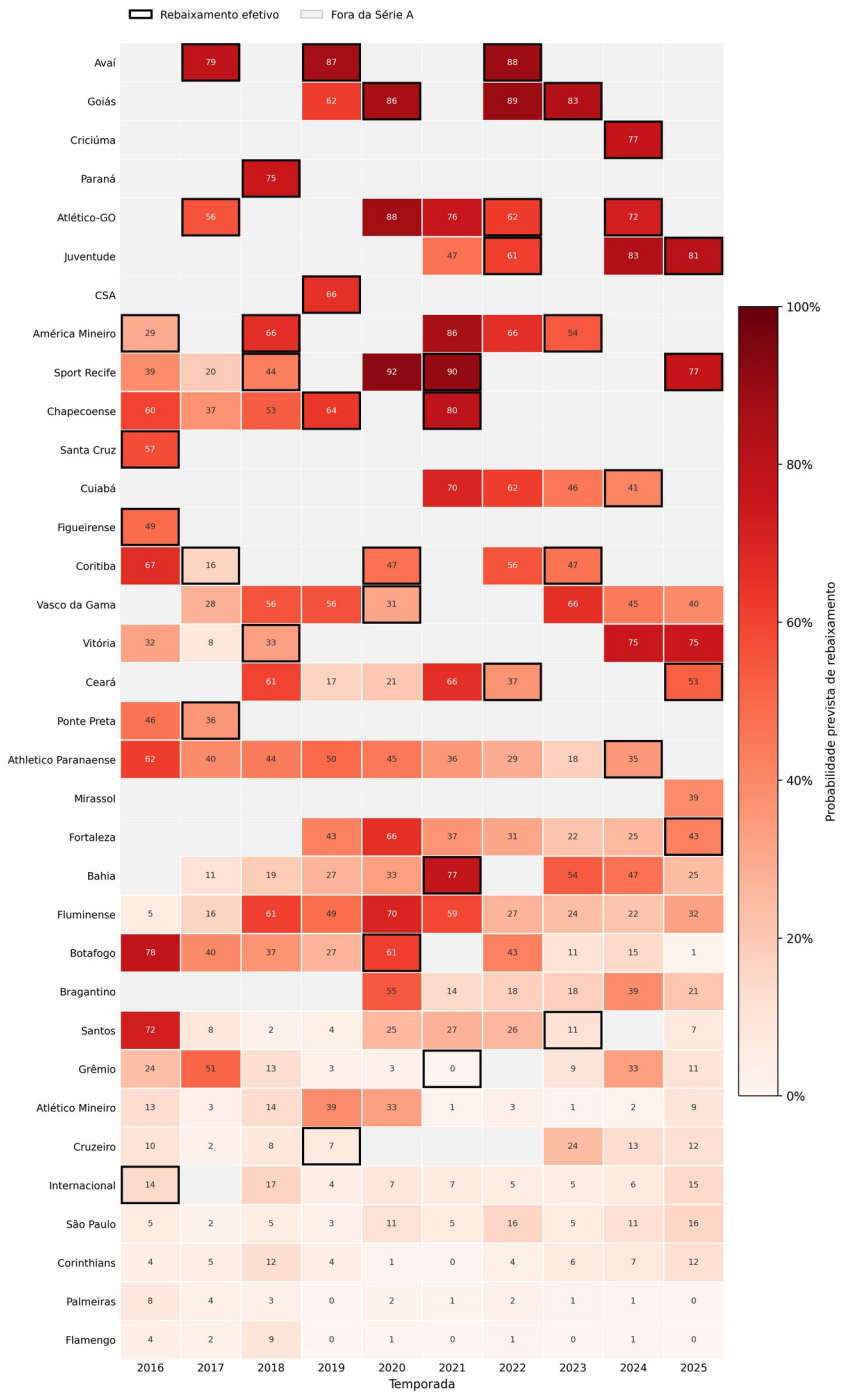

Figura 8 – Mapa de calor longitudinal do risco de rebaixamento (2016–2025): probabilidades previstas pela Regressão Logística retreinada em janela expansiva (apenas temporadas anteriores à prevista). Células contornadas em preto indicam rebaixamentos efetivos; células cinza indicam temporadas fora da Série A. Por adotar protocolo uniforme de retreinamento, os valores de 2025 podem diferir marginalmente dos da Tabela 8.

Fonte: Elaborado pelo autor (2026).

4.10 Discussão Crítica dos Resultados e Limites Metodológicos

Os resultados apresentados nas seções anteriores comportam ressalvas metodológicas que, por transparência, são explicitadas a seguir. Seis pontos merecem exame: a significância estatística das diferenças entre modelos, a estabilidade do ranqueamento entre eles, a escolha do AUC-ROC em detrimento da área sob a curva de precisão-*recall*, a hipótese de independência entre observações, a natureza de ranqueamento do problema e, por fim, a possibilidade de que o modelo esteja apenas reproduzindo uma medida geral de força dos clubes.

(i) Significância estatística das diferenças entre modelos. As diferenças de AUC-ROC relatadas nas Tabelas 4 e 6 não foram submetidas a teste formal de significância, e a leitura dos resultados deve refletir essa ausência. Na validação *walk-forward*, os desvios padrão dos dois melhores modelos se sobrepõem amplamente (Regressão Logística, 0,794 ± 0,058; *Random Forest*, 0,762 ± 0,071), de modo que a vantagem da primeira não é estatisticamente distinguível com cinco *folds*. No conjunto de teste, com apenas 40 observações e 8 rebaixamentos, o erro padrão de qualquer estimativa de AUC-ROC é necessariamente grande — os intervalos de confiança por *bootstrap* reportados na Tabela 9, que chegam a amplitudes superiores a 50 pontos percentuais em nível de clube, são a evidência mais direta desse regime de baixa precisão amostral. A comparação formal de curvas ROC correlacionadas pelo teste de DeLong, ou por *bootstrap* pareado sobre as diferenças de AUC, fica registrada como recomendação metodológica para extensões deste trabalho. Enquanto tais testes não forem conduzidos, as ordens de classificação entre modelos devem ser lidas como indicativas, e não como evidência de superioridade estabelecida.

(ii) Estabilidade do ranqueamento entre modelos. A resposta honesta é negativa: o ranqueamento não se manteve estável entre os períodos avaliados. A Regressão Logística lidera na validação *walk-forward* (0,794) e na temporada 2025 (0,922), ao passo que o LightGBM lidera no conjunto de teste 2023–2024 (0,877) e cai para 0,711 em 2025; o *Random Forest*, segundo colocado no teste, não replica essa posição nos demais recortes. Essa alternância é exatamente o que se espera de comparações conduzidas sobre amostras pequenas, nas quais a variância amostral das estimativas de desempenho é da mesma ordem de grandeza das diferenças entre modelos. Daí a opção, assumida na Seção 4.6, de privilegiar consistência entre períodos e interpretabilidade em lugar do melhor desempenho pontual: a escolha do modelo final não repousa sobre uma diferença de AUC-ROC julgada significativa, mas sobre o comportamento menos volátil da Regressão Logística ao longo das temporadas.

(iii) AUC-ROC e a alternativa da curva precisão-recall. Em classes desbalanceadas, é frequente a recomendação de reportar a área sob a curva de precisão-*recall* (PR-AUC), sob o argumento de que o AUC-ROC pode ser otimista quando a classe positiva é rara. Duas particularidades do problema atenuam essa objeção. A primeira é que a prevalência aqui não é incidental nem variável entre amostras: o regulamento do campeonato fixa em quatro os rebaixamentos entre vinte clubes, de modo que a proporção de positivos é constante em 20% em

todas as temporadas, no treino, no teste e na aplicação. Não há, portanto, o risco de a métrica ser inflada por uma prevalência artificialmente baixa ou instável. A segunda é que o uso pretendido do modelo é de ordenação — apresentar ao gestor a lista de clubes por risco —, e o AUC-ROC é precisamente a probabilidade de que um clube rebaixado receba escore superior ao de um não rebaixado, interpretação explorada na Seção 4.8. Ainda assim, o PR-AUC informaria melhor o desempenho no topo do ranking, região de maior interesse prático, e sua inclusão fica registrada como aprimoramento recomendado, ao lado de métricas sensíveis ao topo da lista.

(iv) Independência entre as observações. A base contém 220 observações provenientes de 34 clubes distintos, com até 11 aparições por clube ao longo do período. As observações não são, portanto, independentes: características não observadas e persistentes de cada clube — estrutura de gestão, torcida, capacidade institucional de recuperação — afetam simultaneamente todas as suas temporadas. A validação *walk-forward* endereça a dependência temporal, impedindo que informação do futuro contamine o treino, mas não a dependência de agrupamento por clube: o mesmo clube aparece no treino e na validação em anos distintos. As consequências são conhecidas: erros padrão subestimados e estimativas de desempenho possivelmente otimistas. Remédios apropriados incluem erros padrão agrupados por clube, esquemas de validação que combinem separação temporal com separação por agrupamento (*GroupKFold*) e modelos de efeitos aleatórios por clube. Registre-se, contudo, que a validação por agrupamento seria de aplicação delicada neste caso, pois excluir do treino um clube presente na validação reduziria substancialmente uma amostra já modesta.

(v) Ranqueamento versus classificação binária. O problema foi formulado como classificação binária independente por clube, ainda que sua estrutura real seja de ranqueamento com restrição: exatamente quatro entre vinte clubes caem, e o que determina o rebaixamento é a posição relativa, não o valor absoluto de um escore. A formulação adotada tem a vantagem de produzir probabilidades interpretáveis e calibradas (Figura 5), mas ignora essa restrição — nada impede, na implementação atual, que o modelo atribua probabilidade superior a 50% a seis clubes ou a nenhum, como de fato ocorre na Tabela 8, em que apenas três clubes ultrapassam o limiar e o quarto rebaixamento é atribuído por posição no ranking. Uma formulação alternativa exigiria: aprendizado direto de ordenação (*learning to rank*), com funções de perda que penalizem inversões de pares; normalização das probabilidades dentro de cada temporada, tornando o escore relativo aos concorrentes daquele ano; imposição da restrição de quatro rebaixamentos por temporada na etapa de decisão; e avaliação por métricas sensíveis ao topo da lista, como precisão em k = 4 e NDCG, complementando o AUC-ROC. As análises das Seções 4.5 e 4.8 sinalizam que essa é a direção natural de evolução do trabalho, pois o valor gerencial demonstrado reside na ordenação, e não no rótulo binário.

(vi) O modelo captura apenas "força do clube"? A objeção é pertinente: se o valor de mercado do elenco domina a importância das variáveis (Seção 4.4) e a heurística baseada apenas nele alcança AUC-ROC de 0,906 em 2025, contra 0,922 do modelo completo, cabe

perguntar se o *pipeline* não é uma reconstrução elaborada de uma medida simples de força financeira. Três evidências sugerem que há mais do que isso, embora nenhuma seja conclusiva. Primeira, o tamanho do plantel entra com coeficiente positivo (+0,549), isto é, aumenta o risco previsto — efeito que não acompanha monotonicamente a força do clube e que só se manifesta quando controlados os demais fatores. Segunda, a média de vitórias nas três temporadas anteriores é o segundo fator de proteção mais relevante (−0,703), contribuição autônoma do histórico em campo que a heurística puramente financeira não incorpora. Terceira, o modelo supera com folga a heurística baseada apenas em desempenho recente (0,922 contra 0,766), o que indica que a combinação das duas dimensões é superior a qualquer uma isoladamente. A verificação decisiva, no entanto, seria um estudo de ablação — reestimar o modelo sem o valor de mercado, sem as janelas de desempenho e com cada grupo isoladamente, comparando o AUC-ROC resultante. Essa análise não foi conduzida e fica registrada como limitação e como pendência imediata para trabalhos futuros. Reconheça-se, por fim, que a margem estreita sobre a heurística financeira em 2025, já assinalada na Seção 4.8, é um dado a ser lido com sobriedade: em uma única temporada, o ganho do aprendizado de máquina sobre uma regra simples foi modesto.

4.11 Aplicação Web: Entrega Tecnológica do Modelo

Em coerência com a natureza tecnológica deste trabalho, o modelo final foi disponibilizado em uma aplicação web de acesso aberto, desenvolvida em Python com a biblioteca Streamlit e hospedada em serviço público de nuvem. A aplicação converte o resultado analítico em ferramenta consultável, permitindo que gestores, jornalistas e pesquisadores acessem o ranking de risco, examinem as métricas de avaliação e explorem a base histórica sem executar código. Sua arquitetura é a de uma camada de visualização sobre o mesmo *pipeline* descrito na Seção 3: os coeficientes e as transformações de pré-processamento são carregados a partir do modelo serializado, de modo que os números exibidos correspondem aos reportados neste artigo. A interface organiza-se em seis telas, descritas a seguir e ilustradas nas Figuras 9 a 14.

A tela de previsão (Figura 9) reúne a identificação do modelo — algoritmo, períodos de treino e teste, acurácia, AUC-ROC de teste e de validação cruzada — e disponibiliza um simulador individual de risco, no qual o usuário informa indicadores de elenco e de desempenho e obtém a probabilidade estimada para um clube hipotético. Registre-se que o simulador opera em versão reduzida, com nove indicadores em que as janelas de 3 e de 5 temporadas recebem o mesmo valor; o ranking oficial da aplicação utiliza o modelo completo, com as 15 *features* e o histórico real de cada clube, razão pela qual os números das duas telas não coincidem exatamente. Essa distinção está sinalizada na própria interface, para evitar leitura equivocada.

Figura 9 – Aplicação web: tela de previsão, com o painel de identificação do modelo e o simulador individual de risco.

Fonte: Elaborado pelo autor (2026).

A tela de ranking (Figura 10) apresenta a projeção do modelo para os 20 clubes da Série A 2025, ordenados pela probabilidade estimada de rebaixamento, com destaque em vermelho para os quatro maiores riscos — a zona de rebaixamento projetada. Indicadores agregados informam o risco médio da liga e o clube de menor risco, e cada linha exibe a barra de probabilidade, o estado de origem, o rótulo de clube promovido, quando aplicável, e o valor de mercado do elenco. Os valores reproduzem os da Tabela 8.

Figura 10 – Aplicação web: ranking de risco dos 20 clubes da Série A 2025, com destaque para a zona de rebaixamento projetada.

Fonte: Elaborado pelo autor (2026).

A tela de desempenho (Figura 11) documenta a avaliação do modelo final no conjunto de teste independente. Além de acurácia, precisão, *recall*, F1-score e AUC-ROC — de teste e de validação *walk-forward* —, exibe a matriz de confusão, a curva ROC e o diagrama de calibração discutidos nas Seções 4.3 e 4.5, acompanhados de nota que adverte, na própria interface, sobre a inadequação da acurácia como métrica isolada em classes desbalanceadas. A tela cumpre, assim, função pedagógica: expõe ao usuário não apenas o que o modelo prevê, mas quão confiáveis são suas previsões.

Figura 11 – Aplicação web: painel de desempenho do modelo, com métricas, matriz de confusão, curva ROC e diagrama de calibração.

Fonte: Elaborado pelo autor (2026).

A tela de dados históricos (Figura 12) dá acesso à base rotulada utilizada no treinamento e na avaliação, com filtros por situação do clube na temporada (rebaixado, permanência na Série A ou classificação entre os quatro primeiros). Os indicadores agregados confirmam a composição descrita na Seção 3.1: 220 registros, distribuídos por 34 clubes distintos ao longo de 11 temporadas, com 44 rebaixamentos — os quatro de cada edição —, o que corresponde à prevalência de 20% da classe positiva. A tabela detalha, por clube e temporada, o tamanho do plantel, o número de estrangeiros e o valor de mercado do elenco.

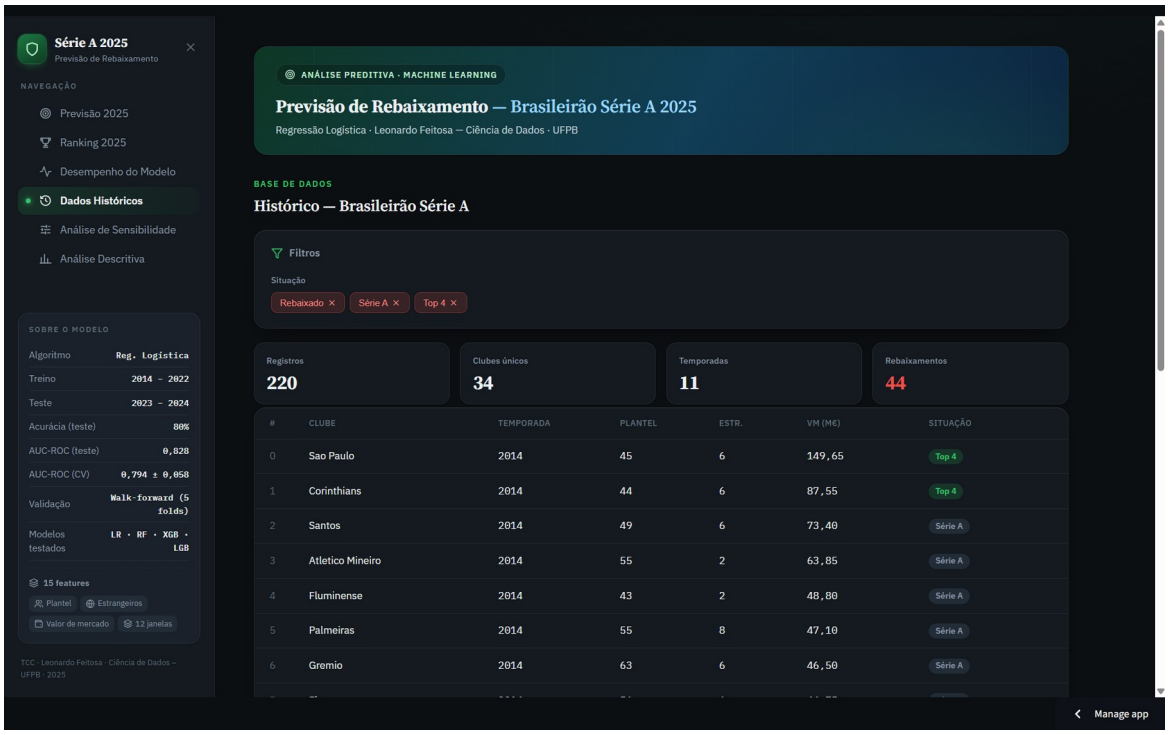

Figura 12 – Aplicação web: base de dados histórica (2014–2024) com filtros por situação do clube na temporada. Fonte: Elaborado pelo autor (2026).

A tela de sensibilidade (Figura 13) traduz os coeficientes da Regressão Logística em curvas de leitura direta: cada gráfico mostra como a probabilidade prevista responde à variação de uma variável, mantidas as demais nos valores médios históricos. A visualização torna evidente o que a Seção 4.4 discute em termos de coeficientes — o valor de mercado é o fator mais decisivo, com queda acentuada do risco à medida que cresce; o número de estrangeiros tem efeito praticamente neutro; e o tamanho do plantel apresenta relação crescente com o risco, o achado contraintuitivo já comentado.

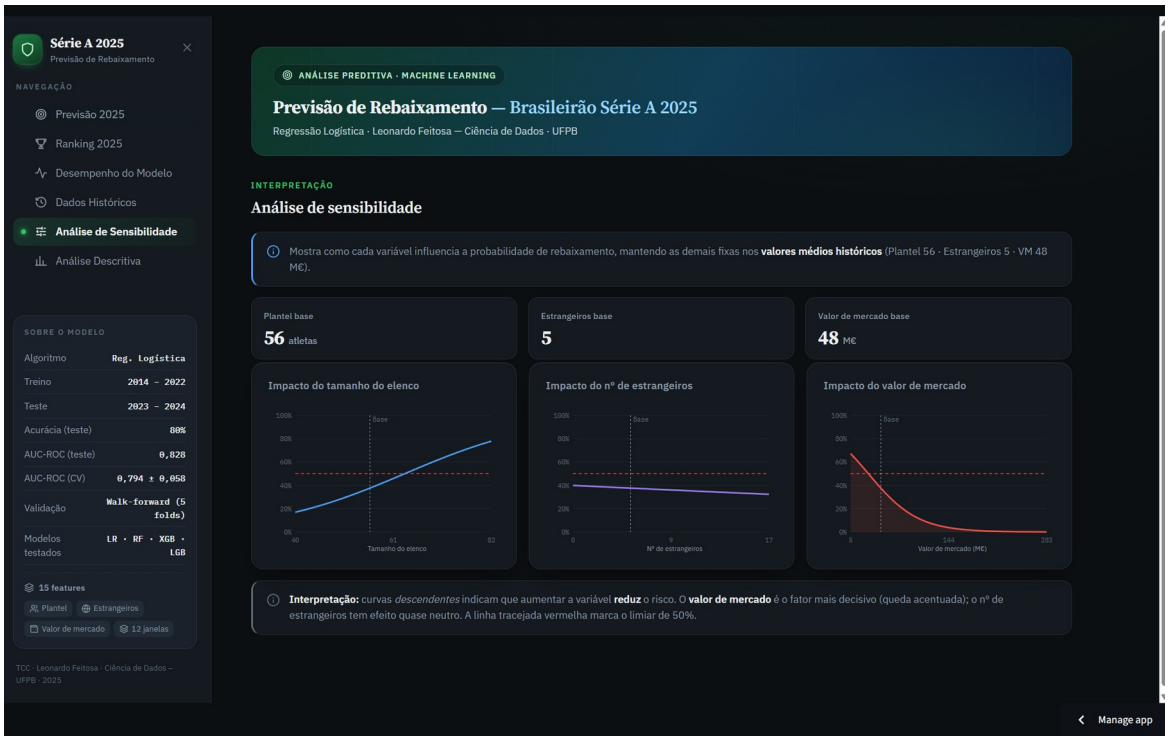

Figura 13 – Aplicação web: análise de sensibilidade da probabilidade prevista às variáveis de elenco.
Fonte: Elaborado pelo autor (2026).

A tela de análise descritiva (Figura 14), por fim, permite selecionar uma temporada e inspecionar as estatísticas descritivas das variáveis — média, desvio padrão, mínimo, quartis e máximo —, além de comparações por clube e da distribuição de cada indicador. Trata-se da contrapartida interativa da Tabela 2, útil para contextualizar um clube específico frente aos demais participantes daquela edição.

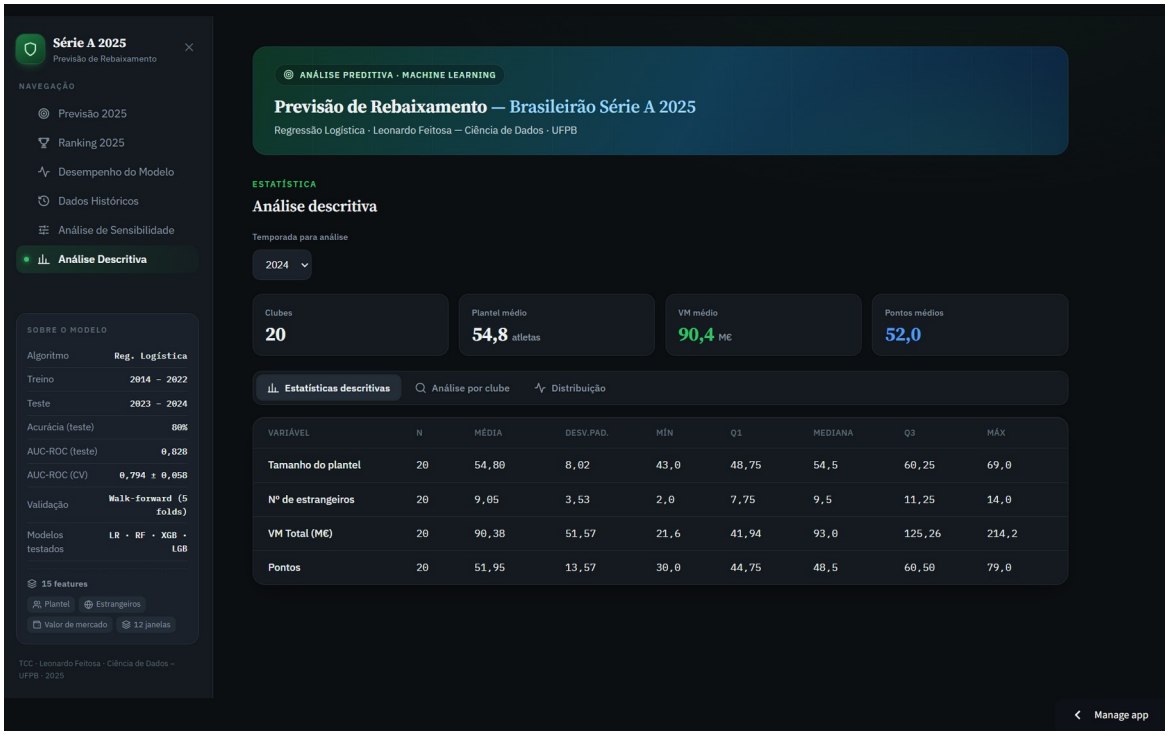

Figura 14 – Aplicação web: análise descritiva das variáveis por temporada selecionada.
Fonte: Elaborado pelo autor (2026).

Cumpre observar que a aplicação herda integralmente as limitações do modelo discutidas na Seção 4.10 e, por isso, apresenta as probabilidades sempre acompanhadas do ranking e das métricas de desempenho — orientando o usuário a interpretar a ordenação, mais robusta, e não os valores pontuais isolados.

5 CONSIDERAÇÕES FINAIS

Este trabalho desenvolveu e comparou quatro modelos de *Machine Learning* para previsão de rebaixamento no Campeonato Brasileiro Série A a partir de informações disponíveis antes do início da temporada. Retomando os objetivos específicos: a base histórica 2014–2025 foi construída com dados do Transfermarkt (objetivo i); a engenharia de *features* com janelas deslizantes foi implementada sem vazamento temporal (objetivo ii); os quatro algoritmos foram treinados e otimizados com validação *walk-forward* e RandomizedSearchCV (objetivo iii); a comparação em teste independente apontou o LightGBM como melhor discriminador (AUC-ROC de 0,877, significativamente acima do 0,500 esperado de um classificador aleatório) e a Regressão Logística como modelo mais estável e interpretável (objetivo iv); e a previsão para 2025 foi gerada, indicando Juventude, Sport Recife, Vitória e Mirassol como os clubes de maior risco (objetivo v). A previsão foi, posteriormente, confrontada com os resultados oficiais do campeonato: dois dos quatro rebaixamentos foram antecipados corretamente, os quatro rebaixados reais figuravam entre os sete maiores riscos do ranking e o AUC-ROC realizado na temporada foi de 0,922 — o melhor entre todos os períodos avaliados, superando ainda heurísticas ingênuas de referência. O objetivo geral foi, portanto, plenamente alcançado.

As contribuições do trabalho são quatro. Primeira, um *pipeline* metodologicamente rigoroso e integralmente reprodutível para previsão esportiva com dados temporais no contexto brasileiro, disponível em repositório público. Segunda, a demonstração empírica da “armadilha da acurácia”: o contraste entre a acurácia de 82,5% e o AUC-ROC de 0,652 do XGBoost oferece um exemplo didático da inadequação da acurácia como métrica única em classes desbalanceadas. Terceira, a evidência de que o histórico recente de desempenho, capturado pelas janelas deslizantes, complementa de forma relevante os indicadores financeiros de elenco na explicação do risco de rebaixamento. Quarta, de natureza tecnológica, a entrega do modelo final como uma aplicação web de acesso aberto, que converte o resultado analítico em ferramenta consultável por gestores e pesquisadores, sem exigir a execução do código. A validação retroativa da previsão 2025, com AUC-ROC de 0,922 e superação das heurísticas de referência, converte a previsão apresentada de promessa em resultado verificável — e os próprios erros do modelo, em especial o caso do Mirassol, delimitam com precisão o alcance da abordagem pré-temporada.

O estudo possui limitações que devem ser reconhecidas. A restrição a dados pré-temporada, embora confira valor prático à previsão, impede que o modelo capture a dinâmica da própria temporada em curso. Fatores sabidamente relevantes — lesões, trocas de treinador, sobrecarga de calendário e movimentações nas janelas de transferências — não são modelados. O volume de dados (~220 observações) é reduzido para os padrões de *Machine Learning*, o que amplia a variância das estimativas, como evidenciado pela oscilação do LightGBM entre *folds* (±0,151), e limita a complexidade dos modelos treináveis. A essas limitações somam-se as ressalvas metodológicas discutidas na Seção 4.10 — ausência de teste formal de significância entre modelos, dependência entre observações do mesmo clube e formulação binária de um problema que é, em essência, de ranqueamento. Por fim, clubes recém-promovidos, sem histórico na Série A, dependem da imputação por mediana, aproximação razoável porém imperfeita de sua realidade competitiva — o caso do Mirassol em 2025, previsto na zona de rebaixamento e concluindo o campeonato em 4.º lugar, é a ilustração mais eloquente dessa limitação. Cabe registrar, por outro lado, que análises de sensibilidade indicaram robustez das principais decisões de projeto: substituir a imputação por mediana pela média não alterou o AUC-ROC no conjunto de teste (0,828 em ambas as configurações), e o uso de janelas de 2 e 4 temporadas produziu resultado equivalente (0,836) ao das janelas de 3 e 5 adotadas.

Como trabalhos futuros, propõe-se: (i) o retreinamento parcial do modelo ao longo da temporada, incorporando a pontuação acumulada para capturar a forma corrente dos clubes; (ii) a inclusão de *features* do mercado de transferências, dado que reforços e vendas impactam diretamente o desempenho esperado; (iii) a experimentação com modelos sequenciais (LSTM, *Transformers*) à medida que mais temporadas ampliem a base de dados; (iv) a reformulação do problema como tarefa de ranqueamento, com métricas sensíveis ao topo da lista e imposição da restrição de quatro rebaixamentos por temporada; (v) a realização de estudo de ablação por grupo de variáveis e de testes formais de significância entre modelos; e (vi) a expansão do escopo para a Série B e para ligas de outros países, aumentando o volume de dados e a testabilidade externa das conclusões.

AGRADECIMENTOS

A conclusão deste trabalho não é obra de uma pessoa só, e por isso agradeço a todos que caminharam comigo até aqui.

Agradeço primeiramente a Deus, por ter me dado a oportunidade de chegar a este momento, por me sustentar nos períodos de cansaço e de dúvida e por colocar no meu caminho as pessoas certas em cada etapa desta jornada.

Aos meus pais, Francisco Pereira Barroso e Luciana de Oliveira Feitosa, que sempre colocaram a minha formação acima de qualquer outra coisa. Cada oportunidade que tive nasceu de um esforço de vocês, e é impossível separar esta conquista de tudo o que abriram mão para que ela fosse possível. Obrigado pelo exemplo, pela paciência e pelo amor de todos os dias — esta vitória também é de vocês.

Aos meus irmãos, Luana Feitosa e Lucas Feitosa, por todo o apoio ao longo desses anos, pelo companheirismo nas horas difíceis e por tornarem o caminho muito mais leve. Agradeço, em especial, ao meu irmão, que me incentivou a cursar Ciência de Dados para Negócios e que segue me incentivando até hoje: a decisão que deu origem a este trabalho começou nas conversas com você.

À minha companheira e namorada, Livia Lopes, por estar comigo durante todo o percurso, pela compreensão nos períodos em que estive ausente por causa deste trabalho e por me ajudar e motivar a cada dia a continuar. Ter alguém que acredita em mim mesmo quando eu duvido fez toda a diferença.

Ao meu orientador, Prof. Dr. Hilton Martins de Brito Ramalho, pela orientação atenta, pelas críticas que tornaram este trabalho melhor e pela disponibilidade ao longo de todo o processo.

Por fim, aos professores do Curso de Bacharelado em Ciência de Dados para Negócios e à Universidade Federal da Paraíba, pela formação que recebi, e aos colegas de curso, pela troca e pela parceria durante a graduação.

REFERÊNCIAS

ARABZAD, S. M. et al. Football match results prediction using artificial neural networks: The case of Iran Pro League. *Journal of Applied Research on Industrial Engineering*, v. 1, n. 3, p. 159–179, 2014.

BABOOTA, R.; KAUR, H. Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, v. 35, n. 2, p. 741–755, 2019.

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, v. 191, p. 192–213, 2012.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, p. 281–305, 2012.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

BUNKER, R. P.; THABTAH, F. A machine learning framework for sport result prediction. *Applied Computing and Informatics*, v. 15, n. 1, p. 27–33, 2019.

CARPITA, M.; CIAVOLINO, E.; PASCA, P. Exploring and modelling team performances of the Kaggle European Soccer database. *Statistical Modelling*, v. 19, n. 1, p. 74–101, 2019.

CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: [s.n.], 2016. p. 785–794.

CONSTANTINOU, A. C.; FENTON, N. E.; NEIL, M. pi-football: A Bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, v. 36, p. 322–339, 2012.

COX, D. R. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 20, n. 2, p. 215–242, 1958.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.

GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3. ed. Sebastopol: O'Reilly Media, 2022.

GIL, A. C. *Como Elaborar Projetos de Pesquisa*. 6. ed. São Paulo: Atlas, 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009.

HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, p. 1263–1284, 2009.

HUBÁČEK, O.; ŠOUREK, G.; ŽELEZNÝ, F. Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning*, v. 108, n. 1, p. 29–47, 2019.

JAMES, G. et al. *An Introduction to Statistical Learning: with Applications in Python*. New York: Springer, 2023.

KE, G. et al. LightGBM: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Long Beach: [s.n.], 2017. p. 3146–3154.

MACIEL, V. B. *Predição de resultados de partidas do Campeonato Brasileiro de futebol*. 2024. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) —

Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: https://lume.ufrgs.br/handle/10183/273233. Acesso em: 14 jul. 2026.

MÜLLER, O.; SIMONS, A.; WEINMANN, M. Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, v. 263, n. 2, p. 611–624, 2017.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

TASHMAN, L. J. Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, v. 16, n. 4, p. 437–450, 2000.

TRANSFERMARKT. *Mercado de transferências, rumores, valores de mercado e notícias*. 2025. Disponível em: https://www.transfermarkt.com.br. Acesso em: abr. 2025.

Emitido em 31/07/2026

DOCUMENTO Nº 1/2026 - CCDN (11.00.52.09)
(Nº do Documento: 1)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 14/09/2026 10:16)
ANDREA ALVES BORBA
ASSISTENTE EM ADMINISTRACAO
1517808

Para verificar a autenticidade deste documento entre em https://sipac.ufpb.br/documentos/ informando seu número: **1**, ano: **2026**, documento (espécie): **DOCUMENTO**, data de emissão: **14/09/2026** e o código de verificação: **645f35e8db**