

Score Predictor: Predição de Resultados no Futebol Usando Programação Lógica

Gabriel B. R. de Oliveira¹, João V. S. Costa¹, Tiago B. e Silva¹, Tatiana A. Simões²

¹Centro de Informática – Curso de Ciência da Computação – Universidade Federal da Paraíba (UFPB) – João Pessoa, PB – Brasil

²Centro de Informática – Departamento de Computação Científica – Universidade Federal da Paraíba (UFPB) – João Pessoa, PB – Brasil

{gabriel.ribeiro,jvsc3,tiago.brito.e.silva}@academico.ufpb.br,
tatiana@ci.ufpb.br

Abstract. *Score Predictor is a system for predicting soccer match results, developed in Prolog and tested during the 2024 Brazilian Championship. Its objective is to support decisions in sports betting, estimating probabilities of victory, draw and defeat based on a logical-statistical model. The system uses data from a specialized database and allows the insertion of additional parameters through an interactive menu. The modeling combines statistical and contextual factors, allowing automated inferences. The tests carried out demonstrated satisfactory performance in predicting results. The proposal highlights the feasibility of using declarative logic in sports analysis applications.*

Resumo. *O Score Predictor é um sistema de previsão de resultados de partidas de futebol, desenvolvido em Prolog e testado durante o Campeonato Brasileiro de 2024. Seu objetivo é apoiar decisões em apostas esportivas, estimando probabilidades de vitória, empate e derrota com base em um modelo lógico-estatístico. O sistema utiliza dados de um banco especializado e permite a inserção de parâmetros adicionais por meio de um menu interativo. A modelagem combina fatores estatísticos e contextuais, permitindo inferências automatizadas. Os testes realizados demonstraram desempenho satisfatório na previsão dos resultados. A proposta evidencia a viabilidade do uso de lógica declarativa em aplicações de análise esportiva.*

Introdução

O futebol, como fenômeno esportivo global, tem despertado crescente interesse na análise de resultados, especialmente no contexto das apostas esportivas. O avanço na disponibilidade de dados estatísticos detalhados sobre o desempenho das equipes possibilita a aplicação de técnicas computacionais para previsões com maior precisão. Neste contexto, este trabalho propõe o desenvolvimento de um sistema de análise estatística para as partidas do Campeonato Brasileiro, que foi testado para o campeonato de 2024, implementado na linguagem de programação Prolog. O sistema utiliza um banco de dados estruturado com informações como gols marcados, posse de bola e histórico recente, aplicando inferência lógica e estatística para calcular as probabilidades de vitória, derrota e empate. A proposta demonstra que, com a seleção adequada de variáveis, definição criteriosa de pesos e uso de programação lógica, é possível gerar

previsões com precisão comparável à de modelos tradicionais de aprendizado de máquina.

Referencial teórico

O estudo de Felix (2019) evidenciou a escassez de pesquisas sobre previsão de resultados no futebol e propôs modelos preditivos aplicados à K League 1, principal liga da Coreia do Sul. A partir de dados estatísticos das doze equipes — como partidas disputadas, gols marcados e sofridos —, foi utilizada a distribuição de Poisson para estimar os resultados das partidas. As previsões foram comparadas com os resultados reais e com a classificação final, demonstrando a eficácia do modelo proposto.

Costa (2021) investigou diversas técnicas de aprendizado de máquina para desenvolver modelos preditivos de resultados de partidas de futebol. Foram testados desde classificadores simples até redes neurais complexas, com o objetivo de realizar previsões tanto antes do início quanto durante os jogos. A avaliação considerou acurácia e lucratividade no mercado de apostas, permitindo analisar tanto os fatores que influenciam os resultados quanto a eficácia das técnicas de machine learning nesse contexto.

Kuhnert (2023) propôs um modelo iterativo que realiza análises probabilísticas de curto e longo prazo no futebol, prevendo resultados de partidas e campeões de campeonatos. O modelo atualiza continuamente os valores numéricos das capacidades ofensivas e defensivas das equipes com base nas partidas disputadas. Para facilitar o acesso às predições, foi desenvolvida uma interface web que permite a interação dos usuários com o sistema.

Diversos estudos, como os de Bunker, Yeung e Fujii (2024), Elmiligi e Saad (2022), Fialho, Manhães e Teixeira (2019), Rico-González et al. (2022) e Yeung et al. (2023), demonstram a aplicação de técnicas de *Machine Learning* no futebol, abrangendo desde a análise tática até a previsão de resultados. Essas pesquisas reforçam a relevância e atualidade do tema, destacando seu impacto na compreensão do esporte e na otimização de estratégias e análises de desempenho.

Embora não existam registros documentados de aplicações diretas de Prolog na predição de resultados de jogos esportivos, os estudos revisados demonstram o potencial das técnicas de aprendizado de máquina e modelagem estatística nesse campo. A integração de Prolog com essas abordagens configura uma área promissora de pesquisa, especialmente no desenvolvimento de sistemas especialistas que combinem regras lógicas e modelos preditivos. Tal abordagem tem ganhado destaque nos últimos anos, impulsionada pelo crescimento do mercado de apostas esportivas.

Programação lógica com Prolog e modelagem

O uso de Prolog neste projeto destaca-se por sua eficiência na manipulação de regras e inferências lógicas, sendo ideal para sistemas baseados em conhecimento. No Score Predictor, os fatos representam os dados das equipes e as regras calculam as

probabilidades, permitindo modelar relações entre variáveis como desempenho e desfalques. Essa abordagem possibilita previsões dinâmicas, integrando dados complexos e aplicando lógica para gerar resultados preditivos consistentes.

A modelagem preditiva se baseia na atribuição de pesos às variáveis, ajustando sua influência no resultado da partida. Modelos ponderados, como indicam estudos (Schneider, 2018), são eficazes para prever resultados esportivos, pois quantificam a contribuição de cada fator. No caso de Pesos Positivos e Negativos, elementos como desfalques (pesos negativos) e vantagem de jogar em casa (pesos positivos) são ajustados conforme seu impacto no desempenho das equipes. A Normalização de Probabilidades garante que as probabilidades de vitória, empate e derrota somem 100%, aspecto crucial para a precisão do modelo, especialmente em esportes como o futebol, onde as margens de erro são mínimas e a exatidão das previsões é vital.

No futebol, variáveis contextuais são fundamentais para a análise preditiva. Este projeto considera: Vantagem de Jogo em Casa (o time mandante tem uma probabilidade maior de vitória devido a fatores como familiaridade com o campo e apoio da torcida); Desfalques (a ausência de jogadores-chave influencia significativamente o desempenho de uma equipe, sendo incorporada ao modelo com ajustes proporcionais) e Clássicos e Rivalidades (partidas entre equipes de status semelhante apresentam maior probabilidade de empate devido ao equilíbrio técnico). Essas variáveis foram selecionadas com fundamentação no referencial teórico estabelecido.

Metodologia

A metodologia de pesquisa adotada neste trabalho é aplicada, com o objetivo de desenvolver e avaliar um sistema de previsão de resultados no contexto do futebol brasileiro. A pesquisa envolve a coleta e análise de dados reais do Campeonato Brasileiro de 2024, permitindo a construção de modelos baseados em evidências empíricas. O desenvolvimento do sistema seguiu um processo sistemático, que incluiu definição de requisitos, modelagem, implementação e testes. A validação foi realizada por meio da comparação entre os resultados previstos pelo sistema, os dados reais das partidas e as previsões de casas de apostas. Métricas como taxa de acerto e margem de erro foram analisadas para aprimorar o modelo, garantindo sua eficácia e confiabilidade em contextos de alta variabilidade, como o mercado de apostas esportivas. A metodologia adotada é uma pesquisa aplicada, com abordagem quantitativa e desenvolvimento experimental, permitindo avaliar tanto a funcionalidade técnica quanto a efetividade prática do sistema em um ambiente de alta incerteza, como o futebol.

Para garantir a confiabilidade e robustez do modelo, utilizou-se uma fonte renomada de estatísticas esportivas, como o SofaScore (2024). Os dados foram processados e estruturados em um banco de dados acessível pelo Prolog, permitindo análises rápidas e precisas.

Detalhamento da Pesquisa

Primeiramente, foi estruturado um banco de dados com informações detalhadas sobre as equipes do Brasileirão 2024, abrangendo estatísticas como gols, posse de bola e chutes. Em seguida, foi implementado um modelo de previsão probabilística, que atribui pesos específicos a variáveis relevantes (estes pesos foram sendo testados e calibrados ao longo da pesquisa) e calcula as probabilidades de cada partida. Por fim, foi criada uma interface interativa que permite ao usuário inserir as equipes e os desfalques, gerando as previsões. O processo foi dividido em três etapas principais:

Etapla 1. Organização e importação dos dados

Foi criado um banco de dados em Prolog, salvo como “dados.pl”, com informações estatísticas detalhadas das 20 equipes do Brasileirão 2024, incluindo status, classificação atual, gols marcados, gols sofridos, posse de bola, chutes a gol, número de vitórias, aproveitamento como mandante/visitante, aproveitamento recente (Figura 1). Esses dados foram extraídos dos principais sites estatísticos sobre futebol, de forma manual, e formatados para serem utilizados no sistema, tornando o modelo mais realista e adaptado ao contexto futebolístico, sem perder a interpretabilidade.

```
time(athletico_pr, medio, 15, 27, 31, 50.4, 5.0, 0.7, 8, 13.33, 41.03, 38.46, 0.2).
time(athletico_pg, grande, 9, 40, 41, 58.8, 4.6, 5.8, 10, 46.66, 53.33, 41.03, 0.2).
time(athletico_go, pequeno, 20, 22, 47, 46.0, 4.5, 5.8, 5, 20.00, 28.57, 20.00, 0.2).
time(bahia, medio, 7, 39, 33, 56.2, 4.6, 5.3, 13, 40.00, 71.11, 30.95, 0.2).
time(botafogo, grande, 1, 47, 25, 52.0, 5.4, 5.0, 18, 86.66, 76.19, 62.22, 0.2).
time(corinthians, grande, 18, 29, 38, 49.9, 4.8, 6.0, 6, 46.66, 54.76, 13.33, 0.2).
time(criciuma, pequeno, 12, 36, 42, 44.4, 4.7, 5.9, 9, 46.66, 48.89, 30.95, 0.2).
time(cruzeiro, grande, 8, 35, 29, 52.8, 4.6, 5.4, 12, 33.33, 71.43, 28.89, 0.2).
time(cuiaba, pequeno, 19, 25, 39, 43.1, 3.4, 4.6, 6, 33.33, 31.11, 30.77, 0.2).
time(flamengo, grande, 4, 45, 32, 54.7, 5.5, 5.1, 15, 46.66, 69.23, 50.00, 0.2).
time(fluminense, grande, 16, 22, 30, 54.5, 4.1, 4.8, 8, 40.00, 47.62, 23.81, 0.2).
time(fortaleza, medio, 3, 38, 29, 42.6, 3.9, 5.0, 10, 46.66, 86.67, 38.10, 0.2).
time(gremio, grande, 11, 33, 37, 46.8, 4.4, 4.4, 10, 46.66, 49.89, 30.95, 0.2).
time(internacional, grande, 6, 37, 26, 52.0, 5.1, 5.5, 12, 73.33, 64.10, 46.67, 0.2).
time(juventude, pequeno, 14, 33, 39, 48.2, 4.2, 5.0, 8, 40.00, 61.90, 17.78, 0.2).
time(palmeiras, grande, 2, 46, 20, 53.8, 5.8, 7.0, 17, 86.66, 75.56, 54.76, 0.2).
time(red_bull_bragantino, medio, 13, 34, 38, 47.1, 4.6, 5.7, 8, 26.66, 59.62, 20.00, 0.2).
time(sao_paulo, grande, 5, 38, 32, 57.0, 4.6, 4.8, 14, 40.00, 71.11, 35.71, 0.2).
time(vasco, grande, 10, 32, 38, 47.8, 3.8, 5.1, 10, 40.00, 59.52, 28.57, 0.2).
time(vitoria, pequeno, 17, 32, 44, 42.4, 4.0, 4.8, 8, 46.66, 40.48, 26.67, 0.2).
```

Figura 1. Exemplo de importação de dados no Prolog

Etapla 2. Modelagem dos parâmetros

Apesar de não utilizar técnicas de aprendizado de máquina, o sistema aplica inferência estatística por meio de pesos numéricos ajustados empiricamente. Esses pesos, incluindo valores negativos para penalizações, representam a relevância de critérios estatísticos na previsão de resultados e foram definidos com base em testes para ajustar e balancear os resultados, sendo armazenados no arquivo “regras.pl” e organizados conforme a seguinte distribuição (ver exemplo na Figura 2):

- Desfalques: variando entre 1 (completo) até 0,7 (muito desfalcado);
- Fator casa: 20% de bônus por estar jogando em casa;
- Pesos de estatísticas dos times: foram atribuídos pesos positivos e negativos (variando de -0.3 a 0.2) sendo elas: média de gols marcados; média de gols sofridos; posse de bola média; chutes ao gol; chutes para fora; classificação no campeonato; vitórias no campeonato; desfalques (dividido entre: muito desfalcado, desfalcado, pouco desfalcado e completo); aproveitamento nas últimas 5 partidas; fator casa ou fora e status de time (divididos em grandes, médios e pequenos).

```
% Definir os pesos ajustados para cada critério
pesos_ajustados(Pesos) :-
    Pesos = [
        gols_pro-0.2,
        gols_contra-(-0.2),
        posse_bola-0.05,
        chutes_gol-0.06,
        chutes_fora-0.03,
        vitorias-0.2,
        classificacao-(-0.25),
        time_grande-0.2,
        time_medio-0.14,
        time_pequeno-0.1,
        aprovrec-0.25
    ].
```

Figura 2. Print do programa com exemplo dos pesos adotados para cada estatística durante o ajuste.

Para o desenvolvimento do modelo probabilístico o sistema utiliza um cálculo dinâmico, ajustando-se a cada atualização do banco de dados, que remodela a força de cada time baseado em suas estatísticas e desfalques, além de ajustar para confrontos diretos entre times de status semelhantes. Ao prosseguir no código, as probabilidades são recalculadas para garantir que a soma das 3 possibilidades (vitória do mandante, empate e vitória do visitante) seja 100%, oferecendo uma previsão equilibrada e realista dos resultados.

Etapa 3. Interação com o usuário

Implementou-se um menu interativo em Prolog para viabilizar a entrada de dados pelo usuário, incluindo a identificação das equipes e a classificação do nível de desfalques. Na Figura 3, segue o fluxograma do processamento, para melhor compreensão do sistema. O Prolog utiliza resolução por unificação para satisfazer objetivos lógicos. Assim, ao executar a consulta “consultar_jogo”, o sistema ativa uma cadeia de inferências.

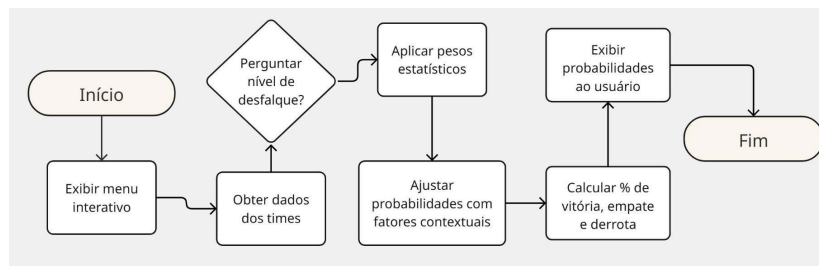


Figura 3. Fluxograma do Sistema Score Predictor

Todo esse encadeamento é feito automaticamente pelo motor lógico do Prolog, garantindo consistência e eficiência. O processo inicia-se com a apresentação de um menu interativo, no qual o usuário insere as equipes de interesse. Posteriormente, o sistema acessa o banco de dados (dados.pl) e solicita ao usuário a classificação do nível de desfalque das equipes, categorizando-as em: completo, pouco desfalcado, desfalcado ou muito desfalcado, com pesos atribuídos de 1.0, 0.95, 0.90 e 0.80, respectivamente. Em seguida, os pesos são aplicados a cada critério estatístico, ajustando as probabilidades com base em fatores contextuais, como vantagem de mando de campo (jogar em casa ou fora) e a classificação da força do time (grande, médio ou pequeno). Com esses ajustes, o sistema realiza o cálculo das probabilidades de vitória, empate e derrota, apresentando os resultados ao usuário de forma estruturada.

Resultados e discussões

A Figura 4 ilustra o sistema em operação, processando os dados, aplicando os critérios estatísticos e retornando as probabilidades de resultados da partida (vitória, empate ou derrota) de forma clara ao usuário. Inicialmente, observam-se resultados com erros, devido à calibração em andamento dos pesos, ajustados manualmente com base em comparações com casas de apostas. Já a Figura 5 apresenta resultados mais consistentes, refletindo uma calibração mais adequada dos parâmetros estatísticos.

```
=====
MENU DA PARTIDA
=====
Digite o time da casa: bahia.
Digite o time de fora: |: flamengo.

=====
O time da casa esta completo, pouco_desfalcado, desfalcado ou muito_desfalcado? |: desfalcado.
O time de fora esta completo, pouco_desfalcado, desfalcado ou muito_desfalcado? |: pouco_desfalcado.
=====
Probabilidade de bahia ganhar: 38.30%
Probabilidade de flamengo ganhar: 29.81%
Probabilidade de empate: 31.89%
=====
```

Figura 4. Exemplo da saída com erro – Sistema ainda estava calibrando os pesos – Simulação da partida Bahia x Flamengo na rodada 29

```
=====
MENU DA PARTIDA
=====
SITUACAO DOS TIMES
completo: [0-1 desfalques]
pouco_desfalcado: [2-3 desfalques]
desfalcado: [4 DESFALQUES]
muito_desfalcado: [5+ desfalques]
=====
Digite o time da casa: sao_paulo.
Digite o time de fora: |: corinthians.

=====
O time da casa esta completo, pouco_desfalcado, desfalcado ou muito_desfalcado? |: pouco_desfalcado.
O time de fora esta completo, pouco_desfalcado, desfalcado ou muito_desfalcado? |: completo.
=====
Probabilidade de sao_paulo ganhar: 57.77%
Probabilidade de corinthians ganhar: 22.57%
Probabilidade de empate: 19.67%
=====
```

Figura 5. Exemplo da saída com acerto – Simulação da partida São Paulo x Corinthians na rodada 28

Em uma análise de 30 jogos (rodadas 27, 28 e 29), o sistema obteve uma média de 53,3% de acertos nas probabilidades, 16,7% de erros dentro da margem (de 5 a 10%) e 30% de erros em relação ao resultado real das partidas. Quando comparado com as previsões das casas de apostas, a taxa de acertos foi de 60%, 33,3% dentro da margem (de 5 a 10%) e apenas 6,7% de erros. Considerando que o futebol é um esporte imprevisível e que o mercado de probabilidades é mais complexo em comparação com outros esportes (como escanteios, gols, cartões, etc.), e levando em conta a ausência de fatores externos, esses resultados podem ser considerados ótimos.

Tabela 1. Tabela com as taxas de acertos dos resultados e da comparação com as casas de apostas.

Rodada	Taxa de acertos (%)	Dentro da margem de erro (%)	Taxa de erros (%)	Casas de apostas (%)	Dentro da margem das casas de apostas (%)	Taxas de erros (%)
27	60%	30%	10%	50%	50%	0%
28	60%	0%	40%	80%	10%	10%
29	30%	30%	40%	50%	40%	10%

Além disso, a escolha pela programação lógica traz diversos benefícios ao sistema como transparência nas regras, explicabilidade dos resultados, facilidade na adição de novos critérios, ausência de comportamentos não explicáveis e baixo custo computacional, tornando o sistema simples, rápido e legível. Por fim, a abordagem em Prolog serve como base para uma evolução com aprendizado automático, mantendo a inteligência simbólica como referência interpretativa.

Análise complementar e trabalhos futuros

Embora os resultados obtidos pelo sistema Score Predictor tenham demonstrado acertos satisfatórios, observa-se algumas limitações, como a amostra reduzida de 30 partidas e a calibração manual dos pesos, o que compromete a consistência estatística, a reprodutibilidade e a generalização do sistema. Amostras maiores e mais diversas poderão fornecer uma base mais confiável para a validação estatística do modelo.

Embora o foco tenha sido demonstrar a viabilidade da abordagem simbólica com Prolog, sugere-se, como trabalho futuro, a integração de algoritmos de aprendizado supervisionado, como regressão logística e árvores de decisão, ao pipeline de análise, para comparação de desempenho com base em métricas como acurácia, precisão e recall. Por fim, ressalta-se que a proposta do Score Predictor se destaca pela interpretabilidade, servindo como base para sistemas híbridos que combinam lógica simbólica e aprendizado automático.

Conclusão

Este trabalho combina estatísticas esportivas e programação lógica para gerar previsões precisas de partidas de futebol. A partir da análise de fatores como desempenho recente, desfalques e histórico, o Score Predictor se estabelece como uma ferramenta eficaz na previsão de resultados do Campeonato Brasileiro de 2024. Apesar da complexidade do contexto, o sistema obteve uma taxa de acerto satisfatória, com resultados alinhados às previsões das casas de apostas. Para o futuro, recomenda-se um aprofundamento na análise estatística para ajustar os pesos dos critérios e aprimorar a precisão do modelo.

REFERÊNCIAS

- BABOOTA, Rahul; KAUR, Harleen. Predictive analysis and modelling football results using machine learning approach for English Premier League. Journal of King Saud University - Computer and Information Sciences, [S.l.], v. 32, n. 7, p. 1–10, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169207018300116>.
- BUNKER, R.; YEUNG, C.; FUJII, K. Machine learning for soccer match result prediction. ArXiv, abs/2403.07669, 2024. Disponível em: <https://doi.org/10.48550/arXiv.2403.07669>.
- BET365 (2024) “Apostas Desportivas Online”. Disponível em: <https://www.bet365.com/#/HO/>.

COSTA, Ígor Barbosa. Modelagem e predição de resultados de futebol antes e durante as partidas usando aprendizagem de máquina. 2021, 115 f. Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Campina Grande, 2021. Disponível em: <
https://bdtd.ufcg.edu.br/jspui/handle/riufcg/20618?utm_source=chatgpt.com>

Departamento de Matemática UFMG (2024) “Série A – Probabilidades no Futebol”. Disponível em: <https://www.mat.ufmg.br/futebol/serie-a/>.

ELMILIGI, H.; SAAD, S. Predicting the outcome of soccer matches using machine learning and statistical analysis. 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), p. 1–8, 2022. Disponível em: <https://doi.org/10.1109/CCWC54503.2022.9720896>.

FELIX, Ingrid De Araújo. Modelagem estatística para previsão de resultados de jogos: uma aplicação ao campeonato da Coreia do Sul 2019. Anais IV CONAPESC. Campina Grande: Realize Editora, 2019. Disponível em: <https://editorarealize.com.br/index.php/artigo/visualizar/56997>

FIALHO, G.; MANHÃES, A.; TEIXEIRA, J. Predicting sports results with artificial intelligence: a proposal framework for soccer games. Procedia Computer Science, p. 131–136, 2019. Disponível em: <https://doi.org/10.1016/j.procs.2019.12.164>.

KUHNERT, Felipe Ventura. Modelo para predição dos resultados de partidas de futebol. 2023, 77 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023. Disponível em: <https://pantheon.ufrj.br/bitstream/11422/21584/1/FVKuhnert.pdf>

RICO-GONZÁLEZ, M.; PINO-ORTEGA, J.; MÉNDEZ, A.; CLEMENTE, F.; BACA, A. Machine learning application in soccer: a systematic review. Biology of Sport, v. 40, p. 249–263, 2022. Disponível em: <https://doi.org/10.5114/biolsport.2023.112970>.

SCHNEIDER, Cristian Felipe Machine learning aplicado na previsão de resultados de partidas de futebol: um estudo de caso para comparação de diferentes classificadores. 2018. Trabalho de conclusão de curso (Curso de Engenharia Elétrica) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

SWI-Prolog Documentation (2024) “Official Documentation”. Disponível em: <https://www.swi-prolog.org/pldoc/>.

SofaScore (2024) “Jogos de hoje e Campeonato Brasileiro 2024”. Disponível em: <https://www.sofascore.com/pt/>.

YEUNG, C.; BUNKER, R.; UMEMOTO, R.; FUJII, K. Evaluating soccer match prediction models: a deep learning approach and feature optimization for gradient-boosted trees. ArXiv, abs/2309.14807, 2023. Disponível em: <https://doi.org/10.48550/arXiv.2309.14807>.