TÉCNICAS DE APRENDIZADO DE MÁQUINA APLICADAS NA PREVISÃO DE RESULTADOS DE JOGOS DE FUTEBOL

LUCAS DOS SANTOS[1]
MATHEUS YOSHIO OGAWA[1]
RONNIE SHIDA MARINHO[1]

RESUMO

Este estudo investiga a aplicação de técnicas de aprendizado de máquina na previsão de resultados de jogos de futebol. Utilizando um conjunto de dados extenso e variado, foram exploradas diferentes abordagens de modelagem preditiva para identificar padrões e tendências significativas. Os resultados demonstram que modelos baseados em redes neurais e Random Forest apresentaram desempenho superior na previsão precisa de resultados de partidas. Além disso, análises adicionais revelaram a importância de variáveis específicas, como desempenho histórico das equipes, na precisão dos modelos preditivos. Este estudo contribui para aprimorar as estratégias de previsão no contexto esportivo, oferecendo insights valiosos para analistas esportivos, apostadores e equipes técnicas, aperfeiçoando suas estratégias e tomadas de decisão com base em dados.

Palavras-chave: Aprendizado de máquina; Previsão de resultados; Redes Neurais; Random forest.

ABSTRACT

This study investigates the application of machine learning techniques in predicting football match outcomes. Using an extensive and varied dataset, different predictive modeling approaches were explored to identify significant patterns and trends. The results demonstrate that models based on neural networks and Random Forest exhibited superior performance in accurately predicting match outcomes. Additionally, further analysis revealed the importance of specific variables, such as the historical performance of teams, in the accuracy of predictive models. This study contributes to enhancing predictive strategies in the sports context, offering valuable insights for sports analysts, bettors, and technical teams, thereby improving their strategies and decision-making based on data.

Keywords: Machine learning; Outcome prediction; Neural networks; Random forest.

[1] Faculdade de Tecnologia de Adamantina (Fatec), Adamantina-SP, Brasil

1 INTRODUÇÃO

A previsão de resultados esportivos tem sido um desafio contínuo e fascinante, intrigando tanto entusiastas quanto profissionais da indústria esportiva ao longo dos anos. Segundo Costa (COSTA, 2021), há um interesse mundial pelo futebol, visto que esse esporte influencia diariamente a vida de aproximadamente 3,5 bilhões de pessoas. A capacidade de antecipar com precisão o desfecho de eventos esportivos não apenas cativa a imaginação, mas também tem implicações significativas em várias áreas, desde apostas esportivas até a tomada de decisões estratégicas por parte de equipes e organizações esportivas. Com os crescentes avanços na área de aprendizado de máquina, uma nova era se desenha, com a promessa de oferecer modelos mais sofisticados e precisos para a previsão de resultados esportivos.

De acordo com a Confederação Brasileira de Futebol (CBF, 2019), no ano de 2019 o investimento em clubes chegou a R$ 52,9 bilhões, aproximadamente 1% do PIB (Produto Interno Bruto) nacional, buscando oportunidades de negócios em um mercado em constante expansão. Apesar dos desafios enfrentados, como infraestrutura limitada e gestão complexa, O crescimento das casas de apostas no Brasil tem sido notável, acompanhando uma tendência global de expansão do mercado de apostas esportivas, especialmente no futebol. Em 2018, a legalização das apostas esportivas no país abriu um novo horizonte para o setor, com projeções de que o mercado brasileiro possa movimentar até R$ 10 bilhões por ano (ABRAPESP, 2019). A aplicação de estatísticas avançadas para prever resultados de partidas tem se tornado uma prática comum entre os apostadores, utilizando dados detalhados sobre desempenho de equipes e jogadores. Ferramentas de análise como essas incluem métricas de posse de bola, finalizações, passes e movimentações em campo, que são cruciais para fazer previsões mais precisas (SOUZA, 2020) o futebol brasileiro continua atraindo investimentos consideráveis, refletindo a paixão e o potencial econômico do esporte no país.

O mundo dos esportes é caracterizado por uma série de fatores imprevisíveis, como histórico de confrontos, desempenho da equipe em casa e fora, lesões de jogadores, mudanças inesperadas nas condições climáticas, entre outros, que podem influenciar decisões e desempenhos.

Em uma era digital, a análise de dados e a inteligência artificial (IA) têm

desempenhado um papel cada vez mais importante na preparação das equipes e na tomada de decisões táticas, como destacado pelo técnico de futebol brasileiro Renato Gaúcho em uma entrevista concedida para o portal de notícias, em que relatou que: "A análise de dados é uma ferramenta essencial no futebol moderno. Ela nos permite compreender melhor o jogo, identificar padrões e tendências, e ajustar nossa estratégia de acordo com as necessidades da partida" (GLOBO, 2022).

A adoção de técnicas de aprendizado de máquina tem sido cada vez mais comum, com clubes e instituições investindo em análises avançadas de dados para melhorar o desempenho das equipes e otimizar a gestão esportiva. Os algoritmos preditivos e modelos estatísticos possibilitam aos treinadores e dirigentes identificar padrões ocultos nos dados, prever resultados. O interesse em antecipar com precisão o desfecho dos jogos não apenas cativa a imaginação dos fãs, mas também tem um impacto significativo em diversas esferas, desde o entretenimento até as estratégias de gestão e investimento dos clubes.

Compreender e prever esses resultados não só alimenta a competitividade entre as equipes, mas também pode proporcionar uma melhor compreensão pode contribuir para uma gestão financeira mais eficiente conforme discutido por Chen (CHEN et al., 2021) os de partidas e tomar decisões mais assertivas (BUNKER; THABTAH, 2019).

A pesquisa de Bunker e Susnjak (BUNKER; SUSNJAK, 2022) examinou como usar o aprendizado de máquina para prever os resultados de partidas em esportes coletivos, analisando estudos realizados de 1996 a 2019. O estudo aplicou algoritmos, análise de dados e avaliações de precisão em cenários de esportes coletivos, como futebol, basquete e rugby. Além disso, retrataram a possibilidade de ser mais difícil prever certos tipos de modalidades esportivas do que outras. Ainda, defenderam uma colaboração mais interdisciplinar entre a análise de desempenho esportivo e a aprendizagem de máquina, e enfatizaram a importância da seleção e engenharia de características adequadas.

De um modo diferente, Sehnem e Frozza (SEHNEM; FROZZA, 2019) analisaram um conjunto de variáveis de partidas anteriores, que podem influenciar na previsão do resultado de uma partida de futebol, utilizando técnicas como cálculos de probabilidade e algoritmos de previsão. Na era da informação, a capacidade de extrair valor de grandes volumes de dados é essencial para o avanço comercial, econômico, social.

Nesse cenário, este trabalho se propõe a aplicar técnicas avançadas de aprendizado de máquina na previsão de resultados de jogos de futebol do Campeonato Brasileiro, buscando contribuir para o aprimoramento contínuo da análise esportiva e da tomada de decisões estratégicas no cenário esportivo

nacional. Além disso, como artefato deste trabalho, foi gerado um protótipo web (versão inicial e simplificada de um site). O qual permite que usuários interajam com o sistema de previsão de resultados, oferecendo uma interface intuitiva para selecionar os times, a arena e outras variáveis relevantes, possibilitando visualizar previsões e estatísticas em tempo real.

2 FUNDAMENTAÇÃO TEÓRICA

Na era da informação, a capacidade de extrair valor de grandes volumes de dados é essencial para o avanço tecnológico e científico. Entre os conceitos que impulsionam essa transformação, destacam-se o aprendizado de máquina e o meta-aprendizado.

2.1 Aprendizado de máquina

O aprendizado de máquina é uma subárea da IA que se concentra no desenvolvimento de técnicas e algoritmos com a capacidade de treinar máquinas para aprender a partir de dados disponíveis. O objetivo principal é desenvolver modelos e sistemas que possam tomar decisões automatizadas e prever resultados com precisão. Isso possibilita a aplicação em áreas variadas, como saúde, mercado financeiro e esportes, ao permitir a identificação de padrões em grandes volumes de dados, tornando-os úteis para a tomada de decisões (REZENDE, 2017).

Existem três categorias no campo do aprendizado de máquina: supervisionado, não supervisionado e por reforço. O aprendizado supervisionado utiliza dados rotulados para treinar modelos a fim de prever resultados específicos, sendo amplamente aplicado em tarefas de classificação, onde o objetivo é atribuir categorias a novos dados, e em tarefas de regressão, que visam prever valores numéricos baseados nos dados disponíveis (SPATTI; ITO; FLAUZINO, 2019).

Por sua vez, o aprendizado não supervisionado difere do supervisionado ao não requerer dados rotulados para treinar modelos. Neste contexto, os algoritmos exploram padrões e estruturas nos dados sem a necessidade de uma resposta pré-determinada. São frequentemente utilizados em tarefas como clusterização, onde o objetivo é agrupar dados similares em conjuntos distintos, e em técnicas de redução de dimensionalidade, que visam simplificar dados complexos preservando suas características essenciais (BARBOSA, 2019).

O aprendizado por reforço, por sua vez, é baseado em um sistema de tentativa e erro, sendo frequentemente utilizado no desenvolvimento de robôs e em jogos, como para encontrar o melhor caminho em um labirinto. Esse método permite que agentes aprendam a tomar decisões sequenciais ao receber recompensas ou penalidades por suas ações, aprimorando suas estratégias ao longo do tempo (DE SOUZA; OLIVEIRA, 2018).

Dentre as categorias de aprendizado de máquina, o aprendizado supervisionado, especialmente a tarefa de classificação, é particularmente relevante para este estudo devido à sua capacidade comprovada de prever resultados com base em dados históricos, oferecendo uma melhor compreensão para análises e estratégias futuras. Destacam-se diversos algoritmos, como Random Forest, Redes Neurais, Regressão Logística e KNN (K-nearest neighbors) (BARBOSA, 2019; DE SOUZA; OLIVEIRA, 2018).

As *Random Forests* são conjuntos de árvores de decisão que combinam múltiplos modelos para melhorar a precisão e a robustez das previsões. Este método é conhecido por sua eficácia em lidar com grandes conjuntos de dados e alta dimensionalidade, sendo aplicado em problemas que exigem tanto classificação quanto regressão (BREIMAN, 2001). Árvores de decisão são modelos de previsão que utilizam uma estrutura hierárquica para tomar decisões baseadas em características dos dados. Cada nó interno representa uma decisão sobre um atributo, e cada ramo representa o resultado dessa decisão, levando eventualmente a um nó folha, que fornece a classificação ou valor de regressão (MITCHELL, 1997).

O *KNN (K-Nearest Neighbors)* é um algoritmo de aprendizado supervisionado que classifica uma nova instância com base nas classes dos seus k vizinhos mais próximos no espaço dos atributos. Este método é simples e eficaz, especialmente em conjuntos de dados menores, onde a relação entre os dados pode ser facilmente determinada pela proximidade (COVER; HART, 1967).

McCulloch e Pitts (MCCULLOCH; PITTS, 1943) apresentaram um modelo matemático inspirado no sistema nervoso biológico, especialmente no cérebro humano, com camadas de neurônios artificiais que processam informações em cascata. Esses modelos são fundamentais em problemas complexos de análise de dados, devido à sua capacidade de aprender padrões hierárquicos nos dados (RUMELHERT et al., 1986).

A regressão logística é um método eficaz para modelar as relações entre variáveis independentes e dependentes (HOSMER et al., 2013).

2.5 Meta aprendizado

O meta-aprendizado busca melhorar a eficácia dos sistemas de aprendizado de máquina, especialmente na seleção de algoritmos baseados no compromisso entre tempo de treinamento e acurácia desejada. Conforme destacado por Vanschoren (VANSCHOREN,2018), "o metaaprendizado visa automatizar a escolha e a configuração de algoritmos, adaptando-os dinamicamente às características específicas dos dados e aos objetivos do usuário". Esta abordagem permite não apenas a identificação do algoritmo mais adequado para um determinado problema, mas também a consideração de medidas múltiplas de desempenho, indo além da acurácia podendo considerar outros objetivos como eficiência computacional interpretabilidade, entre outros. Técnicas como a indicação de algoritmos mais promissores com base em limiares de desempenho, conforme proposto por Kohavi (KOHAVI, 1995), são essenciais para fornecer recomendações úteis.

Uma técnica dentro do meta-aprendizado é a otimização de hiperparâmetros, que são configurações ajustáveis que impactam diretamente o comportamento e a eficácia dos algoritmos de aprendizado de máquina. Assim, uma das abordagens mais comuns para encontrar os melhores hiperparâmetros é a *Grid Search* (busca em grade), onde uma grade de valores pré-definidos para cada hiperparâmetro é especificada. O algoritmo então avalia o desempenho do modelo para cada combinação possível de hiperparâmetros, utilizando validação cruzada para escolher a configuração que produz os melhores resultados. Embora seja abrangente e sistemática, a *Grid Search* pode ser computacionalmente custosa, especialmente com um grande número de hiperparâmetros e valores a serem considerados (BERGSTRA; BENGIO, 2012).

Uma alternativa eficaz à *Grid Search* é a *Random Search* (busca aleatória), que, ao contrário da busca em grade, seleciona aleatoriamente combinações de hiperparâmetros para avaliação. Esta abordagem é mais eficiente em termos computacionais, pois não explora todas as combinações possíveis, focando-se em uma amostragem aleatória que pode ser tão eficaz quanto ou até melhor do que a busca em grade em muitos casos. O *Random Search* é particularmente vantajoso quando a influência de determinados hiperparâmetros é desconhecida ou quando há um grande espaço de busca (BERGSTRA; BENGIO, 2012).

Ambas as técnicas, são essenciais no contexto do meta-aprendizado, permitindo aos desenvolvedores e cientistas de dados ajustar os modelos de aprendizado de máquina de forma eficiente e eficaz, permitindo-se identificar a combinação de hiperparâmetros maximizando o desempenho e a generalização dos sistemas de IA.

As figuras 1 e 2 mostram as fórmulas dos métodos de avaliação acurácia e matriz de confusão. A acurácia é uma métrica simples e amplamente utilizada que mede a proporção de previsões corretas realizadas pelo modelo em relação ao total de previsões. A matriz de confusão complementa a acurácia ao fornecer uma visão detalhada do desempenho do modelo em diferentes classes de previsão.

- VP (Verdadeiros Positivos): corretamente classificadas como positivas pelo modelo.
- VN (Verdadeiros Negativos): corretamente classificadas como negativas pelo modelo.
- FP (Falsos Positivos): erroneamente classificadas como positivas pelo modelo.
- FN (Falsos Negativos): Observações erroneamente classificadas como negativas pelo modelo.

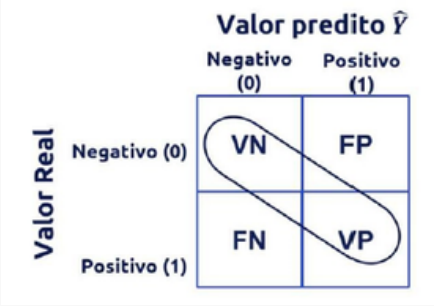
Estas métricas são essenciais para não apenas validar a eficácia do modelo, mas também para identificar áreas que requerem ajustes para melhorar a precisão das previsões (GÉRON, 2019).

Figura 1 – Fórmula de cálculo da acurácia

$$\text{Acurácia} = \frac{VP+VN}{VP+VN+FP+FN}$$

Fonte: Retirada de: https://www.flai.com.br/juscudilio/qual-a-melhor-metrica-para-avaliar-osmodelos-de-machine-learning/

Figura 2 - Fórmula da Matriz de confusão



Fonte: Retirada de https://www.flai.com.br/juscudilio/qual-a-melhor-metrica- para-avaliar-osmodelos-de-machine-learning/

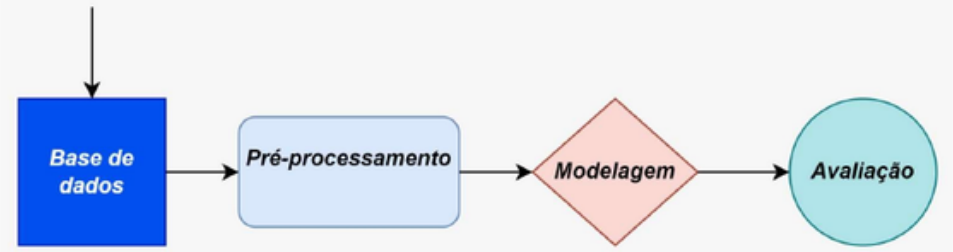
3 METODOLOGIA

Para explicar os procedimentos metodológicos adotados neste estudo, o processo foi organizado em módulos distintos, cada um abordando uma etapa fundamental da análise. Esses módulos são intitulados: base de dados, pré-processamento, modelagem e avaliação figura 3.

No módulo de base de dados, os dados relevantes para o estudo são coletados, organizados e preparados para análise. No estágio de pré-processamento, os dados preparados são submetidos a vários métodos para torná-los adequados para a aplicação de algoritmos de aprendizado de máquina. No módulo seguinte, de modelagem, são aplicados algoritmos de aprendizado de máquina aos dados pré-processados e por fim, no módulo de avaliação, os modelos treinados são avaliados quanto à sua performance e eficácia.

Esta estrutura modular permite uma apresentação clara e detalhada de cada fase, desde a coleta e preparação dos dados até a aplicação dos modelos de aprendizado de máquina e a interpretação dos resultados obtidos.

Figura 3 – Principais etapas do trabalho



Fonte: Autoria Própria (2024)

3.1 Base de dados

Os dados utilizados neste trabalho foram coletados da plataforma Kaggle[1], fornecendo informações abrangentes sobre jogos de futebol disputados no Brasil entre os anos de 2003 e 2023.

[1] https://www.kaggle.com/datasets/adaoduque/campeonato-brasileiro-de-futebol

Para garantir a consistência da predição, utilizou-se dois conjuntos de dados distintos, considerando a identificação única das partidas presentes em ambos. A combinação das bases de dados justifica-se pelo fato de uma apresentar os resultados dos jogos e a outra análises estatísticas dos times.

A primeira base intitulada “Campeonato Brasileiro Estatísticas” apresenta cerca de 13 atributos, como: identificador da partida, rodada, nome do clube, chutes, chutes no alvo, posse de bola, passes, precisão dos passes, faltas, cartão vermelho, cartão amarelo, impedimentos e escanteios. Já a segunda base intitulada “Campeonato Brasileiro Full” possui 16 atributos, sendo: identificador da partida, rodada, data, hora, mandante, visitante, formação mandante, formação visitante, técnico mandante, técnico visitante, vencedor, arena, mandante placar, visitante placar, mandante estado e visitante estado, ainda, para uma melhor análise, foram excluídas os seguintes atributos desta base: formação mandante, formação visitante, técnico mandante, técnico visitante, placar mandante e placar visitante, devido à alta incidência de dados.

3.2 Pré-processamento

Nessa etapa, foram excluídos registros antigos, concentrando-se nos últimos cinco anos, visto que muitos jogadores não fazem mais parte dos elencos atuais. Isso permitiu que se trabalhasse com uma base de dados mais atualizada e relevante para a análise. Além disso, estudos recentes demonstram a importância de utilizar dados atuais para a predição de jogos (GODDARD; ASIMAKOPOULOS, 2004; CLARKE; DYTE, 2017).

Após as exclusões, o conjunto de dados final unificado abrange cerca de 1.900 registros e 28 características, sendo essas características fundamentais para a análise e previsão de resultados de jogos de futebol.

Realizou-se a normalização dos dados, esta etapa foi necessária para evitar que os algoritmos de aprendizado de máquina atribuíssem importância desigual às diferentes categorias.

3.2 Modelagem

No módulo de modelagem, além de selecionar os algoritmos de aprendizado de máquina, exploraram-se técnicas de meta-aprendizado para aprimorar o

desempenho dos modelos. Para isso, aplicaram-se algoritmos de aprendizado de máquina como Redes Neurais, Regressão Logística, *KNN* e *Random Forest.*

Nesta etapa, os dados foram divididos em conjuntos de treinamento e teste na proporção de 85% para treinamento e 15% para teste dos modelos. A escolha dessa divisão da base de dados é justificada pela complexidade de aferir resultados no contexto de jogos de futebol, necessitando de uma maior quantidade de dados para o treinamento dos modelos.

4 RESULTADOS

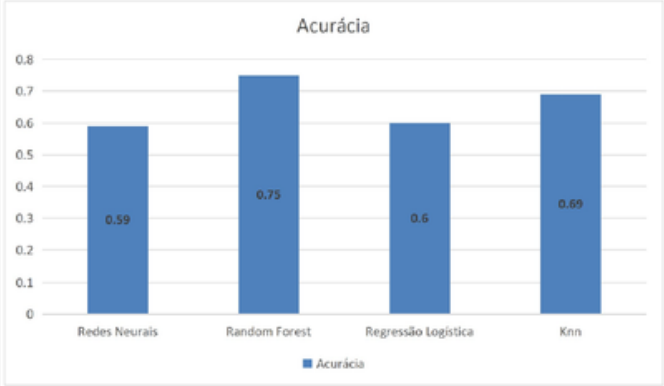
A implementação dos modelos de aprendizado de máquina, realizou-se uma análise comparativa para avaliar o desempenho de diferentes algoritmos na previsão dos resultados dos jogos de futebol. Utilizou-se as métricas de acurácia, que avalia a porcentagem de acertos do modelo.

Por meio de do meta-aprendizado, constatou-se que a melhor configuração neste trabalho foi alcançada com o uso do algoritmo de meta- aprendizado *Random Search,* verificou-se que a combinação de hiperparâmetros que melhorou o desempenho do modelo.

Após a aplicação do meta-aprendizado, observou-se que o algoritmo de aprendizado de máquina *Random Forest* se destacou como a melhor técnica para prever os resultados dos jogos, obtendo uma acurácia de 75%, seguido pelo algoritmo KNN, que alcançou 69%. Este resultado está alinhado a pesquisas anteriores que destacam a eficácia do *Random Forest* em problemas de classificação no contexto de jogos de futebol (BIAU; SCORNET, 2016).

Figura 4 apresentam os valores de acurácia alcançados para os algoritmos investigados neste estudo.

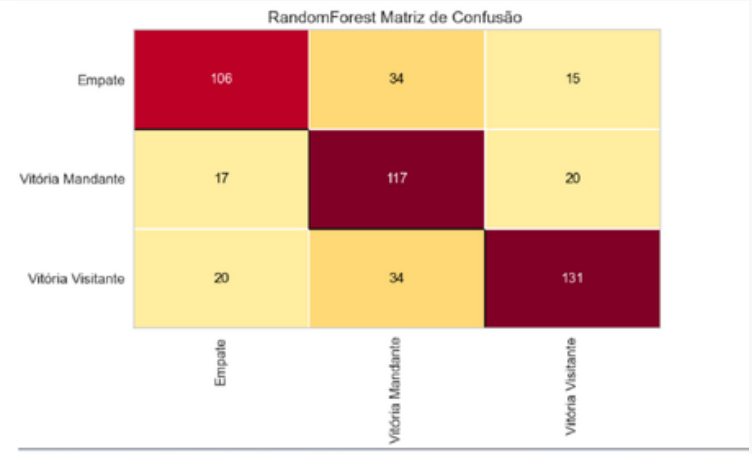
Figura 4 - Comparação de Modelos de Aprendizado de Máquina



Fonte: Autoria própria (2024)

Para obter uma visão mais completa e detalhada do desempenho do algoritmo é apresentada na figura 5 a matriz de confusão referente a classificação oriunda do treinamento utilizando *Random Forest,* onde as classes correspondem a empate (0), vitória mandante (1) e vitória visitante (2).

Figura 5 – Matriz de confusão da classificação utilizando Random Forest



Fonte: Autoria própria (2024)

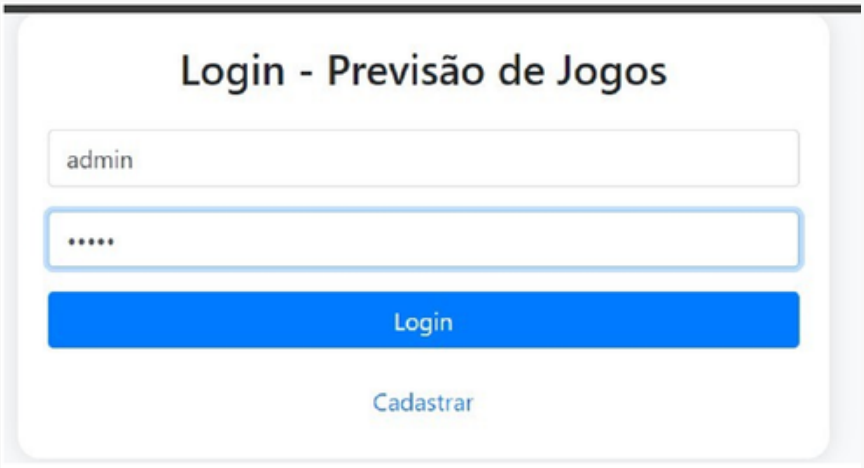
4.1 Protótipo

Para o desenvolvimento do protótipo web (versão inicial e simplificada de um site), foi criada uma interface intuitiva que proporciona uma experiência agradável e eficiente aos usuários. Durante o processo, utilizamos várias bibliotecas essenciais. O modelo de aprendizado de máquina foi salvo no formato *pickle,* permitindo fácil armazenamento e carregamento dos modelos treinados. A biblioteca *scikit-learn*, reconhecida por sua ampla gama de algoritmos de aprendizado de máquina, foi fundamental para treinar e avaliar os modelos. Além disso, a biblioteca flask foi empregada para a criação das páginas web, aproveitando seu poder e flexibilidade para desenvolver aplicações web em *Python.* Essas ferramentas, combinadas, garantem que a plataforma seja funcional, acessível e pronta para auxiliar na previsão de resultados de jogos de futebol.

A partir das avaliações dos algoritmos de aprendizado de máquina, considerando tanto a acurácia quanto a matriz de confusão, desenvolveu-se um protótipo web que auxilia na previsão de resultados de jogos de futebol. Este protótipo utiliza os dados de desempenho do modelo treinado com *Random*

Forest para oferecer previsões precisas e confiáveis, permitindo aos usuários obter insights sobre possíveis resultados das partidas.
Inicialmente, o usuário necessita fazer login na página utilizando credenciais, como nome de usuário e senha. Esse processo garante a segurança e a personalização da experiência, permitindo que cada usuário acesse suas configurações e dados pessoais. Esse processo é ilustrado na figura 6.

Figura 6 - Página “Login”



Fonte: Autoria própria (2024)

Após a autenticação, o usuário poderá navegar pelas diversas páginas da plataforma, como por exemplo a página "Sobre", que relata informações sobre o sistema (Figura 7). Desta forma, o usuário pode utilizar todas as funcionalidades oferecidas pela plataforma, tendo a certeza de que suas informações estão protegidas e acessíveis apenas por ele.

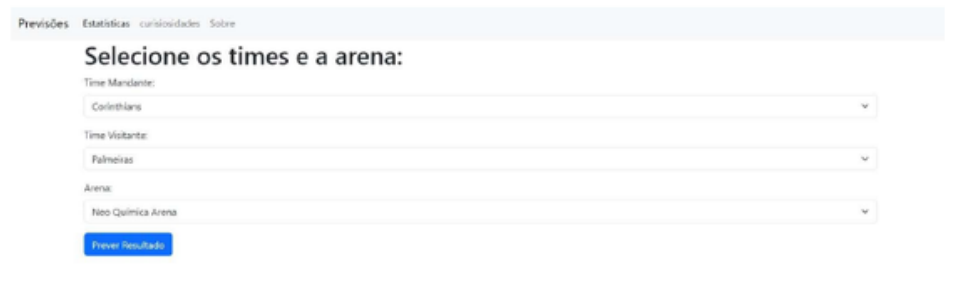
Figura 7 – Página “Sobre”

Fonte: Autoria própria (2024)

A página inclui informações sobre o objetivo do site, a metodologia adotada e as principais funcionalidades disponíveis para os usuários. Ela também destaca a importância da análise de dados e do uso de IA no contexto esportivo.

Na página principal "Previsões" (Figura 8), é possível selecionar os times que jogarão e a arena onde acontecerá a partida, permitindo que o sistema processe as informações fornecidas pelo usuário e apresente uma previsão do resultado do jogo (Figura 9).

Figura 8 – Página “Previsões”



Fonte: Autoria própria (2024)

Figura 9 – “Time vencedor”

Fonte: Autoria própria (2024)

5 CONSIDERAÇÕES FINAIS

Neste estudo, utilizamos técnicas de *machine learning* para prever os resultados das partidas de futebol, alcançando uma precisão de 75%. Essa abordagem, baseada em *Random Forest,* demonstrou um desempenho promissor em relação à previsão de resultados.

Ao comparar nossos resultados com a literatura existente, observamos algumas diferenças notáveis. Pathak e Wadhwa (PATHAK; WADHWA, 2016) relataram uma precisão de 69,5% utilizando regressão logística aplicada à *Premier League* inglesa.

O estudo investiga o objetivo de identificar tendências e fatores determinantes, em reconhecer padrões complexos e oferece perspectivas

promissoras para aprimorar a precisão das previsões no domínio esportivo.

Nosso modelo consiste em uma ferramenta complementar, os resultados obtidos a partir deste trabalho não apenas contribuem para tomada de decisão, implementação do modelo preditivo em plataformas de apostas esportivas, por exemplo, pode melhorar a precisão das probabilidades oferecidas, beneficiando tanto as casas de apostas quanto os apostadores, equipes e técnicos de futebol que podem utilizar esses modelos para planejar estratégias de jogo.

Como trabalho futuro, planeja-se expandir o escopo deste estudo, explorando outras bases de dados que abrangem um maior conjunto de variáveis no contexto esportivo. Essa diversificação de fontes de dados permitirá uma análise mais abrangente e robusta, proporcionando uma visão mais completa dos principais fatores que influenciam os resultados esportivos, especialmente em jogos de futebol.

REFERÊNCIAS

ABRAPESP. **Mercado de apostas esportivas no Brasil deve movimentar R$ 10 bilhões por ano.** 2019. Recuperado de https://www.abrapesp.com.br

Barbosa, M. R. **Aprendizado de máquina na prática.** São Paulo: Novatec.
Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281-305, 2019.

Biau, G., & Scornet, E. **A random forest guided tour.** Statistical Surveys, 2016.

Breiman, L. **Random Forests. Machine Learning,** 45(1), 5-32, 2001.

Bunker, R., & Sušnjak, T. **A machine learning framework for sport result prediction.** Applied Computing and Informatics, 15(1), 27-33, 2019.

Chen, W. J., Jhou, M. J., Lee, T. S., & Lu, C. J. **Hybrid basketball game outcome prediction model by integrating data mining methods for the National Basketball Association.** Entropy, 23, 477, 2021. Recuperado de https://www.mdpi.com/1099-4300/23/4/477.

Clarke, S. R., & Dyte, D. **Modelling football match results and the efficiency of fixedodds betting markets.** International Journal of Forecasting, 33(2), 458-465, 2017.

Confederação Brasileira de Futebol. **CBF apresenta relatório sobre papel do futebol na economia do Brasil.** Recuperado de https://www.cbf.com.br.

Costa, Í. B. da. **Modelagem e predição de resultados de futebol antes e durante as partidas usando aprendizagem de máquina.** Dissertação de mestrado, Universidade Federal de Campina Grande, 2021.

Cover, T., & Hart, P. **Nearest Neighbor Pattern Classification.** IEEE Transactions on Information Theory, 13(1), 21-27, 1967.

Souza, R., & Oliveira, F. Aplicações do aprendizado por reforço: da robótica aos jogos. **Revista de Inteligência Artificial,** 15(3), 67-78, 2018. Recuperado de https://www.revistaia.org.br/aplicacoes-do-aprendizado-por-reforco.

Domingos, P. **A few useful things to know about machine learning.** Communications of the ACM, 55(10), 78-87, 2012.

Globo. **Nova regra do futebol brasileiro.** Portal Globo, 2024. Recuperado de https://g1.globo.com/jornal-nacional/noticia/2022/10/25/inteligencia-artificial-auxiliatecnicos-a-monitorar-condicao-fisica-de-jogadores-que-vao-disputar-a-copa.ghtml.

Goddard, J., & Asimakopoulos, I. Forecasting football results and the efficiency of fixed-odds betting. **Journal of Forecasting,** 23(1), 51-66, 2004. DOI: 10.1002/for.906.

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). **Automated machine learning:** methods, systems, challenges. Springer, 2019. DOI: 10.1007/978-3-030-05318-5.

Marengo, J. A., Nobre, C. A., & Sampaio, G. Clima do Brasil. **Revista Brasileira de Meteorologia,** 24(1), 1-67, 2009.

MCCULLOCH, WS, PITTS, W. **Um cálculo lógico das ideias imanentes na atividade nervosa.** Boletim de Biofísica Matemática 5, 115–133, 1943. https://doi.org/10.1007/BF02478259

PATHAK, A., WADHWA, V. **Logistic regression analysis of Premier League Football Matches.** International Journal of Sports Science, v. 12, n. 1, p. 69-82, 2016.

Rezende, S. O. **Sistemas inteligentes:** fundamentos e aplicações. Barueri: Manole, 2020.

Sehnem, R., & Frozza, R. **Análise de variáveis em partidas de futebol para previsão de resultados.** Anais do Salão de Ensino e de Extensão, 217, 2019.

Souza, P. A importância da análise estatística nas apostas esportivas. **Revista Brasileira de Estatística,** 1(2), 45-56, 2020.

Spatti, D. H., Ito, A. S., & Flauzino, R. A. **Introdução ao aprendizado de máquina.** São Paulo: Edusp, 2019.

Vanschoren, J. **Meta-Learning:** a survey, 2018. arXiv preprint arXiv:1810.03548.