

APLICAÇÃO DE ARIMA E REDES NEURAIIS ARTIFICIAIS NA ANÁLISE ESTATÍSTICA E PREVISÃO DE RESULTADOS DE PARTIDAS DE FUTEBOL DA PREMIER LEAGUE INGLESA 2021-2022

APPLICATION OF ARIMA AND ARTIFICIAL NEURAL NETWORKS IN THE STATISTICAL ANALYSIS AND PREDICTION OF RESULTS OF FOOTBALL MATCHES IN THE ENGLISH PREMIER LEAGUE 2021-2022

Vinícius Doro*^{ID} E-mail: vdoro@ucs.br

Leandro Luís Corso*^{ID} E-mail: llcorso@ucs.br

*Universidade de Caxias do Sul (UCS), Caxias do Sul, RS, Brasil.

Resumo: A utilização de dados estatísticos e a criação de ferramentas de previsão e análise de resultados se apresentam constantemente como parte da realidade encontrada no meio esportivo. No futebol, tais ferramentas vêm sendo tratadas como um modo de adquirir vantagens competitivas, tanto para clubes, quanto para apostadores. Com intuito de obter um eficiente modelo de previsão para partidas de futebol, o estudo buscou realizar o prognóstico de resultados da *Premier League* Inglesa disputada na temporada 2021-2022. Foram coletados dados correspondentes a 70% dos jogos realizados dentro do campeonato em questão e, a partir deles, foi proposta a aplicação e a comparação do modelo Auto Regressivo Integrado de Média Móvel (ARIMA), com o uso de Redes Neurais Artificiais (RNA). Em sua predição, o método ARIMA forneceu um conjunto completo das estatísticas dos confrontos, incluindo o número de gols das equipes em cada partida, alcançando 54,39% de acerto dos vencedores e prevendo 14,04% dos placares com exatidão. O modelo composto por RNA, por sua vez, previu por meio de um código de programação na linguagem *Python* somente os vencedores dos confrontos. Desse modo, obteve-se uma maior acurácia, de 72,81%, visualizada por meio da geração de uma matriz de confusão.

Palavras-chave: Previsão de resultados de futebol. ARIMA. Redes Neurais Artificiais. *Premier League* Inglesa.

Abstract: The use of statistical data and the creation of tools for predicting and analyzing results are constantly part of the reality found in sports field. In football, such tools have been treated as a way to acquire competitive advantages, both for clubs and for bettors. In order to obtain an efficient prediction model for football matches, the study sought to predict the results of the English Premier League played in the 2021-2022 season. Data corresponding to 70% of the games played in the championship in question were collected and, from them, the application and comparison of the Auto Regressive Integrated Moving Average model (ARIMA) with the use of Artificial Neural Networks (ANNs) was proposed. In its prediction, the ARIMA method provided a complete set of clash statistics, including the number of goals scored by the teams in each match, reaching 54.39% of the winners' accuracy and predicting 14.04% of the scores accurately. The model composed of ANNs, in turn, predicted through a programming code in the Python language only the winners of the clashes. In this way, a greater accuracy of 72.81% was obtained, visualized through the generation of a confusion matrix.

Keywords: Soccer results prediction. ARIMA. Artificial Neural Networks. English Premier League.

1 INTRODUÇÃO

Dentre as inúmeras aplicações viáveis da Pesquisa Operacional (PO) existentes atualmente, pode-se mencionar a ascensão da sua utilização em eventos esportivos, realizando a previsão de uma série de variáveis dentro de um intervalo de tempo pré-determinado (Butler; Butler; Eakins, 2020). De acordo com Andrews *et al.* (2021), parte da procura por tais métodos deriva de administradores e analistas dos próprios clubes, que buscam informações a fim de elaborar estratégias para as partidas seguintes, baseados em dados da equipe adversária e resultados anteriores. Outro ramo que alavanca as pesquisas relacionadas é o mercado de apostas esportivas, no qual diversos apostadores se tornaram adeptos de modelos matemáticos visando aumentar seus lucros com palpites mais precisos.

As competições esportivas apresentam constante evolução e são impactadas por inúmeros fatores, sejam eles objetivos ou subjetivos, o que torna a análise preditiva de resultados uma tarefa árdua (Men, 2022). Conforme Bunker e Thabtah (2019), a previsão esportiva é tratada, inicialmente, como um dilema de definição do resultado final de uma partida, seja vitória, empate ou derrota. Porém, recentemente em esportes como o futebol, diversas outras estatísticas estão sendo analisadas, tal como a margem de gols, escanteios ou cartões, incluindo até mesmo a análise de dados relevantes sobre os atletas.

Segundo Koopman e Lit (2019), incontáveis características devem ser analisadas sendo consideradas determinantes para o resultado de uma partida. Dentre elas, é possível citar as forças de ataque e de defesa das equipes e a vantagem de jogar no seu estádio, além dos eventos e agentes específicos correspondentes a cada duelo. Visando obter um resultado com estimativas aceitáveis, é importante que haja uma seleção correta das informações e da forma de análise, já que a qualidade dos resultados está diretamente ligada aos recursos utilizados na modelagem do sistema de previsão (Baboota; Kaur, 2019).

No futebol, os principais torneios nacionais são disputados na Europa. Além de atrair torcedores e investimentos do mundo todo, geram uma elevada quantidade de dados para análise, disponibilizados por sites especializados ou pelo site oficial da competição. Em campeonatos europeus da primeira divisão as equipes duelam em

partidas com objetivos diversos, desde a concorrência pelo título, vagas nas competições continentais e até mesmo para evitar o rebaixamento à divisão inferior (Lahvicka, 2012).

Considerando as informações descritas anteriormente e visando relacioná-las com as técnicas de predição de PO, encontraram-se os seguintes problemas de pesquisa: Como prever corretamente resultados de partidas de futebol a partir de dados disponíveis de partidas já disputadas? E quais métodos estatísticos devem ser utilizados para obter os melhores resultados na previsão?

Como resposta às questões apresentadas e baseado nos dados estatísticos das 26 primeiras rodadas da *Premier League* Inglesa da temporada 2021-2022, o presente estudo apresenta a utilização do modelo Auto Regressivo Integrado de Média Móvel (do inglês, *Autoregressive Integrated Moving Average* (ARIMA)) e da Inteligência Artificial (IA), como ferramentas de análise e previsão de resultados para jogos de futebol do campeonato em questão. Com a aplicação de tais técnicas, tem-se o intuito de prognosticar os vencedores dos confrontos após as simulações realizadas, visando a validação dos métodos para uso por parte de clubes e analistas esportivos, que poderão verificar os resultados necessários para cada equipe alcançar os objetivos esperados na temporada, e apostadores, que podem fazer uso da ferramenta para maximizar seus lucros.

O artigo foi organizado de modo a trazer, de forma sequenciada, um breve referencial teórico, seguido da metodologia utilizada em cada fase do estudo e da apresentação dos métodos de trabalho e sua execução, para, por fim, exibir os resultados obtidos e analisar o desempenho dos métodos aplicados.

2 REFERENCIAL

A pesquisa do referencial foi subdividida e estruturada de acordo com os conhecimentos considerados essenciais para o desenvolvimento dos modelos propostos. Os principais trabalhos realizados na área foram analisados de forma minuciosa e serão detalhados dentro de cada seção, observando também os estudos com modelos ou estratégias diferentes da utilizada no presente artigo.

2.1 Previsão em resultados de futebol

Dentro de um evento esportivo como, por exemplo, uma partida de futebol, inúmeros aspectos podem ser considerados influentes e relevantes para compreender o resultado final. Portanto, previsões dessa magnitude requerem aplicação de diversos métodos de análise, visando tornar possível estabelecer uma solução que forneça a probabilidade de incidência de cada resultado eventual.

Conforme Koopman e Lit (2019), a probabilidade de um resultado ocorrer está implícita na distribuição representada por $P(X = x, Y = y)$, onde os valores de x e y pertencem ao conjunto de números naturais N . As variáveis X e Y representam o número de gols marcados pelas equipes, de modo que se tenha uma hipótese para qualquer resultado imaginável em uma partida. É possível, então, prognosticar a vitória do time da casa, visitante, ou empate, dadas por $P(X > Y)$, $P(X < Y)$ e $P(X = Y)$, respectivamente. Também pode-se verificar a margem de vitória, representada pela diferença de gols no confronto. Para tal, considera-se $Z = X - Y$, com Z pertencendo ao conjunto de números inteiros. Na modelagem dessa variável, são analisadas as probabilidades dadas por $P(Z > 0)$, $P(Z = 0)$ e $P(Z < 0)$. Os indicadores da distribuição podem ser evidenciados em função das forças de ataque e defesa das equipes.

Para Berrar, Lopes e Dubitzky (2018), visando obter uma previsão mais assertiva dos resultados, é preciso descrever uma equipe por meio de aspectos que a caracterizam em relação aos seus pontos fortes e fracos e que são relevantes em um jogo. Com base no conhecimento adquirido em futebol e no domínio da área de modelagem por meio de Inteligência Artificial (IA), entende-se que essas características devem levar em conta alguns aspectos principais, como os desempenhos de ataque e defesa da equipe, o desempenho recente, a força do adversário e a vantagem de jogar em casa.

Seguindo tal linha de observação, Koopman e Lit (2019) estabeleceram três parâmetros como alternativas para realizar a previsão do desfecho de partidas de seis ligas europeias de futebol. O primeiro deles interpreta o número de gols marcados pelas equipes durante o jogo. O segundo diz respeito à diferença entre o número de gols marcados por cada um dos times e, por fim, o último parâmetro representa o resultado final do evento.

Com uma proposta de trabalho diferente das citadas, Butler, Butler e Eakins (2020) coletaram previsões e resultados de 1140 partidas durante três temporadas da *Premier League* Inglesa, no período de 2014 a 2017. Os palpites foram dados por ex-jogadores profissionais, considerados como especialistas, e leigos. Destaca-se uma das formas utilizada para encontrar a eficiência das previsões, por meio do cálculo de proximidade do resultado, representada na Equação (1). Na operação em questão, $Pred_{GD}$ e $Actual_{GD}$ representam a diferença do número de gols prevista e real entre as equipes, enquanto $Pred_{TG}$ e $Actual_{TG}$ correspondem, respectivamente, ao número total de gols marcados e previstos, somados ambos os times.

$$Proximidade = abs(Pred_{GD} - Actual_{GD}) + abs((Pred_{TG} - Actual_{TG})/2) \quad (1)$$

Neste cálculo, é atribuída uma importância maior para a diferença de gols marcados entre as duas equipes, de modo que a resolução do vencedor tenha mais relevância do que a margem vitoriosa. Quando a previsão é igual ao resultado real, o cálculo tem valor igual a zero. Portanto, quanto menor o valor encontrado, mais preciso é o prognóstico (Butler; Butler; Eakins, 2020).

2.2 Modelos de regressão

Para previsão de eventos em geral, analisam-se séries temporais observadas durante um determinado período. Uma série temporal pode ser definida como uma observação, dentre inúmeras possíveis, de um processo estocástico em um tempo t , em função de uma variável qualquer. Conceitualmente, um modelo que explica uma série temporal é aquele capaz de minimizar os erros entre os valores presumidos e os dados observados posteriormente (Xavier, 2018).

Dentre os modelos capazes de realizar esse trabalho está o método ARIMA (p , d , q), construído a partir de um processo de identificação, estimativa, diagnóstico e previsão de dados. Em sua composição, a variável p representa os parâmetros dos números de defasagem, d refere-se ao número de diferenciações para que a série se torne estacionária e q denota a ordem do modelo de média móvel. O método tem

como premissa representar a natureza de uma série temporal gerada por um processo estocástico (Casimiro, 2009).

De acordo com Xavier (2018), o modelo ARIMA pode ser utilizado para delinear uma série não estacionária Y_t , que diferenciada por d vezes consecutivas acaba se tornando estacionária, representada por Z_t na Equação (2). Considerando que Y_t é uma integral de Z_t , o termo pode ser modelado com ARIMA, por intermédio da Equação (3), onde $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ e ε_t representa os números não explicados pelos valores passados da série.

$$Z_t = \nabla^d Y_t \quad (2)$$

$$\Phi(B) \nabla^d Y_t = \theta(B) \varepsilon_t \quad (3)$$

$$\Phi(B) (1 - B)^d Y_t = \theta(B) \varepsilon_t$$

Apesar dos resultados satisfatórios obtidos, uma das poucas implementações do modelo ARIMA para predição esportiva encontrados na literatura é o trabalho de Yiannakis *et al.* (2006). Nele, os dados finais das partidas foram inseridos como variáveis binárias indicando presença ou ausência do resultado. A fim de determinar se a análise de séries temporais correspondia ou não ao resultado real, os autores compararam as distribuições reais com cada uma das três condições de desfecho possíveis em um jogo. Ao final do estudo, alcançaram um sucesso de até 90% na definição de vitória, empate ou derrota de três equipes da *Premier League* Inglesa da temporada 1997-1998, em cada um de seus últimos dez jogos.

Além do sistema de previsão referido, outras técnicas de Regressão podem ser aplicadas para obtenção de resultados. Lasek e Gagolewski (2020) utilizaram modelos de classificação de equipes baseados no método estatístico *Elo rating system*. A proposta verificou a capacidade de predição dos Modelos de Regressão de Poisson (MRP) e de Regressão Logística Ordinal, avaliados quantitativamente a partir de previsões em cinco ligas de futebol europeu, disputadas na Itália, Inglaterra, Alemanha, Espanha e França. Dentre os modelos citados, os sistemas baseados no MRP se mostraram significativamente mais precisos que os demais. Como levam em conta os ajustes temporais de forma automática e não exigem ponderação de

resultados, esses modelos retratam com mais precisão o momento atual de cada equipe na disputa da partida, refletindo teoricamente numa previsão final mais correta.

No estudo do MRP de um parâmetro, que apresentou os melhores resultados da pesquisa, os autores reduziram as forças de ataque e defesa a um único fator r , calculado para a equipe da casa i e, simultaneamente, para a equipe visitante j . Enquanto c representa o intercepto e com um valor igual a zero anula a probabilidade de empate, a variável $h > 0$ ajusta o modelo representando a vantagem de jogar em casa. Os parâmetros das variáveis de Poisson são representados conforme as Equações (4) e (5) mostradas a seguir:

$$\log(u_i) = c + h + r_i - r_j \quad (4)$$

$$\log(u_j) = c + r_j - r_i \quad (5)$$

Ainda dentro do estudo citado, aplicando a Equação (6) ajustada para cada uma das equipes, definem-se as novas forças respectivas após uma iteração, correspondente a uma partida disputada. Os valores de $u_i^{(k)} - u_j^{(k)}$ e $g_i^{(k)} - g_j^{(k)}$ representam as margens de gols esperada e consumada, respectivamente. O termo λ , por sua vez, é definido a partir da Equação (7), onde ρ varia dentro de $[-1, 1]$, sendo uma correlação entre os índices de defesa e ataque.

$$r_i^{(k)} = r_i^{(k-1)} - \gamma \cdot [(u_i^{(k)} - u_j^{(k)}) - (g_i^{(k)} - g_j^{(k)}) + \lambda r_i^{(k-1)}] \quad (6)$$

$$\lambda \cdot \left(\frac{\|r\|_2^2}{2} - \rho \cdot a^\top d \right) = \frac{\lambda}{2} \sum_{i=1}^n (a_i - d_i)^2 \quad (7)$$

No seu estudo, Lahvicka (2012) utilizou simulação de Monte Carlo para calcular a importância das partidas na *Premier League* Inglesa. Em seu cálculo, dois fatores foram considerados fundamentais: a probabilidade de um time alcançar dado objetivo na temporada e como tal probabilidade dependia do resultado da partida analisada. A primeira etapa do método, que mais relaciona-se com o presente artigo, diz respeito à conjectura dos resultados restantes do torneio. Para uma partida, calculam-se inicialmente a pontuação média da equipe da casa nos últimos 19 jogos em seus domínios e da equipe visitante no mesmo período em duelos fora de seu estádio. Na

sequência, apura-se o número de gols marcados em cada um desses jogos pelas equipes, gerando duas variáveis independentes para cada time, expressadas pela distribuição de Poisson, conforme descrito nas Equações (8) e (9) adaptadas.

$$\lambda_{time\ da\ casa} = \frac{média\ de\ gols\ marcados_{casa} + média\ de\ gols\ cedidos_{visitante}}{2} \quad (8)$$

$$\lambda_{time\ visitante} = \frac{média\ de\ gols\ cedidos_{casa} + média\ de\ gols\ marcados_{visitante}}{2} \quad (9)$$

Baseado nos valores encontrados, Lahvicka (2012) estabeleceu uma previsão do número de gols médio de cada time e o retorno esperado para o evento específico objeto de estudo, calculando assim os possíveis resultados da partida e as respectivas probabilidades de vitória das equipes.

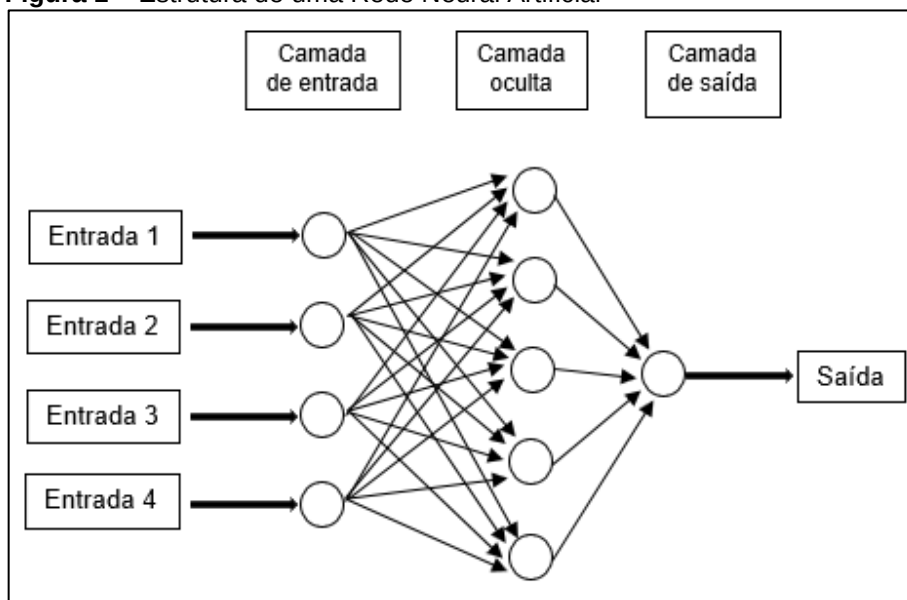
2.3 Inteligência Artificial (IA) e Redes Neurais Artificiais (RNA)

De acordo com Men (2022), para métodos de previsão de resultados esportivos, é válido ressaltar que em situações, nas quais a quantidade de dados utilizada é volumosa, os métodos estatísticos manuais perdem a praticidade e não trazem resultados satisfatórios, o que os difere de soluções que utilizam estratégias de IA. Além disso, o processo de previsão por *Machine Learning* (ML) opera como uma inteligência capaz de gerenciar as várias restrições presentes nos eventos, sendo considerado um modelo de previsão multifatorial com séries temporais. Comparado aos algoritmos de previsão tradicionais, a IA traz vantagens ligadas sobretudo a certos recursos que possuem melhor adaptação às mudanças e variáveis das partidas.

Dentre os métodos de IA utilizados para previsões esportivas destacam-se as Redes Neurais Artificiais (RNAs), que a partir de suas ligações e conexões são capazes de fornecer uma resposta desejada para o problema. Uma RNA é composta por unidades de processamento de funções matemáticas, chamadas de neurônios, e possui uma saída associada aos dados de entrada e outros componentes relacionados. Para cada uma das informações influentes no modelo é atribuído um peso, ajustado continuamente devido a não linearidade da rede, de modo a obter uma saída mais precisa e favorecer uma tomada de decisão final (Bunker; Thabtah, 2019).

Na Figura 1 pode-se verificar um exemplo da estrutura de uma RNA, com suas entradas, suas camadas ocultas que ajustam os pesos das variáveis e a consequente saída do modelo.

Figura 1 – Estrutura de uma Rede Neural Artificial



Fonte: Adaptado de Bunker e Thabtah (2019).

As Redes Neurais podem conter uma ou mais camadas ocultas, conforme a necessidade e o propósito de aplicação. Comumente, as redes com mais de uma camada oculta são chamadas de redes multicamadas (do inglês, *Multi Layer Perceptron* (MLP)). Estas se destacam por algumas características principais, como a presença de uma função de ativação não linear diferenciável em cada neurônio e um grau de conectividade elevado, associado aos pesos sinápticos da rede (Haykin, 2009).

Um exemplo do uso de IA em prognósticos de partidas de futebol de campeonatos europeus está no estudo de Berrar, Lopes e Dubitzky (2018), cujos autores criaram um método que trabalha na previsão de gols com apoio de classificações de desempenho de cada equipe. Após cada jogo realizado, as classificações eram atualizadas de acordo com os resultados esperado e observado. A partir de fatores definidos como forças e fraquezas do ataque e da defesa de cada um dos times, obtiveram-se as Equações (10) e (11) para previsão de gols das equipes da casa e visitante, representadas pelas letras *H* e *A*, respectivamente.

$$\hat{g}_h(H_{hatt}, A_{adef}) = \frac{\alpha_h}{1 + \exp(-\beta_h(H_{hatt} + A_{adef}) - \gamma_h)} \quad (10)$$

$$\hat{g}_a(A_{hatt}, H_{adef}) = \frac{\alpha_a}{1 + \exp(-\beta_a(A_{hatt} + H_{adef}) - \gamma_a)} \quad (11)$$

Dentro das equações apresentadas, tem-se as seguintes variáveis pertencentes ao conjunto de números reais:

- $\hat{g}$ são os gols previstos para a equipe;
- H_{hatt} representa a força ofensiva para jogos em casa;
- H_{hdef} descreve a fraqueza defensiva nos jogos em casa;
- A_{hatt} simboliza a força ofensiva para jogos como visitante;
- A_{adef} representa a fraqueza defensiva em jogos como visitante;
- α é uma constante que delimita o valor máximo para $\hat{g}$;
- β atua como uma constante que define a inclinação das curvas sigmóides;
- γ é uma constante que representa o ponto limite das curvas.

Ainda de acordo com os autores, após cada partida realizada torna-se obrigatória a atualização da classificação dos parâmetros de controle. Para executar tais operações, são utilizadas as Equações (12) e (13), como exemplo para o time mandante.

$$T_{hatt}^{t+1} = T_{hatt}^t + \omega_{hatt}(g_h - \hat{g}_h) \quad (12)$$

$$T_{hdef}^{t+1} = T_{hdef}^t + \omega_{hdef}(g_a - \hat{g}_a) \quad (13)$$

Onde:

- $\hat{g}$ são os gols previstos para a equipe;
- g são os gols marcados pelo time na partida;
- T_{hatt}^{t+1} descreve a nova força de ataque para jogos em casa após uma partida;
- T_{hatt}^t apresenta a força ofensiva nos jogos em casa antes do confronto;
- T_{hdef}^{t+1} reproduz a nova fraqueza defensiva para jogos como mandante após uma partida;

- T_{hdef}^t representa a fraqueza defensiva prevista para jogos em casa;
- ω é um peso atribuído para a atualização das forças calculadas.

Juntamente com esse método de previsão, os autores utilizaram um algoritmo de k-vizinhos mais próximos (do inglês, *k-Nearest Neighbors (kNN)*), com o intuito de selecionar um número determinado de eventos recentes para servir como base na classificação das características determinantes do modelo. A escolha do número de partidas a serem utilizadas é um ponto crítico, visto que uma amostra de tamanho menor pode ser muito sensível a ruídos, todavia, o oposto pode trazer à tona casos irrelevantes e potencialmente prejudiciais ao estudo. O conjunto de recursos aplicados a 31318 jogos, empregando um número ótimo de *kNN* igual a 248 vizinhos, obteve o melhor resultado de previsão, com uma Pontuação de Probabilidade Classificada (do inglês, *Ranked Probability Score (RPS)*) média de 0,2059. No estudo apresentado, quanto menor o *RPS*, melhor o prognóstico.

Em outra pesquisa de cunho esportivo observada nesta área, Pereira (2018) apresentou um modelo considerando inúmeras características influentes em uma partida, a fim de prever resultados dos jogos do Campeonato Brasileiro da Série A de 2018. Foram relacionados fatores como coeficientes de tática, valor do clube, mando de campo, importância do confronto e sorte, proporcionando uma simulação completa das 38 rodadas do campeonato em questão. A partir de um algoritmo criado no *software* MATLAB®, que realizou a leitura de arquivos com os dados coletados e executou códigos específicos para cada um dos atributos citados, foi possível prever diversas situações, com destaque para a assertividade nas escolhas do campeão e vice, e na definição correta de três dos quatro rebaixados à divisão inferior.

Em trabalhos como este, as divergências e erros de predição encontrados dizem respeito a fatores de difícil previsão, como perda ou contratação de novos jogadores importantes, mudanças de técnicos e a priorização pela disputa de outras competições concomitantes.

3 METODOLOGIA

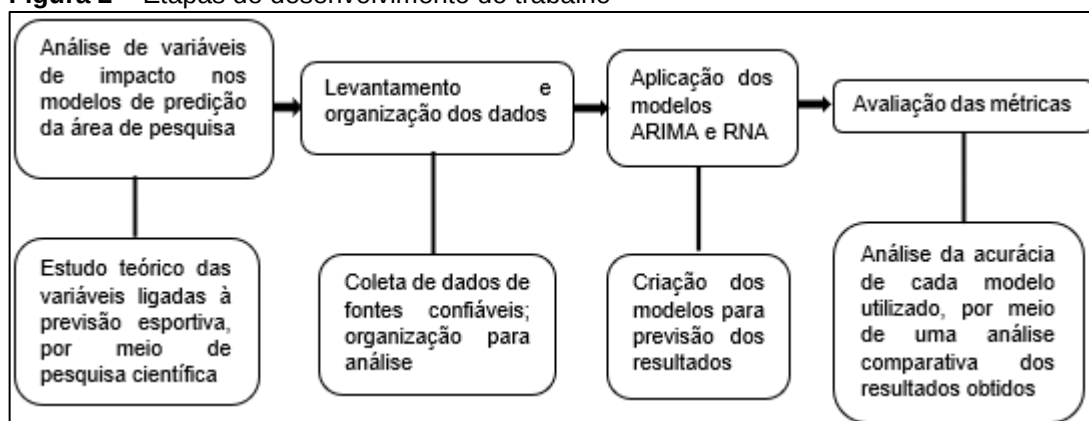
Os estudos relacionados à previsão de partidas de futebol têm ganhado destaque devido à importância dada por apostadores, analistas de clubes e

especialistas de sites e canais esportivos. Desse modo, os trabalhos e criações de modelos de análise se tornaram cada vez mais comuns na área, visto também que as estatísticas dos jogos estão disponíveis em diversas fontes *online*, de forma gratuita e acessível para qualquer indivíduo.

Valendo-se dessa oportunidade, este estudo foi conduzido. Com relação à sua classificação, o trabalho é definido como uma pesquisa aplicada, na medida que possui objetivo de gerar conhecimentos de valor prático para um problema específico já determinado. Além disso, quanto à sua abordagem, a pesquisa é caracterizada como quantitativa e, por fim, quanto aos objetivos, possui caráter explicativa, devido ao seu cunho experimental e realização a partir de fatos passados.

Para melhor andamento das atividades, o projeto foi dividido de modo a propor inicialmente uma abordagem teórica do tema escolhido. Posteriormente, foram coletados os dados utilizados para, por fim, aplicar o modelo ARIMA e de RNA como ferramentas de previsão de resultados das partidas de futebol da *Premier League* Inglesa 2021-2022, com o objetivo de apontar os vencedores dos duelos restantes do torneio com o menor grau de incerteza possível. As etapas do processo são descritas na Figura 2 e detalhadas na sequência.

Figura 2 – Etapas de desenvolvimento do trabalho



Fonte: Autores (2023).

3.1 Análise de variáveis de impacto nos modelos de predição da área de pesquisa

Visando uma elaboração planejada e estruturada, o primeiro passo no desenvolvimento da pesquisa foi a busca por trabalhos semelhantes já produzidos na área. Considerando a complexidade da proposta de trabalho, foi fundamental avaliar e entender o problema objeto de pesquisa para a sequência do projeto.

Realizaram-se então pesquisas relacionadas à previsão de eventos esportivos em geral e implementação de modelos para prognósticos em partidas de futebol, verificando seus resultados alcançados. As buscas foram efetuadas em bases de consulta reconhecidas, como o Portal CAPES e o *ScienceDirect*. A primeira reúne conteúdo científico de alto nível disponível à comunidade acadêmica brasileira. Já a segunda, propriedade da editora *Elsevier*, foi escolhida por conter os textos completos de publicações em renomadas revistas científicas europeias, mais voltadas à área de pesquisa do que as demais.

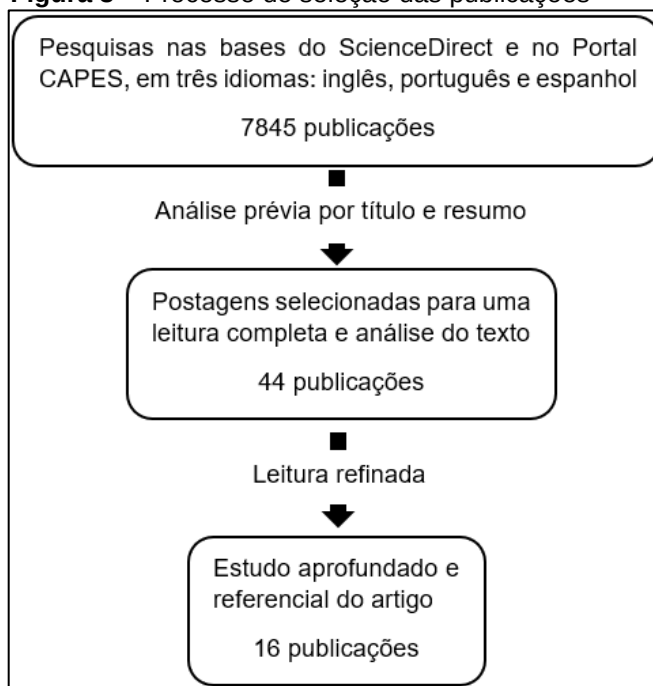
Para tal busca, foram utilizadas expressões chave como “previsão de resultados esportivos”, “ARIMA na previsão de resultados esportivos/de futebol” e “IA na previsão de resultados esportivos/de futebol”. As pesquisas foram conduzidas em três idiomas: português, inglês e espanhol, além de terem sido preferencialmente limitadas a publicações dos últimos cinco anos e obrigatoriamente relacionadas a esportes de cunho profissional e análises quantitativas.

Nas páginas de busca citadas, somadas as pesquisas e seguindo as regras estabelecidas, com cada expressão mencionada sendo pesquisada de forma individual, foram encontrados 7845 artigos relacionados aos tópicos escolhidos e suas variações. No entanto, foi necessário filtrar os resultados, já que alguns itens não correspondiam exatamente ao assunto abordado. Os autores realizaram, então, uma análise prévia pelo título das postagens, de modo a descartar as publicações não relacionadas. Posteriormente, ainda nessa etapa, foram lidos os resumos dos artigos restantes, também com o intuito de eliminar publicações que não faziam referência ao tema escolhido.

Por meio desse processo de seleção das publicações, foram designados para uma leitura refinada os 44 artigos considerados mais relevantes, com elevada

qualificação. Ao final, restaram para um estudo mais profundo um total de 16 publicações. O processo completo de seleção dos referenciais é apresentado de forma visual na Figura 3.

Figura 3 – Processo de seleção das publicações



Fonte: Autores (2023).

Avaliando as 16 publicações selecionadas, foram encontradas estruturas de trabalho semelhantes à desejada, com ideias pertinentes, que serviram de referência para a pesquisa desenvolvida. Nas publicações analisadas, em primeiro momento, observou-se o estudo das características do esporte e a manipulação de suas variáveis, desde o histórico estatístico e de desempenho das equipes, até o mando de campo do duelo e a classificação atual na competição. Também puderam ser vistos, na sequência, os testes feitos pelos autores, em diferentes conjuntos de dados, com a aplicação de IA, ARIMA e demais modelos de Regressão – Linear, Logística e de Poisson.

Dessa forma, baseado nos processos de modelagem e resultados por eles obtidos, foi possível selecionar de modo mais assertivo as técnicas que seriam utilizadas neste artigo, optando pela aplicação de ARIMA e RNA para a previsão dos resultados, a partir da média ponderada das estatísticas dos jogos já realizados.

3.2 Levantamento e organização dos dados

Finalizada a parte teórica e de aprofundamento do conhecimento sobre o assunto abordado, desencadeou-se uma série de pesquisas e análises dos dados. Para o presente estudo, foram coletados e reorganizados dados do site *football-data.co.uk*, correspondentes às trinta e oito rodadas da *Premier League* Inglesa, disputadas entre 13/08/2021 e 22/05/2022. Estavam incluídos números referentes a cada uma das dez partidas disputadas por rodada, tendo sido preservados ao final apenas os utilizados na construção dos modelos. Vale ressaltar que foram levantados somente valores brutos, de modo que não havia análises estatísticas prévias do desempenho das vinte equipes que disputam o torneio.

Para fins de validação, os dados extraídos foram comparados por amostragem com informações verificadas diretamente no site oficial da *Premier League*, de maneira que nenhuma divergência foi encontrada. As estatísticas utilizadas da primeira rodada do campeonato podem ser vistas na Figura 4, como exemplo, após serem ordenadas.

Figura 4 – Dados da primeira rodada do campeonato

Time da casa	Time visitante	Casa Gols Final	Visit. Gols Final	Resultado Final	Casa Gols Intervalo	Visitante Gols Intervalo	Resultado Intervalo	Casa Chutes	Visit. Chutes	Casa Chutes no alvo	Visit. Chutes no alvo	Casa Escanteios	Visitante Escanteios
Brentford	Arsenal	2	0	Casa	1	0	Casa	8	22	3	4	2	5
Man United	Leeds	5	1	Casa	1	0	Casa	16	10	8	3	5	4
Burnley	Brighton	1	2	Visitante	1	0	Casa	14	14	3	8	7	6
Chelsea	Crystal Palace	3	0	Casa	2	0	Casa	13	4	6	1	5	2
Everton	Southampton	3	1	Casa	0	1	Visitante	14	6	6	3	6	8
Leicester	Wolves	1	0	Casa	1	0	Casa	9	17	5	3	5	4
Watford	Aston Villa	3	2	Casa	2	0	Casa	13	11	7	2	2	4
Norwich	Liverpool	0	3	Visitante	0	1	Visitante	14	19	3	8	3	11
Newcastle	West Ham	2	4	Visitante	2	1	Casa	17	8	3	9	7	6
Tottenham	Man City	1	0	Casa	0	0	Empate	13	18	3	4	3	11

Fonte: Autores (2023).

No processo de organização dos dados - que contemplam 14 variáveis assumindo diferentes valores nas 266 partidas iniciais do torneio - realizou-se inicialmente uma preparação para a posterior aplicação do modelo. Foram definidas as variáveis mais influentes a serem analisadas, conforme ilustra a Figura 4, com base no conhecimento adquirido e nos trabalhos já desenvolvidos na área. Em seguida,

para a aplicação de RNA em específico, os dados foram divididos em subconjuntos, direcionando 70% do seu total para treinamento e 30% para teste e validação.

Seguindo a análise das condições de uma partida, de acordo com Yiannakis *et al.* (2006), existem dois tipos de variáveis que afetam um jogo de futebol. Enquanto as descritas como de contexto social têm impacto direto em uma partida, os elementos internos, tal qual as estatísticas das equipes no torneio, representam eventos passados que não afetam o jogo em sua essência, mas trazem indícios para prever o resultado que irá ocorrer. No caso deste trabalho, os dados da *Premier League* extraídos para uso tratam-se somente de variáveis internas, relacionadas a experiências anteriores ocorridas em partidas do torneio, como os históricos de desempenho dos times.

Cabe ressaltar, porém, que inúmeras das variáveis de contexto social que influenciam uma partida de futebol são consideradas incontrolláveis e imponderáveis utilizando os modelos matemáticos aplicados. Alguns exemplos, já citados nas obras de Berrar, Lopes e Dubitzky (2018), Pereira (2018) e Yiannakis *et al.* (2006), são vistos quando uma equipe que disputa torneios continentais simultaneamente escala reservas em partidas da liga, comprometendo seu desempenho, em jogos em que a condição física dos atletas não é adequada, quando ocorrem decisões controversas de arbitragem e se há lesões de jogadores importantes de determinado time.

3.3 Aplicação dos modelos ARIMA e RNA

Após o rearranjo e a preparação dos dados, a próxima etapa executada foi a criação dos modelos matemáticos, que posteriormente seriam solucionados com o apoio de um *software* adequado, considerando as características dos valores e a dimensão da aplicação das análises em estudos de esportes coletivos deste tipo. A partir desse estágio, as verificações realizadas foram de princípio quantitativo, fundamentadas pelas estatísticas disponíveis.

Com base na análise de tais dados estatísticos, foi efetuada a aplicação do modelo ARIMA e de RNA como meios de previsão de resultados das últimas 114 partidas da *Premier League*, baseados nos dados obtidos nos 266 confrontos iniciais. Esses métodos de trabalho foram escolhidos devido à sua capacidade de lidar com

inúmeras variáveis independentes de entrada, além dos resultados satisfatórios obtidos em prognósticos semelhantes.

As análises foram implementadas inicialmente com as variáveis independentes de resultado final, resultado ao intervalo, número de gols, chutes no alvo, finalizações totais e escanteios de cada equipe. Cada previsão foi gerada com base no conjunto total de dados passados, de modo que as estatísticas de novas partidas foram incorporadas a cada rodada.

Para a realização da predição por meio do método ARIMA foi empregado o *software* estatístico SPSS®. O primeiro passo executado foi a verificação do processo visando garantir que este se tornasse estacionário, no qual o conjunto de observações realizadas apresenta um comportamento estável ao longo do tempo. Na sequência, o modelo foi aplicado e refinado por meio dos parâmetros de ajuste, conforme descritos por Zucco (2019), sendo:

- p : ordem máxima de parâmetros de autorregressão simples;
- d : número de diferenciações não sazonais aplicadas para que a série temporal se torne estacionária;
- q : ordem dos parâmetros de média móvel simples;
- P : ordem máxima de parâmetros de autorregressão sazonal;
- D : quantidade de diferenciações sazonais empregadas para que a série temporal se torne estacionária;
- Q : ordem dos parâmetros de média móvel sazonal.

Na previsão realizada com base em RNA fez-se uso de um código de programação na linguagem Python, por meio de uma rede multicamadas. A rede foi configurada com quatro camadas: uma de entrada, com 13 neurônios, duas ocultas, constituídas por 128 e 64 neurônios, respectivamente, e uma de saída, com três neurônios. Para a entrada da rede, foram fornecidos os dados conforme exemplo visto na Figura 4, referentes aos primeiros 266 jogos do torneio. O processo de aprendizado da rede foi supervisionado, visto que ocorreu com o auxílio de um conjunto de exemplos de entradas e saídas, ajustando os pesos das variáveis para alcançar a solução desejada (Santos, 2014).

Para ativação do código foi utilizada a função ReLU (*Rectified Linear Unit*), padrão do *MLPClassifier*. Tal função é aplicada após a combinação linear das

entradas e pesos de cada neurônio nas camadas ocultas e de saída. Para automatizar o processo de ajuste dos parâmetros analisados utilizou-se a ferramenta *GridSearchCV*, visando testar todas as combinações a partir dos valores de entrada e selecionar os parâmetros otimizados com menor percentual de erro.

Conforme descrito anteriormente, o método de previsão por RNA é dividido em três etapas principais, e o processo adotado neste artigo seguiu tal protocolo de divisão de dados, estruturado em treinamento, teste e validação. Inicialmente, no processo de treinamento do modelo, foram fornecidos conjuntos de dados de entrada, juntamente com valores de saídas esperados dos primeiros 266 jogos disputados do torneio, correspondentes à 70% do total. Com a definição do número de camadas ocultas e a quantidade respectiva de neurônios, a rede foi treinada de modo a fazê-la aprender a lidar com os dados.

Nas etapas seguintes, de teste e validação, foram fornecidas para a rede apenas as entradas das últimas 114 partidas realizadas no campeonato, gerando os prognósticos e encontrando o valor de erro. A partir dos resultados alcançados nos testes foi possível estabelecer a melhor configuração para a rede, passando assim para o processo de comparação das saídas obtidas com os valores reais, calculando o erro de previsão conforme método detalhado na próxima seção.

É importante salientar que os prognósticos baseados em RNA foram criados de modo que apenas o vencedor do confronto foi previsto. Já utilizando ARIMA, foram obtidas também as previsões do número de gols de cada equipe, gerando um conjunto de estatísticas mais completo e uma métrica de avaliação distinta.

3.4 Avaliação das métricas

A última etapa do trabalho, ainda relacionada ao desenvolvimento dos modelos, foi a avaliação das técnicas de previsão estabelecidas, de modo a verificar o desempenho dos métodos por meio de métricas definidas, vistas na bibliografia, e definir a melhor técnica de predição dentre as estudadas neste trabalho.

Para o modelo ARIMA, inicialmente foi apurado o acerto do resultado final do confronto, seja vitória, empate ou derrota (Yiannakis *et al*, 2006). Em caso de previsão correta, verificou-se o placar exato da partida. Se necessário, também foi analisada a

diferença de gols e executado o cálculo de proximidade do placar, conforme já detalhado na Equação (1) (Butler; Butler; Eakins, 2020). Já na análise de RNA, utilizando apenas os vencedores de cada confronto, optou-se por aplicar uma matriz de confusão, exibindo a distribuição dos registros em termos de seus valores reais e previstos.

4 RESULTADOS

A partir dos dados do problema e fazendo uso das ferramentas de previsão mencionadas foram avaliados os resultados de 114 jogos, detalhados na sequência dessa seção. Para realizar tal predição e como base de estudo, um banco de dados de 266 confrontos disputados na mesma edição do campeonato foi utilizado. O Quadro 2 apresenta um breve resumo estatístico desse banco, apontando os vencedores dos duelos, juntamente com a média e o desvio padrão de cada uma das variáveis empregadas nos prognósticos.

Quadro 2 – Resumo descritivo da base de dados

	Casa	Empate	Visitante
Resultado Final	108	67	91
Resultado Intervalo	86	112	68
Gols (Média)	1,489	-	1,297
Gols (Desvio padrão)	1,338	-	1,222
Chutes (Média)	14,075	-	11,677
Chutes (Desvio padrão)	5,591	-	5,385
Chutes no alvo (Média)	4,733	-	4,143
Chutes no alvo (Desvio padrão)	2,685	-	2,637
Escanteios (Média)	5,650	-	4,917
Escanteios (Desvio padrão)	2,878	-	2,839

Fonte: Autores (2023).

Analisando brevemente os dados apresentados, observa-se uma proximidade nos valores de mandantes e visitantes em todos os quesitos relativos às estatísticas do jogo. Já referente ao resultado final da partida, nota-se que há um maior número de vitórias do time da casa, visto que são 40,60% de triunfos dos mandantes, contra 34,21% dos visitantes e 25,19% de empates, indicando que jogar em seu estádio pode proporcionar vantagem no confronto. Baseados nessas informações, serão exibidos nas próximas seções do artigo os valores obtidos pelos modelos ARIMA e RNA, com suas respectivas etapas de avaliação.

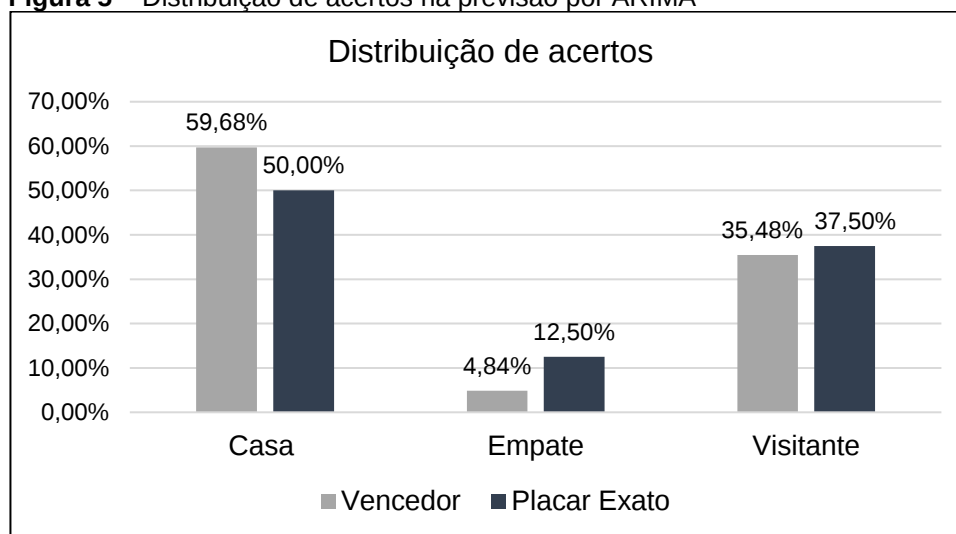
4.1 ARIMA

O primeiro método a ser averiguado é o ARIMA, com seus parâmetros estabelecidos em (1,0,0), que retornaram os resultados de forma mais coerente com o esperado pelo estudo e para uma partida de futebol. Os demais parâmetros testados apresentaram valores negativos ou padrões de vitórias do time mandante, com uma margem de vitória de quatro ou mais gols em todos os duelos. A avaliação seguiu o conjunto de etapas descritas na seção da Metodologia, possibilitando a análise dos valores obtidos.

A primeira das métricas de avaliação aplicada foi a verificação do vencedor do confronto. Nas partidas analisadas, o modelo alcançou um total de 62 acertos, correspondentes a 54,39%. Essa pode ser considerada a etapa mais importante de avaliação, visto que os resultados previstos incorretamente foram descartados da análise e podem ter atribuídos a si um erro total, enquanto os corretos seguiram para as demais etapas de verificação.

Como segunda etapa de apuração de eficácia do modelo, analisou-se o placar exato das partidas. Em caso de acerto nessa fase de avaliação, considera-se o equivalente a um erro zero. Foram obtidos na previsão 16 placares exatos, correspondendo a mais de um por rodada e 14,04% do total de resultados possíveis. A partir dos dados dessas duas primeiras etapas é possível visualizar, conforme a Figura 5, a proporção dos acertos de vencedores e do placar exato em relação ao mando de campo da partida. Nota-se que o modelo prevê resultados assertivos predominantemente apontando vitórias de uma das equipes, sendo que as taxas de acerto de empates são de apenas 4,84% e 12,50%, respectivamente.

Figura 5 – Distribuição de acertos na previsão por ARIMA



Fonte: Autores (2023).

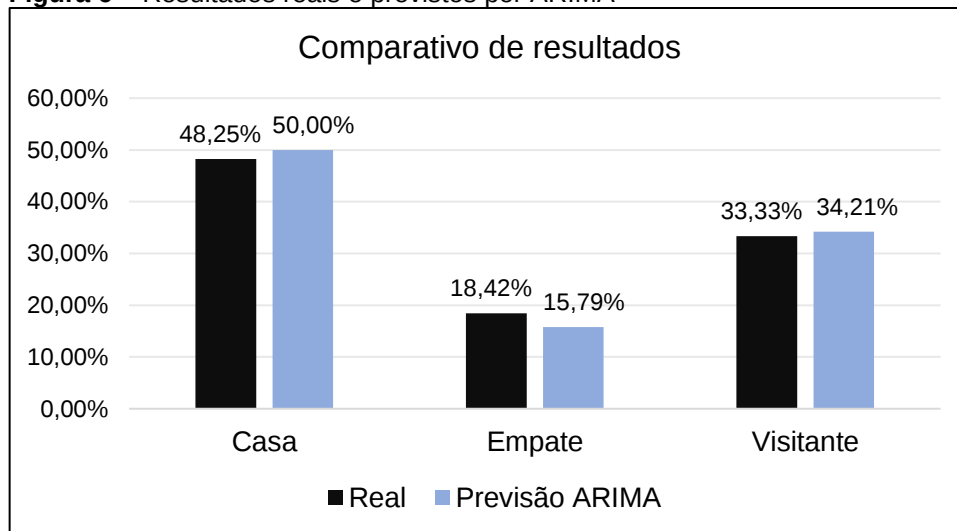
Considerando as partidas que tiveram seus vencedores previstos de maneira correta, porém com placar incorreto, houve duas últimas etapas de verificação de erro – na qual a porcentagem de acerto não foi calculada sobre os 114 jogos, e sim sobre o número respectivo possível. A primeira delas corresponde à conferência da diferença de gols entre a equipe vencedora e a perdedora. Um exemplo é a partida entre *Manchester City* e *Aston Villa*, que teve como resultado final a vitória do time da casa pelo placar de 3 a 2. A previsão apontou uma vitória de 2 a 1 da equipe de *Manchester*, de modo que o placar não foi exato, mas a diferença de gols calculada foi correta. Como resultado dessa etapa de verificação, dentro dos 46 confrontos passíveis de avaliação, o método previu de maneira correta a diferença de gols em 10 deles, equivalente a 21,74% de acerto;

Por fim, a última fase aplicada na avaliação do modelo foi o cálculo da proximidade do placar. Essa análise só ocorreu em situações específicas, nas quais o vencedor foi previsto corretamente, porém o placar exato e a diferença de gols entre os times não estavam corretos. Das 36 partidas que se encaixaram nos requisitos comentados, 12 delas tiveram um valor calculado de 1,5 gols ou menos, o que representa um placar com diferença incorreta de um gol em relação ao número real.

Além das avaliações dos resultados obtidos pode-se verificar também que, proporcionalmente, o método previu os vencedores dos confrontos de modo majoritariamente assertivo. Conforme pode ser visto na Figura 6, os prognósticos

apontaram para um número de vitórias maior da equipe da casa em relação à visitante e aos empates, com valores que se assemelham aos dados reais.

Figura 6 – Resultados reais e previstos por ARIMA

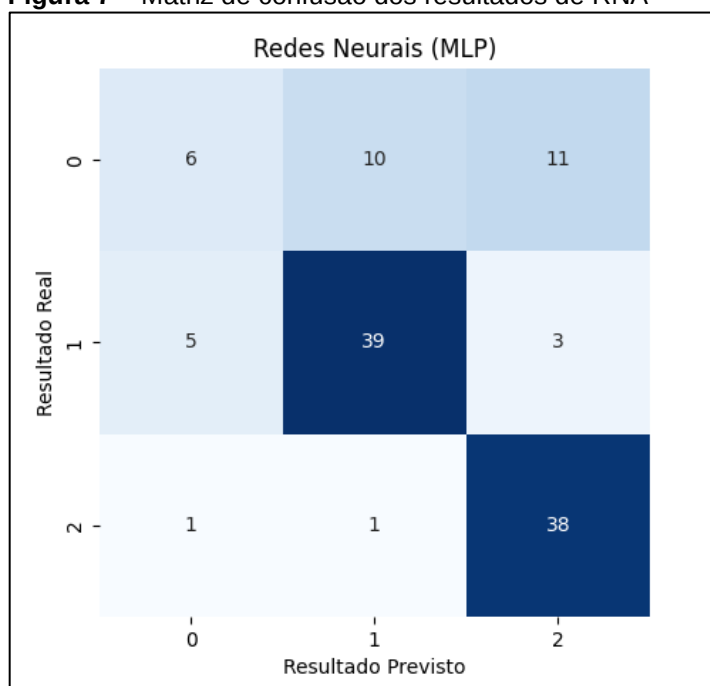


Fonte: Autores (2023).

4.2 RNA

O modelo de RNA seguiu uma métrica de avaliação diferente do método ARIMA, devido a sua também distinta forma de geração dos resultados. Considerando a análise somente dos vencedores, a matriz de confusão utilizada exibe de forma prática e visual os valores obtidos, sendo que os verdadeiros positivos são os valores corretos na previsão. Na construção do modelo, a variável “1” representa vitória do time da casa, “2” atribui o triunfo para a equipe visitante e “0” representa o empate no confronto. As previsões relacionadas com os resultados reais e suas respectivas ocorrências estão dispostas conforme matriz apresentada na Figura 7.

Figura 7 – Matriz de confusão dos resultados de RNA

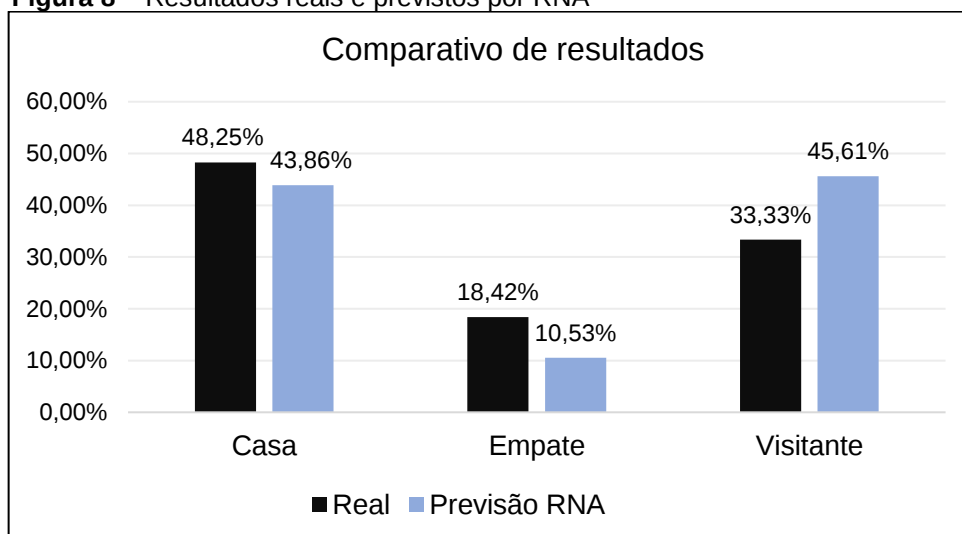


Fonte: Autores (2023).

No estudo de RNA, somados os prognósticos corretos, obteve-se um total de 83 acertos nas 114 partidas analisadas. Verificando o aproveitamento do modelo, encontra-se uma acurácia de 72,81%, equivalente a pouco mais de sete dos dez jogos disputados em cada rodada. Os acertos obtidos foram vistos quase que em sua totalidade apontando a vitória de uma das equipes, sendo 39 do time da casa e 38 do visitante, que somados equivalem a 92,77% dos resultados gerados corretamente. Apenas seis acertos foram alcançados a partir da previsão de um empate.

Conforme é possível verificar na matriz, o resultado das previsões apontou para 50 vitórias dos mandantes (43,86%), 52 dos visitantes (45,61%) e 12 empates (10,53%). Esses resultados divergem uma tendência inicialmente avaliada na base de dados, cujas equipes jogando em seu estádio obtinham mais vitórias e a taxa de empates era cerca de 15% maior. Comparando com os dados reais dos confrontos das últimas rodadas, percebe-se que realmente há uma queda no número de empates, porém as equipes mandantes continuam somando mais vitórias do que as visitantes. A Figura 8 traz essa análise comparativa entre os resultados efetivos e previstos por RNA em um gráfico, facilitando a visualização de tal diferença.

Figura 8 – Resultados reais e previstos por RNA



Fonte: Autores (2023).

5 CONSIDERAÇÕES

O estudo propôs a criação de métodos de previsão para resultados de partidas de futebol baseado no histórico estatístico de parte de uma temporada disputada. Levando em conta os objetivos da pesquisa, estabelecidos no início do processo de construção do artigo, considera-se que foram alcançados de modo satisfatório, visto que completaram com êxito as etapas planejadas. Inicialmente, por meio de pesquisa científica, as variáveis relacionadas à predição esportiva foram estudadas e analisadas de modo teórico. Na sequência, fazendo uso dos métodos escolhidos, de diferentes formas, foram realizados os prognósticos dos resultados dos confrontos da *Premier League* 2021-2022 para posterior análise de desempenho de cada modelo.

Com relação aos resultados obtidos, verificou-se que, para apontar o vencedor dos jogos, o modelo de RNA apresenta maior acurácia, cerca de 18,40% mais eficiente do que o método ARIMA. No entanto, em estudos deste tipo pode-se considerar relevante também observar o placar previsto em cada partida, tornando a análise mais completa. Nesse ponto, constatou-se que a técnica ARIMA se mostra mais eficiente, pois entrega de maneira mais assertiva e coerente as estatísticas integrais dos confrontos. A escolha do método a ser utilizado, desse modo, se relaciona diretamente com o cliente em questão, seja um analista que examine minuciosamente os dados, ou um apostador que pretende apenas encontrar o vencedor de um duelo.

O trabalho realizado implicou em algumas descobertas e gerou contribuições que poderão ser empregadas em novas publicações referentes ao assunto tratado. Em relação à aplicação dos métodos, foi possível encontrar a melhor configuração dos parâmetros ARIMA para estudos desse tipo, além de verificar que o conjunto de métricas de avaliação, que ainda podem ser aperfeiçoadas, se mostra eficiente e capaz de realizar um apanhado geral do desempenho dos modelos.

Alguns fatores podem ser citados como justificativa para os resultados previstos incorretamente e apontados como limitações do trabalho. O primeiro deles é que apenas dados estatísticos das partidas foram utilizados como base para previsão. Desse modo, a influência de todos os fatores externos é desconsiderada nesse tipo de modelo, como o intervalo de descanso que cada um dos times teve desde o seu último jogo e se as equipes estão com seu elenco completo para o confronto. Além disso, é preciso ressaltar que mesmo em situações ótimas, nem sempre o melhor time e com maior probabilidade irá vencer uma partida, elevando ainda mais a possibilidade de erro.

Juntamente das limitações observadas, também é válido citar a complexidade deste tipo de problema. Para o funcionamento adequado dos métodos é necessário realizar testes e adequações que envolvem todos os parâmetros apresentados nas seções iniciais do artigo, otimizando-os, a fim de obter o resultado desejado. Esses ajustes, inclusive, diferem conforme o modelo de previsão. Somando isso ao controle dos fatores de erro já citados, o estudo se apresentou laborioso, porém executado de acordo com o desejado dentro do planejamento estabelecido.

Para pesquisas futuras, sugere-se o aperfeiçoamento dos modelos a partir da inclusão de novas variáveis que possam trazer mais profundidade ao estudo. Além dos fatores citados anteriormente, pode-se utilizar o valor de mercado dos elencos para atribuir sua força e o histórico estatístico de outras temporadas da liga como parte da base de dados. Na predição por meio de RNA, o código de programação pode ser ajustado de modo a gerar também o número de gols de cada equipe, o que possibilitaria uma comparação mais adequada entre os modelos. Na análise dos resultados, pode ser adicionada a classificação final do campeonato depois dos resultados previstos, comparando-a com a tabela real. Essa ação poderá expor os reais impactos dos erros obtidos para o campeonato, como possíveis alterações no

campeão, classificados para competições continentais e rebaixados, trazendo assim maior respaldo e confiabilidade para a pesquisa.

REFERÊNCIAS

ANDREWS, S Kevin *et al.* Analysis on Sports Data Match Result Prediction Using Machine Learning Libraries. **Journal Of Physics: Conference Series**, [S.L.], v. 1964, n. 4, p. 1-9, 1 jul. 2021. IOP Publishing. <http://dx.doi.org/10.1088/1742-6596/1964/4/042085>. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1964/4/042085/meta>. Acesso em: 26 ago. 2022.

BABOOTA, Rahul; KAUR, Harleen. Predictive analysis and modelling football results using machine learning approach for English Premier League. **International Journal Of Forecasting**, [S.L.], v. 35, n. 2, p. 741-755, abr. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.ijforecast.2018.01.003>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169207018300116?via%3Dihub>. Acesso em: 21 ago. 2022

BERRAR, Daniel; LOPES, Philippe; DUBITZKY, Werner. Incorporating domain knowledge in machine learning for soccer outcome prediction. **Machine Learning**, [S.L.], v. 108, n. 1, p. 97-126, 7 ago. 2018. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10994-018-5747-8>. Disponível em: <https://link.springer.com/article/10.1007/s10994-018-5747-8>. Acesso em: 30 set. 2022.

BUNKER, Rory P.; THABTAH, Fadi. A machine learning framework for sport result prediction. **Applied Computing And Informatics**, [S.L.], v. 15, n. 1, p. 27-33, jan. 2019. Emerald. <http://dx.doi.org/10.1016/j.aci.2017.09.005>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2210832717301485>. Acesso em: 13 ago. 2022.

BUTLER, David; BUTLER, Robert; EAKINS, John. Expert performance and crowd wisdom: Evidence from English Premier League predictions. **European Journal Of Operational Research**, [S.L.], v. 288, n. 1, p. 170-182, jan. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.ejor.2020.05.034>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S037722172030480X?via%3Dihub>. Acesso em: 03 set. 2022.

CASIMIRO, María Pilar González. **Análisis de series temporales: Modelos ARIMA**. Bizkaia: Universidad del País Vasco, 2009. 165 p. Disponível em: <https://addi.ehu.es/handle/10810/12492>. Acesso em: 08 out. 2022.

HAYKIN, Simon. **Neural networks: a comprehensive foundation**. 3rd. ed. Upper Saddle River: Prentice Hall, 2009. 938p. Disponível em: <https://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf>. Acesso em: 22 set. 2024.

KOOPMAN, Siem Jan; LIT, Rutger. Forecasting football match results in national league competitions using score-driven time series models. **International Journal Of Forecasting**, [S.L.], v. 35, n. 2, p. 797-809, abr. 2019. Elsevier BV.

<http://dx.doi.org/10.1016/j.ijforecast.2018.10.011>. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0169207018302048?via%3Dihub>. Acesso em: 22 ago. 2022.

LAHVICKA, Jiri. Using Monte Carlo simulation to calculate match importance: the case of English Premier League. **Munich Personal Repec Archive**, Munich, p. 1-20, 01 set. 2012. Disponível em: <https://mpra.ub.uni-muenchen.de/40998/>. Acesso em: 21 ago. 2022.

LASEK, Jan; GAGOLEWSKI, Marek. Interpretable sports team rating models based on the gradient descent algorithm. **International Journal Of Forecasting**, [S.L.], v. 37, n. 3, p. 1061-1071, jul. 2021. Elsevier BV.

<http://dx.doi.org/10.1016/j.ijforecast.2020.11.008>. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0169207020301849>. Acesso em: 10 set. 2022.

MEN, Yanhua. Intelligent sports prediction analysis system based on improved Gaussian fuzzy algorithm. **Alexandria Engineering Journal**, [S.L.], v. 61, n. 7, p. 5351-5359, jul. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.aej.2021.08.084>.

Disponível em:

<https://www.sciencedirect.com/science/article/pii/S1110016821006001>. Acesso em: 26 ago. 2022.

PEREIRA, Roberto Augusto Lazzarotto. **Modelagem matemática para previsão esportiva: uma aplicação no futebol nacional**. 2018. 46 f. TCC (Graduação) - Curso de Licenciatura em Ciências Exatas Com Habilitação em Matemática, Universidade Federal do Paraná, Pontal do Paraná, 2018. Disponível em:

<https://acervodigital.ufpr.br/handle/1884/60661>. Acesso em: 17 set. 2022.

SANTOS, Tatiana Fernanda Mousquer dos. **Aplicação de séries temporais e redes neurais em um ambiente de computação em nuvem**. 2014. 92 f.

Dissertação (Mestrado) - Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção, Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria, 2014. Disponível em: <https://repositorio.ufsm.br/handle/1/8316>. Acesso em: 04 nov. 2022.

XAVIER, Thainá Santos. **Previsão de séries temporais utilizando modelos clássicos e redes neurais artificiais**. 2018. 46 f. TCC (Graduação) - Curso de

Bacharel em Ciências no Domínio da Engenharia Elétrica, Centro de Engenharia Elétrica e Informática – Ceei, Universidade Federal de Campina Grande – UFCG, Campina Grande, 2018. Disponível em:

<http://dspace.sti.ufcg.edu.br:8080/xmlui/handle/riufcg/18949>. Acesso em: 08 out. 2022.

YIANNAKIS, Andrew *et al.* Forecasting in Sport. **International Review For The Sociology Of Sport**, [S.L.], v. 41, n. 1, p. 89-115, mar. 2006. SAGE Publications.

<http://dx.doi.org/10.1177/1012690206063508>. Disponível em:
<https://journals.sagepub.com/doi/abs/10.1177/1012690206063508>. Acesso em: 09 out. 2022.

ZUCCO, Luiz Henrique. **Implementação da previsão de demanda por meio de modelos matemáticos clássicos e de inteligência artificial na gestão de estoque de uma empresa importadora de componentes automotivos da linha pesada**. 2019. 55 f. TCC (Graduação) - Curso de Engenharia de Produção, Universidade de Caxias do Sul, Caxias do Sul, 2019.

Biografia dos Autores

Vinícius Doro

Bacharel em Engenharia de Produção pela Universidade de Caxias do Sul (UCS). Atualmente compõe o departamento de Engenharia da Caderode, atuando no desenvolvimento de novos produtos no ramo mobiliário corporativo, bem como na melhoria contínua de processos fabris, com foco na redução de tempos e custos. Possui experiência em processos de otimização e em gerenciamento de projetos.

Leandro Luís Corso

Possui Pós-doutorado na Monash University/AUSTRÁLIA na área de Otimização, pós-doutorado na Naval Postgraduate School, California/EUA em Otimização Global considerando incertezas. Mestrado e doutorado em Engenharia com foco em otimização pela Universidade Federal do Rio Grande do Sul (UFRGS). Graduação em Engenharia Mecânica pela Universidade de Caxias do Sul (UCS). Pesquisador/professor de áreas de Otimização em graduação e pós-graduação, professor do Mestrado de Engenharia Mecânica e no Doutorado da Área de Ciências da (UCS). Membro do Comitê Científico do Estado do Rio Grande do Sul, recebendo menção honrosa na atuação. Revisor de periódicos nacionais/internacionais nas áreas de otimização. Esta presente no ranking Alper-Döğler Scientific Index como um dos melhores cientistas da América Latina. Possui experiência nas áreas de Otimização, Modelagem Matemática, Inteligência Artificial, Otimização Estrutural, Otimização considerando incertezas, Pesquisa Operacional, matemática aplicada e computacional.



Artigo recebido em: 31/07/2023 e aceito para publicação em: 12/05/2025

DOI: <https://doi.org/10.14488/1676-1901.v25i3.4965>