

Aplicação da Lógica Fuzzy na Previsão do Campeonato Brasileiro de Futebol de 2024

1° Marcelo Andrade de Jesus Filho
Engenharia de Sistemas
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
marcelo.anje@outlook.com

2° Vinicius Cardoso Antunes
Engenharia de Sistemas
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
vinicius.mct17@gmail.com

Abstract—Esse artigo apresenta uma abordagem matemática para a previsão dos resultados do Campeonato Brasileiro de Futebol de 2024 utilizando lógica fuzzy. A lógica fuzzy é uma técnica que lida com a incerteza e imprevisão, aplicando modelos e prevendo resultados dos confrontos entre times a partir de diferentes variáveis. O estudo envolve a coleta e a análise de dados de partidas dos anos anteriores disponíveis na internet, fornecendo dados sobre condições dos times, atributos dos jogadores e informações sobre cada partida. Através da definição de funções de pertinência e regras, é criado um sistema de inferência fuzzy, sendo possível gerar um resultado de cada partida, e por fim, o resultado do campeonato. Os resultados obtidos demonstram a eficácia da lógica fuzzy na previsão de cenários, sendo uma ferramenta útil para usos científicos e comerciais.

Index Terms—Campeonato, Futebol, Fuzzificação, Modelagem, Previsão, Sistema, Esporte, Lógica

I. INTRODUÇÃO

Desde os primórdios da civilização, os seres humanos possuem um desejo intrínseco de competir, característica impulsionada por uma combinação de diversos fatores. Biologicamente, a competição pode ser atrelada à busca por sobrevivência e à reprodução, pois os indivíduos competiam por recursos escassos e por parceiros sexuais. Psicologicamente, competir é uma maneira de obter reconhecimentos dos outros indivíduos do grupo e ser aceito socialmente, além de validar habilidades. Por fim, a competição pode ser considerada uma característica que molda o comportamentos dos indivíduos e influencia as estruturas sociais e econômicas da nossa sociedade.

Com o passar dos anos, as apostas esportivas crescem cada vez mais no Brasil, com casas de apostas presentes como patrocinadoras de praticamente todos as equipes da série A do Campeonato Brasileiro de 2024. Porém, as apostas não são um fenômeno recente, existindo registros que remontam à antiguidade, quando indivíduos de civilizações greco-romanas apostavam em resultados dos Jogos Olímpicos e corridas de bigas. Ao longo dos séculos houveram mudanças na maneira que as apostas eram feitas, sendo incorporadas em diferentes esportes e eventos. No século XIX, o Reino Unido ficou conhecido por estabelecer uma base de como as apostas são feitas nos dias atuais, onde haviam casas que geravam os valores de retorno para os diferentes tipos de apostas disponíveis. Atualmente, o crescimento do uso da internet possibilitou maior facilidade para que usuários pudessem realizar os seus

palpites sem precisar sair de casa.

Com o grande crescimento do poder computacional e avanços tecnológicos, foram desenvolvidos diversos algoritmos que tentam prever os resultados de partidas e campeonatos. Ferramenta de machine learning e análise de grandes bases de dados, ou big data, permitem criar modelos com alta precisão. Uma destas diversas ferramentas é a lógica fuzzy, particularmente eficaz em modelar casos complexos e subjetivos, diferentes da lógica binária. Este artigo explora o uso dessa ferramenta na tentativa de prever os resultados do Campeonato Brasileiro de 2024, que com o tempo, pode provar ou não a eficácia do modelo.

II. REVISÃO DA LITERATURA

A. Sistemas Difusos

Os sistemas difusos, também conhecidos como sistemas nebulosos ou sistemas fuzzy, são sistemas de controle que possuem como objetivo criar modelos matemáticos baseados em informações imprecisas ou subjetivas, intrínsecas aos seres humanos. O conceito de fuzzy surgiu a partir da Teoria de Conjuntos Fuzzy, desenvolvida em 1965 pelo matemático estadunidense Lotfi Asker Zadeh.

B. Lógica Fuzzy

Alternativamente à lógica Booleana, em que os valores de variáveis são somente 0 ou 1, em um sistema difuso as variáveis podem ter valores intermediários entre 0 e 1, que representam graus de pertinência. Na lógica Fuzzy, as proposições podem ser parcialmente verdadeiras ou parcialmente falsas, refletindo a natureza ambígua dos fenômenos do mundo real. Atualmente, a lógica Fuzzy possui diversas aplicações, sendo utilizada em sistemas de controle, sistemas de predição, sistemas de reconhecimento de padrões e sistemas de tomadas de decisão.

III. OBJETIVOS

Esse artigo tem como principais objetivos desenvolver um modelo de previsão baseado em lógica fuzzy para prever os resultados do Campeonato Brasileiro de 2024 utilizando dados disponibilizados na internet, analisar a eficácia desse modelo comparando com os últimos campeonatos, identificar as variáveis mais interessantes de serem utilizadas, validar o

modelo através de testes, alterando variáveis, comparando os resultados e explorando diferentes cenários.

A modelagem desse problema pode ser dividida em quatro partes, sendo o tratamento dos dados utilizados através da busca por variáveis mais relevantes e combinação de dados de diferentes anos através de ferramentas estatísticas, criação das funções de pertinência a partir dos dados gerados, criação das regras com base nas variáveis e funções de pertinência e simulação e computação dos resultados. Por fim, com a finalização do projeto, é possível que a forma de modelar o sistema de previsão de resultados possa ser expandido para diversas outras áreas, como a medicina, engenharia, biologia ou mesmo em outros esportes.

IV. METODOLOGIA

Este trabalho tem como objetivo criar um sistema de previsões para um campeonato. Para isso, será necessário criar uma maneira de analisar os dados disponíveis e criar a estrutura do sistema fuzzy necessária para que os resultados sejam obtidos de maneira correta.

A. Fonte de Dados

Como fonte de dados, foi utilizado um banco de dados disponibilizado na internet com o nome 'Base dos Dados' em um arquivo de formato CSV (Comma-separated Values), que contém as seguintes colunas:

- Ano
- Data
- Rodada
- Estádio
- Árbitro
- Público
- Público Máximo
- Time Mandante
- Time Visitante
- Técnico Mandante
- Técnico Visitante
- Colocação Mandante
- Colocação Visitante
- Valor Equipe Titular Mandante
- Valor Equipe Titular Visitante
- Idade Média Titular Mandante
- Idade Média Titular Visitante
- Gols Mandante
- gols Visitante
- Gols 1º Tempo Mandante
- Gols 1º Tempo Visitante
- Escanteios Mandante
- Escanteios Visitante
- Faltas Mandante
- Faltas Visitante
- Chutes Bola Parada Mandante
- Chutes Bola Parada Visitante
- Defesas Mandante
- Defesas Visitante

- Impedimentos Mandante
- Impedimentos Visitante
- Chutes Mandante
- Chutes Visitante
- Chutes Fora Mandante
- Chutes Fora Visitante

B. Variáveis Utilizadas

A partir desses dados, as variáveis de entrada definidas serão as seguintes:

- Valor da Equipe Mandante
- Valor da Equipe Visitante
- Idade Média da Equipe Mandante
- Idade Média da Equipe Visitante
- Média de Gols a Favor da Equipe Mandante
- Média de Gols Contra da Equipe Visitante
- Média de Gols Contra da Equipe Mandante
- Média de Gols a Favor da Equipe Visitante
- Percentual de Vitórias da Equipe Mandante
- Percentual de Derrotas da Equipe Mandante
- Percentual de Vitórias da Equipe Visitante
- Percentual de Derrotas da Equipe Visitante
- Percentual de Vitórias da Equipe Mandante
- Percentual de Vitórias da Equipe Visitante

Como variável de saída temos apenas o número de gols de cada equipe na partida simulada.

C. Funções de Pertinência

Após as definições das variáveis, é necessário criar as funções de pertinência, que serão responsáveis por representar a incerteza na modelagem. Os valores escolhidos foram:

1) Valor da Equipe Mandante:

- Baixo: [0, 30, 70]
- Médio: [40, 100, 110]
- Alto: [90, 150, 220]

2) Valor da Equipe Visitante:

- Baixo: [0, 30, 70]
- Médio: [40, 100, 110]
- Alto: [90, 150, 220]

3) Idade Média do Time Mandante e Visitante:

- Jovem: [20, 20, 23]
- Média: [23, 25, 27]
- Experiente: [26, 29, 31]

4) Média de Gols Marcados e Sofridos pelo Time Mandante:

- Baixa: [0, 0.6, 1.2]
- Média: [1, 1.5, 1.8]
- Alta: [1.5, 1.85, 2.2]

5) Média de Gols Marcados e Sofridos pelo Time Visitante:

- Baixa: [0, 0.4, 1]
- Média: [0.85, 1, 1.1]
- Alta: [1, 1.2, 1.4]

6) Percentual de Vitórias e Derrotas do Time Mandante:

- Baixo: [0, 32, 46]
- Médio: [35, 50, 60]
- Alto: [52, 60, 75]

7) Percentual de Vitórias e Derrotas do Time Visitante:

- Baixo: [0, 15, 20]
- Médio: [18, 25, 30]
- Alto: [28, 34, 40]

8) Percentual de Vitórias por Colocação do Time Mandante:

- Baixo: [0, 15, 40]
- Médio: [38, 50, 65]
- Alto: [60, 85, 100]

9) Percentual de Vitórias por Colocação do Time Visitante:

- Baixo: [0, 15, 24]
- Médio: [17, 24, 32]
- Alto: [30, 69, 75]

10) Gols Marcados no Jogo:

- Poucos: [0, 0, 2]
- Moderados: [0, 2, 4]
- Muitos: [3, 4, 7]

D. Regras Definidas

As regras foram definidas de modo a estimar a quantidade de gols em uma partida de futebol com base nas várias variáveis de entrada apresentadas, como o valor da equipe mandante e visitante, a idade média dos times mandante e visitante, a média de gols marcados e sofridos pelos times, o percentual de vitórias e derrotas dos times mandante e visitante, e o percentual de vitórias por colocação dos times mandante e visitante.

Cada regra fuzzy é definida com uma série de condicionais e consequências. Exemplificando, a regra 1 estabelece que se o valor da equipe mandante for alto e o valor da equipe visitante também for alto, a quantidade de gols será considerada baixa, pois são equipes similares em um aspecto. Assim, as regras abrangem diversas combinações possíveis entre as variáveis de entrada para prever a quantidade de gols esperada na partida.

E. Simulação

Por fim, criadas as condições para a simulação do sistema, podemos realizar a execução, que trará como resultado a quantidade de gols de cada equipe em cada partida. Ao todo serão 380 partidas divididas em 38 rodadas, que terão como possíveis resultados a vitória do time mandante ou visitante, resultando em 3 pontos para o time vencedor e 0 pontos para o time perdedor ou um empate, que resultará em 1 ponto para cada equipe. No final do campeonato, a classificação é dada pelo número de pontos, havendo como critério de desempate o número de vitórias e o saldo de gols.

V. RESULTADOS

A. Resultados dos Campeonatos Anteriores

Antes de analisar os resultados, podemos ter como referência os 4 primeiros colocados do Campeonato Brasileiro nos últimos 5 anos, presentes na tabela abaixo:

TABLE I
COLOCAÇÕES DOS 4 PRIMEIROS TIMES NOS ÚLTIMOS 5 CAMPEONATOS

Tabela Pos.	Ano			
	2023	2022	2021	2020
1°	Palmeiras	Palmeiras	Atlético-MG	Flamengo
2°	Grêmio	Internacional	Flamengo	Internacional
3°	Atlético-MG	Fluminense	Palmeiras	Atlético-MG
4°	Flamengo	Corinthians	Fortaleza	São Paulo

A partir desses dados, é possível observar que times como Palmeiras, Atlético-MG e Flamengo estão presentes frequentemente nas melhores colocações nos últimos anos, de modo a se esperar que a simulação tenda a continuar classificando estes clubes nas melhores colocações.

B. Resultado da Simulação

Classificação final para o Campeonato Brasileiro de 2024:

TABLE II
RESULTADO DA SIMULAÇÃO (PREVISÃO)

Posição	Clube	Pontos
1°	Flamengo	110
2°	Palmeiras	100
3°	Atlético-MG	88
4°	Corinthians	88
5°	Internacional	78
6°	Athletico-PR	73
7°	Fluminense	68
8°	São Paulo	67
9°	RB Bragantino	56
10°	Cruzeiro	55
11°	Grêmio	52
12°	Botafogo	34
13°	Cuiabá-MT	32
14°	Atlético-GO	29
15°	EC Bahia	26
16°	Vasco da Gama	26
17°	Fortaleza	23
18°	Criciúma EC	15
19°	EC Vitória	9
20°	Juventude	7

Como citado anteriormente, a presença de clubes como Palmeiras, Atlético-MG e Flamengo era esperada pois eram clubes que estão com resultados consistentes nos últimos anos.

C. Possíveis Problemas

Devido ao fato do modelo não ser perfeito e ao futebol brasileiro ser muito imprevisível, é possível que os resultados sejam muito divergentes dos resultados reais. O modelo não considera fatores como troca de técnicos, problemas financeiros das equipes, lesões de jogadores ou desastres naturais, tornando-o impreciso por natureza.

Outra consideração a ser feita é a de que em todo campeonato, os 4 últimos clubes na classificação são rebaixados para a divisão abaixo, tornando os dados dessas equipes inexistentes em alguns anos. Uma solução seria ter os dados de todas as partidas de todas as divisões do Campeonato Brasileiro, porém essa documentação é escassa devido ao fato das divisões menores terem menos relevância para o público em geral.

VI. CONCLUSÃO

Este trabalho explorou a aplicação da lógica fuzzy em um sistema de previsão dos resultados do Campeonato Brasileiro de 2024, mostrando ser um modelo eficaz de previsão e provando a aplicação da matemática para a modelagem de sistemas que envolvem a incerteza e subjetividade presentes nas ações humanas.

A partir da análise de dados coletados, foi possível gerar variáveis, funções de pertinência, regras e uma simulação que resultou na previsão da classificação dos clubes da série A do campeonato.

Os resultados obtidos foram os esperados, mostrando similaridade com os dos últimos anos e próximo do que se espera do ano em curso, contudo, é importante reconhecer que o modelo possui imprecisões e pode estar distante da realidade que pode haver no final do campeonato.

Por fim, este estudo teve grande contribuição para o entendimento do funcionamento dos sistemas difusos, assim como para as aplicações desse sistema em problemas atuais.

Código fonte disponível em [4]

REFERENCES

- [1] Trawiński, Krzysztof, A Fuzzy Classification System for Prediction of the Results of the Basketball Games, July 2010.
- [2] Vaidehi .V, Monica .S , Mohamed Sheik Safeer .S, Deepika .M, Sangeetha .S, A Prediction System Based on Fuzzy Logic, October 2008.
- [3] Omomule T. G., Ibinuolapo A. J., Ajayi O. O., Fuzzy-Based Model for Predicting Football Match Results, February 2020.
- [4] V. C. et al., "fuzzy-championship-predictor", 2024. código fonte disponível no GitHub. [Online]. Disponível em: <https://github.com/vinicius-cardoso/fuzzy-championship-predictor>