

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
CENTRO DE PÓS-GRADUAÇÃO E PESQUISAS EM ADMINISTRAÇÃO

FELIPE ALEXANDRE DE SOUZA FÉLIX NUNES

**MODELOS PREDITORES DE CONTRATAÇÃO DE PROFISSIONAIS DO  
FUTEBOL BASEADO EM INTELIGÊNCIA ARTIFICIAL POR MECANISMOS  
*MACHINE LEARNING***

Belo Horizonte

2025

Felipe Alexandre de Souza Félix Nunes

**Modelos preditores de contratação de profissionais do futebol baseado em Inteligência Artificial por mecanismos *Machine Learning***

Tese apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Administração

Área de concentração: Estratégia, Mercadologia e Operações

Orientador: Prof. Dr. Carlos Alberto Gonçalves

Co-orientador: Prof. Dr. Jonathan Simões Freitas

Belo Horizonte

2025

Ficha catalográfica

N972m  
2025  
Nunes, Felipe Alexandre de Souza Félix.  
Modelos preditores de contratação de profissionais do futebol baseado em inteligência artificial por mecanismos Machine Learning [manuscrito] / Felipe Alexandre de Souza Félix Nunes – 2025.  
1v.: il.

Orientador: Carlos Alberto Gonçalves.  
Coorientador: Jonathan Simões Freitas.

Tese (doutorado) – Universidade Federal de Minas Gerais, Centro de Pós-Graduação e Pesquisas em Administração.  
Inclui bibliografia.

1. Inteligência artificial – Teses. 2. Futebol – Teses. 3. Administração – Teses. I. Gonçalves, Carlos Alberto II. Freitas, Jonathan Simões. III. Universidade Federal de Minas Gerais. Centro de Pós-Graduação e Pesquisas em Administração. IV. Título.

CDD: 658.4062

Elaborado por Leonardo Vasconcelos Renault CRB-6/2211  
Biblioteca da FACE/UFMG. – /2025



UNIVERSIDADE FEDERAL DE MINAS GERAIS

## ATA

FACE - COLEGIADO DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO - SECRETARIA

ATA DA DEFESA DE TESE EM ADMINISTRAÇÃO do Senhor **FELIPE ALEXANDRE DE SOUZA FÉLIX NUNES** REGISTRO Nº 335/2025. No dia 17 de Julho 2025, às 14:00 horas, reuniu-se de forma remota e presencial, na Faculdade de Ciências Econômicas da UFMG, a Comissão Examinadora de Tese, indicada pelo Colegiado do Centro de Pós-Graduação e Pesquisas em Administração do CEPEAD, em 25 de junho 2025, para julgar o trabalho final intitulado "Modelos preditores de contratação de profissionais do futebol baseado em Inteligência Artificial por mecanismos Machine Learning". Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Dr. Carlos Alberto Gonçalves, após dar conhecimento aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do seguinte resultado final:

APROVAÇÃO

REPROVAÇÃO

O resultado final foi comunicado publicamente ao candidato pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Senhor Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 17 de julho de 2025.

Prof. Dr. Carlos Alberto Gonçalves - Orientador/UFMG  
Prof. Dr. Thadeu Miranda Gasparetto – West Virginia University  
Prof. Dr. Wilson Toshiro Nakamura – PPGA Mackenzie  
Prof. Dr. Fernando de Andrade Franco Malagrino – Trevisan  
Prof. Dr. Pedro Olmo Stancioli Vaz de Melo - UFMG



Documento assinado eletronicamente por **Thadeu Miranda Gasparetto, Usuário Externo**, em 27/08/2025, às 11:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fernando de Andrade Franco Malagrino, Usuário Externo**, em 28/08/2025, às 16:17, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Pedro Olmo Stancioli Vaz de Melo, Professor do Magistério Superior**, em 01/09/2025, às 08:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Alberto Goncalves, Membro de comissão**, em 01/09/2025, às 12:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Wilson Toshiro Nakamura, Usuário Externo**, em 04/09/2025, às 17:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4497362** e o código CRC **6F75AF91**.

---

Aos meus pais, minha irmã e meus amores  
Dani, João Miguel e Miguel Alexandre

## AGRADECIMENTO

Agradeço a Deus pelo dom da vida e por me manter forte nos momentos mais duros.

Ao entrar na sala virtual da arguição oral do projeto proposto no processo seletivo do programa de pós-graduação em nível de doutorado no CEPEAD ouvi de um dos avaliadores: “Olha lá o rapaz do futebol!”. Não era a primeira vez e continuou um comentário constante quando ingressava em alguma disciplina da linha de Estratégia, Mercadologia e Operações e encontrava professores conhecidos.

Minha primeira grande memória atrelada ao futebol remonta à Copa do Mundo de 1994, eu tinha meus cinco anos de idade e guardei com carinho os festejos que foram aqueles sete jogos da Seleção Brasileira até o tetracampeonato mundial. Depois, no ano 2000, me lembro de assistir a cobrança de falta de Geovanni atravessar a barreira de defensores do São Paulo e sacramentar a conquista do tricampeonato da Copa do Brasil pelo Cruzeiro. Talvez tenha sido o momento que eu definitivamente abracei a torcida por esse time.

Embora sempre tenha sido apaixonado por acompanhar esse esporte eu quase nunca me dei bem praticando-o. Tentava, é verdade, mas minhas contribuições para o Irreal Madrid, nosso time de amigos, eram muita mais prolíficas na parte organizacional que como atleta amador. Por outro lado, nos jogos de futebol eletrônico eu era mais competitivo, sempre comandando a Juventus Football Club nos campeonatos de *Winning Eleven 8* e *Pro Evolution Soccer 6*.

Naquela época tínhamos como hábito fazer negociações internas, trocando jogadores entre os amigos para montar equipes mais ligadas aos estilos de cada um. Na versão pirateada (Bomba Patch) do *Winning Eleven* tocava a música *Crawling*, do Linkin Park, a cada vez que entrávamos no modo de edição, que permitiam as nossas negociações. Talvez tenha sido entre aqueles momentos, negociando a troca de Pavel Nedved por Eric Djemba-Djemba e Ruud Van Nistelrooy, que eu tenha desenvolvido o gosto por compreender as etapas de negociações e montagens de elencos no futebol.

É, portanto, por aí que começo meus agradecimentos formais aos amigos Bruno Marques, Gabriel Lobo, Joel Maia, Johnson Cunha, Marcelo Mourão e Pedro D’Angelo que participaram daquela etapa comigo e pela amizade que perpasa duas décadas. Ainda no campo das amizades o meu muito obrigado aos irmãos da República Giramundo, em especial ao querido Antônio Clemente, o Lumbriga, pelo suporte fundamental nesta caminhada.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concedida a mim, que proporcionou condições para que eu pudesse desenvolver o projeto. Muito obrigado à Armatore, lugar em que pude desenvolver algumas pesquisas. Preciso reconhecer a importância de alguns professores que tiveram participação na minha trajetória. Da graduação, quando alguns deles observaram que o aluno disperso poderia se interessar pelo campo das pesquisas, em especial Juliana Teixeira, Euclides Couto, Jorge Nei Brito e Luciana Walther. Muito obrigado, vocês foram fundamentais.

Muito obrigado ao CEPEAD/UFMG, aos professores da instituição, secretários e demais funcionários que me proporcionaram a oportunidade de cursar o doutorado. Obrigado aos meus colegas de doutoramento, pelas trocas de ideias, correções, conselhos e indicações. Não são poucos os agradecimentos devidos ao meu orientador, o professor Carlos Alberto Gonçalves, pelo diálogo, pelas contribuições, pelos debates, ideias e cobranças. Este trabalho não seria o mesmo sem a participação do senhor. Agradeço ao meu coorientador, professor Jonathan Simões Freitas, que me acompanha desde o mestrado, por toda a paciência, dedicação e confiança reservados à minha pessoa.

O final desta seção é direcionado às pessoas mais importantes da minha vida. Meus pais, Rubens e Delza, que sempre estiveram ao meu lado e nunca mediram esforços para que eu alcançasse essa conquista. Vocês são meus dois maiores exemplos de luta, perseverança, fé e amor. Minha irmã, Rafaela, por ser o suporte essencial na minha caminhada, sendo exemplo de persistência e obstinação. Dedico aos meus avós, primos, tios, padrinhos e madrinhas. À minha esposa Danielle, meu ponto de apoio, eu primeiro peço desculpas pelas noites trocadas, por certa ausência nos jantares e por ter parado de te acompanhar nos filmes e séries, sem todo o suporte que você me deu eu jamais teria conseguido. Obrigado pelo seu amor, minha eterna namorada. Ao meu filho João Miguel, que ainda é muito novo para compreender por que o papai não conseguia brincar em alguns momentos, você iluminou os meus dias para alcançar esse momento.

*“Football is a game about feelings and intelligence.”*

(José Mourinho)

## RESUMO

Esta tese investiga as dinâmicas estratégicas envolvidas no processo de recrutamento de jogadores de futebol, com foco na construção de um modelo interpretável de apoio à decisão baseado em algoritmos de aprendizado de máquina, lógica *fuzzy* e análise de redes sociais. A partir de uma abordagem interdisciplinar que articula estudos organizacionais, ciência do esporte e modelagem matemática, a pesquisa propõe uma solução metodológica para avaliar o desempenho técnico de atletas, considerando as especificidades contextuais do futebol profissional contemporâneo. O estudo estrutura-se em três eixos analíticos. No primeiro, uma revisão bibliométrica revelou a emergência de uma literatura convergente entre performance esportiva, modelos *fuzzy* e inteligência artificial. No segundo eixo, aplicou-se a análise de redes sociais a mais de 100 mil transações internacionais de jogadores, identificando padrões de centralidade, intermediação e formação de comunidades no mercado global de transferências, com destaque para o papel estratégico de países como Bélgica, Portugal e Brasil. No terceiro eixo, desenvolveu-se e testou-se um modelo de avaliação de performance baseado em lógica *fuzzy*, alimentado por estatísticas extraídas da plataforma *Wyscout* e validadas com atributos do *Football Manager 2024*. O modelo demonstrou robustez na comparação funcional de atletas, respeitando o princípio do *design* representativo e integrando os fundamentos da abordagem ecológica da performance esportiva. Os resultados apontam para a utilidade prática da ferramenta no suporte à tomada de decisão, especialmente na prospecção de talentos e na gestão eficiente de recursos. A tese propõe uma inovação teórica e aplicada, ao oferecer um sistema de pontuação adaptável, que confronta a subjetividade do *scouting* tradicional com critérios analíticos, promovendo maior racionalidade estratégica na composição de elencos. As contribuições do trabalho se estendem também ao campo da pesquisa organizacional, ao demonstrar como capacidades analíticas podem se tornar ativos estratégicos na busca por vantagem competitiva sustentável no futebol global.

**Palavras-chave:** Análise de redes sociais; Futebol; *Machine learning*; *Fuzzy*; Estratégia.

## ABSTRACT

This thesis investigates the strategic dynamics involved in the recruitment process of professional football players, with a focus on developing an interpretable decision-support model based on machine learning algorithms, fuzzy logic, and social network analysis. Drawing on an interdisciplinary approach that integrates organisational studies, sports science, and mathematical modelling, the research proposes a methodological solution for evaluating athletes' technical performance while considering the contextual specificities of contemporary professional football. The study is structured around three analytical axes. The first comprises a bibliometric review, which revealed the emergence of a converging body of literature between sports performance, fuzzy models, and artificial intelligence. The second axis applies social network analysis to over 100,000 international player transfers, identifying patterns of centrality, intermediation, and community formation within the global transfer market—highlighting the strategic roles of countries such as Belgium, Portugal, and Brazil. The third axis presents the development and testing of a fuzzy logic-based performance evaluation model, fuelled by statistics from the Wyscout platform and validated through attributes from Football Manager 2024. The model demonstrated robustness in functional comparisons of players, adhering to the principle of representative design and integrating the tenets of the ecological approach to sports performance. The results indicate the practical usefulness of the tool in supporting decision-making, particularly in talent scouting and efficient resource management. The thesis offers both a theoretical and applied innovation by introducing a flexible scoring system that confronts the subjectivity of traditional scouting methods with analytical criteria, thereby promoting greater strategic rationality in squad composition. Furthermore, the contributions of this work extend to the field of organisational research, by demonstrating how analytical capabilities can become strategic assets in the pursuit of sustainable competitive advantage in global football.

**Keywords:** Social network analysis; Football; Machine learning; Fuzzy; Strategy.

## LISTA DE FIGURAS

Figura 1 - Evolução do número de publicações .....	35
Figura 2 - Agrupamento por rede de co-ocorrência .....	42
Figura 3 - Evolução do número de trabalhos publicados .....	43
Figura 4 - <i>Three field plot</i> entre termos, autores e periódicos.....	46
Figura 5 - Agrupamento bibliográfico.....	48
Figura 6 - Rede de co-ocorrência .....	49
Figura 7 - Mapa temático.....	50
Figura 8 - Evolução dos trabalhos sobre o tema.....	53
Figura 9 - Rede de co-ocorrência .....	55
Figura 10 - Mapa temático.....	56
Figura 11 - Framework estratégia (KHALIFA, 2008) .....	62
Figura 12 - Rede global das transferências em definitivo (2010–2023).....	69
Figura 13 - Tríades centrais com Inglaterra como destino .....	74
Figura 14- Tríades centrais com França como destino.....	75
Figura 15 - Tríades centrais com Itália como destino .....	75
Figura 16 - Tríades centrais com Alemanha como destino .....	76
Figura 17 - Tríades centrais com Espanha como destino .....	77
Figura 18 - Atributos do jogo <i>Football Manager</i> .....	115
Figura 19 - Importância dos fatores para variável faltas sofridas.....	146
Figura 20 - Importância das estatísticas na variável gol.....	147
Figura 21 - Importância das estatísticas na variável minutos jogados .....	148
Figura 22 - Importância das estatísticas na variável interceptações.....	149
Figura 23 - Importância das estatísticas na variável $x_A$ .....	150
Figura 24 - Importância das estatísticas na variável $x_G$ .....	151
Figura 25 - Importância das estatísticas na variável assistências .....	152
Figura 26 - Importância das estatísticas na variável faltas cometidas.....	153
Figura 27 - Pontuação do escore de silhueta .....	155
Figura 28 - Comparação entre Arrascaeta e Raphael Veiga.....	162
Figura 29 - Comparação entre Arrascaeta e Raphael Veiga.....	163
Figura 30 - Evolução dos indicadores para as estatísticas de Jean Lucas .....	164

## LISTA DE QUADROS

Quadro 1 - Análise dos metadados .....	34
Quadro 2 - Principais fontes da amostra.....	35
Quadro 3 - Fatores de impacto dos periódicos .....	36
Quadro 4 - Artigos mais citados .....	40
Quadro 5 - Afiliação dos autores .....	41
Quadro 6 - Fontes mais relevantes .....	44
Quadro 7 - Autores mais proeminentes sobre o tema.....	45
Quadro 8 - Países com maior número de autores filiados .....	45
Quadro 9 - Trabalhos mais citados .....	48
Quadro 10 - Contagem de palavras mais citadas nos resumos.....	51
Quadro 11 - Análise de tópicos .....	51
Quadro 12 - Dez artigos mais citados.....	55
Quadro 13 - Transferências mais caras da história corrigidas pela inflação .....	59
Quadro 14 - Indicadores estatísticos oriundos do <i>Wyscout</i> .....	95
Quadro 15 - Resultados obtidos nas regressões.....	116
Quadro 16 - Estatísticas utilizadas na <i>fuzzificação</i> .....	118
Quadro 17 - Resultado para valores de transferência .....	120
Quadro 18 - Resultados para valores de transferência positivo.....	121
Quadro 19 - Resultados para valor de mercado.....	123
Quadro 20 - Resultados para variáveis financeiras .....	126
Quadro 21 - Resultado para variáveis de performance.....	127
Quadro 22 - Comparação resultados para variável valor de transferência .....	129
Quadro 23 - Comparação de resultados para a variável valor de mercado .....	131
Quadro 24 - Comparação de resultados para a variável valor de transferência positivo.....	133
Quadro 25 - Fatores estatísticos e variáveis de interesse para modelo final .....	137
Quadro 26 - Compilação de resultados para variável de performance.....	144
Quadro 27 - Agrupamento de estatísticas a partir de funções em campo.....	160

## LISTA DE EQUAÇÕES

Equação 1 - Especificação alternativa de calibração <i>fuzzy</i> .....	101
Equação 2 - Função de pertinência da <i>Integrated Fuzzy and Relative</i> .....	101

Equação 3 - Função equivalente da <i>Integrated Fuzzy and Relative</i> .....	102
Equação 4 - Regressão Ridge .....	107
Equação 5 - Regressão LASSO .....	108
Equação 6 - Regressão <i>Elastic Net</i> .....	108

## SUMÁRIO

1.	Introdução.....	16
1.1.	Pergunta de pesquisa .....	21
1.2.	Objetivo geral .....	22
1.3.	Objetivos específicos.....	22
1.4.	Justificativa.....	23
2.	<i>Machine learning</i> aplicado aos esportes: uma revisão bibliométrica.....	27
2.1.	Introdução.....	27
2.2.	A aprendizagem de máquina .....	28
2.3.	O desempenho do atleta e equipes.....	29
2.4.	Definições e origens da aprendizagem de máquina.....	30
2.5.	Método.....	31
2.6.	Análise .....	32
2.7.	Considerações finais .....	56
3.	Fluxos de transferência e estratégia no futebol global .....	58
3.1.	Introdução.....	58
3.2.	Formulação e implementação de estratégia nas organizações.....	60
3.3.	Capacidades Dinâmicas .....	63
3.4.	Metodologia.....	65
3.5.	Resultados.....	66
3.6.	Fluxo de Transferência entre Países – Todas as Transferências .....	66
3.7.	Fluxo de Transferência entre Países (Transferências Definitivas) .....	68
3.8.	Fluxo de transferência entre países – Todas as transferências por intervalos de temporada .....	70
3.9.	Fluxo de transferência entre países – Apenas transferências em definitivo por intervalos de temporada .....	72
3.10.	Conclusão .....	78

4.	Um modelo difuso para pontuar estatísticas de jogadores .....	80
4.1.	Introdução .....	80
4.2.	Arcabouço Teórico: A jornada do atleta.....	84
4.3.	Análise Técnica: a habilidade dos atletas .....	91
4.4.	Metodologia.....	93
4.5.	Modelagem de Incerteza: distinções fundamentais na lógica <i>fuzzy</i> .....	96
4.6.	Transformações <i>Fuzzy</i> Aplicadas .....	100
4.7.	Base de regras e inferência .....	104
4.8.	Regressões lineares penalizadas .....	106
4.9.	Clusterização hierárquica .....	108
4.10.	<i>Fuzzy C-Means</i> .....	108
4.11.	<i>Random Forest</i> .....	109
4.12.	Descrição do procedimento .....	110
4.13.	Análise .....	113
4.14.	Outros modelos <i>fuzzificados</i> testados e resultados obtidos .....	134
4.15.	Análise de agrupamentos – Hierárquico e <i>Fuzzy C-Means</i> .....	153
4.16.	Possíveis aplicações do modelo.....	161
4.17.	Discussão .....	165
5.	Conclusão .....	170
	Referências Bibliográficas .....	177

## 1. Introdução

Em determinado momento o número de investigações científicas realizadas sobre a gestão de clubes de futebol cresceu, da mesma maneira que o esporte passou a ser tratado como um negócio bem mais que apenas recreação (SANTOS; GARCÍA, 2011). Uma área de relativa importância diz respeito aos fatores envolvidos nas negociações que envolvem as transferências dos atletas, uma vez que muitos recursos são alocados para a execução desses negócios entre os clubes e cuja importância é justificada pelo comprometimento à estrutura financeira da organização (PAYYAPPALLI; ZHUANG, 2019).

As transferências são um acordo feito de maneira livre, cujas negociações incluem representantes de, ao menos, dois clubes (proprietários, CEO's e diretores executivos) além de intermediários e os próprios jogadores (POLI *et al.*, 2022). Alguns preços pagos no mercado de transferência de jogadores têm sido criticados como ineficientes, mas isso pode ser explicado pela presença de processos de negociação de preços (MARTÍN *et al.*, 2022). Os preços excessivos podem ser causados por rendas monopolísticas dos jogadores, riscos do comprador e condições comerciais intensas. A dificuldade de recuperar as taxas de transferência superestimadas aumenta a possibilidade de insolvência do clube (MARTÍN *et al.*, 2022). Desta maneira é preciso diminuir as margens de erro e aproximar-se da excelência, evitando que os custos da aquisição de novos atletas conduzam à resultados desportivos ruins (SZYMANSKI, 2017).

Dentro desse contexto, compreender quais habilidades são necessárias em um plantel que compreende, em média, 30 jogadores é uma linha de ação fundamental. Nesta perspectiva tem-se definido que para um jogador atuar em alto desempenho não lhe basta apenas nascer com talento, mas uma intensa carga de estímulos e treinos, uma vez que diversos fatores vão impactar esse desenvolvimento (GARGANTA, 2004; MACDONALD, 2011). Diante disso, o processo de recrutamento e seleção de atletas assume papel estratégico e requer mais do que percepção intuitiva. Exige mapeamento objetivo das habilidades relevantes para o modelo de jogo da equipe, além de um alinhamento com a cultura organizacional e com a filosofia de longo prazo do clube. Como demonstrado por Costa *et al.* (2009), os princípios táticos e operacionais do futebol variam conforme o sistema adotado, influenciando diretamente os requisitos técnicos e cognitivos esperados dos atletas.

Cada equipe deve constituir o seu modelo de jogo e determinar uma maneira de jogar, tendo por orientação os princípios táticos, a partir das escolhas do treinador (CASARIN *et al.*, 2011). Por exemplo, um dos treinadores com maior número de títulos desde 2009, Josep Guardiola, é conhecido por adotar um modelo de jogo de conservação da posse de bola e recuperação sob pressão a partir da perda da posse (PERARNAU, 2015). Portanto, as habilidades necessárias para o recrutamento de jogadores para formar um plantel que jogará sob o comando do treinador catalão deve cumprir, por indução lógica, com características que serão aproveitadas em seu modelo de jogo.

Esse cenário estimula um questionamento, o que fazer se, em média, o tempo de um treinador no clube é inferior ao contrato oferecido ao novo atleta? A solução então será estabelecer uma filosofia do próprio clube que oriente, não apenas a seleção de atletas, mas também a escolha pelos treinadores (BORGES; SANTOS, 2021; GUIMARÃES; OLIVEIRA; PAOLI, 2020). Assim, identificar as características condizentes com a cultura do clube para selecionar as habilidades requeridas e desenvolver indivíduos talentosos é uma importante vantagem competitiva do futebol como negócio.

A partir desse entendimento é possível avaliar lances dentro de uma partida de futebol que sejam baseadas nas habilidades dos atletas. Uma partida típica, com duração de 90 minutos, consiste em  $1.682 \pm 101$  eventos englobando passes, duelos (ofensivos e defensivos), finalizações, entre outros (PAPPALARDO *et al.*, 2019). Além dos eventos com a bola é preciso compreender todas as movimentações necessárias aos princípios táticos, de modo que cada um dos 22 atletas participantes da contenda é capaz de tomar mais de 2.500 decisões em um jogo (COSTA *et al.*, 2021).

Essa quantidade de tomadas de decisão envolvidas no jogo é que determinam as capacidades alcançadas por jogadores extraclasse. Muitos fatores – tempo, competição, atitude, riscos – são capazes de influenciar o processo de tomada de decisão e uma combinação desses elementos é o que determinará a escolha (PALAFOX, 2016). Os melhores tomadores de decisão são pessoas cujo processo de racionalidade são baseadas em análise das decisões anteriores e da complexidade do ambiente. Sob incertezas e conflitos a tomada de decisão é determinada com um risco calculado e ocorrem quando há mais de uma opção em aberto para o ganho ótimo (PALAFOX, 2016). Nesse contexto alguns conceitos e cenários podem ser analisados pela Teoria dos Jogos.

A Teoria dos Jogos é o estudo matemático dividido em várias linhas, geralmente é definida como combinatória ou clássica, sendo o primeiro tipo os estudos sobre combinações e permutações de conjuntos e relações matemáticas que os caracterizam (PALAFOX, 2016). A implementação da Teoria dos Jogos na compreensão de processos decisórios no esporte é frequente, principalmente no que tange aos esportes que são jogos sequenciais, em que os jogadores tomam suas decisões estratégicas em uma sequência definida anteriormente (YAMABE; NAKAJIMA, 2013).

No contexto do futebol uma simulação feita na Cornell University é de que atletas fora de série alteraram o paradigma da tomada de decisão. Na simulação em questão foram utilizados dados de dois atletas, um ofensivo (Lionel Messi) e o outro defensivo (Virgil Van Dijk) (CORNELL UNIVERSITY, 2022). Foi observada a existência de um Equilíbrio de Nash assim como um Equilíbrio Misto de Nash. Os resultados apontaram que Messi joga uma estratégia aleatória que tem seus respectivos retornos, embora ele não se beneficie da escolha, enquanto Van Dijk não consegue receber uma recompensa melhor devido à aleatoriedade da escolha de Messi (CORNELL UNIVERSITY, 2022).

Desta maneira a busca por jogadores de característica extraordinárias é um processo crítico para obtenção de vantagem competitiva ao longo de uma competição entre clubes de futebol. E identificar atletas com potencial demanda entendimento extenso das obrigações da performance, além de uma avaliação das capacidades de um atleta em relação a estas exigências (DAY, 2011). Foi com base nessa premissa que esta tese se desenvolveu, propondo uma abordagem integrada com foco na seleção de atletas e no mapeamento estratégico do mercado global de transferências.

O conceito de recorrer aos dados estatísticos históricos para avaliar investimentos em jogadores de futebol mostra-se importante para minimizar riscos financeiros associados aos montantes gastos. Assim, alguns clubes constituíram departamentos de análise de desempenho, ou *scouting*, e metodologias surgiram para identificar, explicar e recomendar os diversos elementos que compõem o desempenho de um jogador de futebol (SANTOS, 2012).

Os avanços tecnológicos alimentaram o desenvolvimento de algoritmos de análise baseados em modelos de aprendizado de máquina (HOLZINGER, 2018). Essa evolução também englobou o campo da gestão esportiva, com cientistas e profissionais da engenharia, administração e educação física investigando seu potencial (ALMULLA; ALAM, 2020). Embora a análise de

dados tenha ganhado espaço no futebol no suporte à tomada de decisão ainda encontra resistência dentro do esporte (RØNNINGEN, 2021), em parte motivada pela resistência cultural à implementação de processos inovadores (RATTEN, 2019). Alguns temem que a tomada de decisões tendo por base a utilização de dados impacte na ausência de jogadores com liberdade criativa que manteriam o esporte interessante e imprevisível (BEIDERBECK *et al.*, 2023).

A imprevisibilidade dos resultados no futebol é relacionada, por vezes, à aleatoriedade envolvida nas partidas (ANDERSON; SALLY, 2013; WUNDERLICH; SECK; MEMMERT, 2021). A aleatoriedade não é um elemento necessariamente ruim na estrutura do jogo, uma vez que provoca incertezas ao resultado, embora as habilidades dos jogadores envolvidos ainda sejam elementos relevantes para o desfecho do jogo (SUMPTER, 2016, 2017). A incerteza do resultado, *per se*, é um dos elementos que atraem o público para acompanhar grandes eventos esportivos, incluindo as partidas de futebol (NEALE, 1964; ROTTENBERG, 1956).

Ainda que as partidas de futebol carreguem uma alta proporção de aleatoriedade em termos de resultados, para os clubes envolvidos o mais importante é o desfecho de uma temporada. Neste caso as consequências das habilidades passam a se sobrepor sobre a imprevisibilidade. Logo, são as habilidades dos jogadores os fatores mais relevantes na formação de um elenco de atletas para os times, de maneira que bilhões de euros são gastos anualmente no processo de recrutamento de novos atletas (JEDELHAUSER; FLEPP; FRANCK, 2023; POLI; BESSON; RAVENEL, 2022).

O recrutamento de atletas também é um elemento que deve ser considerado para a atratividade das partidas para os torcedores. As audiências das partidas são maiores quando envolvem jogadores talentosos e midiáticos, chamados de estrelas (BURAIMO; SIMMONS, 2015), ao passo que os resultados esportivos contam com a contribuição dos demais atletas, mesmo que possam não ser reconhecidos como as estrelas do espetáculo (JEDELHAUSER; FLEPP; FRANCK, 2023). Os processos de recrutamento na composição de um elenco segmentaram o futebol em um esporte cujo poder de investimento impacta nos resultados de uma temporada, de maneira que os clubes com maior somatório dos valores de mercado dos atletas tendem a vencer os campeonatos nacionais (GASPARETTO; BARAJAS, 2022).

Outra hipótese para justificar resistência em realizar contratações de atletas baseadas em dados pode ser relacionada aos observadores cujos anos de experiência na função geram descrédito aos métodos que utilizam análise de dados para deliberar sobre contratações. Muitos

recrutadores usam práticas padrões para avaliar jovens jogadores promissores com base em seu comportamento em relação às demandas futuras previstas (LÜDIN; DONATH; ROMANN, 2023). Os olheiros usam, com alguma frequência, preditores de talentos de forma holística e intuitiva. Isto pode levar a armadilhas psicológicas, como preferências pessoais, intuição, enquadramento, efeitos de dotação, ilusão de confiança, viés de confirmação e o efeito de primazia, que podem contribuir para os baixos índices de precisão na seleção de talentos (LÜDIN; DONATH; ROMANN, 2023).

Este conceito, de atalhos mentais que reduzem os níveis de ansiedade e simplificam o processo decisório, introduz vieses atrelados aos erros de julgamento (YAZDIPOUR; CONSTAND, 2010). Parte dessas heurísticas são usadas em ambientes hipercompetitivos, que provocam os tomadores de decisão e influenciam sua ação de maneira emocional e não racional (MAITLAND; SAMMARTINO, 2015). Há casos em que o gestor influenciado pelo ambiente problemático se utiliza de heurísticas quando as decisões são excessivamente complexas (GABAIX *et al.*, 2006).

Um dos cenários considerados como hipercompetitivos é o futebol. Os sistemas de competição, com equipes relegadas à segunda divisão em caso de maus resultados desportivos, provocam uma atmosfera de concorrência acirrada (SZYMANSKI, 2017). E há uma percepção que grande parte dos casos de insolvência no futebol são movidos por clubes em uma exuberância irracional, se comportando de maneira irresponsável no investimento dedicado a recrutar novos jogadores para a equipe (SZYMANSKI, 2017). A presente tese vai de encontro a essa realidade, propondo uma ferramenta de apoio à decisão, utilizando técnicas de aprendizado de máquina. Essa abordagem permitiu mensurar com as variáveis que compõem a performance de atletas, como gols, minutos jogados e interceptações.

Entretanto, nas sociedades empresariais, a falta de viabilidade econômica implica na falência e descontinuação da organização, haja vista casos como o da Fiorentina, do Napoli e do Parma, na Itália, e o do Glasgow Rangers, na Escócia, entre outros (BARAJAS; RODRÍGUEZ, 2010; SALGADO-BARANDELA *et al.*, 2017; SZYMANSKI, 2017). O também chamado formato de “clubes-empresa” possibilitou o aparecimento de um novo fenômeno, as corporações de clubes – *multi-club ownership* – como os conglomerados da Red Bull e do City Football Group (PASTORE, 2018).

Ao propor uma análise de redes sociais em uma base de mais de 100 mil transações extraídas da plataforma *Transfermarkt*, esta tese investiga os fluxos estratégicos entre países. A aplicação da *Social Network Analysis* (SNA) indicou padrões de interdependência e formação de comunidades, mapeando a centralidade de países como Bélgica, Portugal e Brasil nas cadeias globais de transferência, tanto como origem, destino e intermediários. Esses resultados oferecem insumos aos clubes para decisões eficazes sobre o perfil de talentos com potencial de revenda. Com isso espera-se que os departamentos de análise de dados se tornem elementos chave na estrutura organizacional dos clubes de futebol, dando suporte no recrutamento de jogadores. Consequentemente, os modelos de avaliação de performance devem acompanhar a evolução e se tornarem vantagens competitivas entre os clubes.

Portanto, esta tese se insere na interseção entre a estratégia organizacional, a gestão esportiva e as metodologias de apoio à decisão. Propõe uma inovação teórica ao combinar análise de redes e lógica *fuzzy* como lentes para interpretar o futebol como campo estratégico. Apresenta uma contribuição prática, ao oferecer um sistema interpretável de pontuação de atletas, fundamentado em dados objetivos, mas adaptável ao contexto subjetivo de cada clube. Em síntese, ao compreender o fenômeno da seleção de atletas e confrontar decisões subjetivas com modelos baseados em dados, a tese propõe uma ferramenta de apoio capaz de reduzir riscos, maximizar retornos e fortalecer as capacidades gerenciais dos clubes, propondo um diferencial competitivo no futebol profissional contemporâneo.

### 1.1. Pergunta de pesquisa

Desta forma, esta pesquisa atuará no campo da estratégia, mais especificamente a compreensão do planejamento estratégico dos clubes de futebol no que tange aos processos de seleção e aquisição de jogadores de futebol de maneira eficiente. Esta pesquisa irá se orientar tendo por base a pergunta: **Como clubes de futebol podem utilizar modelos baseados em aprendizado de máquina, lógica *fuzzy* e análise de redes sociais para estruturar estratégias mais eficientes de recrutamento de atletas, minimizando riscos e maximizando o desempenho esportivo ao longo de uma temporada?**

De modo que, identificando os elementos de desempenho relevantes para o impacto do resultado em uma temporada desportiva, almeja-se estabelecer esses fatores como decisórios

para influenciar a contratação de atletas. Ao determinar fatores críticos à performance, este trabalho servirá para orientar gestores de clubes na formulação de estratégias mais adequadas ao contexto em que se encontram.

## 1.2. Objetivo geral

O principal objetivo deste trabalho é desenvolver e validar uma ferramenta de apoio à decisão estratégica para clubes de futebol, utilizando algoritmos de aprendizado de máquina, lógica *fuzzy* e análise de redes sociais, com o propósito de identificar padrões de desempenho de jogadores e dinâmicas do mercado de transferências que contribuam para decisões mais eficientes de recrutamento. Esta ferramenta visa maximizar a eficiência dos recursos, melhorar o desempenho desportivo e atingir objetivos de sucesso ao longo de uma temporada. Os altos montantes envolvidos nas contratações implicam a utilização eficiente de recursos como determinante para o sucesso de uma equipe de futebol para cumprir com os objetivos organizacionais.

De modo que será necessário trabalhar com algoritmos de aprendizado de máquina (*machine learning*) alimentados com os dados *fuzzificados* a partir do histórico das partidas, utilizando dados da plataforma *Wyscout* e outros dados relevantes para esta análise. Por meio deste processo será possível obter uma ferramenta que dê suporte à tomada de decisão dos gestores com base nos cenários anteriores.

## 1.3. Objetivos específicos

Entre os objetivos específicos podem ser listados:

- 1) Revisar a literatura sobre seleção de atletas, desempenho técnico e o uso de inteligência artificial no esporte, com foco na lógica *fuzzy* e aprendizado de máquina;
- 2) Mapear os padrões de transferência de jogadores entre países e clubes por meio de análise de redes sociais, identificando rotas, centralidades e comunidades estratégicas;
- 3) Construir e testar modelos preditivos baseados em *machine learning*, capazes de identificar os fatores de desempenho com maior poder explicativo sobre resultados como gols, assistências, minutagem e valor de mercado;
- 4) Desenvolver uma modelagem *fuzzy* para classificação e pontuação de atletas, integrando atributos técnicos e indicadores estatísticos relevantes ao contexto competitivo;

- 5) Comparar diferentes versões do modelo para avaliar a robustez e a aplicabilidade da ferramenta em diferentes cenários;
- 6) Propor recomendações estratégicas para gestores de clubes, com base nos achados dos modelos e das redes, com vistas a minimizar riscos financeiros e maximizar o retorno técnico das contratações

A investigação em curso visa responder à questão de como os algoritmos de aprendizagem automática podem ser utilizados para otimizar o processo de seleção e contratação de jogadores de futebol, identificando os fatores de desempenho mais influentes nas equipes desportivas. O objetivo é fornecer aos gestores dos clubes uma ferramenta que considere tanto dados históricos como cenários futuros, melhorando a eficiência do investimento e melhorando o desempenho desportivo da equipe ao longo de uma temporada.

#### **1.4. Justificativa**

A construção desta tese foi guiada por um percurso marcado por reorientações teóricas, revisões metodológicas e um contínuo exercício de refinamento analítico. Ao longo do processo de doutoramento, o projeto amadureceu ao articular três domínios distintos. Os estudos organizacionais, a análise de desempenho no futebol e a modelagem matemática orientada à decisão, com o objetivo de compreender, com profundidade e aplicabilidade, os mecanismos que regem o recrutamento e a valorização de atletas profissionais no mercado global do futebol.

O ponto de partida da tese remonta a uma proposta de investigação inicial sobre o caso do Cruzeiro Esporte Clube, que viveu um processo agudo de fracasso organizacional entre 2019 e 2021. A proposta, centrada em entender as causas da crise e os impactos da transformação jurídica em Sociedade Anônima do Futebol (SAF), foi descontinuada por limitações de acesso a dados internos. No entanto, a pergunta daquela pesquisa – (“como organizações esportivas falham, sobrevivem ou se reinventam?”) – permaneceu ao longo da trajetória. Esse olhar foi revitalizado e ampliado por meio do acompanhamento de outros processos de reestruturação no futebol brasileiro, tanto a partir da observação direta quanto do contato com orientandos que pesquisaram, com foco empírico, a transição institucional de clubes como o próprio Cruzeiro.

Concomitantemente, o projeto passou a incorporar uma abordagem mais estrutural, orientada à compreensão do futebol como um sistema global de trocas e decisões interdependentes. Esse movimento foi influenciado pelos debates contemporâneos em estratégia, sobretudo aqueles

ancorados na teoria das capacidades dinâmicas e na estratégia como prática. A partir dessa inflexão, o foco da tese deslocou-se da explicação de crises isoladas para o entendimento das redes que conectam clubes, países e atletas, revelando como certas posições e competências organizacionais se convertem em vantagem competitiva durável.

Esta tese se enquadra na abordagem de *Design Science Research* (DSR), cuja ênfase recai na proposição, construção e avaliação de um artefato orientado à solução de um problema. Assim, a proposta de ferramenta é aplicada na avaliação e comparação do desempenho de atletas de futebol de modo informativo para apoio à decisão de recrutamento, voltado para a montagem de elenco e gestão de ativos esportivos.

Em DSR, diferentemente de estudos exclusivamente exploratórios, descritivos ou explicativos, o objetivo central é prescritivo, isto é, projetar e justificar um artefato capaz de melhorar decisões em um domínio prático (HEVNER *et al.*, 2004; MARCH; SMITH, 1995; SIMON, 1996). O conceito de artefato proposto aqui, cuja avaliação da banca examinadora destacou um processo ainda em estágios iniciais, consiste em um procedimento de transformação *fuzzy* que integra estatísticas de jogo e validações cruzadas para gerar escores compostos de performance (*e.g.*, Finalização) coerentes com a lógica do jogo e com as posições em campo.

O trabalho evoluiu em ciclos iterativos de diagnóstico e requisitos do problema, projeto e construção do artefato, demonstração e avaliação e refinamento. O artefato produziu escores coerentes com a lógica do jogo, interpretabilidade e alinhamento com avaliações externas. O leitor notará que cada capítulo segue estrutura completa (introdução, método, resultados e discussão). Isso reflete subciclos de design dentro do programa de DSR da tese. Para evitar redundâncias, não há um capítulo único de metodologia, em vez disso, este destaque apresenta o método global (DSR), enquanto cada capítulo detalha os métodos específicos de sua fase.

Esse método nos permitiria utilizar as Matrizes Epistêmicas de Paes de Paula (2016) para abordar diversos temas em cada capítulo que interrelacionam os atores do ecossistema do futebol brasileiro. Esta abordagem epistemológica procura compreender os fenômenos dentro das suas matrizes e promove o diálogo entre abordagens sociológicas, reconhecendo que nenhuma abordagem única pode abranger todo o fenômeno.

Desta maneira, etapas do processo incluem tentativas exploratórias com menor desempenho. Para clareza e foco nos achados principais, optou-se por preservar apenas alguns destes ensaios,

dando ênfase aos resultados finalistas. Essa organização reforça a transparência sem comprometer a fluidez do argumento central. À luz dessa abordagem, apresenta-se a seguir a estrutura e a função de cada capítulo desta tese.

O primeiro eixo empírico da tese, apresentado no Capítulo 2, consistiu em uma revisão bibliométrica da literatura internacional, que evidenciou uma crescente convergência entre as áreas de *machine learning*, performance esportiva e modelos *fuzzy* de tomada de decisão. Essa revisão serviu não apenas como ancoragem teórica, mas também como radar de lacunas: embora abundem estudos sobre predição de desempenho e valor de mercado, são escassos os trabalhos que integram lógica *fuzzy* à análise interpretável em larga escala, com foco aplicado à gestão de clubes.

No Capítulo 3, a tese mergulha na análise das transferências internacionais de jogadores por meio da Análise de Redes Sociais (SNA), modelando mais de 100 mil negociações registradas na plataforma *Transfermarkt*. A partir de métricas de centralidade, intermediação e formação de comunidades, o capítulo revela rotas estáveis, intermediários estratégicos e blocos geográficos de transferência. Demonstra como países como Bélgica e Portugal atuam como porta de entrada para o futebol europeu, assumindo papéis centrais na circulação de ativos esportivos. O Brasil, por sua vez, aparece como um grande exportador, mas com baixa presença em posições de intermediação. Isso sinaliza uma oportunidade estratégica ainda não plenamente explorada pelos clubes nacionais.

Por fim, no Capítulo 4, a tese propõe e testa um modelo de pontuação baseado em lógica *fuzzy*, orientado à avaliação de desempenho técnico e potencial de atletas. A modelagem, alimentada por estatísticas extraídas da *Wyscout* e validadas por dados do jogo *Football Manager 2024*, é aplicada a diferentes variáveis de interesse (minutagem, gols, assistências), com e sem ponderação por posição e habilidade. Apesar das limitações na explicação de valores de mercado, reflexo da natureza especulativa e multifatorial da variável de interesse, os modelos *fuzzy* demonstraram alta capacidade explicativa para variáveis objetivas de performance, indicando sua utilidade prática para comitês técnicos, analistas e gestores de futebol.

Ao integrar os níveis de análise macro (fluxos internacionais), meso (dinâmicas organizacionais) e micro (avaliação individual de atletas) a tese oferece uma ferramenta interpretável para suportar decisões estratégicas em um mercado volátil, competitivo e orientado por dados. Ao fazer isso, contribui para o campo de pesquisa multidisciplinar,

incluindo um trabalho que executa a intersecção entre gestão estratégica, ciência de dados e futebol profissional.

Os desdobramentos futuros, indicados ao final da tese, contemplam tanto a aplicação prática do modelo *fuzzy* em contextos reais (clubes e categorias de base). Também há capacidade para a expansão analítica de fenômenos como a propriedade cruzada de clubes (*multi-club ownership*), cuja articulação estratégica ainda carece de exploração teórica e empírica no campo dos Estudos Organizacionais.

O cenário do futebol tem características únicas que exigem uma proposta de trabalho mais elaborada, com fundamentos teóricos cuidadosamente desenvolvidos. O método de estruturação da tese escolhido vai priorizar a elaboração dos capítulos com uma organização fechada, isto é, o capítulo composto por introdução, fundamentação teórica, método, resultados e conclusão. Esta proposta pretende abranger um amplo conjunto de aspectos relacionados com a gestão do futebol no país, com o objetivo de aumentar a eficiência do investimento e melhorar o desempenho esportivo da equipe ao longo de uma época.

## 2. *Machine learning* aplicado aos esportes: uma revisão bibliométrica

### 2.1. Introdução

A aprendizagem de máquina (*machine learning* – ML) é uma técnica de análise de dados que automatiza o desenvolvimento de modelos analíticos (WARING *et al.*, 2020). Os avanços tecnológicos facilitaram o surgimento de algoritmos de análise que adotam modelos de aprendizagem de máquina e inteligência artificial. Originada a partir da pergunta seminal de Alan Turing, "As máquinas podem pensar?", e fundamentada na aprendizagem *bayesiana*, otimização e predição, a aprendizagem de máquina ganhou destaque por realizar análises complexas ao adotar fatores, processos para trabalhar variáveis, parâmetros que permitem predição (HOLZINGER, 2018; TURING, 1950). A aplicação dessas técnicas, sistemas de apoio a decisão de gestão esportiva tem atraído a atenção de pesquisadores das áreas de engenharia, administração e ciências do esporte (ALMULLA; ALAM, 2020; PARIATH *et al.*, 2018).

O ambiente esportivo e, particularmente, o futebol, apresenta variáveis que não se manifestam em estados absolutos, mas sim em gradientes contínuos de manifestação, frequentemente dependentes de contexto, interpretação e subjetividade. Nesse sentido, optou-se por integrar ao presente estudo a teoria dos conjuntos *fuzzy*, por sua capacidade singular de modelar imprecisões e incertezas que escapam aos modelos probabilísticos tradicionais. Ao considerar que atributos como capacidade defensiva ou capacidade ofensiva não se reduzem a categorias dicotômicas, tais como sim e não ou, ainda, alto e baixo, a lógica *fuzzy* permite representar essas dimensões por meio de funções de pertinência, fornecendo variáveis de entrada alinhadas à realidade técnica e contextual dos atletas.

Essa abordagem se mostrou relevante diante da escassez de estudos que combinem *machine learning* com modelagens *fuzzy* no domínio do futebol, evidenciando uma lacuna metodológica ainda pouco explorada na literatura internacional. Como será demonstrado na análise bibliométrica apresentada na próxima seção deste capítulo, embora o uso de *machine learning* em contextos esportivos venha se expandindo, há uma predominância de enfoques voltados à estatística tradicional, com menor representatividade de abordagens capazes de lidar com imprecisão, vagueza e ambiguidade, características inerentes ao jogo de futebol. Assim, a escolha pela lógica *fuzzy* se sustenta tanto por fundamentos teóricos quanto pela necessidade de representar com maior fidelidade os fenômenos analisados nesta pesquisa, reforçando a originalidade e adequação metodológica da proposta.

## 2.2.A aprendizagem de máquina

A aprendizagem de máquina opera dentro da dimensão da inteligência artificial, baseada na premissa de que sistemas computacionais podem aprender a partir de dados, discernir padrões, apoiar, tomar decisões (como Agentes autônomos de Inteligências Artificiais [IA]) conforme o grau de autonomia atribuída ao modelo para a intervenção humana (WARING *et al.*, 2020). Embora a modelagem estatística tradicional tenha sido por muito tempo a base da análise e interpretação de dados, os avanços recentes no poder computacional e o sucesso notável das tecnologias de inteligência artificial em diversos domínios impulsionaram as aplicações de *machine learnings* (ML) e as trouxeram para o centro das atenções (BARNETT-ITZHAKI *et al.*, 2020; HUNG, 2019; REEL *et al.*, 2021). O principal objetivo dos métodos de aprendizagem de máquina é extrair do histórico de dados – o “conhecimento intrínseco” – que está contido nas suas combinações, interpretações por modelos qualitativos e quantitativos para elaborar, emular empiricamente cenários diversos, predizer, orientar caminhos para a tomada de decisões e, por limitação, “permanecendo tudo o mais invariante” (*Ceteris Paribus*) (REEL *et al.*, 2021).

Algumas subdivisões ganharam destaque, como a aprendizagem profunda (*deep learning*), que aprimorou o desempenho de classificadores em relação a algoritmos tradicionais de ML, especialmente em cenários que envolvem grandes volumes de dados de alta dimensionalidade (REEL *et al.*, 2021). No entanto, esses métodos são intensivos em termos computacionais, exigindo hardware de alto desempenho (REEL *et al.*, 2021). A aplicação eficaz da aprendizagem de máquina demanda compreensão à natureza do problema analisado e à compatibilidade entre os dados disponíveis e o tipo de abordagem escolhida. Existem três categorias de aprendizagem: (i) a aprendizagem baseada em algoritmos determinísticos ou probabilísticos, que seguem fórmulas matemáticas estabelecidas; (ii) a aprendizagem por correlação heurística, aplicada em contextos de *big data*, na qual padrões são extraídos de grandes volumes de dados por meio de modelagens empíricas e estatísticas; e (iii) os modelos que simulam comportamentos, como agentes autônomos e *chatbots*, programados a partir de interações e simulações baseadas em regras (RUSSELL; NORVIG, 2021; WARING *et al.*, 2020).

Entretanto, ainda que estes sistemas se mostrem competentes na resolução de problemas complexos, não há um algoritmo universalmente eficaz para todos os tipos de desafios (*no free lunch theorem*) (ADAM *et al.*, 2019). Portanto, isso demanda esforço humano na curadoria, interpretação e modelagem dos dados, da mesma maneira que exige esforço computacional para

processar grandes volumes de informação com consistência e escalabilidade. Destarte, a eficácia da aprendizagem de máquina depende da escolha do tipo de abordagem, da preparação dos dados e da contextualização dos resultados. Mesmo pesquisadores de gestão esportiva com experiência em análise de dados frequentemente não possuem a expertise necessária para aplicar essas técnicas em grandes conjuntos de dados (WARING *et al.*, 2020). Esse processo demanda um investimento substancial de tempo e esforço, e o acesso a dados e conhecimento especializado nem sempre está prontamente disponível.

A aplicação de aprendizado de máquina no futebol emergiu à medida que pesquisadores buscaram modelar quantitativamente fatores que influenciam o desempenho de atletas e equipes. Essa abordagem oferece possibilidade para a utilização de algoritmos preditivos, como os de regressão, e integrar com os dados obtidos em modelos mais flexíveis e interpretáveis, como aqueles baseados em lógica de conjuntos (*fuzzy*).

A combinação desses modelos possibilitaria uma análise híbrida, em que modelos estatísticos possam fornecer pesos objetivos para variáveis-chave, enquanto a lógica *fuzzy* permite a transformação desses dados em escalas mais intuitivas, adaptáveis a diferentes contextos esportivos. Assim, a integração entre aprendizagem de máquina e métodos *fuzzy* sugere novas possibilidades para avaliações e previsões, aproximando a modelagem quantitativa das nuances da tomada de decisão no futebol.

Há razões para considerar o uso de transformações *fuzzy*. Essas técnicas têm demonstrado eficácia na construção de modelos de classificação que se destacam tanto em desempenho quanto em interpretabilidade (ELKANO *et al.*, 2019). Além disso, classificadores *fuzzy* podem lidar com a imprecisão e incerteza inerentes a conjuntos de dados do mundo real (RAJESWARI; DEISY, 2019). Os classificadores *fuzzy* mantêm a transparência e a interpretabilidade em seus processos decisórios. Ao se basear em declarações estocásticas *fuzzy* em vez de valores precisos ou categóricos, essas representações geram modelos mais aderentes ao fenômeno preditivo (RAJESWARI; DEISY, 2019).

### **2.3.O desempenho do atleta e equipes**

Pesquisadores na área de ciências do esporte intensificaram seus esforços para desenvolver modelos que prevejam com eficácia o desempenho de atletas no futebol. A influência de técnicas de modelagem provenientes de outros esportes coletivos, como voleibol, basquete e beisebol, pode ter estimulado a criatividade dos pesquisadores do futebol. Assim, este trabalho

analisa as publicações sobre a integração entre aprendizagem de máquina, esporte e futebol, assim como publicações ligadas à classificação de performance no futebol utilizando lógica *fuzzy*. O objetivo é compreender de que maneiras técnicas de aprendizagem de máquina e abordagens utilizando bancos de dados *fuzzy* possam ser adequadas para classificar jogadores de futebol.

Ao compreender o processo de seleção de jogadores de futebol, abrangendo critérios utilizados por tomadores de decisão e comparando-os com modelagem quantitativa, torna-se viável desenvolver técnicas que permitam que esse processo seja conduzido de maneira eficiente. Dessa forma, este estudo busca propor inovações teóricas no planejamento estratégico por meio da adoção da aprendizagem de máquina, bem como inovações gerenciais na administração do futebol como uma modalidade preditiva eficiente. O desenvolvimento de modelos que identifiquem fatores pertinentes ao sucesso esportivo de um time de futebol tem o potencial de conferir vantagens estratégicas aos clubes de futebol.

#### **2.4. Definições e origens da aprendizagem de máquina**

A aprendizagem de máquina, por definição, é uma programação voltada para compreender regras a partir de dados, adaptar-se a mudanças e melhorar o desempenho por meio da aprendizagem empírica (BLUM, 2007). O conceito reflete o desejo inicial do campo da Ciência da Computação, com suas origens remontando ao problema do Jogo da Imitação proposto pelo matemático Alan Turing (TURING, 1950; MUGGLETON, 2014).

Em seu artigo seminal, Turing propôs que a busca pela Inteligência Artificial (IA) passaria de um empreendimento puramente filosófico para uma tarefa de engenharia, impulsionada por três abordagens distintas: IA por programação, IA por aprendizagem de máquina *ab initio* e IA baseada em lógica, probabilidades, aprendizagem e conhecimento prévio (MUGGLETON, 2014; TURING, 1950). Nesse contexto, as máquinas podem adquirir conhecimento de diferentes formas, a depender do tipo de aprendizagem aplicada: (i) quando se trata de aprendizagem supervisionada, o modelo é treinado com dados rotulados; (ii) na aprendizagem não supervisionada, busca-se identificar padrões ou estruturas nos dados; e (iii) na aprendizagem por reforço, o agente aprende com base em recompensas provenientes de interações com o ambiente (ALZUBI *et al.*, 2018). De modo geral, a aprendizagem de máquina envolve um programa de computador que aprende a partir da experiência com relação a uma tarefa e sua respectiva métrica de desempenho (SAMUEL, 1959).

A Teoria da Aprendizagem de Máquina busca compreender os princípios fundamentais da aprendizagem como um processo computacional, visando interpretar informações e dados com precisão matemática para melhorar o desempenho. Essa busca segue o Princípio da Parcimônia, ou Navalha de Occam, favorecendo explicações mais simples em lugar das mais complexas (BLUM, 2007). Esse princípio encontra um equivalente técnico na teoria da aprendizagem de máquina, fornecendo bases matemáticas formais para defender a redução da complexidade dos modelos (MJOLSNESS; DECOSTE, 2001). Ao selecionar um modelo, é fundamental minimizar as métricas de complexidade, garantindo ao mesmo tempo um bom desempenho no conjunto de treinamento e assegurando uma generalização eficaz para conjuntos de teste derivados da mesma ou de distribuições relacionadas (MJOLSNESS; DECOSTE, 2001).

Dessa forma, a aprendizagem de máquina envolve o desenvolvimento de modelos matemáticos que encapsulam os princípios fundamentais da aprendizagem, a criação de algoritmos de aprendizagem que atendam a critérios desejáveis e a análise matemática de questões subjacentes (BLUM, 2007). Como um campo multidisciplinar, a aprendizagem de máquina encontra aplicação em diversas áreas de pesquisa (ALZUBI *et al.*, 2018).

O reconhecimento de padrões e a inferência estatística estão intrinsecamente interligados na aprendizagem de máquina (ML). A evolução dessas técnicas expandiu sua utilidade prática para áreas diversas, incluindo engenharia, negócios e ciências da saúde (MJOLSNESS; DECOSTE, 2001). Esses avanços possuem o potencial de revolucionar a investigação científica, com a expectativa de automação parcial das etapas do método científico, desde a formulação de hipóteses até a construção de modelos e experimentação crítica (MJOLSNESS; DECOSTE, 2001).

Os métodos de classificação empregados em ML são geralmente categorizados em dois grupos. Modelos generativos buscam capturar a probabilidade conjunta das variáveis em um sistema físico e podem abranger a aprendizagem de mecanismos causais, como redes bayesianas ou modelos gráficos (MJOLSNESS; DECOSTE, 2001). Por outro lado, modelos discriminativos concentram-se em distinguir entre exemplos positivos e negativos, em vez de modelar toda a densidade de probabilidade conjunta, como exemplificado pelas máquinas de vetor de suporte (SVMs) (MJOLSNESS; DECOSTE, 2001).

## **2.5.Método**

A metodologia de análise bibliométrica abrange o exame quantitativo da produção, disseminação e utilização da informação, registrado por meio de técnicas bibliométricas. Essas técnicas têm como objetivo identificar tendências, a expansão do conhecimento dentro de um campo, os periódicos mais relevantes em uma disciplina e as colaborações entre autores, entre outros fatores (MARQUES *et al.*, 2018). Dentro desse escopo, estudos bibliométricos consistem na análise quantitativa da literatura científica, empregando métodos matemáticos e estatísticos. O propósito é delinear redes intelectuais entre pesquisadores, identificar tendências dentro de um campo específico, esclarecer as principais teorias e autores prolíficos, bem como mapear a estrutura intelectual de uma disciplina ou área de estudo (FERREIRA *et al.*, 2016).

Essas análises são amplamente utilizadas em diversos domínios e servem para avaliar contribuições acadêmicas específicas, medir o impacto da pesquisa, identificar autores e instituições influentes e examinar a evolução da literatura dentro de um campo de estudo específico (DIODATO; GELLATLY, 2013). Assim, a análise bibliométrica proporciona uma abordagem sistemática e quantitativa para compreender o panorama acadêmico e científico, fornecendo insights valiosos para a pesquisa e a tomada de decisões.

A análise bibliométrica é sustentada por duas abordagens principais: a análise de desempenho, que emprega métricas quantitativas para avaliar a produção e o impacto dos estudos, e o mapeamento científico, que visualiza e traça a literatura dentro de domínios específicos utilizando dados bibliográficos. O advento de ferramentas avançadas de software revolucionou a bibliometria, permitindo que pesquisadores obtenham insights valiosos a partir de grandes volumes de dados acadêmicos. Entre as ferramentas mais notáveis estão o *CiteSpace*, o *SciMAT* e o *Bibliometrix*, reconhecidos por suas funcionalidades e interfaces amigáveis (ROBLEK *et al.*, 2022). Destaca-se, em particular, o *Bibliometrix*, um pacote abrangente desenvolvido na linguagem R, que fornece funções de análise bibliométrica que aprimoram o processamento de dados e geram visualizações de alta qualidade, mapas de rede e relatórios detalhados (DONTHU *et al.*, 2021; ECHCHAKOUI, 2020).

## **2.6. Análise**

Nos próximos tópicos serão apresentadas as análises bibliométricas executadas para cada um dos temas de interesse apresentados.

### **2.6.1. Aprendizagem de máquina e esporte: uma revisão bibliométrica**

Para elucidar o cenário atual sobre a aplicação de técnicas de aprendizagem de máquina no campo esportivo, realizamos uma revisão bibliométrica, com o objetivo de examinar a evolução dos estudos científicos sobre esse tema, utilizando técnicas matemáticas e estatísticas (DIODATO; GELLATLY, 2013). A fase inicial consistiu em uma busca bibliográfica sistemática abrangendo o período de janeiro de 2013 até dezembro de 2024, direcionada a artigos relevantes sobre aprendizagem de máquina. Utilizando um conjunto de palavras-chave predefinidas, realizamos a pesquisa em duas das principais bases de periódicos acadêmicos, a *Scopus* e a *Web of Science*.

As consultas de pesquisa incluíram os termos em inglês "*machine learning*" e "*sports*" no resumo (*abstract*). Foram incluídos artigos, trabalhos de conferências, revisões, revisões de conferências, capítulos de livros e livros. A busca resultou em 2.226 documentos provenientes da base *Scopus* e 1.364 documentos da *Web of Science*. A integração dos achados de ambas as bases, utilizando o *Bibliometrix*, resultou em um corpus de 2.330 trabalhos únicos. Posteriormente, a depuração do conjunto de dados excluiu registros sem atribuição de autoria, culminando na análise de 2.247 artigos. Também foram excluídos da amostra entradas relacionadas a retratações e erratas, com resultado em 2.209 trabalhos

Para a análise, utilizamos o software RStudio em conjunto com o algoritmo *Bibliometrix* (ARIA; CUCCURULLO, 2017). Desenvolvido na linguagem R, o pacote *Bibliometrix* facilita análises bibliométricas, fornecendo visualizações intuitivas e insights sobre a produção científica, incluindo a frequência e centralidade de artigos (ARIA; CUCCURULLO, 2017). As visualizações geradas incluem mapas científicos e resultados de diversas análises, como modelos geométricos, redes temáticas e mapas de proximidade entre itens (COBO *et al.*, 2011; SKUPIN, 2009; VAN ECK; WALTMAN, 2010).

A incorporação de artigos de anais de conferências em revisões bibliométricas oferece diversas vantagens, principalmente permitindo que os pesquisadores acompanhem os avanços no campo, disciplinas correlatas e autores contribuintes (GONZÁLEZ-ALBO; BORDONS, 2011). Essa inclusão proporciona uma perspectiva abrangente sobre o estado da arte da pesquisa. Além disso, o processo acelerado de publicação em conferências oferece aos autores uma via mais rápida para disseminação de seus trabalhos em comparação com periódicos tradicionais. Ademais, o processo de revisão por pares inerente aos trabalhos de conferência garante padrões de qualidade (GONZÁLEZ-ALBO; BORDONS, 2011).

Os artigos de conferências permitem que os autores refinem seus conteúdos com base no feedback recebido durante o evento, aprimorando assim sua contribuição acadêmica (GONZÁLEZ-ALBO; BORDONS, 2011). Enquanto artigos de conferências costumam apresentar um número maior de autores por publicação, os artigos de periódicos tendem a ser mais extensos e a receber mais citações, indicando pesquisas mais aprofundadas e influentes (GONZÁLEZ-ALBO; BORDONS, 2011). Embora estatisticamente insignificante, a presença de documentos altamente citados entre artigos de conferências excede as expectativas, evidenciando sua relevância no discurso acadêmico (GONZÁLEZ-ALBO; BORDONS, 2011).

Uma síntese dos principais dados obtidos a partir dessa compilação revela que os 2.209 documentos foram provenientes de 1.301 fontes distintas e envolveram 7.164 autores. A taxa de crescimento anual observada ao longo do período é de 37,6%. A amostra inclui 1.065 (48,2%) artigos de periódicos, 929 (42,1%) trabalhos de conferências e uma proporção menor de outros tipos de publicações, como capítulos de livros, revisões e livros (9,7%). A análise dos metadados identificou dados suficientes para realizar testes para 12 indicadores, conforme detalhado no quadro 1.

<b>Metadado</b>	<b>Descrição</b>	<b>Células vazias</b>	<b>% (total)</b>	<b>Status</b>
AB	Resumo (Abstract)	0	0	Excelente
AU	Autor (Author)	0	0	Excelente
DT	Tipo de documento (Document Type)	0	0	Excelente
SO	Periódico (Journal)	0	0	Excelente
LA	Idioma (Language)	0	0	Excelente
PY	Ano de publicação (Publication Year)	0	0	Excelente
TI	Título (Title)	0	0	Excelente
TC	Total de citações (Total Citation)	0	0	Excelente
C1	Afiliação (Affiliation)	18	0,80	Bom
DI	DOI	184	8,16	Bom
DE	Palavras-chave (Keywords)	296	13,13	Aceitável
RP	Autor correspondente (Corresponding Author)	344	15,26	Aceitável

Quadro 1 - Análise dos metadados

O crescimento das publicações é representado na figura 1, apontando a crescente importância atribuída ao assunto ao longo do período avaliado.

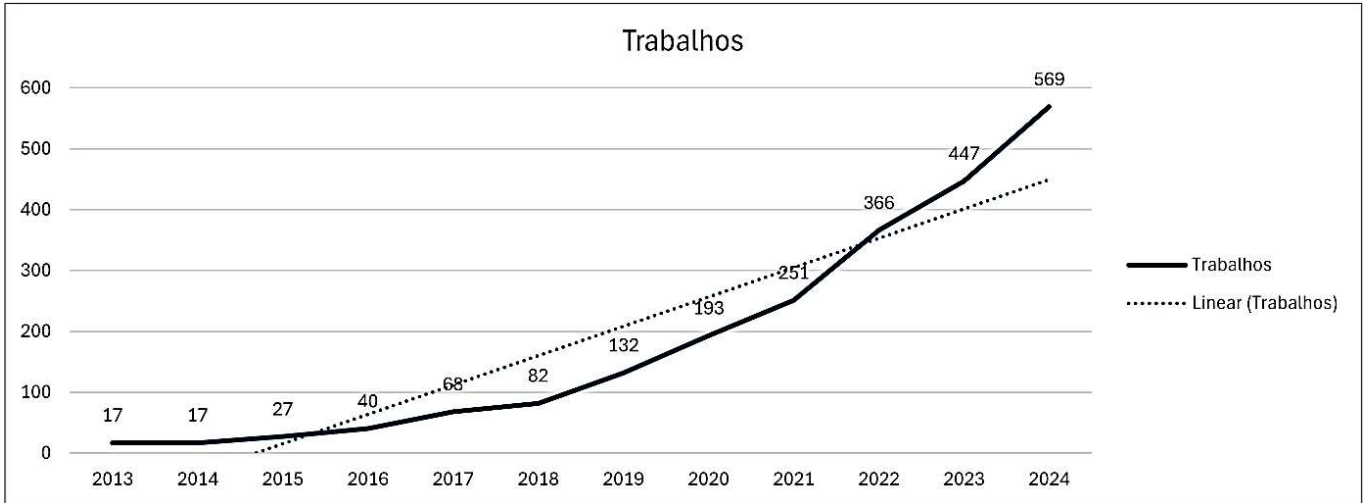


Figura 1 - Evolução do número de publicações

As principais fontes da amostra apontam o periódico *Sensors*, da editora MDPI, como principal meio propagador do tema. A série de livros da Springer *Lecture Notes in Networks and Systems* surge como segunda principal fonte de publicações envolvendo aprendizado de máquina e esportes. Os periódicos podem ser observados no quadro 2.

Fontes mais relevantes	Publicações
<i>Sensors</i>	60
<i>Lecture Notes in Networks and Systems</i>	36
<i>IEEE Access</i>	29
<i>ACM International Conference Proceeding Series</i>	28
<i>Communications in Computer and Information Science</i>	26
<i>PLOS One</i>	23
<i>Applied Sciences-Basel</i>	22
<i>Computational Intelligence and Neuroscience</i>	17
<i>Advances in Intelligent Systems and Computing</i>	16
<i>Lecture Notes in Computer Science</i>	16

Quadro 2 - Principais fontes da amostra

Entre os fatores de impacto dos meios de publicação das pesquisas envolvendo aprendizado de máquina e esportes, estão o índice  $h$  ( $h\_index$ ), calculado com base no número de publicações que foram citadas pelo menos o mesmo número de vezes; o índice  $g$  ( $g\_index$ ), baseado na

distribuição de citações recebidas pela publicação; e o índice  $m$  ( $m\_index$ ), que avalia o volume e a produtividade acadêmica. Além dos índices estão o total de citações (TC) e o ano de lançamento do periódico (Ano). Esses dados podem ser vistos no quadro 3.

<b>Periódico</b>	<b>h_index</b>	<b>g_index</b>	<b>m_index</b>	<b>TC</b>	<b>Ano</b>
<i>Sensors</i>	18	33	2.250	1227	2018
<i>IEEE Access</i>	10	20	1.250	422	2018
<i>American Journal of Sports Medicine</i>	8	12	1.333	184	2020
<i>Frontiers in Sports and Active Living</i>	8	12	1.600	159	2021
<i>Applied Sciences-Basel</i>	8	14	1.000	207	2019
<i>International Journal of Sports Physiology and Performance</i>	6	6	0.667	168	2017
<i>PLOS One</i>	6	10	0.600	128	2016
<i>Advances in Intelligent Systems and Computing</i>	5	8	0.625	82	2018
<i>CMC-Computer Materials &amp; Continua</i>	5	8	0.833	71	2020
<i>Computational Intelligence and Neuroscience</i>	5	7	1.000	60	2021

Quadro 3 - Fatores de impacto dos periódicos

Entre os trabalhos avaliados nessa amostra, os 10 com maior número de citações globalmente versam quase todos sobre desenvolvimento de técnicas de detecção de movimento e outras atividades humanas que podem ser usados para o desenvolvimento de acessórios (*wearables*). Ademais, o trabalho de Rein e Memmert (2016), sobre a utilização de *big data* no futebol, também lista entre os dez mais citados. Outro trabalho relevante versou sobre uma estrutura (*framework*) para gerar modelos de predição de resultados esportivos, publicado por Bunker e Thabtah (2019). Estes trabalhos podem ser vistos no quadro 4.

<b>Artigo</b>	<b>Periódico</b>	<b>Autores</b>	<b>Ano</b>	<b>Total Citações</b>	<b>TC por ano</b>	<b>TC Normalizada</b>	<b>Relação ML/Espportes</b>
<i>Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications</i>	Science Advances	Zhu <i>et al.</i>	2020	519	86,50	27,51	Possível aplicação de interface para treinamento esportivo
<i>A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System</i>	Sports Medicine - Open	Colyer <i>et al.</i>	2018	304	38,00	16,29	Artigo que apresenta modelos em pesquisa de captura de movimento sem marcadores a partir de visão computacional
<i>Deep learning-enabled triboelectric smart socks for IoT-based gait analysis and VR applications</i>	npj Flexible Electronics	Zhang <i>et al.</i>	2020	296	49,33	15,69	Metodologia para tratar dados relacionados a velocidade de marcha humana
<i>Intelligent Wearable Systems: Opportunities and Challenges in Health and Sports</i>	ACM Computing Surveys	Yang <i>et al.</i>	2024	295	147,50	187,97	Fornece uma visão de dispositivos vestíveis inteligentes ( <i>wearebles</i> ) e suas aplicações em saúde e esportes.

<i>Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science</i>	SpringerPlus	Rein e Memmert	2016	291	29,10	13,66	Artigo sobre a utilização de big data nas análises táticas de futebol
<i>Triboelectric nanogenerator based self-powered sensor for artificial intelligence</i>	Nano Energy	Zhou et al.	2021	236	47,20	18,58	Revisão sobre a aplicação de nanogeradores triboelétricos que coletam informações por sensores, aplicados a esportes inteligentes e outras áreas.
<i>A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications</i>	Artificial Intelligence Review	Pareek e Thakkar	2021	195	39,00	15,35	O artigo discute as características de dados usados para Human Action Recognition (HAR), junto com as aplicações em resumo de vídeo baseado em esportes entre outros.

<i>A machine learning framework for sport result prediction</i>	Applied Computing and Informatics	Bunker e Thabtah	2019	184	26,29	12,14	O artigo fornece uma análise da literatura em ML, com foco na aplicação de Rede Neural Artificial à previsão de resultados esportivos.
<i>Deep learning for human activity recognition: A resource efficient implementation on low-power devices</i>	2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks	Ravi <i>et al.</i>	2016	174	17,40	8,17	O Reconhecimento de Atividade Humana fornece informações contextuais para aplicações em esportes
<i>Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey</i>	IEEE Access	Demrozi <i>et al.</i>	2020	173	28,83	9,17	Esta pesquisa se concentra no desenvolvimento de aplicações HAR baseadas em sensores inerciais em conjunto com sensores fisiológicos e ambientais em vários domínios, incluindo

							esportes, saúde e aplicações de bem-estar.
--	--	--	--	--	--	--	--

Quadro 4 - Artigos mais citados

A análise permitiu observar as afiliações dos autores, o que fornece uma indicação dos principais centros de pesquisa dedicados ao tema. Entre os dez centros de pesquisa mais contabilizados, aparecem quatro instituições dos Estados Unidos da América, quatro universidades europeias e duas escolas asiáticas. Um dado relevante sobre a listagem de afiliações é a menção ao centro de pesquisa do *Hospital for Special Surgery*, uma das afiliações com mais artigos publicados na área. A lista está descrita no quadro 5.

<b>Afiliação</b>	<b>País</b>	<b>Artigos</b>
<i>National University of Singapore</i>	Singapura	23
<i>Stanford University</i>	Estados Unidos da América	23
<i>Hospital for Special Surgery</i>	Estados Unidos da América	21
<i>Shanghai Jiao Tong University</i>	China	20
<i>Northeastern University</i>	Estados Unidos da América	17
<i>Oxford University</i>	Reino Unido	17
<i>Università di Pisa</i>	Itália	17
<i>International Hellenic University</i>	Grécia	16
<i>Penn State</i>	Estados Unidos da América	16
<i>University College Dublin</i>	Irlanda	16

Quadro 5 - Afiliação dos autores

O agrupamento por rede de coocorrência, considerando palavras-chave extraídas de títulos de artigos citados (*Keywords Plus*) indicou um *cluster* principal, que concentra uma maioria de termos relacionados aos modelos de aprendizado de máquina. Entre dois pontos que aparecem com a mesma cor estão *performance* e *football*, indicando uma possível relação entre esses termos. Outras palavras associadas são *sport*, *prediction*, *exercise* e *human*. Esse agrupamento pode ser observado na figura 2.



Para esse levantamento situou-se no mesmo período, entre 2013 até 2024, utilizando termos de busca nas bases *Web of Science* e *Scopus* “*machine learning*” e variações do termo usado para buscar o esporte futebol “*football*” e “*soccer*” (e.g.: “*machine learning*” AND (“*football*” OR “*soccer*”)). Foram encontrados 399 títulos na *Web of Science* e 674 retornos na *Scopus*. Após remoção de resultados duplicados restaram 685 trabalhos. Optou-se por remover retratações e trabalhos sem indicação de autores, resultando em 606 documentos. Outro filtro foi remover os artigos que continham em seus resumos os termos *robots* e *robocup* relacionados às competições de futebol de robôs e, portanto, fora do escopo deste trabalho. Após esses filtros mantiveram-se 567 documentos.

O tópico de pesquisa tem uma taxa de crescimento anual ao longo do período observado de 36,94%, observando o crescimento do tema de maneira semelhante ao notado no levantamento da produção bibliométrica sobre *machine learning* e esportes. Foram observadas 394 fontes e 1.959 autores. A evolução da quantidade de trabalhos publicados pode ser vista na figura 3.

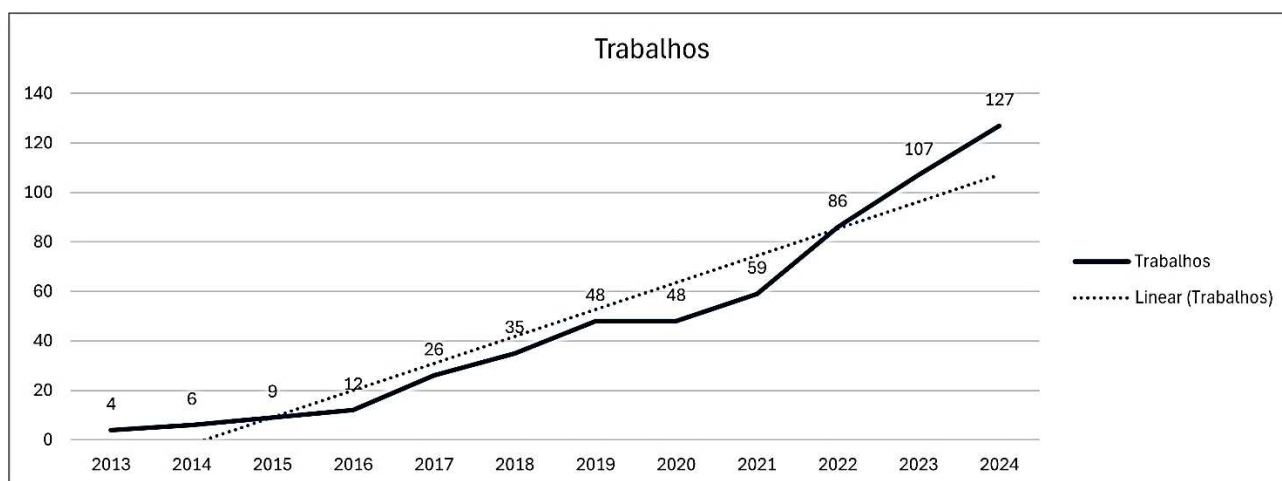


Figura 3 - Evolução do número de trabalhos publicados

Entre as fontes mais relevantes para publicação da temática estão o periódico *PLOS One* e o repositório aberto *CEUR Workshop Proceedings*. A listagem das principais fontes pode ser observada no quadro 6, mantendo seis fontes semelhantes às observadas no levantamento feito anteriormente.

Fontes mais relevantes	Publicações
<i>PLOS One</i>	15

<i>CEUR Workshop Proceedings</i>	12
<i>Applied Sciences-Basel</i>	9
<i>IEEE Access</i>	9
<i>Lecture Notes in Networks and Systems</i>	9
<i>Lecture Notes in Computer Science</i>	8
<i>Communications in Computer and Information Science</i>	7
<i>International Journal of Computer Science in Sport</i>	7
<i>International Journal of Sports Physiology and Performance</i>	7
<i>Machine Learning</i>	6

Quadro 6 - Fontes mais relevantes

Entre os dez autores mais proeminentes sobre o tema de aprendizado de máquina aplicado ao futebol quem aparece como autor com maior quantidade de publicações é Alessio Rossi, da *Uninversit degli Studi di Milano*, com dez artigos publicados no período. Depois aparece a professora Aida Mustapha, da *Universiti Malaysia Pahang Al-Sultan Abdullah*, com oito trabalhos publicados. Na sequência seis pesquisadores se apresentam com seis *papers* difundidos em duas das principais bases científicas do mundo. Os autores mais relevantes podem ser observados no quadro 7.

<b>Autor</b>	<b>Artigos</b>	<b>Centro de Pesquisa</b>	<b>País do centro de pesquisa</b>
Alessio Rossi	10	<i>Uninversit degli Studi di Milano</i>	Itália
Aida Mustapha	8	<i>Universiti Malaysia Pahang Al-Sultan Abdullah</i>	Malásia
Mohamad Razali Abdullah	6	<i>Universiti Sultan Zainal Abidin</i>	Malásia
Jesse Davis	6	<i>KU Leuven</i>	Bélgica
Pål Halvorsen	6	<i>University of Oslo</i>	Noruega
Fedon Marcello Iaia	6	<i>Uninversit degli Studi di Milano</i>	Itália
Tolga Kaya	6	<i>Istanbul Technical University</i>	Turquia
Daniel Memmert	6	<i>German Sport University Cologne</i>	Alemanha

Yuzhe Liu	5	<i>Beihang University/Stanford University/Tsinghua University</i>	China/Estados Unidos da América
David Camarillo	5	<i>Stanford University</i>	Estados Unidos da América

Quadro 7 - Autores mais proeminentes sobre o tema

O país cujos autores estão afiliados aos programas de pesquisa e que possui maior número de publicações são os Estados Unidos da América. Entre os outros dez países estão três asiáticos (China, Índia e Japão), cinco europeus (Alemanha, Itália, Reino Unido, Espanha e Portugal) e a Austrália. A relação pode ser observada no quadro 8.

<b>País</b>	<b>Frequência</b>
Estados Unidos da América	199
Itália	134
Alemanha	133
Reino Unido	97
China	92
Espanha	86
Índia	84
Austrália	70
Portugal	69
Japão	49

Quadro 8 - Países com maior número de autores filiados

Um gráfico gerado para complementar a análise bibliométrica foi o *three-field plot*, que considera interações entre três campos da amostra. Foi escolhida na posição da esquerda as palavras-chave extraídas de títulos de artigos citados (*Keywords Plus*), no centro os autores, e na coluna da direita as principais fontes. Esse fluxo indica que os nomes das técnicas de aprendizado de máquina são frequentes entre os títulos dos trabalhos. O termo performance se liga com mais frequência a Rossi, Abdullah e Liu. O gráfico pode ser observado na figura 4.

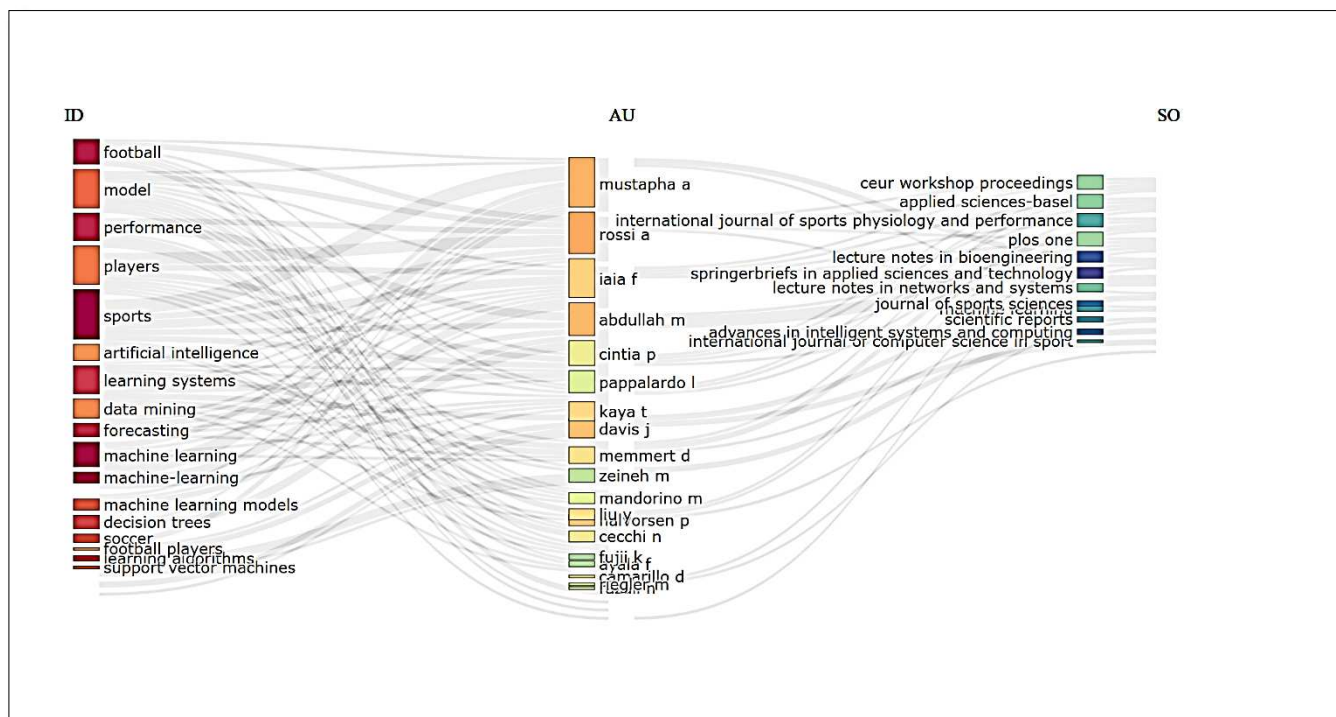


Figura 4 - Three field plot entre termos, autores e periódicos

Entre os dez trabalhos mais citados da amostra, três se concentram na previsão e prevenção de lesões no futebol utilizando modelagem em *machine learning*, e há outro trabalho na área de saúde cujo foco foram as lesões cerebrais em atletas aposentados (GOSWAMI *et al.*, 2016; LÓPEZ-VALENCIANO *et al.*, 2018; ROMMERS *et al.*, 2020; ROSSI *et al.*, 2018). O trabalho mais citado é sobre técnica de captação de dados em larga escala que pode ser aplicado em várias áreas, incluindo a captura de dados visuais de campos de futebol (HACKEL *et al.*, 2017).

O tópico sobre análise de performance e suas aplicações envolvendo *machine learning* e futebol, exatamente o tópico de pesquisa na qual se encontra a proposta de estudo desta tese, detém quatro trabalhos entre os dez citados. Em ordem temporal o primeiro é o de Rein e Memmert (2016), em um trabalho que debatia os desafios e oportunidades de utilizar técnicas de modelagem de dados no campo dos esportes. O segundo trabalho da área de performance é o de Bartlett *et al.* (2017), que avaliou o impacto de cargas de treino na performance utilizando ML. Outras aplicações, incluindo análise preditiva de resultados foram propostas por Baboota e Kaur (2019) e por Herold *et al.* (2019).

Por fim, há um trabalho envolvendo análise textual de comentários de torcedores de futebol nas redes sociais utilizando *machine learning* (BURNAP *et al.*, 2015). Esse trabalho explicita a variação de aplicações possíveis de modelagem de aprendizado de máquina e, igualmente, a

diversidade de campos envolvendo pesquisas com futebol, neste caso sobre o comportamento dos torcedores. A relação de trabalhos mais citados está disponível no quadro 9.

<b>Título</b>	<b>Autores</b>	<b>Citações</b>	<b>TC por ano</b>	<b>TC normalizada</b>
Semantic3D.Net: A New Large-Scale Point Cloud Classification Benchmark	Hackel <i>et al.</i> , (2017)	552	61,33	13,94
Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science	Rein; Memmert, (2016)	291	29,1	7,8
Effective injury forecasting in soccer with GPS training data and machine learning	Rossi, A. <i>et al.</i> (2018)	162	20,25	10,67
Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach	Bartlett J. <i>et al.</i> , (2017)	130	14,44	3,28
Detecting tension in online communities with computational Twitter analysis	Burnap P. <i>et al.</i> , (2015)	103	9,36	6,24
Predictive analysis and modelling football results using machine learning approach for English Premier League	Baboota, R.; Kaur, H. (2019)	91	13	6,19
Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions	Goswami, R. <i>et al.</i> (2016)	90	9	2,41
A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players	Rommers N. <i>et al.</i> , (2020)	88	14,67	6,29
A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms	López-Valenciano, A. <i>et al.</i> , (2018)	69	8,63	4,55
Machine learning in men's professional football: Current	Herold <i>et al.</i> , (2019)	66	9,43	4,49

applications and future directions for improving attacking play				
---	--	--	--	--

Quadro 9 - Trabalhos mais citados

Foi feito um acoplamento bibliográfico (*clustering by coupling*) agrupando os tópicos a partir das *Keywords plus*. Entre os agrupamentos há um envolvendo os termos *performance – players – reliability*, permitindo inferir que sejam trabalhos que lidaram com confiabilidade no processo de avaliação da performance de jogadores. Outro acoplamento associado ao tema central desta tese foi com os termos *football – model – performance* que, de maneira análoga, permite inferir uma linha de trabalhos dedicados a modelagem de performance no futebol. O gráfico do agrupamento pode ser observado na figura 5.

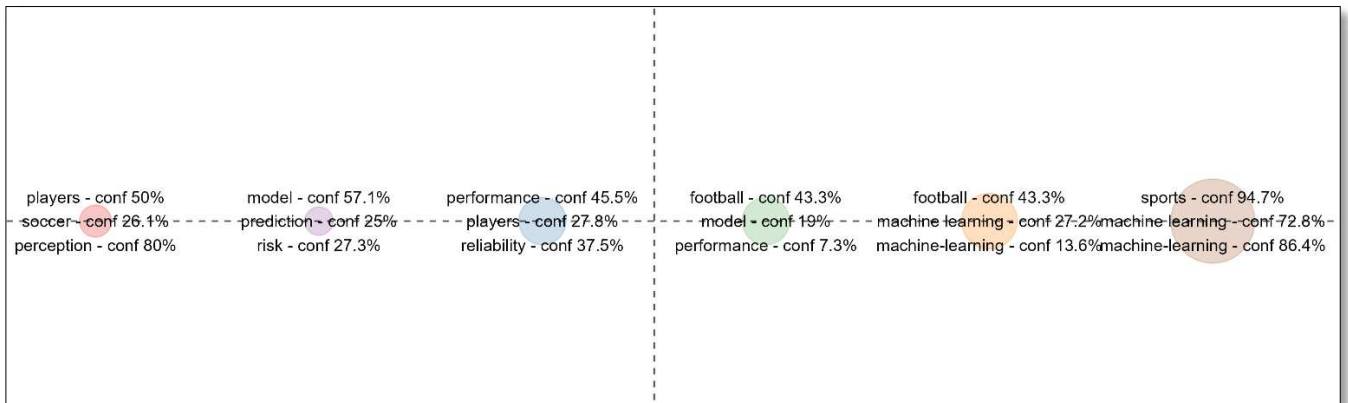


Figura 5 - Agrupamento bibliográfico

Para executar o agrupamento por rede de co-ocorrência optou-se por restringir algumas palavras da análise (*sports, sport, machine learnig, algorithms, machine learning algorithms, machine learning methods, football, soccer*). Ainda assim alguns destes termos constaram no mapeamento geral de termos considerando palavras-chave extraídas de títulos de artigos citados (*Keywords Plus*).

O resultado indicou um *cluster* principal, com termos relacionados ao aprendizado de máquina e seus principais algoritmos. Alguns termos destacados se conectam ao principal agrupamento, como *performance, athletic performance, prediction* entre outros. A rede de coocorrência pode ser observada na figura 6.

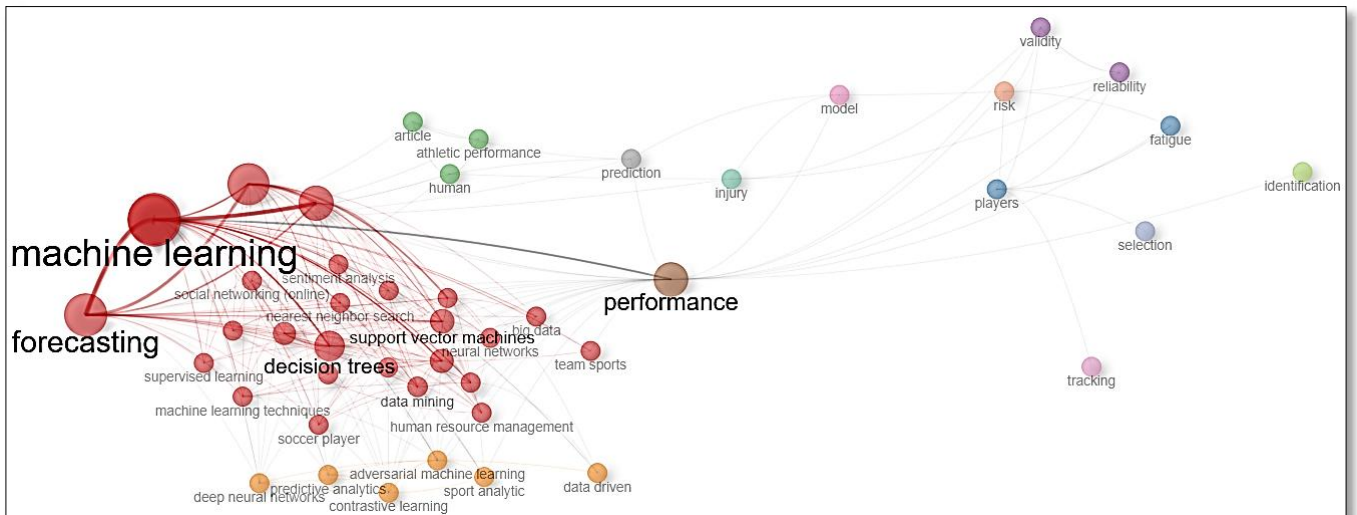


Figura 6 - Rede de co-ocorrência

O mapa temático realizado com a mesma amostra-base de palavras que a análise anterior identificou algumas das associações de termos que compõem as pesquisas da área. O risco envolvido na seleção de jogadores aparece como um dos temas, assim como classificação de performance além de pesquisas com fatores que impactam a marcação de gols nas partidas. Entre outros agrupamentos estão a prevenção de lesões. O mapa temático pode ser observado na figura 7.

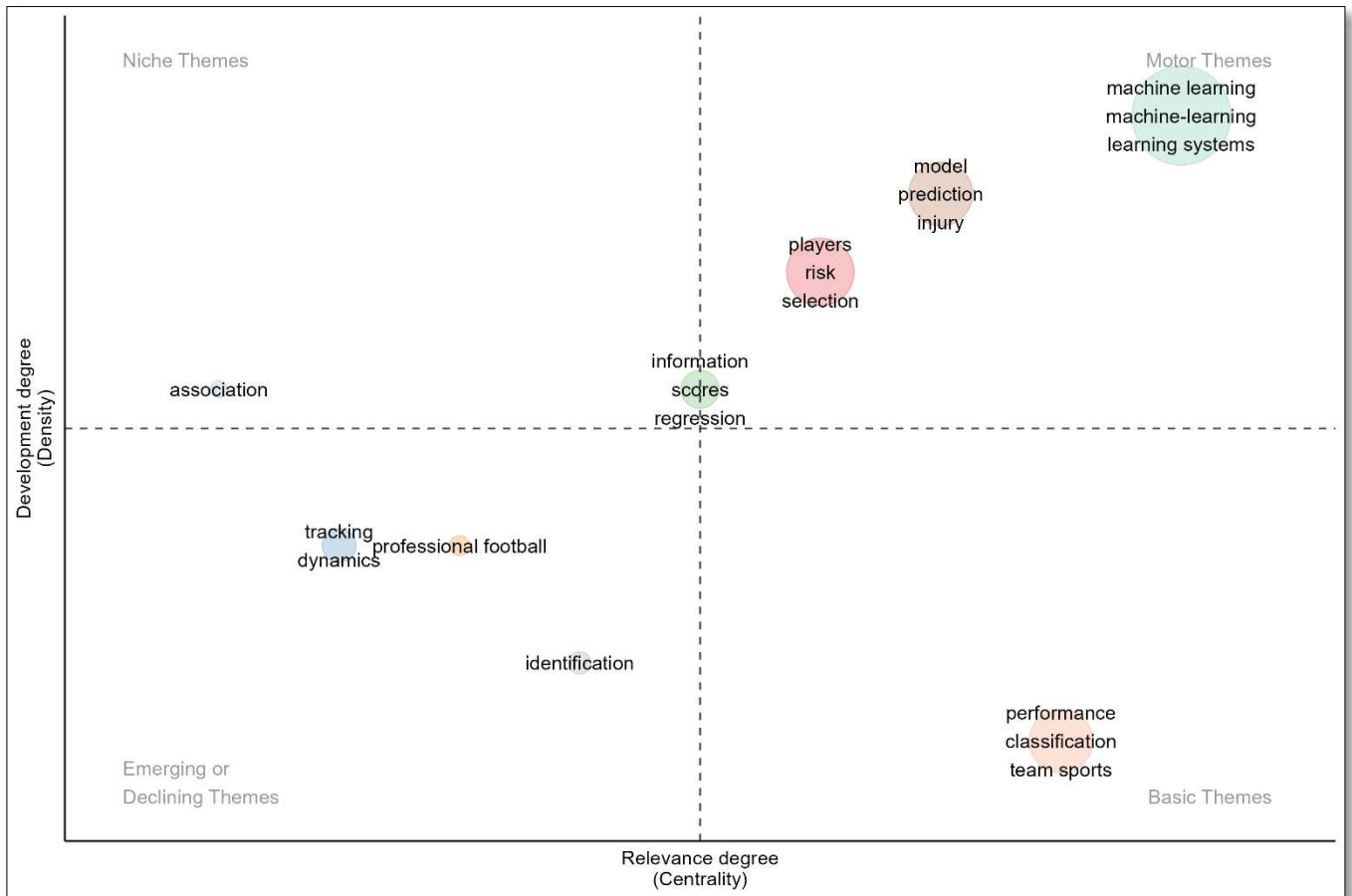


Figura 7 - Mapa temático

A observação sobre os agrupamentos temáticos da área tendo por filtro palavras-chave extraídas de títulos de artigos citados (*Keywords Plus*) indicou um grande *cluster*, aparentemente focado no desenvolvimento de técnicas e metodologia para aplicação de ML no futebol. A partir dos estudos dos modelos e algoritmos é que surgem trabalhos voltados à prevenção de lesões, seleção de atletas e informações sobre fatores relevantes nas marcações de gols. Essas propostas e, também, modelos de classificação de performance, parecem ser os principais temas de pesquisa publicados na área.

Alguns dos autores publicaram trabalhos em mais de um dos temas identificados na análise de *clusters*. Um exemplo são o conjunto de trabalhos liderados pelo departamento de ciência da computação da *Università di Pisa*, que divulgou trabalhos tanto no campo da prevenção de lesão, quanto de dados de performance dos atletas (PAPPALARDO *et al.*, 2019; PAPPALARDO; CINTIA, 2018; ROSSI *et al.*, 2018, 2019).

Executou-se uma análise sobre todos os resumos disponíveis dos 567 trabalhos, excluindo termos que poderiam ser os mais repetidos exatamente por serem os termos de busca e algumas

de suas variações. Em uma contagem de palavras as vinte mais citadas nos trabalhos foram listadas no quadro 10.

<b>Termos</b>	<b>Frequência</b>
<i>data</i>	914
<i>model</i>	586
<i>performance</i>	539
<i>players</i>	535
<i>based</i>	491
<i>results</i>	470
<i>models</i>	434
<i>match</i>	420
<i>analysis</i>	381
<i>player</i>	367

Quadro 10 - Contagem de palavras mais citadas nos resumos

Foram realizados dois ensaios para modelagem de tópicos utilizando algoritmo LDA, um teste classificando em dois grupos e o outro classificando em três grupos. O que parece distinguir as duas classificações é em relação a um tópico analisar a performance de equipe (*team*) enquanto outro agrupamento de pesquisas indica um foco em avaliar a performance individual (*players*), ao passo que classificar em dois conjuntos de tópicos comprime a análise de performance em um único grupo. O outro conjunto de tópicos observado nos dois ensaios se direciona a predição de resultados a partir dos dados de entrada. Os agrupamentos da modelagem de tópicos podem ser observados no quadro 11.

<b>1° agrupamento realizado</b>			<b>2° agrupamento realizado</b>		
<b>Tópico 1</b>	<b>Tópico 2</b>		<b>Tópico 1</b>	<b>Tópico 2</b>	<b>Tópico 3</b>
<i>data</i>	<i>performance</i>		<i>data</i>	<i>data</i>	<i>performance</i>
<i>based</i>	<i>models</i>		<i>match</i>	<i>model</i>	<i>analysis</i>
<i>results</i>	<i>data</i>		<i>prediction</i>	<i>models</i>	<i>data</i>
<i>match</i>	<i>model</i>		<i>results</i>	<i>team</i>	<i>based</i>
<i>prediction</i>	<i>players</i>		<i>based</i>	<i>performance</i>	<i>players</i>

Quadro 11 - Análise de tópicos

A partir dessas observações sobre as pesquisas que tratam sobre a aplicação de *machine learning* e futebol compreende-se que análise de performance e modelagens aplicáveis a estes ensaios é uma das principais subáreas de pesquisa. Neste contexto, para concluir a observação do campo de pesquisa da tese é necessário a compreensão da utilização de classificações difusas (*fuzzy*) e sua aplicação na subárea envolvendo a performance de atletas de futebol.

### **2.6.3. Avaliação de performance no futebol utilizando lógica *fuzzy***

No sentido de completar a análise bibliométrica sobre as áreas de pesquisa da tese, foi realizada um levantamento nas bases da *Web of Science* e *Scopus*. Procuraram-se os resumos (*abstracts*) que contivessem em seu texto os termos *fuzzy* e *performance* combinados às palavras relacionadas ao futebol ("*fuzzy*" AND "*performance*") AND ("*football*" OR "*soccer*"). Foram encontrados 78 documentos na base *Scopus* e outros 42 trabalhos na base *Web of Science*. Na junção das duas buscas excluiu-se 42 *papers* duplicados resultando em 78 entradas.

Por opção foram removidos da amostra os artigos que tratam de futebol de robôs (que continham os termos *robot*, *robots* e *robocup*), além dos trabalhos sem indicação de autoria (NA N no campo Author) e uma única retratação. Outro filtro aplicado após o início da análise foi remover quatro artigos que tratavam de aplicação de lógica *fuzzy* ao desenvolvimento de tecnologia para energia solar e distribuição energética. Desta maneira, a análise bibliométrica sobre a utilização de lógica *fuzzy* na avaliação de performance no futebol restringiu-se a 34 documentos, nos quais foram executadas as análises descritas a seguir.

Em uma visão geral sobre o tópico, foram 126 autores envolvidos na publicação dos 34 trabalhos em 33 fontes. A taxa de crescimento anual do tema é de 10,5%, como pode ser observado na figura 8. Com a publicação dos trabalhos ocorrendo de maneira dissipada, apenas uma fonte publicou mais de uma vez sobre o tema, o periódico *Neural Computing & Applications* da editora Springer.

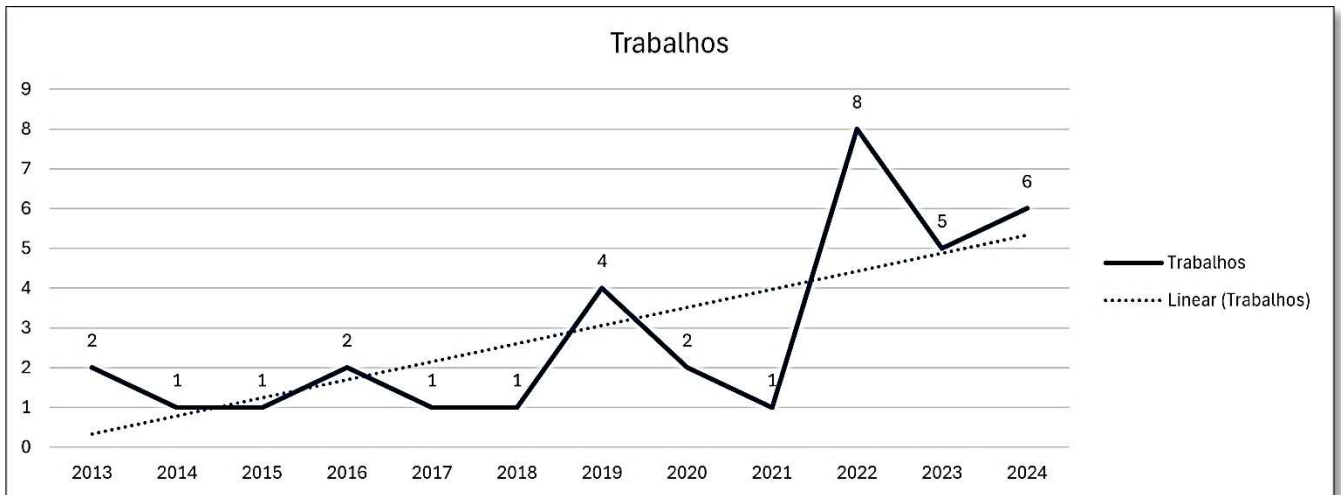


Figura 8 - Evolução dos trabalhos sobre o tema

Dos 126 autores envolvidos nas pesquisas, onze publicaram por duas vezes ao longo do período analisado. Em relação às universidades e centros de pesquisa aos quais os autores são afiliados uma única instituição possui quatro artigos publicados, é a *La Salle University* nos Estados Unidos da América. Outras onze universidades possuem três artigos publicados por autores afiliados a elas.

Os dez artigos mais citados obtidos na busca relacionando performance, *fuzzy* e futebol apontam para a utilização do sistema de inferência *fuzzy* (*fuzzy inference system*) em dois modelos para verificação na seleção e avaliação de atletas. Ainda há trabalhos que aplicam redes neurais utilizando abordagem *fuzzy*, assim como outros trabalhos propondo modelagem que utilizam a lógica difusa na análise de problemas relacionados ao futebol. Os trabalhos mais citados podem ser observados no quadro 12.

<b>Título</b>	<b>Autores</b>	<b>Citações</b>	<b>TC por ano</b>	<b>TC normalizada</b>
<i>A fuzzy inference system with application to player selection and team formation in multi-player sports</i>	Tavana <i>et al.</i> (2013)	61	4,69	3,54

Derivation of personalized numerical scales from distribution linguistic preference relations: an expected consistency-based goal programming approach	Tang <i>et al.</i> (2019)	28	4	2,87
Prediction of Attendance Demand in European Football Games: Comparison of ANFIS, <i>Fuzzy</i> Logic, and ANN	Şahin e Erol (2018)	25	3,13	2
A <i>Fuzzy</i> Inference System for Players Evaluation in Multi-Player Sports: The Football Study Case	Saġabun <i>et al.</i> (2020)	25	4,17	1,97
New group-based generalized interval-valued q-rung orthopair <i>fuzzy</i> soft aggregation operators and their applications in sports decision-making problems	Hayat <i>et al.</i> (2022)	24	8	3,69
Orchestrating the flow of human resources: Insights from Spanish soccer clubs	Fainshmidt <i>et al.</i> (2017)	22	2,44	1
Optimization analysis of football match prediction model based on neural network	Guan e Wang (2022)	15	3,75	3,87
A conditional <i>fuzzy</i> inference approach in forecasting	Hassanniakalager <i>et al.</i> (2020)	11	1,83	0,61

FD-LSTM: A <i>Fuzzy</i> LSTM Model for Chaotic Time-Series Prediction	Langeroudi <i>et al.</i> (2022)	8	2	2.06
A new similarity function for Pythagorean <i>fuzzy</i> sets with application in football analysis	Li <i>et al.</i> (2024)	8	4	4.8

Quadro 12 - Dez artigos mais citados

Executou-se a análise de coocorrência nas palavras contidas nos resumos obtidos, gerando um gráfico de rede para esses termos. Observou-se um núcleo que lida com a lógica *fuzzy* e algumas de suas técnicas possíveis de aplicação. Outro aglomerado contém termos relacionados a redes neurais e árvores de decisão, enquanto há um *cluster* com termos atrelados a análise de vídeos sobre futebol, possivelmente para sua aplicação na avaliação dos atletas. Por fim, um desses agrupamentos de termos aparentemente se relaciona com métodos configuracionais comparativos (MCC), com o mais proeminente dessas abordagens sendo a Análise Qualitativa Comparativa (QCA). A análise de coocorrência pode ser visualizada na figura 9.

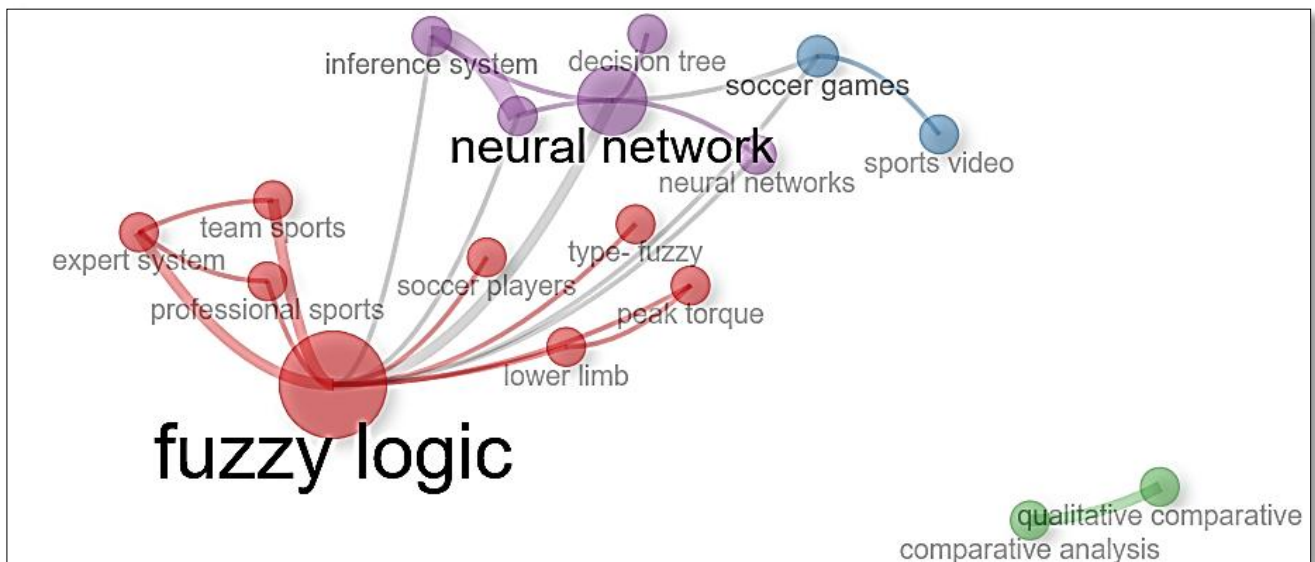


Figura 9 - Rede de co-ocorrência

Em uma observação sobre os temas agrupados, a partir das palavras contidas nos resumos, foi possível observar os grupos em relação aos sistemas de inferência *fuzzy*, a aplicação de redes neurais e aos algoritmos de clusterização *c-mean* aplicado em uma abordagem difusa. Outros

tópicos observados são o processamento de imagem e a análises qualitativas comparativas. O mapa temático pode ser observado na figura 10.

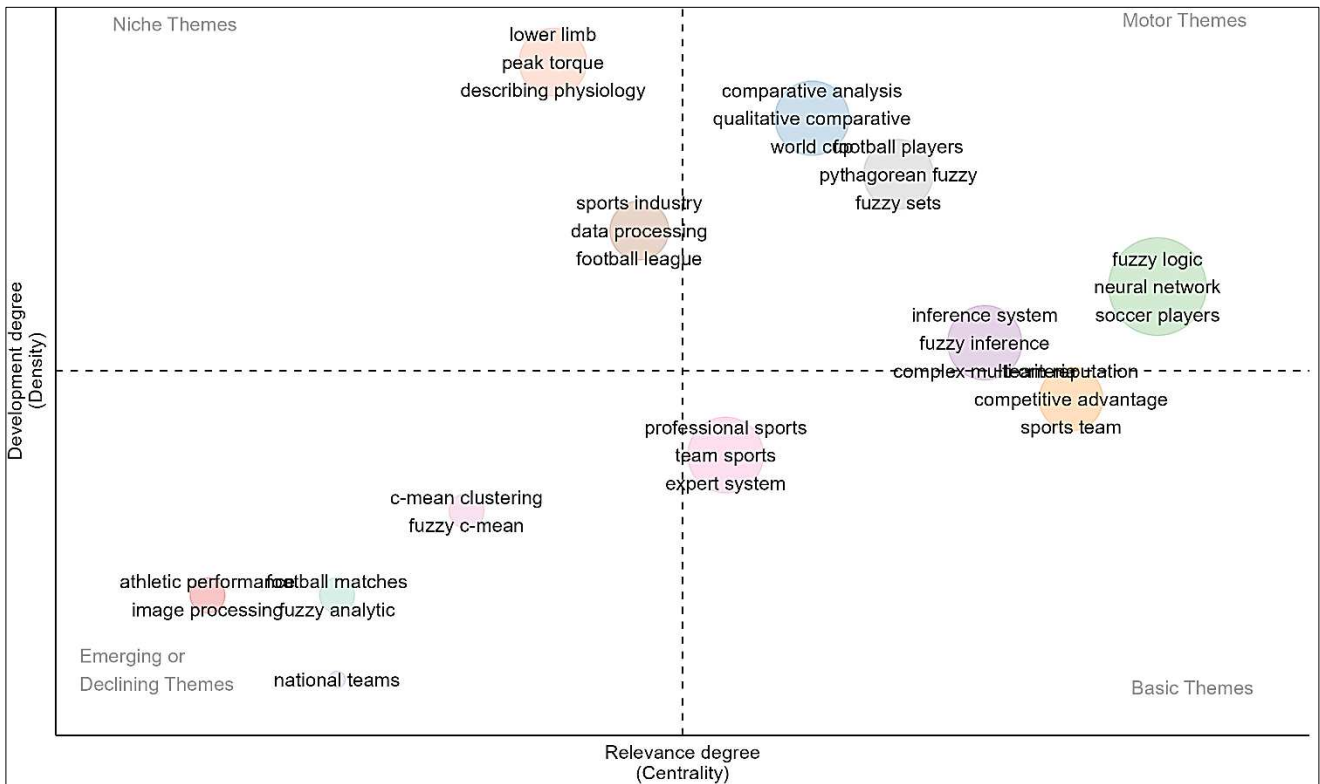


Figura 10 - Mapa temático

## 2.7.Considerações finais

Este capítulo teve como objetivo compreender a utilização da modelagem estatística baseada em aprendizado de máquina aplicada, primeiramente no setor esportivo, na sequência aplicada na análise de performance de jogadores de futebol. Por fim, de maneira a complementar o contexto de pesquisa desta tese, investigou-se a aplicação da lógica *fuzzy* no futebol, identificando artigos que tratam dessa vertente específica da modelagem de performance. Utilizou-se o RStudio e o algoritmo *Bibliometrix* para realizar uma revisão bibliométrica sobre o tema entre 2013 e 2024, publicados nas bases *Web of Science* e *Scopus*.

Observou-se com essa revisão que essa improvável mistura oriunda da teoria computacional com o campo da ciência do esporte vem adquirindo relevância à medida que os trabalhos vão se tornando mais frequentes. Ainda assim há uma demanda potencial por pesquisas que envolvam aspectos ainda não abarcados, quanto à sistematização da lógica *fuzzy* como um

componente fundamental na avaliação da performance de jogadores, especialmente no que diz respeito à sua integração com algoritmos preditivos de aprendizado de máquina.

Estes estudos demonstraram que técnicas como regressões penalizadas (*LASSO*, *Ridge*, *Elastic Net*) e modelos baseados em árvores de decisão (*Random Forest*, *XGBoost*) podem identificar padrões a partir de estatísticas de jogadores, permitindo a extração de pesos interpretáveis que podem ser aplicados na transformação *fuzzy* de dados de desempenho. Desta maneira, há possibilidade para emergirem pesquisas empíricas que explorem como a lógica *fuzzy* pode ser calibrada e integrada a modelos preditivos, aprimorando a classificação de jogadores e a projeção de seu potencial esportivo.

No contexto específico do futebol, a combinação de técnicas *fuzzy* com aprendizado de máquina demonstra potencial para reduzir a subjetividade na avaliação da performance dos atletas. Identificou-se um potencial para investigações futuras que abordem a definição de funções de pertinência e regras de agregação *fuzzy* específicas para diferentes posições em campo, assim como a possibilidade de modelar a evolução do desempenho ao longo do tempo. A integração de variáveis táticas e físicas em modelos híbridos de classificação *fuzzy* pode representar um avanço que permitiria aos clubes e analistas esportivos tomarem decisões fundamentadas para a prospecção e contratação de jogadores.

Ademais, os estudos analisados sugerem que a utilização de bancos de dados *fuzzy* no futebol pode aprimorar a capacidade de previsão de modelos preditivos, particularmente ao permitir que métricas de desempenho sejam interpretadas de maneira contínua e probabilística, ao invés de categórica e determinística. Isso se alinha com a crescente tendência de adaptação de métodos computacionais avançados para contextos dinâmicos, onde o desempenho esportivo depende de múltiplos fatores interrelacionados.

Dessa forma, esta revisão aponta direções estratégicas para futuras pesquisas, cujos achados podem contribuir para a evolução dos métodos de análise de performance no futebol, além de fornecer suporte para o desenvolvimento de ferramentas computacionais avançadas que auxiliem no *scouting* e na gestão esportiva. A mescla entre aprendizado de máquina e lógica *fuzzy* aplicada no futebol se apresenta como uma vertente analítica para compreender a performance dos atletas e para otimizar processos de avaliação e tomada de decisão no futebol profissional.

### 3. Fluxos de transferência e estratégia no futebol global

#### 3.1. Introdução

As transferências de atletas entre clubes de futebol é uma das decisões de alto significado para a área da gestão do futebol e suas estratégias. A negociação a respeito da cessão de jogadores movimenta significativos montantes em recursos financeiros e recursos de permuta anualmente, sendo uma das principais fontes de receita de alguns clubes e base para determinados modelos de negócio no futebol (BUCK; IFLAND, 2023). Em seu estudo global sobre expatriados no futebol, o *International Centre for Sports Studies* (CIES) apontou para o Brasil como país com mais jogadores atuando em outras nações durante o período (2017-2023) (POLI; RAVENEL; BESSON, 2023). Ainda assim, trabalhos mais recentes com o foco em entender o mercado de transferências poucas vezes abordam a liga brasileira e outras competições da América Latina e Ásia entre os dados analisados (MATESANZ *et al.*, 2018; WAND, 2022).

A transação envolvendo direitos econômicos dos atletas foi responsável por, em média, 20,8% da composição dos recursos dos clubes de futebol mais relevantes do Brasil nos últimos cinco anos (GRAFIETTI, 2024). Compreender maneiras de explorar a venda dos direitos econômicos como parte da estratégia da organização torna-se relevante, uma vez que os valores envolvidos nessas negociações são alguns elementos que influenciam na cotação das ações dos clubes de futebol na bolsa de valores (MAJEWSKI, 2016).

As cinco ligas com maiores receitas geradas com transferências, chamadas *Big Five* do futebol, são: Inglaterra, Espanha, Alemanha, Itália e França (WAGNER; PREUSS; KÖNECKE, 2021). Também são nas competições que participam as equipes que detêm a atenção de uma grande parte do público em todo o mundo e cujos clubes estão entre 20 maiores geradores de receitas do mundo, excluindo nesse cálculo a receita oriunda das transferências (DELOITTE, 2023).

Para ilustrar os montantes envolvidos nas negociações de direitos econômicos de jogadores, recorreu-se à plataforma *Transfermarkt* (<https://www.transfermarkt.de/>), que embora usualmente seja utilizada como fonte de trabalhos acadêmicos detém algumas falhas no seu regime de avaliação de valores de mercado dos jogadores (DETZEN; LÖHLEIN, 2023). Apesar disso, o banco de dados da plataforma não pode ser desprezado como fonte de informação mesmo diante de algumas dificuldades (DETZEN; LÖHLEIN, 2023). Alguns dados da plataforma oferecem apoio ao argumento de que os clubes brasileiros têm margem para evoluir nas suas estratégias de transferência de jogadores.

Por exemplo, o Brasil ainda é o país com o atleta cuja transferência movimentou maior montante (Neymar, do *Futbol Club Barcelona* para o *Paris Saint-Germain* em 2017 por 222 milhões de euros) (ver quadro 13), que também é o atleta que somou a maior quantidade de dinheiro movimentada em transferências (€ 400 milhões, até o fechamento desta tese). Em relação às transferências que deram maior lucro, os jogadores brasileiros são os mais frequentes entre as primeiras 25 transações. Esses são alguns indicativos de que os clubes brasileiros podem buscar alternativas para aumentar as receitas com transferências. A ineficiência de alguns clubes brasileiros nas operações que envolvem negociar a cessão ou aquisição de um novo atleta são parte das justificativas para os problemas financeiros dessas organizações que acumulam dívidas superior a R\$ 10 bilhões no Brasil (GRAFIETTI, 2024; RODRIGUES *et al.*, 2022).

<b>Atleta</b>	<b>Naturalidade</b>	<b>Valor da transferência (em milhões de euros) – valores corrigidos de acordo com a inflação pelo índice Euroland em novembro de 2024</b>
Neymar	Brasil	276,48
Kylian Mbappé	França	218,81
Ousmane Dembélé	França	167,80
Philippe Coutinho	Brasil	166,50
João Félix	Portugal	152,20

Quadro 13 - Transferências mais caras da história corrigidas pela inflação

Ainda na plataforma *Transfermarkt* é possível comparar as receitas geradas apenas com transferências de jogadores. Entre as temporadas 2018-2019 até a metade do ano 2023 (*i.e.*, 2022-2023) a *Série A* italiana liderava entre as receitas geradas (€ 4,52 bilhões), seguida pela *Premier League* inglesa (€ 3,80 bilhões), a *Ligue 1* francesa (€ 3,66 bilhões), *La Liga* espanhola (€ 3,46 bilhões) e a *Bundesliga* alemã (€ 2,73 bilhões). Ainda à frente do Campeonato Brasileiro estão a Liga Portuguesa (€ 2,08 bilhões), a *Championship* (segunda divisão do futebol inglês) (€ 1,53 bilhões) e a *Eredivise* holandesa (€ 1,39 bilhões). Alcançar o patamar do futebol português pode ser uma proposta de referência para o futebol brasileiro. Portugal é o país com mais jogadores expatriados do Brasil, detém a liga de futebol com o maior saldo entre receitas e despesas com a transferência de jogadores e o único a superar um bilhão de euros nesta balança (€ 1,21 bilhões).

Nas últimas três décadas a dinâmica das propriedades de times de futebol mudou (KIRSCH *et al.*, 2024; SÁNCHEZ *et al.*, 2021). Diante disso uma maior compreensão dos efeitos dessas estruturas na dinâmica interna dos clubes e no mercado de transferência de jogadores é necessária. Estudos como os de Sánchez *et al.* (2021), Pastore (2018) e Buck e Ifland (2023) apontam para a lacuna sobre governança e táticas organizacionais dentro dos clubes. A esse contexto aplicou-se uma metodologia como a análise de redes sociais (*Social Network Analysis*) (SNA) para mapear os movimentos de jogadores entre clubes e entre países. Este capítulo explora o padrão de transações de jogadores de futebol no fluxo de transferências entre países. Com essa análise será possível inferir quais rotas já estão estabelecidas.

O referencial teórico deste capítulo discorrerá sobre o tema, buscando arcabouço teórico no campo das estratégias organizacionais para direcionar a sua vantagem competitiva em relação a outros modelos de negócio. A partir desse entendimento será apresentada uma análise sobre o tema, a evolução desse procedimento para delimitar as teorias do campo e uma proposta de abordagem deste fenômeno no campo das estratégias. A compreensão das dinâmicas de transferências fornecerá argumentos para outras estratégias de negócios, relacionadas ao recrutamento e seleção de atletas, objetivo principal desta tese.

### **3.2. Formulação e implementação de estratégia nas organizações**

No compasso em que as transações comerciais se expandiram em nível global, o advento de tecnologias 4.0 com a expansão de produtos/mercados, diferentes modelos e configurações corporativas e de negócios, o campo acadêmico de estudos de estratégia foi se expandindo, se tornou mais complexo. Entre os modelos do desenvolvimento teórico do campo da administração estratégica, há o modelo pendular de evolução teórica apresentado Hoskisson *et al.* (1999). Na estrutura proposta por estes autores, o “olhar” do estrategista segue a analogia de um pêndulo em permanente oscilação em que se convencionou, por exemplo, para a direita, olhar para o ambiente interno da organização e para a esquerda, para o ambiente externo. Nessa oscilação do olhar, do exame, se analisam oportunidades e ameaças em cada estágio. O pêndulo é analógico e nas suas trajetórias para os limites se dá ênfase nas análises dos ambientes. O desenvolvimento da teoria em estratégia portanto oscila entre o ambiente interno e o ambiente externo à firma (*inside out e outside in*), em que esse comportamento permanente tem poder de explicação da heterogeneidade do desempenho organizacional.

O modelo pareceu captar algo da realidade do desenvolvimento teórico em estratégia, pois pouco depois de sua publicação, a perspectiva das capacidades dinâmicas começou a se consolidar, mantendo ênfase no ambiente interno, mas destacando a mudança dos ambientes externos e a sua nova dinamicidade (EISENHARDT; MARTIN, 2000). Em outra proposta de integração para a analogia pendular, Guerras-Martin; Madhok e Montoro-Sanchez (2014) inseriram um segundo movimento pendular, que oscila entre visões macro e micro, tal que, na posição macro se inserem os trabalhos que entendem a formulação da estratégia como um processo de planejamento formal orientado para a organização, enquanto a configuração micro compreende a estratégia como uma sequência contínua das micro atividades dos atores organizacionais. Adicionalmente, pode-se dizer que essas visões orientam as decisões e, portanto, mantêm os “motores” estratégicos da organização em processo sustentável para o crescimento e rentabilidade.

Uma outra proposta é o chamado quadro de estratégia (*strategy frame*) (KHALIFA, 2008). Essa proposta, de cunho mais gerencial, também organizou as principais perspectivas da área em termos de ambiente interno e externo; de maneira análoga, dividiu a área entre a orientação para o futuro versus orientação para o presente. A figura 11 apresenta uma sugestão de combinação das propostas desses dois modelos em um único *framework*.

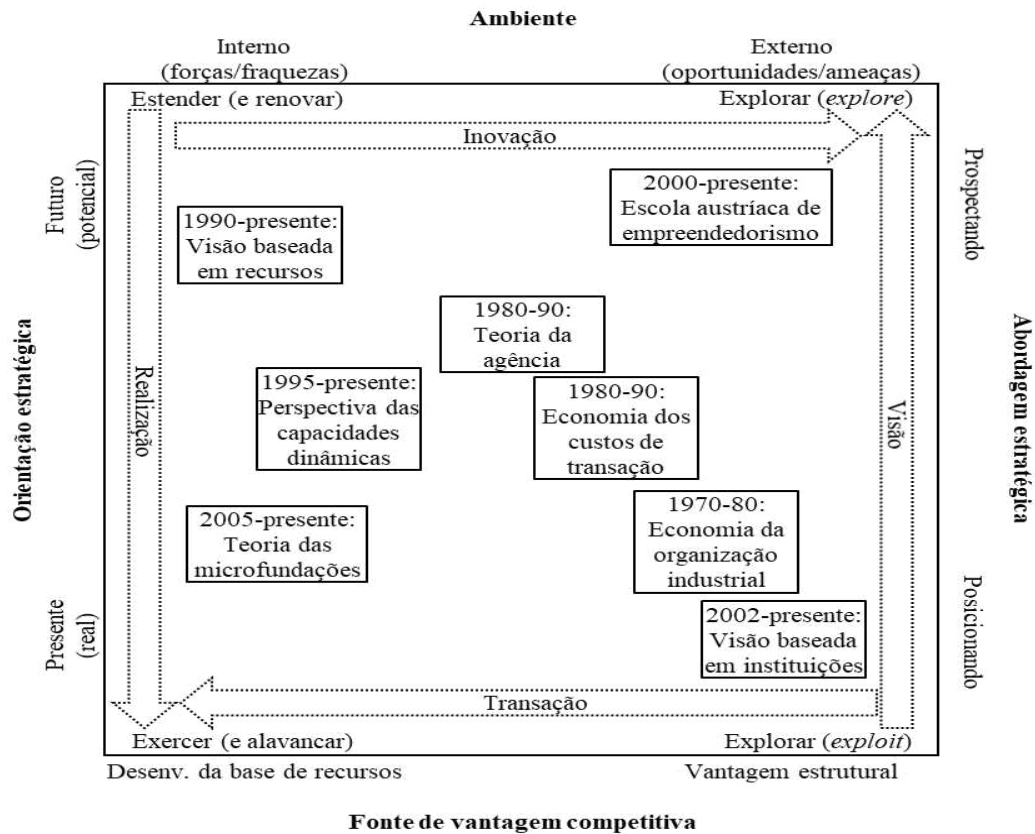


Figura 11 - Framework estratégia (KHALIFA, 2008)

Considerando esse contexto, duas plataformas teóricas são as bases da coesão da pesquisa em estratégia: (i) a pesquisa sociológica e sociocognitiva; e (ii) a microeconomia aplicada ao estudo da competição e a teoria penroseana do crescimento da firma (FURRER; THOMAS; GOUSSEVSKAIA, 2008; KENWORTHY; VERBEKE, 2015). Fundamentada nessa divisão estão os paradigmas da estratégia-como-conteúdo, aplicada nas investigações de variáveis críticas para interpretação da heterogeneidade de desempenho organizacional. A outra vertente é a estratégia-como-processo, com enfoque nas pesquisas sociológicas e psicológicas da dinâmica e atores da *estrategização (strategizing)* (MELLAHI; SMINIA, 2009). Desta segunda vertente é que se origina o conceito de estratégia-como-prática, linha que vem ganhando espaço por conectar-se melhor a outros fluxos de pesquisa, da gestão estratégica à teoria institucional (BURGELMAN *et al.*, 2018).

Há que se mencionar as estratégias corporativas e a de negócios. Enquanto a corporativa está mais focada na composição dos vários negócios e suas contribuições ao conglomerado, a estratégia de negócios está focada no portfólio de produtos, nos mercados, fornecedores e clientes, resultados por cada negócio em operação. No nível gerencial, essa estratégia é usada

para desempenhar uma variedade de funções, incluindo suporte à tomada de decisões e coordenação e comunicação de objetivos ou propósitos estratégicos (FUERTES *et al.*, 2020). Uma estratégia organizacional deve estabelecer a visão de futuro da empresa e avaliar objetivamente sua posição atual a fim de determinar como alcançar os objetivos estabelecidos (PEPPARD; WARD, 2016). Uma boa estratégia é um conjunto coeso de análises, conceitos, políticas, argumentos e ações que respondem a objetivos, metas para alto desempenho, ou respondem desafios de alto risco ou incerteza (RUMELT, 2012). As organizações passam por transformações e reestruturação de recursos e capacidades, em que é preciso decidir sobre a melhor estrutura organizacional para alcançar uma vantagem competitiva (BENAVIDES ESPINOSA; MERIGÓ LINDAHL, 2016).

A formulação de estratégia combina perspectivas variadas, incluindo processos cognitivos conscientes, deliberados e formais, bem como processos inconscientes, intuitivos e informais. A estratégia resultante é imbuída de formulação cognitiva racional e refletida, assim como conteúdo emergente relacionado à percepção empírica do estrategista e do ambiente externo. O equilíbrio dos diferentes elementos é essencial para obter a melhor estratégia de negócios para um caso específico. Por assumir uma abordagem empresarial, a estratégia deve ser uma combinação equilibrada de processos cognitivos conscientes e intuitivos.

### **3.3. Capacidades Dinâmicas**

No processo da evolução do campo da administração estratégica a maioria das pesquisas antes da década de 1990 foram voltadas para obtenção de vantagem competitiva focada na identificação de oportunidades e ameaças de mercado. Nesse contexto, Barney (1991) elaborou o seu arcabouço teórico com a Visão Baseada em Recurso (*Resource Based View*) (RBV), estabelecendo o modelo VRIN em que uma organização possuiria uma vantagem competitiva sustentada a partir do atendimento de quatro critérios: valor, raridade, inimitabilidade e insubstituibilidade.

A evolução dessa abordagem conduziu ao desenvolvimento dos conceitos de Capacidades Dinâmicas (TEECE; PISANO; SHUEN, 1997) e, mais especificamente, das Capacidades Gerenciais Dinâmicas (DMC), que exploram fatores como o capital humano, o capital social e a cognição gerencial (ADNER; HELFAT, 2003). Um recurso só confere vantagem competitiva se for raro, e se empresas – ou clubes – compartilham recursos de valor semelhante, esses

recursos podem evitar uma desvantagem competitiva, mas não podem garantir vantagem por si só.

No futebol, especificamente, isso significa que um clube precisa ser mais eficaz que seus concorrentes na seleção de atletas, seja no cenário doméstico ou no cenário internacional. Isto é, o clube precisa se destacar na previsão consistente do valor futuro desses jogadores, em um ambiente competitivo, em que os clubes disputam talentos em nível global. As organizações que possuem melhores capacidades de seleção conseguirão distinguir com mais precisão os jogadores que terão impacto positivo, assegurando uma alocação mais eficiente dos seus investimentos.

Nesse sentido, as capacidades são um tipo especial de recurso cujo propósito é aprimorar a produtividade dos outros recursos disponíveis à organização. Elas não podem ser transferidas ou adquiridas, são inerentes à organização, integradas aos processos e rotinas internas (KERO; BOGALE, 2023; TEECE, 2019). Como argumentado por Teece *et al.* (1997), essas capacidades são enraizadas de maneira que sua transferência só é possível com a própria organização ou com uma unidade suficientemente autônoma dela. Ademais, a capacidade de uma organização será economicamente valiosa se ela for capaz de adquirir ou desenvolver recursos cuja produtividade será ampliada pelas suas capacidades. No contexto do futebol, clubes que conseguem maximizar a performance dos seus jogadores podem gerar retornos no mercado de transferência (HOLZMAYER; SCHMIDT, 2020).

Por outra perspectiva, clubes que não detêm essas capacidades se tornam dependentes da sorte para obter uma vantagem competitiva, isto é, obtêm retornos inesperados sobre jogadores estratégicos. Ou seja, sem uma capacidade de previsão consistente, o sucesso nesses casos se deve ao acaso mais que a um processo deliberado (HOLZMAYER; SCHMIDT, 2020). Quanto mais eficaz for a capacidade preditiva de um clube no recrutamento de novos atletas, menor será o papel da sorte. Clubes que não detêm as capacidades necessárias podem ser beneficiados pelo acaso no mercado de transferências, mas só terão uma vantagem competitiva sustentada se forem capazes de reproduzir esses ganhos de forma sistemática (MUSTAFI; BAYLE; TERRIEN, 2024).

Jogadores formados internamente são um dos recursos valiosos de um clube. São atletas que se destacam pelo impacto direto na identidade e no desempenho do time, por consequência são importantes no mercado de transferências (RUBIO MARTIN *et al.*, 2022). Nesse sentido, torna-

se relevante compreender como os clubes possuem as capacidades de recrutamento e desenvolvimento que os distinguem dos seus concorrentes, criando vantagem competitiva no mercado de transferências. A projeção e mensuração dessa vantagem é o objetivo deste trabalho.

Entende-se que há oportunidades de investigação sobre as trocas de ativos (jogadores de futebol) entre clubes. Também é relevante compreender como isso afeta os contextos migratórios nas ligas de futebol, assim como a implicação na importância de clubes e ligas. Outra questão se dá na projeção das características gerenciais de CEO's e as equipes de alta gerência (*top management team*) (TMT). Em relação à avaliação de atletas já se destacou em outro capítulo a importância das avaliações de desempenho e atributos individuais. No entanto há um campo para ser explorado em relação a como as Capacidades Gerenciais Dinâmicas (DMC's) dos proprietários e gerentes de clubes afetam as capacidades de atração de novos atletas. Esse conceito é uma extensão das Capacidades Dinâmicas que foca no papel dos gestores, tanto individualmente quanto em equipes (HELFAT; MARTIN, 2015).

### **3.4. Metodologia**

A análise de redes sociais (ARS) ou em inglês *Social Network Analysis* (SNA) é uma técnica para fazer diagnósticos em vários contextos, desde o estudo de territórios até o mapeamento de redes organizacionais (FIALHO, 2015). A abordagem é relevante para identificar estruturas dentro das redes. Ao mapear as interações e a densidade das redes sociais é possível compreender as dinâmicas em jogo dentro do contexto avaliado (FIALHO, 2015).

A SNA se insere na teoria dos grafos para investigar estruturas sociais, consistindo em métodos relacionais que permitem a compreensão sistemática das conexões entre diferentes entidades (BORGATTI *et al.*, 2013). Essas entidades são tipicamente representadas por pessoas, grupos ou organizações, e suas interações são o ponto central dessa abordagem. Portanto, as redes sociais podem ser entendidas como estruturas formadas por participantes que interagem em busca de um objetivo comum, de maneira colaborativa ou individual (WASSERMAN; FAUST, 1994).

As redes sociais são compostas por nós (entidades) e laços (relações), que podem ser analisadas em termos de intensidade, direção, centralidade e reciprocidade (BRACKE, 2016). A SNA oferece um arcabouço teórico para compreender a estrutura das redes visualizando a sua complexidade (HEWA WELEGE; PAN; KUMARASWAMY, 2021). A análise de redes

sociais proporciona uma compreensão das interações que formam as redes de interdependência, explicando tanto o comportamento individual quanto as estruturas sociais (FIALHO, 2015).

A análise compreende sistemas sociais em diferentes níveis. A díade examina as relações par-a-par entre os atores. No nível do nó estudam-se as características dos atores em relação a como a quantidade de laços que eles possuem interferem nos seus resultados. Em um nível da rede observa-se o comportamento da rede como um todo, como a capacidade de disseminação de informações em redes conectadas (BORGATTI *et al.*, 2013).

Nos ensaios realizados para este capítulo trabalhou-se com os conceitos de centralidade, em que o grau de centralidade avalia a quantidade de conexões de cada nó. Neste conceito também é possível compreender a centralidade de intermediação, identificando nós que atuam como intermediários nas redes, o que se aplica à análise para identificar os clubes que facilitam a movimentação dentro da rede (CAMACHO *et al.*, 2020). A transitividade é a formação de triângulos, o que indica como os países tendem a formar *clusters* ou blocos coesos, identificando ligas que compartilham jogadores com frequência (CAMACHO *et al.*, 2020).

### 3.5. Resultados

O autor desta tese iniciou o trabalho de obtenção de dados a partir de um *script* para a interface RStudio/2023.06.0+421 na versão 4.4.1 do R. O objetivo do código era realizar *web scraping* na plataforma Transfermarkt, e essa função extraiu dados de transferência, com limitações em relação ao banco de dados da plataforma, que por vezes não detinha os registros para alguns países. No total foram observadas 212.830 transações, desde a temporada 2010 até a temporada 2023, incluindo as saídas (*departures*) e chegadas (*arrivals*) de jogadores.

Para remover as duplicações e otimizar a capacidade de processamento, optou-se por trabalhar apenas com as informações de chegadas de jogadores, contabilizando 103.790 negociações. As informações sobre os negócios incluem chegadas em definitivo e as contratações por empréstimo (*on loan*), os valores envolvidos (*transfer fee*), a idade, o nome, a posição do atleta, o país e clube em que o jogador chegou e o país e o clube de onde o atleta se desligou.

### 3.6. Fluxo de Transferência entre Países – Todas as Transferências

A primeira etapa da análise visou identificar os principais padrões de movimentação internacional de jogadores de futebol entre países no período de 2010 a 2023. Por meio da

Análise de Redes Sociais (ARS), utilizou-se métricas como grau de centralidade (envio e recepção de jogadores), centralidade de intermediação (papel como elo entre outras transações), centralidade de proximidade (distância geodésica na rede) e proporção de transferências internas, além da identificação de tríades, que são estruturas de três países conectados por transferências sucessivas.

Os resultados apontaram para Bélgica e Portugal como os principais eixos do sistema internacional de transferências. Ambos os países se destacaram não apenas no número de transferências recebidas e enviadas, mas também como nós intermediários recorrentes em diversas tríades. Tais posições reforçam o papel dessas ligas como pontos de passagem estratégicos para o desenvolvimento, valorização e reexportação de atletas. Esse contexto possivelmente é relacionado às políticas de recrutamento flexíveis, acordos bilaterais e redes de múltipla propriedade.

Emirados Árabes Unidos, Estados Unidos e Suíça também figuram entre os países com maior centralidade e intermediação, indicando envolvimento ativo nas redes internacionais. Considera-se, ainda, que suas proporções de transferências internas estejam entre as mais altas da amostra, sugerindo mercados também voltados para o desenvolvimento doméstico.

No que diz respeito à proporção de transferências internas, ligas como a da Tchêquia, China, Polônia e México apresentaram mais de 85% de movimentações dentro do próprio país. Esse movimento contrasta com mercados internacionalizados, tais como França, Espanha, Alemanha e Inglaterra, onde as transferências internas não ultrapassam os 60%.

A análise das tríades revelou padrões de circulação mais densos envolvendo Portugal, Bélgica, Emirados Árabes, EUA e Suíça funcionando como etapas para outros mercados. Notou-se, por exemplo, que Inglaterra, Alemanha, França, Itália e Espanha são frequentemente o destino das transações, com diversos países intermediando a chegada dos jogadores.

O Brasil apresentou um comportamento misto. Quando figura como país de origem, suas transações frequentemente passam por países intermediários antes de alcançar mercados mais valorizados, sendo ele próprio também um elo intermediário em diversas redes. Isso reforça a posição de país exportador e transitório, mas cuja capacidade de captação direta de valor nos principais mercados ainda é mediada por terceiros.

Essa primeira análise revelou também existência de sub-redes, chamadas na análise de redes sociais de comunidades, dentro do sistema global de transferências, com alguns países

ocupando funções estruturais distintas. São padrões de desenvolvedores, vitrines intermediárias e receptores da elite do esporte. A compreensão desses papéis fornece a base para compreender o mercado de transferências como fluxo de atletas e como estratégia organizacional e econômica.

Por fim, os resultados sugerem hipóteses para as seções seguintes, como o impacto de redes de propriedade cruzada, acordos preferenciais e estruturas de governança local na inserção internacional de atletas. A próxima etapa, isolando as transferências definitivas, permitirá compreender de maneira mais apurada os padrões estruturais das negociações e como eles evoluíram ao longo do tempo.

### **3.7. Fluxo de Transferência entre Países (Transferências Definitivas)**

Ao isolar as transferências em definitivo, a análise da rede global de movimentações de jogadores de futebol entre países revelou alterações relevantes na estrutura da rede e na posição relativa dos países quando comparadas às transferências totais. A centralidade das nações, medida por indicadores como grau de centralidade, centralidade de intermediação e proximidade, manteve alguns padrões já observados anteriormente, com destaque para Bélgica, Portugal, Estados Unidos da América e Emirados Árabes Unidos. Esses países reafirmaram seu papel no ecossistema das transferências internacionais pelo volume de transações e por seu papel de conectores entre outras regiões.

Contudo, a separação das transferências definitivas evidenciou movimentações relevantes na hierarquia da rede. A China, por exemplo, perdeu espaço em alguns indicadores de centralidade e frequência nas tríades, o que pode indicar uma redução de protagonismo nas transferências definitivas, sinalizando a mudança de estratégia dos clubes chineses em relação às políticas de transferência. Em contrapartida, os Países Baixos e a Áustria emergem como países com ganho de relevância, especialmente como pontos de origem nas cadeias de transferências.

Outro ponto de destaque refere-se à centralidade de intermediação, que novamente posiciona Portugal, Bélgica, EUA e Emirados Árabes como facilitadores recorrentes nas rotas de movimentação de atletas, inclusive nas transações definitivas. Países como a Suíça e a Polônia mantêm papel relevante, mas com flutuações mais modestas entre as análises. O gráfico com a rede global de transferências em definitivo pode ser observado na figura 12.

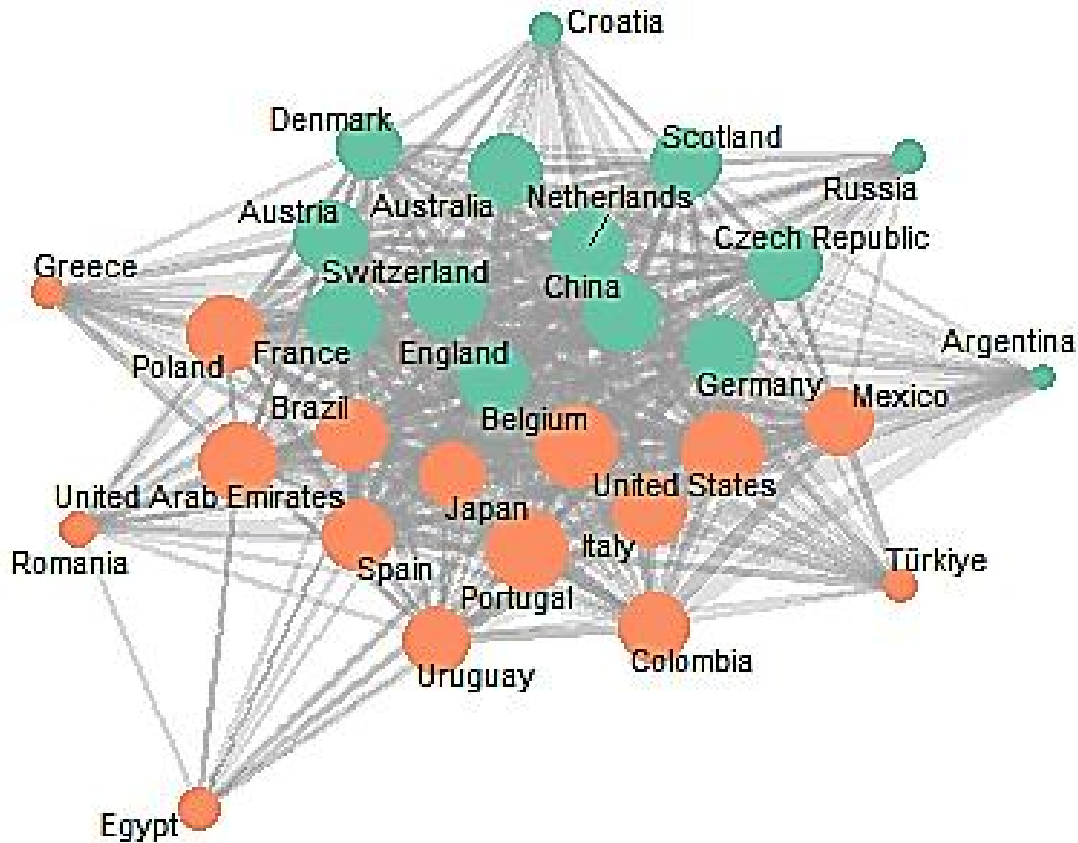


Figura 12 - Rede global das transferências em definitivo (2010–2023)

A análise das tríades confirmou a tendência de concentração de rotas estratégicas e bem definidas, em relação aos países do *Big-5* europeu (Alemanha, Espanha, França, Inglaterra e Itália). Nesses casos, embora os destinos se mantenham, houve alterações nos países de origem e de intermediação. Observa-se, por exemplo, a maior frequência dos Países Baixos e da Áustria nas rotas de origem das transferências definitivas para esses países, ao passo que China e Brasil perderam espaço como intermediários.

Quando se analisam especificamente as tríades em que o Brasil surge como país de origem, intermediário ou destino, nota-se um decréscimo no protagonismo nas transferências definitivas, com a perda de frequência em várias cadeias. É possível observar, por exemplo, que Espanha, Alemanha e Escócia deixam de ser intermediários nas rotas originadas no Brasil. Quando assume a posição de intermediário, o Brasil também perde posições em favor de países como Países Baixos e Colômbia, refletindo uma possível limitação da sua atuação como eixo de exportação em processos estruturados de negociação internacional.

Ainda assim, o Brasil permanece como um nó importante, principalmente pelo volume absoluto de transações e sua presença recorrente em diferentes etapas da rede. A redução de sua

participação nas tríades definitivas, no entanto, pode sinalizar que parte das movimentações envolvendo clubes brasileiros ainda se baseia em empréstimos, ou que há maior dispersão e informalidade no processo de negociação com clubes de fora da rede de intermediação dominante.

Um dado observado foi a consolidação de blocos de influência regional, mesmo que a análise de comunidades ainda não tenha indicado subgrupos robustos nesta etapa (em parte pela concentração excessiva de conexões em poucos *clusters*). A observação por intervalo de temporadas permitiu uma leitura mais refinada sobre esses blocos emergentes.

Por fim, o recorte para as transferências em definitivo, ao eliminar as transações por empréstimos que são utilizados como mecanismo de adaptação ou vitrine, revelou as relações estruturais entre os mercados com maior capacidade de absorção e redistribuição de atletas. Tal como sugerido no arcabouço teórico da tese, estes países se configuram como sistemas com capacidades organizacionais superiores de captação e redistribuição de recursos escassos (*i.e.*, talentos esportivos), refletindo suas capacidades dinâmicas e estratégias posicionais no mercado internacional de jogadores.

### **3.8. Fluxo de transferência entre países – Todas as transferências por intervalos de temporada**

A estrutura global de transferências de jogadores de futebol apresenta, ao longo do tempo, uma estabilidade em seus padrões relacionais. Para investigar os detalhes dessa estabilidade, optou-se por uma análise em recortes temporais, agrupando as transferências bianualmente entre as temporadas de 2010 e 2023. Cada agrupamento contemplou dois anos consecutivos, totalizando sete intervalos distintos. Essa divisão permitiu verificar a persistência de determinadas configurações, bem como observar eventuais mudanças na composição da rede e nas posições ocupadas pelos países dentro dela.

Ao longo de todos os intervalos, os resultados mantiveram consistência, em que as métricas do grau de centralidade, centralidade de intermediação e centralidade de proximidade apresentaram padrões praticamente inalterados, com variações envolvendo países periféricos na rede, como a inversão pontual entre Tailândia e Cazaquistão em alguns dos períodos observados. Os países centrais se mantiveram constantes, implicando em uma estrutura da rede de transferências que opera com uma lógica consolidada.

Nesse arranjo, o Brasil é o país com o maior grau de centralidade normalizado, indicando sua relevância tanto como origem quanto como destino de transações. A posição brasileira reforça sua importância como um dos principais celeiros de talentos no futebol mundial, com participação representativa no fluxo internacional de jogadores. Na sequência, Estados Unidos da América, Itália, Colômbia e França completam a lista dos países mais centrais em termos de conectividade geral com os demais nós da rede.

Apesar da proeminência brasileira em centralidade, a intermediação das transferências está concentrada em outros países. A Bélgica, por exemplo, ocupa consistentemente a posição de maior centralidade de intermediação, o que caracteriza o país como ator estratégico na configuração do fluxo de talentos. Bélgica e Portugal são, frequentemente, os elos que conectam jogadores oriundos de diferentes países aos principais mercados europeus. Emirados Árabes Unidos e Estados Unidos também desempenham esse papel de intermediários, sugerindo que seus clubes e redes de agenciamento operam com capilaridade nas dinâmicas de mercado. Essa posição de ambos os países, que historicamente não detêm grande relevância entre as ligas de futebol do mundo, possivelmente reflete a estratégia dos casos de atletas que optam por deixar o mercado europeu por interesses financeiros.

A centralidade de proximidade, por sua vez, reforça a ideia de uma rede altamente conectada. Um grupo considerável de países apresenta valores idênticos ou muito próximos, denotando que os caminhos de transferência entre os nós são curtos, e a distância geodésica entre os clubes é reduzida. Esse cenário representa uma malha interligada, com pouca assimetria estrutural entre os principais atores do futebol internacional.

Complementando a análise quantitativa, o estudo das comunidades revelou subgrupos recorrentes de países cujas interações ocorrem com mais frequência entre si do que com o restante da rede. Seis comunidades binacionais se destacaram por sua persistência ao longo dos intervalos, refletindo laços históricos, geográficos ou estratégicos. Entre elas, a relação entre Inglaterra e Escócia implica em uma integração sistêmica reforçada por sua proximidade territorial e cultural. Analogamente, Portugal e Brasil mantêm uma conexão evidente, sustentada por fatores linguísticos, histórico-coloniais e interesses de mercado. As demais comunidades, tais como França e Bélgica, Argentina e Uruguai, Tchêquia e Eslováquia, Estados Unidos e Canadá também derivam de vínculos próximos, sejam eles geopolíticos, históricos ou organizacionais.

A análise das tríades reforçou essa impressão. Países como Bélgica, Portugal, Estados Unidos e Emirados Árabes Unidos aparecem frequentemente como origens de trajetórias de transferência, enquanto Alemanha, Inglaterra, Brasil, Espanha e Itália consolidam-se como intermediários recorrentes. Já os destinos, majoritariamente representados pelas cinco grandes ligas europeias, confirmam a atratividade dos clubes dessas nações como ponto final da jornada de muitos atletas. Neste conjunto, destaca-se a consistência com que a Bélgica aparece como ponto de partida das transferências, reforçando seu papel como plataforma de lançamento para jogadores em direção a centros mais prestigiados.

Ao observar as trajetórias em que o Brasil aparece como país de origem, intermediário ou destino, nota-se uma dinâmica menos intermediária do que outros países da rede. O Brasil surge como exportador direto de talentos, tendo o Uruguai, Japão e Dinamarca aparecendo como principais países intermediários para atletas oriundos do futebol brasileiro. Por outro lado, quando posicionado como intermediário, o Brasil se vincula a uma diversidade de países de origem, como Portugal, Bélgica e Estados Unidos, o que pode indicar operações de reencaminhamento de atletas dentro do próprio continente americano ou por meio de parcerias pontuais com clubes estrangeiros. Quando analisado como destino, o Brasil também mantém relações com países como Inglaterra, Alemanha, Espanha e Itália, sugerindo repatriações ou retorno de atletas que atuaram por períodos fora do país.

Esses resultados sugerem que os fluxos de transferência de atletas, embora sujeitos a circunstâncias econômicas e contextos regulatórios, operam sob uma estrutura relacional consolidada. A rede de transferências funciona como um sistema adaptativo complexo, em que alguns países mantêm posições centrais estratégicas e outros exercem funções de suporte, seja como originadores, receptores ou articuladores logísticos dos deslocamentos. Nesse sentido, entender o posicionamento dos países não apenas em termos absolutos, mas também em relação às posições que ocupam nas tríades (origem, intermediação e destino), permite compreender as lógicas de circulação global de talento que sustentam o mercado futebolístico atual.

### **3.9. Fluxo de transferência entre países – Apenas transferências em definitivo por intervalos de temporada**

Com o intuito de compreender as dinâmicas das transferências definitivas de jogadores ao longo do tempo, esta seção recorreu à análise em recortes bianuais, agrupando as temporadas entre 2010 e 2023. Essa abordagem temporal permitiu observar a estabilidade dos padrões centrais da rede e, ao mesmo tempo, identificar nuances na evolução das relações entre os países.

Os resultados apontaram para a consistência nos valores e nas ordens de grandeza das medidas de centralidade, intermediação e proximidade. Em todos os intervalos analisados, o Brasil manteve a posição de país com maior grau de centralidade, indicando sua inserção nas redes internacionais de transferência, seguido por Estados Unidos, Itália, Colômbia e França. Isso reforça o protagonismo do país como nó relevante no mercado global de transferência de atletas.

No que diz respeito à centralidade de intermediação, a Bélgica aparece como principal elo entre diferentes regiões da rede, atuando como facilitadora do fluxo de atletas entre mercados. A ela se somam Emirados Árabes Unidos, Portugal, Estados Unidos e Austrália, cuja atuação como intermediários revela o papel estratégico dessas nações em promover a mobilidade internacional de jogadores.

Já a análise de centralidade de proximidade, que mede o quão acessível um país é em relação aos demais, reafirma a relevância da Bélgica também nesse aspecto. Outros países europeus, como Portugal, Suíça, Polônia e Países Baixos, também se destacam, sugerindo estruturas densas de conexão e uma alta capacidade de se conectar rapidamente com diferentes mercados.

A análise de comunidades, ou subgrupos formados por países que possuem mais interações entre si do que com os demais, revelou formações que refletem laços históricos, culturais e regionais. Foram identificados agrupamentos como Inglaterra-Escócia, Brasil-Portugal, França-Bélgica, Alemanha-Países Baixos, além de Argentina, Uruguai e México formando um bloco latino-americano de interações. Também surgem pares históricos como Tchêquia-Eslováquia e a comunidade entre Estados Unidos e Canadá, que remonta à proximidade geopolítica e institucional de suas ligas. Esses agrupamentos implicam a ideia de que, embora o futebol seja um mercado global, ele opera por meio de redes regionais interligadas, onde a cultura, a geografia e a história moldam os caminhos possíveis para as transferências de atletas.

Quando observadas as tríades nas transferências definitivas, essas estruturas que são formadas por um país de origem, um intermediário e um destino, os padrões anteriores foram mantidos, com algumas variações que merecem destaque. A Bélgica segue como o país com maior frequência na posição de origem, seguida de Portugal, Estados Unidos e Emirados Árabes Unidos. Tais países têm se consolidado como pontos de partida importantes nas rotas globais de transferência.

Os países que mais ocupam a posição de intermediários (*Country 2*) são Alemanha, Inglaterra, Brasil, Espanha e Itália. Esse padrão é idêntico ao padrão anterior, o que sugere que a frequência



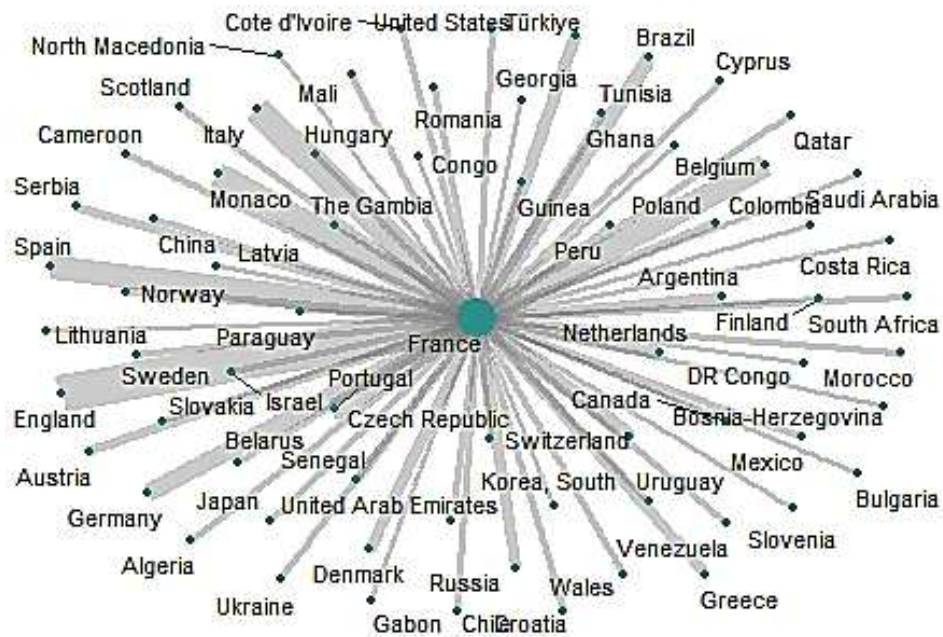


Figura 14- Triádes centrais com França como destino

O gráfico das triádes tendo a Itália como destino está representado na figura 15.

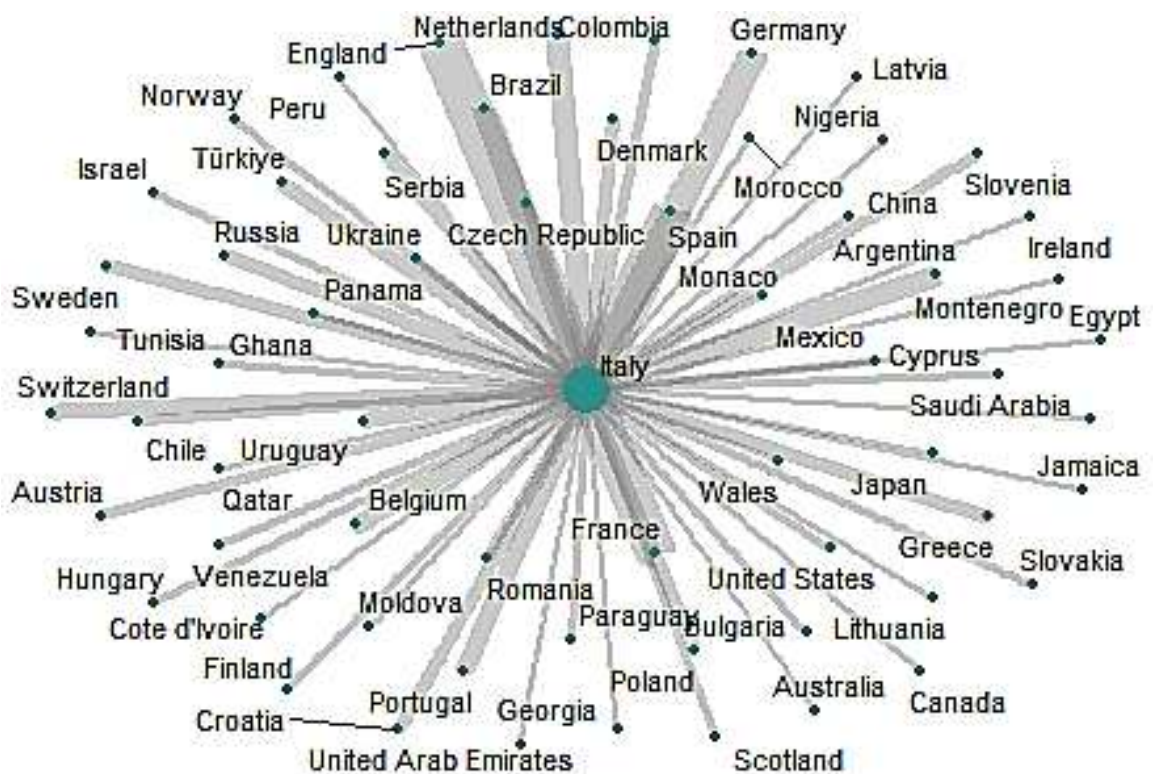


Figura 15 - Triádes centrais com Itália como destino

No caso da Alemanha e da Espanha, destaca-se a ascensão da Colômbia como intermediária, indicando o país sul-americano com um papel de ponte entre mercados locais e o futebol europeu. A redução do Brasil nas posições intermediárias, observada especialmente nas



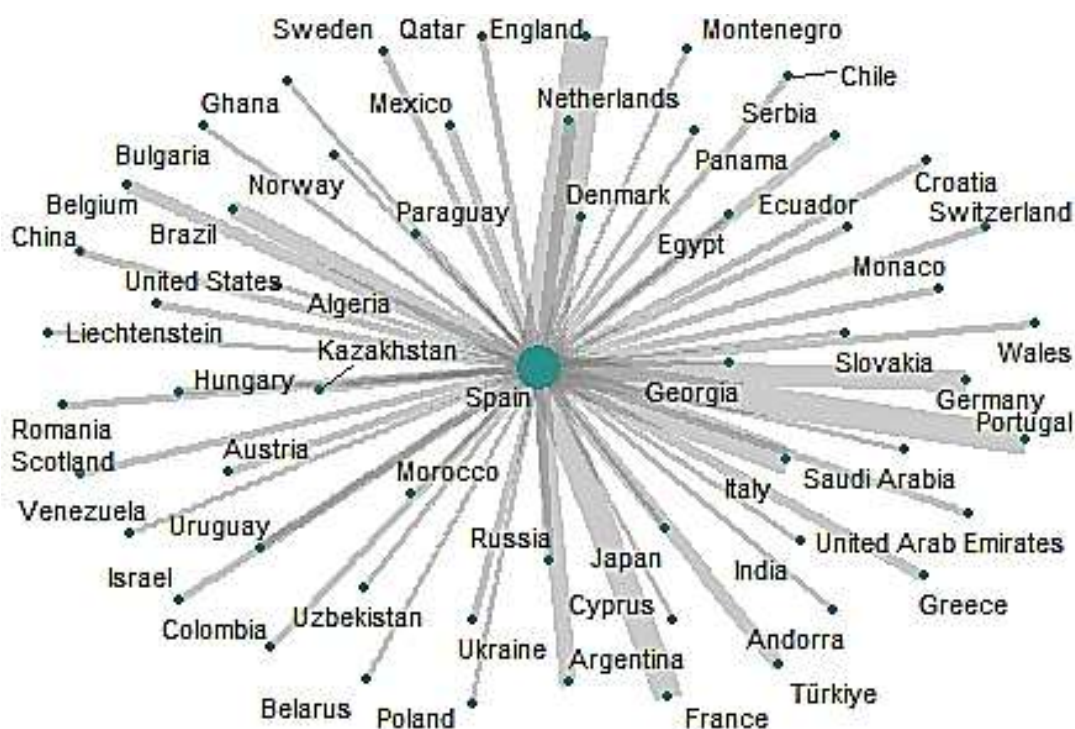


Figura 17 - Tríades centrais com Espanha como destino

Ainda em relação ao Brasil, ao assumir a posição de origem nas tríades, os principais intermediários são México, Uruguai, Japão e Dinamarca, que são países que operam como escalas para a exposição e adaptação dos atletas brasileiros. A relação do Uruguai como intermediário para jogadores saídos do Brasil envolve a atuação do Club Deportivo Maldonado, responsável por receber 44,3% das transações. Este clube vem sendo utilizado como ponto estratégico de negociações por direitos econômicos de atletas, motivado pela legislação uruguaia que mitiga a carga tributária imposta nessas transações (EXAME, 2014).

Quando o Brasil atua como intermediário, recebe jogadores oriundos de Portugal, Bélgica, EUA e Emirados Árabes, servindo como ponte para diversos outros destinos, demonstrando sua capacidade de articulação e influência nas rotas internacionais. Por fim, como país de destino, o Brasil recebe jogadores especialmente de países como Bélgica, Portugal, Estados Unidos e Emirados Árabes Unidos, com intermediação frequente de Inglaterra, Alemanha e Espanha, reiterando sua inserção no fluxo de repatriação ou negociação de atletas cuja carreira potencialmente terminou ou necessita de um novo processo de alavancagem.

Em síntese, mesmo com a exclusão das transferências por empréstimo, os padrões gerais da rede permanecem relativamente estáveis, reforçando a ideia de uma estrutura consolidada e eficiente. No entanto, as transferências em definitivo revelam detalhes, como a ascensão de intermediários alternativos, a reorganização das rotas e o papel de mercados tradicionalmente

considerados secundários. Tais informações ajudam a compreender os caminhos estabelecidos pelos jogadores e fornecem informações relevantes para decisões estratégicas em recrutamento, negociação e planejamento de carreira, elementos essenciais da gestão contemporânea do futebol.

### **3.10. Conclusão**

Os ensaios empreendidos neste capítulo permitiram observar, por meio da análise de redes sociais (ARS), a estrutura que define os fluxos de transferências internacionais de jogadores de futebol ao longo das últimas treze temporadas. A partir da modelagem de mais de 100 mil transações, o estudo revelou padrões de centralidade, intermediação e proximidade entre países, sugerindo que o mercado global de jogadores opera com base em conexões relativamente estáveis e altamente estratégicas. Clubes e ligas não atuam de forma isolada, mas como nós interdependentes de uma rede global que articula interesses esportivos, econômicos e organizacionais.

Sob a perspectiva da estratégia organizacional, esses fluxos demonstram como determinados países, tais como Bélgica e Portugal, consolidaram posições centrais na rede, atuando como eixos logísticos e simbólicos nas cadeias globais de valor do futebol. Outros países com posição de relevância neste cenário são os Estados Unidos da América e o Emirados Árabes Unidos, embora com algumas características distintas. O cenário estadunidense provoca a interpretação de uma liga em crescimento, que adotou reposicionamento estratégico na rede global de transferências, se movendo do ponto de destino de jogadores em final de carreira para um ambiente propenso à prospecção e desenvolvimento de jovens talentos. Por sua vez, a liga dos Emirados Árabes Unidos e sua posição na rede global de transferências sugere uma competição que atrai jogadores no estágio final de suas carreiras, que abraçam a possibilidade de um contrato financeiramente valioso antes de seguirem para as ligas em que irão encerrar suas carreiras.

Essa posição pode ser interpretada como resultado da combinação de capacidades dinâmicas organizacionais com recursos intangíveis específicos, como sistemas de prospecção, redes de relacionamento institucional e expertise gerencial. Em termos da teoria da estratégia, tais países se aproximam do conceito de vantagem competitiva sustentada, ao operarem de forma sistemática e repetida como intermediários ou portais de entrada para mercados mais valorizados.

Os dados também dialogam diretamente com os princípios da estratégia como prática, uma vez que a eficácia desses fluxos está ancorada na estrutura das ligas e nas decisões microgerenciais que envolvem a interpretação de dados, negociação de contratos e antecipação de oportunidades. A presença consistente de determinados países como intermediários indica que o futebol reproduz padrões de racionalidade econômica, ainda que inseridos em um ambiente de incertezas e subjetividades, características típicas dos mercados simbólicos.

Nesse cenário, o Brasil ocupa uma posição ambivalente. Por um lado, apresenta centralidade elevada como ponto de origem, refletindo sua tradição como exportador de talentos. Por outro, revela vulnerabilidades estratégicas ao depender de rotas externas para agregar valor aos seus ativos. A baixa presença do Brasil como país intermediário, especialmente nas transferências em definitivo, sugere uma fragilidade no aproveitamento das capacidades gerenciais dinâmicas que poderiam permitir maior captura de valor ao longo do ciclo de vida dos jogadores.

Essas evidências reforçam a necessidade de ferramentas estratégicas que auxiliem os clubes na execução das transações e na avaliação dos atletas, os principais ativos envolvidos nessas negociações. Se a vantagem competitiva está na capacidade de prever o valor futuro de um jogador, conforme sugerido pela abordagem das capacidades dinâmicas, então a existência de instrumentos analíticos capazes de qualificar essa previsão torna-se um ativo organizacional em si. Em outras palavras, clubes que dominam a avaliação de atletas sob múltiplos critérios (técnico, físico, financeiro e estratégico) tendem a transformar incerteza em vantagem.

Dessa forma, os resultados apresentados neste capítulo mapeiam rotas e fluxos de transferências, evidenciando a necessidade de integrar essas informações com instrumentos refinados de avaliação de jogadores. A compreensão das redes foi a primeira etapa, e a próxima é utilizar esse conhecimento para fortalecer os mecanismos internos de tomada de decisão dos clubes. Essa transição será explorada no próximo capítulo, em que se discutirá o desenvolvimento de um modelo de avaliação estratégica de atletas, com base em indicadores objetivos e subjetivos, orientado pelas necessidades de clubes inseridos em diferentes posições da rede global do futebol.

## 4. Um modelo difuso para pontuar estatísticas de jogadores

### 4.1. Introdução

Muitos clubes de futebol estabeleceram departamentos de análise de desempenho, ou *scouting*, e metodologias foram desenvolvidas para avaliar os aspectos do perfil de um jogador (SANTOS, 2012). Parte dos objetivos do *scouting* é fornecer informações sobre treino, desempenhos individuais de jogadores e identificar jovens atletas com potencial de desenvolvimento (ALMEIDA, 2010).

A definição de indicadores de performance significativos e combinados com os planos, modelos e estratégias de jogo são a base da análise de desempenho no futebol (ZAMBOM-FERRARESI; RIOS; LERA-LÓPEZ, 2018). As interações decorrentes do equilíbrio e desequilíbrio da organização da equipe são fatores críticos para compreender e influenciar o desempenho nos esportes coletivos (GARGANTA, 2009). Combinar métodos de pesquisa qualitativa com modelagem matemática, simulação e técnicas computacionais é fundamental para avaliar a contribuição de cada jogador no sistema de jogo proposto e obter uma compreensão abrangente e relevante da prática esportiva (GARGANTA, 2009; REIN; MEMMERT, 2016).

Identificar atletas com potencial de alto desempenho demanda um entendimento extenso das obrigações da performance e das capacidades do atleta (DAY, 2011). As investigações sobre indicadores de desempenho no esporte de alto rendimento buscam o "Santo Graal" do futebol, em busca de tornar as contratações mais eficazes e econômicas, utilizando modelagem baseada em método científico (HUGHES *et al.*, 2012; WILLIAMS, 2009). Essa motivação é associada ao conceito do *Moneyball*, trabalho de Billy Beane e Paul DePodesta no time de beisebol Oakland Athletics, que utilizou estatísticas de jogos anteriores para identificar características essenciais em cada uma das posições (WILLIAMS; WARD, 2017).

A geração de dados relacionados ao futebol é grande e trabalhos alertam para a necessidade de uma metodologia criteriosa na avaliação de desempenho dos atletas. Os principais indicadores de desempenho foram identificados em um estudo que demonstrou o potencial do aprendizado de máquina para ser utilizado no processo de prospecção e recrutamento no futebol profissional (HEROLD *et al.*, 2019). Embora esses aspectos sejam considerados preditivos do sucesso na partida e fundamentais na identificação de talentos, alguns clubes ainda os avaliam de forma intuitiva e pessoal (YI *et al.*, 2018; ZHOU *et al.*, 2018; MARTINS, 2019). Uma abordagem

comparativa entre jogadores de uma mesma posição poderia gerar contribuições importantes para definir escalas e treinamentos em equipes de variados escalões.

O futebol requer que um atleta se destaque não apenas em um aspecto, seja ele físico, técnico, tático, psicológico, clínico, administrativo e social, mas pela combinação deles (HUGHES *et al.*, 2012; MEYLAN *et al.*, 2010; PAOLI, 2007; SARMENTO *et al.*, 2018; ZHOU *et al.*, 2018). Apesar disso, o elemento técnico é base fundamental para atingir uma performance superior (SARMENTO *et al.*, 2018). A proficiência técnica é a característica mais importante que os treinadores e olheiros procuram, porque acredita-se que ela preveja o desempenho (WALDRON; WORSFOLD, 2010). Vários estudos descobriram que habilidades técnicas avançadas, como dribles, passes curtos/longos e chutes a gol, fornecem informações úteis para sistemas de identificação de talentos (WALDRON; WORSFOLD, 2010). Assim, várias pesquisas buscam entender e codificar esses indicadores (SARMENTO *et al.*, 2014, 2018; SHERWIN; CAMPBELL; MACINTYRE, 2017; YI *et al.*, 2018; ZHOU *et al.*, 2018).

Segundo a estrutura teórica da dinâmica ecológica, o comportamento é entendido por meio da relação dinâmica entre o praticante e o ambiente (DAVIDS *et al.*, 2013). Uma variedade de elementos molda o desenvolvimento dos atletas, a macroestrutura e a microestrutura de histórias e práticas contextualizadas devem ser consideradas na aquisição de habilidades esportivas, desempenho especializado e desenvolvimento de talentos (DAVIDS *et al.*, 2017).

Além disso, jogadores mais habilidosos parecem superar jogadores menos habilidosos em termos de conhecimento declarativo (saber o que fazer) e procedimental (como fazer) (ARAÚJO *et al.*, 2019). No processo de seleção de jogadores para as categorias do futebol juvenil a velocidade e qualidade na tomada de decisão são determinantes quando comparados àqueles que não são escolhidos (MACHADO; GONZÁLEZ-VÍLLORA; TEOLDO, 2023).

Um exame da relação entre cognição e desempenho no esporte enfatiza a interdependência entre atleta e ambiente, incorporando ideias da teoria dos sistemas dinâmicos, da ecologia e da psicologia sociocognitiva (VAUGHAN *et al.*, 2019). Além disso, as restrições socioculturais devem ser consideradas como um componente crítico no desenvolvimento de um atleta e um catalisador para a aprendizagem contínua no desporto (VAUGHAN *et al.*, 2019). Alavancar o potencial humano no desporto requer uma abordagem multidisciplinar para abordar as intrincadas interações entre indivíduos, contextos e desenvolvimento humano (VAUGHAN *et al.*, 2019).

A abordagem dos sistemas de desenvolvimento humano vem recebendo atenção nas ciências do comportamento. Observar fenômenos por esta ótica significa dar ênfase a ideia fundamental de que o desenvolvimento é um sistema complexo composto de múltiplos níveis, incluindo biológico, psicológico, sociológico e cultural, que interagem de maneira complexa e interdependente (BAUER; SHANAHAN, 2007). Incluindo a Teoria Ecológica (LERNER, 1998). No entanto, há divergências sobre a melhor abordagem para analisar sistemas de desenvolvimento humano. A utilização de “análises específicas para pessoas” (*person-specific*) e “análises centradas nas pessoas” (*person-centered*) são abordagens possíveis para estudar o desenvolvimento (HOWARD; HOFFMAN, 2018).

Os métodos específicos da pessoa analisam a variação intraindividual, enquanto os métodos centrados na pessoa podem funcionar tanto com a variação interindividual como com a variação intraindividual (WOO *et al.*, 2018). Nesse sentido, as análises centradas na pessoa oferecem um ajuste entre a parcimônia da abordagem centrada na variável e a abordagem específica da pessoa (HOFMANS; WILLE; SCHREURS, 2020). Essas análises buscam identificar configurações de variáveis que possam explicar processos comportamentais (REY-MARTÍ *et al.*, 2022; SAUNDERS; BUCKMAN; PILLING, 2020).

Estas abordagens têm uma vantagem no estudo de aspectos complexos, particularmente aqueles que envolvem interações confusas entre variáveis (GABRIEL *et al.*, 2018; MORIN; BUJACZ; GAGNÉ, 2018). Métodos emergentes, como o agrupamento difuso (*fuzzy*), oferecem maneiras de superar algumas limitações, tornando a análise mais flexível e adaptável (GABRIEL *et al.*, 2018; MORIN; BUJACZ; GAGNÉ, 2018; REY-MARTÍ *et al.*, 2022). Em casos em que as condições são derivadas de variáveis quantitativas, a análise de conjuntos *fuzzy* é ideal (FILIPPONE; CHELI; D’AGOSTINO, 2001). Esses valores iniciais das variáveis são então calibrados para valores no intervalo fechado entre 0 e 1, sendo que esses valores extremos representam a ausência total e, inversamente, a presença total da condição em questão (DUSA, 2018). Nessa lógica, um determinado caso pode pertencer a um conjunto em graus variados, ao contrário do modelo binário, pois no cenário *fuzzy* podem existir graus intermediários de pertinência (pontuações) entre 0 e 1. A ideia de que certos objetos com características que pertencem apenas parcialmente a determinados conjuntos aproximam o estudo da realidade (RAGIN, 2009), permitindo a expressão de incerteza na atribuição de pertencimento.

Segundo Emmenegger (2011), a utilização da abordagem *fuzzy* em determinadas áreas implica calibração da forma mais transparente possível, com o objetivo de minimizar erros provenientes

da atribuição subjetiva de valores. Encontrar associações de um valor original  $x$  com um valor dentro do intervalo  $[0,1]$  pode, assim, representar a pertinência relativa do caso correspondente dentro do conjunto representado pelo fator em questão. Calibração é o processo de obtenção de valores para variáveis a partir de dados de base (dados brutos). A abordagem de diferenciação difusa permite estabelecer diferenças entre casos que parecem semelhantes em espécie, ou seja, conjuntos *fuzzy* expressam diferenças de grau entre casos que estão do mesmo lado da âncora em 0,5 (SCHNEIDER; ROHLFING, 2013).

Desta maneira, uma proposta de abordagem multidimensional para definir pontuações técnicas e táticas para jogadores de futebol passa pela utilização da lógica *fuzzy* para criar indicadores calibrados mediante a frequência em que determinados eventos, como número de passes, chutes e cruzamentos corretos, acontecem ao longo de uma partida. Na linguagem da teoria dos conjuntos, um grau de pertencimento ao conjunto dos jogadores com bom desempenho nos indicadores técnicos considerados. A álgebra Booleana se mostra particularmente adequada como uma estrutura matemática capaz de lidar com conceitos vagos.

A utilização da lógica difusa já foi proposta para representar a complexidade e os níveis de incerteza na significação dos indicadores técnicos de performance (WIEMEYER, 2003; BAZMARA, 2014). Embora seja uma linguagem consistente para aplicações computacionais, a lógica difusa requer certo critério para codificar dados complexos da atividade humana (HEALD, 2018). Para atribuir os valores adequados aos gestos técnicos de uma partida de futebol em um intervalo entre "0" para totalmente ausente e "1" para totalmente presente, será necessário estabelecer limiares (FILIPPONE; CHELI; D'AGOSTINO, 2001).

Três outros trabalhos estão na concepção do índice proposto por este trabalho, o primeiro – uma monografia de pós-graduação – tratou da perspectiva configuracional utilizando abordagem *fuzzy* para determinar indicadores técnicos relevantes na performance de goleiros (NUNES, 2020). Outras duas produções científicas – ambas apresentadas no Congresso Brasileiro de Gestão do Esporte – abordaram a comparação de pontuações atribuídas por meio da TFRa de atacantes do futebol brasileiro e combinação entre esses fatores técnicos que fossem necessários e suficientes para o alto montante envolvido na transferência desses jogadores para outros times (NUNES; ANON; FREITAS, 2021; NUNES; FLEURY; GONÇALVES, 2021).

Portanto, este capítulo da tese pretende demonstrar a comparação de fatores técnicos cujas pontuações serão atribuídas utilizando a função de pertinência alternativa (FILIPPONE; CHELI; D'AGOSTINO, 2001). Serão executadas atribuições de pontuação para atletas da fase

ofensiva e a agregação entre os fatores em índices significativos. Ao passo que a exemplificação dos fatores técnicos poderá ser demonstrada, pretende-se levantar opções para a execução do processo considerando os elementos táticos e físicos, de forma a produzir um índice que considere elementos multidimensionais para a comparação de atletas de futebol, no sentido de proporcionar uma métrica fiável para esse propósito.

Embora os analistas estejam interessados em identificar as características individuais que contribuem para o desempenho da equipe, o fator humano às vezes pode superar as estatísticas derivadas dos dados disponíveis. Afinal, a física da bola e a mentalidade do jogador que chuta determinam a probabilidade de um gol e entender todos os fatores determinantes para um bom resultado é uma tarefa difícil (SARMENTO *et al.*, 2018). No contexto desportivo, a medição é frequentemente complexa, desta maneira quando os pesquisadores desenvolvem um modelo teórico para esportes, eles estão abstraindo a realidade.

O conceito de abstração está presente em outros aspectos da vida cotidiana e simplifica o mundo complexo. A própria proposta de modelagem remete à abstração (ROSENBLUETH; WIENER, 1945). A realidade possui muitas variáveis e, por vezes, a combinação aleatória de algumas delas é o que de fato define o resultado. De maneira que, ao codificar os eventos de uma partida de futebol, opta-se por uma representação abstrata dos eventos determinantes para o resultado da partida. Certos elementos devem ser eliminados para simplificar um processo complexo, já que capturar o mundo inteiro é impraticável. A abstração resultante torna-se o modelo de realidade, influenciado por crenças e preconceitos do pesquisador. Desta maneira, por mais assertivo e confiável que o índice decorrente desta pesquisa se torne, ainda assim não será possível compreender todos os elementos que constituem a realidade de uma partida de futebol.

#### **4.2. Arcabouço Teórico: A jornada do atleta**

Algumas sociedades estimulam seus indivíduos a compreender a vida como um processo linear, cujo objetivo pessoal é alcançar seu potencial (WOODS; DAVIDS, 2023). Esta questão tem sido abordada na tradição do pensamento científico ocidental por meio do paradigma da complementaridade, que divide o ser humano em partes distintas, contudo interdependentes, abrangendo o biológico e o cultural (INGOLD, 2021). Destarte, por potencial se define como um conjunto de capacidades inatas localizadas na mente e no corpo, transmitidas por meio de um mecanismo de herança genética, portanto pertencente ao âmbito biológico (INGOLD, 2021). A sua completude, isto é, a realização se traduz como as informações transmitidas por

meio de regras, símbolos, representações e sistemas de classificação que definem o que significa saber (INGOLD, 2021; WOODS; DAVIDS, 2023).

A conceitualização do comportamento centrada no potencial e na realização segue trajetória crítica similar. O realismo ecológico de Gibson rejeita a ideia de que o conhecimento reside na mente, impondo estruturas a um mundo estático. Em vez disso, enfatiza a noção de encontrar ativamente uma estrutura, chamada de conhecimento do ambiente, em um campo de relações dinâmicas e em constante evolução (GIBSON, 2014; HEFT, 2013). De maneira que a vida não é uma progressão linear mecânica, cujas capacidades inatas são preenchidas com conteúdo transmitido por predecessores, mas um processo ativo, como uma jornada em busca de crescimento ao longo desse caminho (INGOLD, 2021; WOODS; DAVIDS, 2023).

Neste contexto, o conhecimento não é interpretado como um conjunto de informações transmitidas de geração em geração, a espera de ser aplicado na prática. De maneira oposta, ele é cultivado através da prática, principalmente por meio da experiência de coisas que emergem enquanto se está em relação com o ambiente (INGOLD, 2021; WOODS; DAVIDS, 2023). Essa perspectiva ecologicamente dinâmica implica que o indivíduo não se move em direção a um ponto predefinido de realização (um fim) por meio da absorção de informações secundária. De fato, é uma jornada de observação participante e relacionamento com outras pessoas (INGOLD, 2021; WOODS; DAVIDS, 2023).

A dinâmica ecológica tem origem e se destaca como interdisciplinar; no contexto das ciências do esporte, envolve disciplinas de aprendizagem motora, biomecânica, psicologia do desporto, pedagogia do desporto e análise do desempenho (ARAÚJO *et al.*, 2019). Esse paradigma estabeleceu uma lógica para a tomada de decisão constituindo ligação ao conceito de design representativo (ARAÚJO *et al.*, 2019; BRUNSWIK, 1956). Este termo, design representativo, refere-se aos métodos de treinamento dos atletas em que se busca reproduzir ambientes de treino que repliquem apropriadamente o contexto de performance (ARAÚJO *et al.*, 2019).

O comportamento é definido no nível ecológico de análise, que é o nível de interações entre um organismo e seu ambiente, sendo que ambos se moldam continuamente (GIBSON, 2014). Como resultado deste conceito, o comportamento só pode ser compreendido simetricamente em relação às características de um ambiente de desempenho. Essa orientação foi demonstrada por meio de padrões comportamentais do atleta resultantes da coordenação emergente entre o atleta e o ambiente de alto rendimento (ARAÚJO *et al.*, 2019; BRUNSWIK, 1956). Não existe uma estrutura interna de conhecimento ou gerador de padrões central dentro do organismo que

controle a ação na dinâmica ecológica. Em vez disso, durante a regulação da ação, todas as partes do sistema (cérebro, corpo, ambiente) são integradas dinamicamente (ARAÚJO *et al.*, 2019).

A unidade de análise para compreender o desempenho é o sistema executor-ambiente. Esse sistema não é uma entidade fixa e evolui à medida que o foco da ação do performer muda (SARMENTO; ARAÚJO, 2021). Isso significa que os processos psicológicos não se limitam à mente do indivíduo, se estendendo ao ambiente de maneira que os limites entre o jogador e o ambiente são permeáveis, e a ação é influenciada por essa interação constante (SARMENTO; ARAÚJO, 2021). O comportamento é uma reorganização dinâmica do sistema organismo-ambiente (DAVIDS *et al.*, 2017). Portanto, para entender o desenvolvimento de habilidades em contextos de desempenho, considera-se tanto o indivíduo quanto o ambiente em que ele atua.

A abordagem da dinâmica ecológica reconhece que o ambiente é repleto de *affordances*. Este termo foi citado pela primeira vez por Gibson e se refere ao ajuste entre as capacidades do organismo e as oportunidades de interação que o ambiente suporta, sejam essas interações boas ou más (GIBSON, 2014). Em tradução literal para o português, *affordances* são disponibilidades ambientais, que incluem não apenas objetos, mas recursos do ambiente, incluindo aspectos benéficos e prejudiciais (NORMAN, 2002).

As *affordances* existem fora da dicotomia entre propriedades objetivas e subjetivas, confundindo a distinção entre características físicas e psicológicas. Eles se preocupam tanto com o meio ambiente quanto com o comportamento, atuando como um elo entre os dois (CHEMERO, 2006). São processos semióticos enraizados em características naturais, facilitando a aquisição através de ações ou instrumentos que incorporam essas características (KOCKELMAN, 2013). Em suma, a percepção e a ação são intrinsecamente ligadas. Esta abordagem integrada à percepção e à ação é essencial para compreender como os performers interagem com o mundo ao seu redor (GIBSON, 2014; HEFT, 2013).

Tendo por base essa perspectiva, a jornada do atleta até alcançar o potencial perpassa uma abertura para *affordances* que surgem (WOODS; DAVIDS, 2023). Habitar uma paisagem repleta de *affordances* diversificadas oferece oportunidades múltiplas, em contraposição a uma paisagem de *affordances* limitadas, onde apenas caminhos pré-determinados levam a resultados preestabelecidos (WOODS; DAVIDS, 2023).

O desempenho, no contexto da dinâmica ecológica, é uma manifestação de auto-organização sob restrições. Isso significa que o desempenho emerge dentro das limitações existentes (SARMENTO; ARAÚJO, 2021). Dentro dessas restrições, existem um número limitado de soluções estáveis que podem ser alcançadas para atingir objetivos específicos (ARAÚJO *et al.*, 2019). Um modelo com três categorias de restrições – organismo, ambiente e tarefas – que operam em diferentes escalas de tempo – uma partida, uma temporada ou uma carreira – e que são vistas como limites ou características que moldam as soluções comportamentais (NEWELL, 1986). Quando um sistema, *e.g.* ambiente-executor, estabelece um estado de organização por meio apenas de interações dinâmicas entre seus componentes individuais, esse estado é considerado auto-organizado (SARMENTO; ARAÚJO, 2021).

A abordagem da dinâmica ecológica destaca que o desempenho esportivo especializado é resultado das soluções emergentes de um sistema auto-organizado do atleta em resposta a diferentes restrições, não desconsiderando o papel das características pessoais, como hereditariedade e genética (SARMENTO; ARAÚJO, 2021). O desempenho humano é mais bem compreendido como uma expressão das tendências naturais de auto-organização sob restrições, em vez de ser meramente imposto de dentro ou de fora do indivíduo (ARAÚJO; DAVIDS, 2018).

No contexto das *affordances* é fundamental que o praticante esteja afinado com percepções aguçadas e ações habilidosas para identificar e aproveitar as oportunidades raras que podem levar a uma carreira bem-sucedida. Não envolve apenas detectar informações ambientais, mas também requer habilidades e capacidades de ação que possibilitem a ação em relação a essas *affordances* (ARAÚJO *et al.*, 2019). É importante notar que o potencial de um indivíduo não é estático, mas sim dinâmico e influenciado por uma série de fatores, como genes, motivação, prática deliberada, fadiga e disponibilidade de recursos, incluindo instalações e apoio de treinadores (SARMENTO; ARAÚJO, 2021). É fundamental compreender que nem toda adaptação é bem-sucedida, e as oportunidades para mudanças bem-sucedidas não são necessariamente abundantes. Em resumo, a abordagem que considera o sistema atleta-ambiente como a unidade de análise destaca a importância de entender as interações complexas entre as características pessoais de um atleta, as restrições ambientais e as oportunidades de carreira (SARMENTO; ARAÚJO, 2021).

Fundamentada nestes pressupostos, as pesquisas recentes que envolvem modelos de avaliação de desempenhos de atletas buscam a complementaridade entre o que as diversas medidas podem

fornecer (TRAVASSOS *et al.*, 2013). Em um mundo em que a expansão de dados coletados permite a observação de um número de eventos muito grande, compreender o contexto dos dados é importante no processo de mensuração e extração de relações reais de causa e efeito (PATEL; SHAH; SHAH, 2020). A diversidade de contextos, dimensões e eventos gerados por observações de empresas especializadas, como a *Hudl*, proporciona condições para o desenvolvimento de ferramentas que auxiliem aos treinadores, observadores e gestores a tomarem decisões relacionadas ao desempenho e performance dos jogadores de futebol (KUSMAKAR *et al.*, 2020; PAPPALARDO *et al.*, 2019).

Examinar o desempenho de um jogador pode exigir dimensões distintas, como o desempenho geral do jogador, a identificação de subgrupos com base no desempenho e a compreensão do desempenho em jogadores individuais. Cada uma dessas perspectivas necessita de uma abordagem, que pode ser centrada na variável, centrada na pessoa ou específica na pessoa (HOWARD; HOFFMAN, 2018). Os métodos fornecem informações para decisões subsequentes, de maneira que não existe uma abordagem universal, ao contrário disso a abordagem escolhida é determinada pela natureza da questão de investigação (HOWARD; HOFFMAN, 2018).

A literatura científica contém numerosas conceituações da abordagem centrada na pessoa (WOO *et al.*, 2018). Alguns pesquisadores definem estes estudos como pesquisas que enfocam características individuais. Um estudo é considerado centrado na pessoa quando enfoca características como personalidade, perícias ou habilidades. Outra conceituação utiliza o termo centrado na pessoa para descrever uma variedade de métodos que classificam os indivíduos com base na similaridade de suas pontuações em diversas variáveis (HOWARD; HOFFMAN, 2018). Este método é consistente com o conceito de examinar indivíduos com base em perfis específicos através de múltiplas variáveis ou atributos (WOO *et al.*, 2018).

Estes métodos são frequentemente tipológicos, em consonância com a conceituação da investigação centrada na pessoa como o estudo do agrupamento de indivíduos (MORIN; BUJACZ; GAGNÉ, 2018). Essa natureza é atribuída ao uso de um sistema de classificação que divide os indivíduos em subpopulações distintas com base em diferentes conjuntos de parâmetros do modelo (MORIN; BUJACZ; GAGNÉ, 2018). A pesquisa centrada na pessoa é frequentemente descrita como prototípica, o que significa que cada indivíduo da amostra está associado até certo ponto a cada um dos perfis estimados (HOFMANS; WILLE; SCHREURS, 2020). O grau em que a configuração de pontuação única de um indivíduo nas variáveis do

estudo se assemelha à configuração de pontuação específica de um perfil determina essa probabilidade. Os indivíduos não são simplesmente atribuídos a um único perfil neste contexto probabilístico, mas são avaliados como sendo mais ou menos semelhantes a cada perfil prototípico (HOFMANS; WILLE; SCHREURS, 2020).

Os modelos centrados na pessoa são inerentemente de natureza exploratória. Devido à falta de informações sobre a qualidade do ajuste para avaliar diretamente a adequação do modelo, esses modelos são normalmente derivados da comparação de soluções com números variados de perfis ou *clusters* (HOFMANS; WILLE; SCHREURS, 2020). É importante notar que a utilização de modelos centrados na pessoa para fins confirmatórios não é excluída por estas características metodológicas (HOFMANS; WILLE; SCHREURS, 2020).

Vários métodos, embora não se enquadrem estritamente na definição acima, alinham-se com os objetivos das abordagens centradas na pessoa porque visam estudar perfis ou padrões de pontuações dos indivíduos (HOFMANS; WILLE; SCHREURS, 2020). Embora não sejam baseados em classificações, estes métodos centram-se no padrão de pontuações dos indivíduos, considerando assim a pessoa como um todo (BERGMAN; MAGNUSSON, 1997). Como resultado, classificam-se como centrados na pessoa (HOFMANS; WILLE; SCHREURS, 2020).

Os métodos centrados na pessoa desviam-se do pressuposto de homogeneidade populacional ao agrupar as pessoas em subgrupos ou subpopulações. Esta abordagem, conhecida como abordagem idiográfica ou individual, visa criar um modelo único para cada indivíduo, com base na ideia de que um modelo individualizado é melhor para descrever e compreender cada pessoa (HOWARD; HOFFMAN, 2018). Os dados utilizados como entrada revelam as diferenças filosóficas entre as abordagens específicas da pessoa e as centradas na pessoa. Em contraste com os métodos específicos de pessoa, que funcionam com matrizes de variáveis de evento, os métodos centrados na pessoa funcionam com matrizes de variáveis de pessoa (HOFMANS; WILLE; SCHREURS, 2020).

A análise de partidas de futebol tem sido utilizada para avaliar eventos comportamentais e buscar identificar padrões de jogo e comportamento em ações de diferentes equipes e jogadores (PAIXÃO *et al.*, 2015). Elementos contextuais, como a localização da partida, a qualidade do adversário e o resultado da partida, podem afetar a performance dos atletas dentro do campo (KONEFAŁ *et al.*, 2018). Estudos destacam a importância da análise sequencial de ações do jogo e individual de times em uma competição para melhor compreensão da performance, já

que o futebol é um esporte dominado por aspectos estratégicos e táticos (COSTA *et al.*, 2009; PAIXÃO *et al.*, 2015). Alguns trabalhos focam nas atividades técnicas dos jogadores em relação ao resultado da partida, enquanto outros analisam as ações técnicas e sua influência no resultado (KONEFAŁ *et al.*, 2018). Essas análises baseadas nos resultados dos jogos são úteis para explicar as diferenças que os impactos contextuais tiveram sobre a performance dos atletas.

Indicadores de performance são essenciais para definir aspectos do desempenho esportivo e são usados para prever comportamento futuro em atividades esportivas, especialmente no futebol (ZHOU *et al.*, 2018). O desempenho de jogo é um resultado da interação de atividades físicas, técnicas e táticas apresentadas pelos jogadores durante a partida, cada um desempenhando funções específicas de acordo com sua posição no campo (YI *et al.*, 2018). Portanto, a avaliação da performance dos jogadores deve levar em consideração as diferenças entre as posições, com critérios individualizados identificados para caracterizar o que cada uma demanda (YI *et al.*, 2019). O uso de indicadores de performance pode refletir a identidade de cada atleta e equipe, além de prever comportamentos futuros dentro de um contexto específico (PAIXÃO *et al.*, 2015).

O trabalho de Hughes e Probert (2006) tratou análise de desempenho no futebol em cinco categorias principais: física, tática, técnica-defesa, técnica-ataque e psicológica. Por outra perspectiva, considera-se que o atleta apresenta características físicas, técnicas, táticas, psicológicas, clínicas e administrativas (PAOLI, 2007). Em trabalho de revisão sistemática, foram classificados os determinantes de performance em quatro dimensões com suas subdimensões correspondentes: (1) limitações da tarefa: (a) especificidade e volume de prática; (2) limitações dos atletas: (a) fatores psicológicos; (b) habilidades técnicas e táticas; (c) fatores antropométricos e fisiológicos; (3) limitações ambientais: (a) efeito da idade relativa; (b) influências socioculturais; e (4) análise multidimensional (SARMENTO *et al.*, 2018).

Outra revisão da literatura apontou caminhos para novos estudos: entre as possibilidades para expansão do campo, estão as aplicações de análises utilizando aprendizagem de máquina (WAKELAM; STEUBER; WAKELAM, 2022). Entre os levantamentos foram a identificação de 1.518 atributos diferentes já utilizados na literatura sobre futebol, com 81% que podem ser mensurados numericamente (WAKELAM; STEUBER; WAKELAM, 2022). Para a execução de trabalhos de análise é valioso que se execute métodos de redução de dimensionalidade, como a análise de componentes principais, para identificar conjuntos de atributos que formam determinadas dimensões (WAKELAM; STEUBER; WAKELAM, 2022).

Por fim, o que se busca em parte dos trabalhos sobre desempenho em ciência do esporte é conseguir adequar as mais diversas dimensões de análise para compreender a performance dos atletas (PAPPALARDO *et al.*, 2019; PATEL; SHAH; SHAH, 2020; SARMENTO *et al.*, 2018). Quantas mais dimensões sejam possíveis mensurar e quantos mais eventos em contextos diferentes puderem ser adequados ao modelo, mais se avança na compreensão das influências de determinadas conjunturas sobre os desempenhos em campo. Englobar os aspectos técnicos e táticos, por exemplo, em um instrumento calibrado para permitir interações e comparações multidimensionais pode auxiliar na observação do desempenho dos atletas.

### **4.3. Análise Técnica: a habilidade dos atletas**

O desempenho no futebol é determinado pela interação das atividades técnicas, táticas e físicas realizadas pelos jogadores durante uma partida (JAMIL, 2020; HUGHES *et al.*, 2012; YI *et al.*, 2018). Cada jogador contribui para o sucesso da equipe em uma partida, desempenhando funções específicas com base na sua posição em campo (HUGHES *et al.*, 2012). Como resultado, ao avaliar o desempenho de um jogador em campo, a posição do jogador deve ser levada em consideração, sendo necessário critérios individualizados para identificar as características e requisitos específicos de cada função, a fim de não ocultar informações valiosas que possam limitar a compreensão do desempenho dos jogadores (YI *et al.*, 2018).

Alguns estudos investigaram demandas técnicas específicas de cada função, destacando características e requisitos técnicos típicos (DI SALVO *et al.*, 2008; ERMIDIS *et al.*, 2019; HUGHES *et al.*, 2012; WIEMEYER, 2003; YI *et al.*, 2018; YI *et al.*, 2019). Estudos sobre as influências das ações e eventos de jogo sobre o desempenho técnico de jogadores em múltiplas posições revelou distinções na frequência de variáveis sobre gols, passes e organização do jogo, bem como variáveis sobre defesa (ERMIDIS *et al.*, 2019; YI *et al.*, 2018), demonstrando as diferenças nos desempenhos técnicos e táticos das diversas posições de jogo.

Yi *et al.* (2018) agrupam os gestos técnicos por resultado ou objetivo das ações. As posições dos jogadores exigem diferentes atributos, com os meio-campistas cobrindo maiores distâncias e atuando com alta intensidade, enquanto os laterais e atacantes realizam mais sprints durante a partida, exigindo agilidade e reflexos, e os zagueiros centrais cobrem uma distância menor e fazem menos sprints (YI *et al.*, 2018).

Estas capacidades são fundamentais para distinguir jovens jogadores de futebol talentosos daqueles menos talentosos (LARKIN; O'CONNOR, 2017; LARKIN; REEVES, 2018;

SARMENTO *et al.*, 2018). Habilidades técnicas são gestos que resultam em ações, como desarmes, bem como passes, controle de bola, dribles e chutes que são aspectos críticos do desempenho avaliados por treinadores e olheiros em sistemas de identificação de talentos (HUGHES *et al.*, 2012; JAMIL, 2020).

A execução bem-sucedida de ações técnicas é crítica para o desempenho, assim como fatores socioculturais influenciam os jogadores a desenvolverem determinados gestos (HUGHES *et al.*, 2012; JAMIL, 2020; LIU, YI *et al.*, 2015; ZHOU *et al.*, 2018). Os gestos técnicos diferem até mesmo entre atletas que atuam nas mesmas faixas do campo de jogo, como defensores centrais e defensores laterais, bem como entre meio-campistas centrais e meio-campistas laterais (ERMIDIS *et al.*, 2019).

Comparados aos zagueiros centrais e até mesmo os meio-campistas laterais, os defensores laterais apresentam maior frequência de cruzamentos (YI *et al.*, 2018). Isto enfatiza a importância do cruzamento como uma habilidade fundamental para a posição, já que frequentemente estes jogadores são instigados a avançar para o terço de ataque e elevar bolas para a área adversária. Estas distinções destacam a importância dos papéis específicos da posição no campo de jogo, que têm um impacto significativo nas exigências técnicas e no desempenho dos jogadores de futebol em diversas áreas do jogo.

No contexto geral, zagueiros costumam realizar menos ações ofensivas, como tentativas e arremates de dentro da área, quando comparados às demais posições (ERMIDIS *et al.*, 2019; YI *et al.*, 2018). Isto faz sentido, dado que a posição de defesa está normalmente associada a funções defensivas e ações em lances de bola parada. Laterais, meio-campistas laterais e atacantes, por outro lado, cobrem distâncias totais maiores durante o jogo do que outras posições (ERMIDIS *et al.*, 2019; YI *et al.*, 2018), devido ao fato de que essas posições frequentemente exigem movimentação por todo o campo, bem como participação ativa em jogadas defensivas e ofensivas (ERMIDIS *et al.*, 2019; YI *et al.*, 2018).

Por sua vez, a posição de goleiro é considerada uma das mais determinantes para o resultado da partida, uma vez que um único erro pode mudar todo o placar do jogo. Di Salvo *et al.* (2008) estudaram as distâncias e níveis de intensidade dos goleiros na *Premier League* inglesa, relacionando atividades de alta velocidade às habilidades técnicas e ao resultado da partida. Outra investigação sobre se deu sobre o desempenho técnico e tático de goleiros em diferentes níveis de equipes e situações na Liga Espanhola de Futebol Profissional da Primeira Divisão, encontrando relações diferentes para o desempenho dos goleiros de acordo com o nível das

equipes em que atuavam (LIU; GÓMEZ *et al.*, 2015). Os atributos mais necessários para essa posição incluem a porcentagem de bolas defendidas e desviadas, a distribuição da bola para os companheiros, passes certos e a remoção da bola da zona de perigo (RUSU; STOICA; BURNS, 2011).

Diferenças regionais e de estratégias de jogo podem resultar em diferenças nas estatísticas técnicas (ERMIDIS *et al.*, 2019; JAMIL, 2020). Entre as diferenças culturais, jogadores sul-americanos são mais proficientes em remates à baliza do que europeus ou asiáticos, assim como sul-americanos e europeus superam os jogadores africanos, asiáticos e norte-americanos em termos de eficiência de passe (JAMIL, 2020). As infrações são mais comuns entre atletas sul-americanos, resultando em um número significativamente maior de cartões amarelos e vermelhos (JAMIL, 2020).

De modo semelhante, futebolistas europeus demonstraram maior proficiência nas ações de passe, que pode ser atribuído às influências culturais que moldam os estilos de jogo dos jogadores e das equipes europeias, como o “Futebol Total” e o “*tiki-taka*” (INGERSOLL; MALESKY; SAIEGH, 2017; JAMIL, 2020; MITROTASIOS *et al.*, 2019). Além disso, é fornecido um apoio substancial aos jogadores europeus, particularmente em termos de formação e equipes juvenis, com ênfase na especialização precoce (LARKIN; REEVES, 2018). Atletas norte-americanos, por outro lado, demonstraram menos proficiência nas ações de passe do que os jogadores europeus ou sul-americanos, principalmente nas áreas de ataque (JAMIL, 2020; LARKIN; REEVES, 2018). Variáveis contextuais influenciam os estilos de jogo e devem ser tidas em conta na revisão dos indicadores de performance (FERNANDEZ-NAVARRO *et al.*, 2018).

#### **4.4. Metodologia**

Os principais dados utilizados neste estudo foram coletados a partir do portal *Wyscout*, que era uma das principais plataformas de coleta de dados do futebol, com milhões de registros relacionados a partidas de diversos campeonatos (PAPPALARDO *et al.*, 2019). A coleta de informações de eventos relacionados aos jogos foi baseada nessa plataforma, que foi adquirida pelo grupo *Hudl*, junto com a *InStat*, uma outra plataforma de dados estatísticos de partidas de futebol (POLI; RAVENEL; BESSON, 2023).

Estes dados foram extraídos da plataforma entre fevereiro e abril de 2024, por meio da assinatura da licença básica da plataforma. O projeto de qualificação da tese utilizou a análise

de componentes principais (PCA), oriunda de sugestões de trabalhos anteriores (WAKELAM *et al.*, 2022) para agregar as habilidades dos jogadores. Entretanto, como sugestão de melhoria a ser aplicada na versão final, optou-se por outros processos que foram aplicados para geração do modelo e análise de resultados. Foram obtidas estatísticas de jogadores que compunham os elencos das equipes das Séries A e B do futebol brasileiro, englobando todos os dados técnicos disponíveis na plataforma daquele determinado atleta.

No total foram captadas 205.319 linhas de informações, que descreviam 1.266 atletas diferentes. Entre colunas descritivas estavam as que indicavam a partida, a competição disputada, a temporada e a posição em que o atleta atuou naquela partida. Depois havia outras 66 colunas com indicadores de performance. O quadro 14 apresenta estes indicadores.

Minutos jogados	Cruzamentos	Recuperações no meio-campo do adversário	Duelos ofensivos	Passes para a grande área	Saídas
Ações totais	Cruzamentos certos	Duelos defensivos	Duelos ofensivos ganhos	Passes para a grande área precisos	Carrinhos GK
Ações totais bem-sucedidos	Dribles	Duelos defensivos ganhos	Toques na área	Passes recebidos	Carrinhos bem-sucedidos GK
Golos	Dribles com sucesso	Duelos de bola livre	Foras de jogo	Passes para a frente	Pontapés de baliza
Assistências	Duelos	Duelos de bola livre ganhos	Corridas seguidas	Passes para a frente certos	Pontapés de baliza curtos

Remates	Duelos ganhos	Carrinhos	Faltas sofridas	Passes para trás	Pontapés de baliza longos
Remates à baliza	Duelos aéreos	Carrinhos bem-sucedidos	Passes em profundidade	Passes para trás certos	
$xG$	Duelos aéreos ganhos	Alívios	Passes em profundidade certos	Golos sofridos	
Passes	Intercepções	Faltas	$xA$	$xCG$	
Passes certos	Perdas	Cartões amarelos	Segundas assistências	Remates sofridos	
Passes longos	Perdas no seu meio-campo	Cartões vermelhos	Passes para terço final	Defesas	
Passes longos certos	Recuperações	Assistências para remate	Passes para terço final certos	Defesas com reflexos	

Quadro 14 - Indicadores estatísticos oriundos do *Wyscout*

Duas outras fontes de dados foram utilizadas para o estudo. A primeira delas é derivada do videojogo *Football Manager 2024*, produzido pela SEGA e desenvolvido pela *Sports Interactive*. Para isso foram selecionadas ligas que proporcionaram dados de 188.794 jogadores disponíveis no jogo com os atributos visíveis e, a partir disso, a pesquisa feita dentro do *game* permitindo extrair os dados dos atletas.

Por fim, a outra fonte de dados foi oriunda da plataforma de registro de transferências *Transfermarkt.de*. Foi necessária a utilização de um pacote disponível para R na interface *RStudio*, conhecido como *worldfootballR*. Esse pacote permitiu capturar por meio de um processo de raspagem (*web scraping*) dados dos atletas, como o valor de mercado estimado pelos usuários do *website*, assim como o valor registrado das transferências dos atletas.

#### 4.5. Modelagem de Incerteza: distinções fundamentais na lógica *fuzzy*

O sucesso comercial do esporte está vinculado à capacidade de se mostrar enquanto entretenimento. Grandes eventos, entre eles a Copa do Mundo da FIFA, atraem centenas de milhões de espectadores estimulados pela incerteza do resultado. Essa ideia é oriunda dos trabalhos seminais de Rottenberg (1956) e Neale (1964), consolidando-se como um dos pilares da economia do esporte.

Nesse contexto, as previsões desempenham um papel na dinâmica estratégica das equipes. O capítulo 2 apontou para a crescente adoção de técnicas de aprendizado de máquina no esporte a partir do avanço da coleta e disponibilidade de dados. Ainda que se utilize de técnicas para aprimorar a predição de variáveis atreladas aos desempenhos esportivos, o futebol detém seu apelo a partir do entendimento que o resultado dos jogos permaneça incerto, isto é, cada equipe envolvida possa ser percebida com possibilidades reais de vitória.

A teoria de conjuntos *fuzzy* foi introduzida por Zadeh (1965) e expandiu a lógica matemática clássica por possibilitar aos elementos pertencerem a um conjunto em graus contínuos de associação entre 0 e 1. Desta maneira, a lógica *fuzzy* proporcionou base teórica para representar incerteza, distinguida entre incerteza epistêmica e incerteza ontológica (PALMEIRA, 2021). A primeira refere-se à ausência de dados confiáveis, à imprecisão nas medições ou à própria limitação dos instrumentos utilizados para mensuração. Um exemplo ocorre em modelos de dinâmica populacional, nos quais coeficientes de equações diferenciais ou condições iniciais não são conhecidos com exatidão, gerando incertezas nos parâmetros do sistema (MIZUKOSHI, 2004). Nesse contexto, a teoria *fuzzy* reconhece que variáveis observadas por vezes apresentam níveis de imprecisão que compromete a qualidade e a confiabilidade dos dados (PALMEIRA, 2021).

Por sua vez a incerteza ontológica está vinculada à própria natureza dos conceitos analisados. Nesse cenário, dado que alguns termos utilizados no futebol são dependentes de interpretações contextuais e da subjetividade do modelador, a lógica *fuzzy* torna-se útil por permitir formalizar variáveis linguísticas por meio de funções de pertinência, classificando as interpretações subjetivas em estruturas matemáticas manipuláveis (BARROS; ESMI *et al.*, 2021).

A incerteza epistêmica pode resultar em conjuntos *fuzzy* com funções de pertinência que refletem uma menor confiança nos dados subjacentes. Por sua vez, a incerteza ontológica é expressa na definição própria dos conjuntos *fuzzy*, como no caso de atributos como finalização,

cuja caracterização varia de forma gradual e dependente do contexto (BARROS; WASQUES *et al.*, 2021).

Há distinção entre a lógica *fuzzy* e as abordagens probabilísticas tradicionais. Enquanto a probabilidade trata da ocorrência de eventos futuros, quantificando sua chance antes da realização, a lógica *fuzzy* lida com graus de verdade de proposições que já ocorreram, ainda que os resultados sejam ambíguos ou difíceis de mensurar (BARROS; ESMI *et al.*, 2021).

Por permitir a manipulação matemática de termos linguísticos subjetivos, a lógica *fuzzy* proporciona aos modelos capturar nuances do comportamento humano e interações complexas. Desta maneira, ao integrar a incerteza epistêmica e ontológica em uma mesma estrutura formal, a lógica *fuzzy* se apresenta como um recurso analítico para representar fenômenos como a performance de atletas no futebol de alto rendimento.

Adicionalmente, a aplicação da lógica *fuzzy* nesta pesquisa se justifica não apenas pela capacidade dessa abordagem em lidar com incertezas epistêmicas e ontológicas, mas também por sua aderência aos conceitos imprecisos e subjetivos presentes no domínio futebolístico. No esporte, atributos como cansaço ou habilidade são, por natureza, vagos e dependentes do contexto. Como destacam Betti *et al.* (2023) e Alkire *et al.* (2015), a lógica *fuzzy* permite tratar tais predicados vagos sem reduzi-los a categorias rígidas, mas, ao contrário, expressando-os em graus de pertencimento a um conjunto, por meio de funções de pertinência.

No presente estudo, essa abordagem foi operacionalizada por meio da *fuzzificação* de indicadores estatísticos dos jogadores, gerando estruturas de dados que mantêm a complexidade e subjetividade do jogo ao mesmo tempo em que possibilitam análise matemática. Ao traduzir variáveis linguísticas como excelente finalizador, muito bom finalizador e razoável finalizador em valores contínuos entre 0 e 1, os modelos *fuzzy* capturam nuances que seriam perdidas em abordagens tradicionais de categorização binária. Como sugerido por Cheli *et al.* (2021) e Tavares e Betti (2024), a lógica *fuzzy* pode aprimorar a entrada de dados para modelos preditivos ao capturar a intensidade e a nuance de características que não se reduzem a dicotomias.

Essa transformação mostrou-se eficaz quando integrada aos algoritmos de aprendizado de máquina, como o *Random Forest*. Conforme será demonstrado nos resultados empíricos, os resultados da análise neste contexto decorrem devido à robustez frente a dados complexos e

não-lineares e da granularidade dos dados *fuzzy*, que fornecem ao algoritmo entradas mais ricas e informativas do que variáveis categóricas tradicionais.

Além disso, a combinação entre lógica *fuzzy* e *Random Forest* demonstrou ser apropriada para o contexto investigado por esta tese, em razão de três fatores: (i) a alta variabilidade dos dados empíricos do futebol, que demanda modelos capazes de lidar com incertezas e não-linearidades; (ii) a natureza subjetiva de muitas variáveis utilizadas, cuja classificação binária seria artificial; e (iii) o interesse em fornecer uma ferramenta interpretável, que auxilie decisores em clubes na leitura de indicadores de performance de maneira contínua e comparável entre atletas.

Nesse sentido, a utilização da lógica *fuzzy* neste trabalho não se limita a um recurso técnico de transformação de dados, mas constitui um fundamento epistemológico que permite representar de forma mais realista os conceitos analisados. A *fuzzificação* dos dados atuou como um meio de preservar a complexidade do fenômeno esportivo, ao mesmo tempo em que preparava os dados para análise computacional via *machine learning*. Essa integração metodológica é, portanto, coerente com a proposta desta tese de desenvolver modelos aplicáveis à tomada de decisão no contexto do futebol profissional, respeitando as particularidades do campo empírico.

#### **4.5.1. Etapas da Inferência *Fuzzy***

A inferência *fuzzy* opera a partir de cinco fases: (i) *fuzzificação*; (ii) base de regras; (iii) máquina de inferência; (iv) agregação e; (v) *defuzzificação*, sendo esta nem sempre necessária (BARROS; ESMI *et al.*, 2021; PALMEIRA; 2021).

Na primeira etapa, os valores das variáveis de entrada são transformados em graus de pertinência associados a um conjunto *fuzzy*. Esta etapa é chamada de *fuzzificação*, e sua função é representar variáveis sob a ótica da teoria dos conjuntos difusos (BARROS; ESMI *et al.*, 2021). Para cada variável de entrada é definido um intervalo de valores possíveis, e sobre esse intervalo são aplicadas funções de pertinência (BARROS; ESMI *et al.*, 2021). A função de pertinência é responsável por atribuir a cada valor um grau de associação entre 0 e 1. Essa representação contínua é uma das vantagens da lógica *fuzzy* para contextos esportivos, onde o desempenho não se resume a dicotomias, mas à intensidade relativa com que um jogador manifesta determinadas características em campo.

A segunda etapa consiste na estruturação de uma base de regras do tipo *se/então*, que codifica o conhecimento prévio sobre o domínio do problema. Essa base de conhecimento é formada

por proposições linguísticas que relacionam variáveis de entrada e saída em formações do tipo: se remates à baliza é alto e o  $xG$  é alto, então a finalização é muito alta (BARROS; ESMI *et al.*, 2021).

Os antecedentes e consequentes das regras são definidos com base nos conjuntos *fuzzy* estabelecidos na fase de *fuzzificação*. Os operadores lógicos *fuzzy* (AND, OR, NOT) são aplicados para combinar condições múltiplas dentro de uma mesma regra. Esses operadores, diferentemente da lógica booleana clássica, operam com funções mínimas, máximas ou produtos, de acordo com o grau de pertinência de cada variável (BARROS; ESMI *et al.*, 2021).

Essa estrutura permite incorporar tanto conhecimento empírico quanto expertise do domínio, sendo útil em contextos cujas decisões qualitativas precisam ser quantificadas, como na análise de atletas com base em múltiplas variáveis de performance.

A terceira etapa constitui o núcleo lógico do sistema, responsável por combinar as informações das regras e dos valores de entrada *fuzzificados* para produzir conclusões. Ela executa três sub etapas, a avaliação dos antecedentes, implicação e agregação (BARROS; WASQUES *et al.*, 2021). A avaliação dos antecedentes determina o grau de ativação de cada regra, aplicando os operadores *fuzzy* às pertinências dos valores de entrada. O grau de ativação do antecedente é então aplicado ao consequente da regra (BARROS; ESMI *et al.*, 2021).

Se várias regras são ativadas ao mesmo tempo, seus respectivos consequentes *fuzzy* precisam ser combinados em uma única saída *fuzzy* agregada. Isso é feito por meio de operadores de agregação, como o máximo, média ponderada ou soma probabilística. Essa agregação sintetiza a informação proveniente de múltiplas regras em uma representação única, que será utilizada para interpretar a saída do sistema (PALMEIRA, 2021).

O resultado da inferência *fuzzy* é, em essência, um conjunto *fuzzy* de saída. Dependendo da aplicação, esse conjunto pode ser utilizado diretamente, em classificações, ordenações ou comparações entre atletas, conforme a distribuição de pertinência dos valores *fuzzy*. Também é possível a sua transformado em um valor nítido (*crisp*) por meio da *defuzzificação* (BARROS; ESMI *et al.*, 2021).

No caso desta tese, a aplicação de métodos de inferência *fuzzy* se deu por meio da transformação dos dados estatísticos em funções de pertinência, utilizando diferentes abordagens de calibração, como TFRa, IFR e IT2FS. Esses dados transformados foram então utilizados tanto para análise descritiva quanto como insumo para modelos preditivos com aprendizado de

máquina. A integração entre os valores *fuzzificados* e técnicas como *Random Forest* permitiu testar a viabilidade e a performance dos modelos. Destarte, não apenas se representou a incerteza dos dados, como também se testou a aplicabilidade desses graus de pertencimento na classificação de atributos técnicos no futebol na observação de variáveis de interesse como gols, assistências, minutos jogados, valor de mercado e taxa de transferência.

#### **4.6. Transformações *Fuzzy* Aplicadas**

O modelo proposto neste capítulo utilizou seis tipos distintos de transformação dos dados brutos para valores de pertinência a conjuntos *fuzzy*. Utilizou-se uma proposta de transformação já abordada pelo autor da tese em trabalhos anteriores, conhecida como *Totally Fuzzy and Relative* alternativa e, como comparação de resultados, outra técnica de transformação relativa, a *Integrated Fuzzy and Relative*. Como outro parâmetro de comparação, optou-se por uma transformação *fuzzy* intervalar conhecida como *Interval Type-2 Fuzzy Sets*, considerando graus de pertinência inferior (*lower*), superior (*upper*) e central (*center*).

##### **4.6.1. *Totally Fuzzy and Relative* alternativo**

O método *Totally Fuzzy and Relative* alternativo (TFRa), apresentado por Filippone *et al.* (2001) foi o primeiro modelo *fuzzy* implementado. Ele se baseia em uma abordagem relativa para definir a pertinência dos valores observados, considerando a distribuição dos próprios dados. O método já havia sido utilizado pelo autor desta tese em seu trabalho de mestrado assim como os artigos derivados (NUNES, 2020; NUNES *et al.*, 2022; NUNES *et al.*, 2024).

O objetivo da TFRa é atribuir valores de pertinência *fuzzy* de forma relativa, ou seja, a pertença de um jogador a uma categoria de desempenho é avaliada em relação ao desempenho dos demais jogadores da amostra. A transformação utiliza a função de distribuição empírica acumulada e a frequência relativa dos valores. Cada evento foi calibrado sendo a frequência relativa à menor incidência daquela ação (ordenada crescentemente), a frequência acumulada até o ponto  $x$  e o nível (*i.e.*, *fuzzy*) de membresia, para aquela incidência “ $x$ ” avaliada (FILIPPONE; CHELI; D’AGOSTINO, 2001). A Equação 1 apresenta os cálculos realizados para executar a calibração:

$$g(xi) = \begin{cases} \frac{1}{2} h(x(1)) & \text{se } xi = x(1) \\ H(x(k-1)) + \frac{1}{2} h(x(k)) & \text{se } xi = x(k) \text{ com } (k > 1) \end{cases}$$

Equação 1 - Especificação alternativa de calibração *fuzzy*

A lógica de utilizar a distribuição de eventos como maneira de relativizar a atribuição de valores de membresia permite comparações diretas entre valores distintos. Os valores de pertinência atribuídos ficaram concentrados em um intervalo lógico entre 0 e 1, com boa distribuição e com a média dos atributos transformados concentrada no valor de 0,5. O TFRa demonstra ser eficiente para capturar variações sutis entre desempenhos semelhantes e sua escolha se deu devido à sua capacidade de lidar com dados de diferentes escalas e sua robustez em cenários com variabilidade moderada (FILIPPONE *et al.*, 2001).

#### 4.6.2. *Integrated Fuzzy and Relative*

Na esteira da abordagem TFRa, a literatura sobre o tema evoluiu no sentido de desenvolver métodos que integram mais amplamente as diferentes dimensões da privação, especialmente ao lidar com dados socioeconômicos ou contextos marcados por desigualdades estruturais. Nesse sentido, destaca-se o trabalho seminal de Betti *et al.* (2006), que propôs a abordagem *Integrated Fuzzy and Relative* (IFR) como um refinamento da TFR original (CHELI; LEMMI, 1995) e de sua versão alternativa (FILIPPONE *et al.*, 2001). O IFR combina duas funções de pertinência, sendo essas a função baseada na posição relativa de um indivíduo na distribuição ( $1 - F$ ), conforme proposta por Cheli e Lemmi (1995), e a função baseada na concentração da renda entre os indivíduos menos privados ( $1 - L$ ), tal como sugerida por Betti e Verma (1999), incorporando a curva de Lorenz como elemento estruturante.

A IFR se baseia na atribuição de escores *fuzzy* para cada indivíduo  $j$ , utilizando uma função de pertinência definida da seguinte maneira:

$$\mu_{j,k} = \left( \frac{\sum_{\gamma=j+1}^n \omega_{\gamma} |X_{\gamma} > X_j}{\sum_{\gamma=2}^n \omega_{\gamma} |X_{\gamma} > X_1} \right)^{\alpha_{k-1}} \left( \frac{\sum_{\gamma=j+1}^n \omega_{\gamma} X_{\gamma} |X_{\gamma} > X_j}{\sum_{\gamma=2}^n \omega_{\gamma} X_{\gamma} |X_{\gamma} > X_1} \right)$$

Equação 2 - Função de pertinência da *Integrated Fuzzy and Relative*

Com  $j = 1, \dots, n - 1$  e  $\mu_{j,k} = 0$ , sendo:

- $X_j$  o valor da variável para o indivíduo na posição  $j$ ;
- $\omega_{\gamma}$  peso amostral do indivíduo  $\gamma$ ;

- $\alpha_k$  parâmetro de calibração associado à dimensão  $k$ ;
- $F_{j,k}$  função da distribuição acumulada (CDF) normalizada;
- $L_{j,k}$  curva de Lorenz associada à variável  $X$ .

Essa função é equivalente a:

$$\mu_{j,k} = (1 - F_{j,k})^{\alpha_k - 1} (1 - L_{j,k})$$

Equação 3 - Função equivalente da *Integrated Fuzzy and Relative*

A combinação dos termos permite representar simultaneamente a posição relativa do indivíduo na distribuição e o escore acumulado até a sua posição, conferindo à função um caráter duplamente relativo.

A metodologia IFR distingue-se por aplicar simultaneamente essas duas dimensões dentro de uma função única de pertinência *fuzzy*. Essa generalização permite que a modelagem da propensão à privação seja não apenas sensível à posição do indivíduo em relação ao restante da amostra, mas também à distribuição entre os demais casos analisados. Trata-se de um avanço metodológico, pois responde a limitações anteriormente apontadas na TFRa no que diz respeito à sensibilidade referente à desigualdade (FILIPPONE *et al.*, 2001; BETTI *et al.*, 2015).

Adicionalmente, o trabalho subsequente de Betti *et al.* (2020) estabeleceu a aplicação da IFR como base para construção de índices compostos *fuzzy*, combinando variáveis de diferentes naturezas em uma estrutura agregativa coerente. Esses índices são úteis em estudos com fenômenos multidimensionais, como é o caso da avaliação de habilidades esportivas no futebol, em que aspectos técnicos, físicos, táticos e contextuais influenciam simultaneamente o desempenho dos atletas.

A abordagem IFR tem sido utilizada em estudos longitudinais sobre pobreza (VERMA *et al.*, 2017) e metodologias para estimação de erros padrão e análise de sensibilidade de medidas *fuzzy* (BETTI *et al.*, 2018), demonstrando sua aplicabilidade em cenários que exigem robustez estatística e refinamento teórico. Com base nessas contribuições, evidencia-se o potencial da IFR para ser adaptada ao contexto esportivo, onde a desigualdade de desempenho e a variabilidade contextual desempenham papéis na análise de dados.

#### 4.6.3. *Interval Type-2 Fuzzy Sets*

Dentre as estratégias de transformação *fuzzy* implementadas nesta tese, destaca-se o uso dos *Interval Type-2 Fuzzy Sets* (IT2FS), originalmente propostos como extensão da teoria dos conjuntos *fuzzy* de tipo 1 elaborada por Zadeh (1975). Enquanto os conjuntos *fuzzy* tipo 1 trabalham com funções de pertinência bidimensionais, cuja incerteza está restrita à variabilidade no grau de pertencimento de um valor a um conjunto, os conjuntos *fuzzy* de tipo 2 introduzem um terceiro eixo que representa a incerteza associada à própria função de pertinência (MENDEL; JOHN, 2002). No caso dos IT2FS, essa incerteza é modelada por um intervalo para cada valor do domínio, o que torna essa abordagem especialmente útil em contextos em que a variabilidade ou ambiguidade dos dados não pode ser negligenciada.

A adoção da abordagem IT2FS tem se consolidado em áreas onde a precisão dos dados é limitada ou onde há incertezas tanto epistêmicas quanto ontológicas associadas às decisões, como demonstrado por Meniz e Ozkan (2023). Embora sua aplicação em modelos de decisão multicritério, como o AHP *fuzzy*, tenha sido amplamente explorada, seu potencial vai além dessa classe de problemas. Em análises onde a interpretação de desempenho técnico está sujeita à variabilidade intercontextual, como no futebol profissional, a modelagem por intervalos *fuzzy* representa uma forma robusta de incorporar tais incertezas na avaliação dos atributos.

No presente estudo, a transformação IT2FS foi utilizada para atribuir a cada observação um intervalo de pertinência, com limites inferior e superior determinados a partir de funções de pertinência ajustadas com base na distribuição empírica dos dados. A amplitude desse intervalo representa a incerteza associada ao grau de pertencimento de um atleta a uma dada categoria de desempenho, permitindo que casos ambíguos, como jogadores que oscilam entre desempenhos médios e altos, sejam tratados com maior sensibilidade. Essa abordagem é particularmente relevante para modelar variáveis que sofrem influência de fatores táticos, físicos ou contextuais, muitas vezes difíceis de quantificar diretamente, mas que afetam a interpretação dos dados disponíveis.

Estudos como os de Milošević *et al.* (2021) já demonstraram que os IT2FS são superiores aos conjuntos do tipo 1 na representação de sistemas incertos, justamente por acomodarem a variabilidade na atribuição de valores de pertinência sem comprometer a interpretabilidade do modelo. A adoção dessa abordagem nesta pesquisa, portanto, responde a uma demanda metodológica de capturar zonas de transição e sobreposição entre categorias *fuzzy*, reforçando a aderência teórica da transformação ao fenômeno investigado: o desempenho técnico contextualizado de jogadores de futebol.

#### 4.7. Base de regras e inferência

A construção de regras linguísticas *fuzzy* no presente estudo foi orientada por uma abordagem híbrida que integrou modelos supervisionados de aprendizado de máquina à modelagem baseada em lógica difusa. Foram extraídos os coeficientes dos modelos de regressão penalizada LASSO, Ridge e *Elastic Net*, utilizando-os como fatores de influência (*degree of influence factors* – DIF) incorporados diretamente no processo de inferência como pesos de ativação das regras *fuzzy*.

Essa escolha metodológica encontra respaldo na literatura sobre sistemas *fuzzy* orientados por dados (*data-driven fuzzy systems*), em pesquisas que incorporaram algoritmos de aprendizado de máquina para estruturar, calibrar e modular sistemas de regras (NAUCK; KRUSE, 1999). Sistemas adaptativos como o *Adaptive Neuro-Fuzzy Inference System* (ANFIS) utilizam os parâmetros numéricos oriundos de algoritmos supervisionados para ajustar funções de pertinência, assim como para determinar a força de ativação das regras *fuzzy* (JANG, 1993).

Na perspectiva adotada nesta tese, os pesos obtidos a partir das regressões penalizadas foram utilizados para modular a influência de cada variável no processo de inferência *fuzzy*. Esta estratégia de atribuição ponderada de pertinência *fuzzy* tendo por base coeficientes de modelos supervisionados foi explorada em estudos aplicados a problemas de classificação, previsão e suporte à decisão em contextos de alta incerteza (CHIU, 1994; SETIONO; LIU, 1995).

Outros artigos sobre sistemas *fuzzy* orientados por regressão sugerem que a direcionalidade do coeficiente pode ser interpretada como indicativa da relação causal entre variável e resposta, enquanto o peso do coeficiente pode ser proporcional ao grau de ativação da regra (CORDÓN, 2001). Assim, atributos com maior coeficiente absoluto foram ponderados com maior intensidade na saída *fuzzy*, influenciando a composição do índice de performance.

Ao adotar essa abordagem, buscou-se incorporar a evidência empírica no desenho das regras *fuzzy* e promover um sistema capaz de traduzir padrões estatísticos extraídos dos dados em linguagens de decisão compatíveis com a realidade do futebol profissional. A preocupação com a interpretabilidade é compartilhada por Nauck e Kruse (1999), que argumentam que os pesos derivados de modelos lineares ou penalizados pode favorecer a auditabilidade dos sistemas *fuzzy*.

A geração de regras *fuzzy* a partir de dados é uma das bases teóricas do campo de sistemas inteligentes interpretáveis, buscando integrar aprendizado supervisionado com modelagem linguística. No presente estudo, buscou-se aproveitar os coeficientes extraídos de modelos de regressão penalizada (LASSO, Ridge, *Elastic Net*) como critérios orientadores para a estruturação automatizada de regras *fuzzy*. Essa abordagem se insere no escopo da chamada *fuzzy rule induction*, cuja proposta central é transformar relações estatísticas latentes em regras linguísticas com valor interpretativo e operacional.

Conforme destacam Setiono e Liu (1995), pesos extraídos de modelos estatísticos ou neurais podem servir como base heurística para a identificação das variáveis mais relevantes para inferência, permitindo construir regras respeitando a direção (sinal) e intensidade (magnitude) da relação observada. Essa relação entre coeficiente e lógica de regra é explorada em técnicas de extração de regras *fuzzy* de modelos de regressão e redes neurais (*fuzzy rule extraction from regression models*), onde se busca a transição de modelos de caixa preta para representações compreensíveis (*white-box modeling*) (SETIONO; LIU, 1995).

A utilização de pesos como critérios de construção de premissas *fuzzy* também se manifesta em abordagens híbridas como as árvores de decisão *fuzzy*, nas quais os atributos com maior importância, quantificada por medidas como Gini, ganho de informação ou SHAP *values*, são priorizados em nós de decisão *fuzzy* (CINTRA *et al.*, 2007). Nessas árvores, a lógica *fuzzy* é incorporada para tratar incertezas nas divisões, permitindo regras com ativação gradual.

Portanto, ao se utilizar os coeficientes como sinalizadores das variáveis com maior poder explicativo, a modelagem *fuzzy* se beneficia de uma camada adicional de fundamentação empírica, aproximando-se de métodos evolutivos ou heurísticos de construção de base de conhecimento *fuzzy* (CHEN; MYNETT, 2003; DIMITRIOU *et al.*, 2008). Esta prática possibilita a elaboração de regras linguísticas coerentes com os padrões de dados históricos, ao mesmo tempo em que se preserva a interpretabilidade exigida em domínios como a análise de performance esportiva.

No contexto desta pesquisa, os coeficientes positivos de variáveis como *xG* médio histórico, presença na área e remates por 90 minutos foram incorporados como antecedentes nas regras *fuzzy*, em que a intensidade da relação foi usada para modular sua contribuição relativa. Essa estratégia contribui para a inteligibilidade do sistema e para sua adequação prática, aliando rigor quantitativo com representação linguística compatível com a prática decisória no futebol profissional.

Esta abordagem híbrida é fundamentada em literatura que defende o uso de métodos estatísticos e de aprendizado supervisionado como guias para o design de sistemas *fuzzy* mais robustos e interpretáveis (CINTRA *et al.*, 2007; JANG *et al.*, 1997). A proposta se baseia na premissa de que, ao invés de definir arbitrariamente os parâmetros das funções de pertinência, estes podem ser informados por evidência empírica proveniente da modelagem preditiva, em particular a partir de uma modelagem conduzida com técnicas de regularização que detectam variáveis mais relevantes (ZHOU *et al.*, 2002).

Essa técnica favorece a interpretação dos conjuntos *fuzzy* como representações linguísticas moduladas por evidência estatística, em contraste com abordagens puramente heurísticas. A calibragem orientada por dados confere aderência do sistema *fuzzy* à realidade observada, ao mesmo tempo em que preserva sua flexibilidade semântica (JANG *et al.*, 1997; ZHOU *et al.*, 2002).

Em termos metodológicos, esse tipo de integração aproxima-se das propostas de sistemas *neuro-fuzzy* ou *fuzzy-weighted learning*, em que os pesos estatísticos são utilizados para refinar os limites dos conjuntos *fuzzy* durante processos de treinamento ou de modelagem supervisionada (JANG *et al.*, 1997; ZHOU *et al.*, 2002). No contexto desta tese, o procedimento de *fuzzificação* incorporou os coeficientes como elementos de escalonamento para cada variável selecionada, proporcionando que a modelagem *fuzzy* refletisse a estrutura de importância das estatísticas do jogo derivada da modelagem com regressão penalizada.

Portanto, a utilização dos *Degree of Influence Factors* como critério para ajuste das funções de pertinência representa uma etapa entre articulação da inferência *fuzzy* e aprendizado de máquina. Essa estratégia contribui para a criação de um sistema *fuzzy* empírico e adaptado ao domínio, que mantém a lógica interpretativa das funções linguísticas e se ancora em relações estatísticas previamente validadas.

#### **4.8. Regressões lineares penalizadas**

Dado o número de variáveis oriundas da plataforma do *Wyscout* e do videogame *Football Manager*, com risco de sobreajuste e multicolinearidade optou-se pela utilização dos algoritmos de regressão penalizada Ridge, Lasso e *Elastic Net* (FRIEDMAN; TIBSHIRANI; HASTIE, 2010). Essas técnicas aplicam penalidades ao processo de estimativa de coeficientes na intenção de reduzir a variância, incentivando a seleção das variáveis relevantes para atribuição dos pesos na calibração dos modelos *fuzzy* (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

A regressão Ridge introduz a norma quadrada dos coeficientes (L2), para minimizar uma função de perda composta pela soma dos erros quadráticos e do termo de penalização (GARCÍA-NIETO *et al.*, 2021). Essa característica reduz a variância dos coeficientes e melhora o condicionamento do problema, acrescentando uma constante positiva ( $\lambda$ ) à matriz de covariância  $X^T X$  à custa de um ligeiro viés. Esse método funciona quando há multicolinearidade ou quando há mais preditores do que observações (FRIEDMAN *et al.*, 2010; HASTIE *et al.*, 2009). Entretanto, sua desvantagem principal é que todos os preditores são incluídos no modelo finalizado, embora com coeficientes menores o que pode dificultar a interpretação.

A regressão *Least Absolute Shrinkage and Selection Operator* (LASSO), em contraste, emprega a norma absoluta dos coeficientes (L1), que permite a seleção de variáveis e a regularização (TIBSHIRANI, 1996). A função de penalidade promove a redução dimensional automatizada, direcionando coeficientes menos relevantes ao valor zero conforme o parâmetro de regularização  $\lambda$  aumenta. Destarte, LASSO é utilizada em situações nas quais se deseja construir modelos compactos, mas explicativos e interpretativos (GARCÍA-NIETO *et al.*, 2021).

A regressão *Elastic Net* integra as penalidades L1 (LASSO) e L2 (Ridge) em uma única função de maneira a reduzir desvantagens relacionadas aos métodos anteriores (ZOU; HASTIE, 2005). A *Elastic Net* introduz dois hiperparâmetros, o  $\lambda$ , que controla a intensidade geral da penalidade, e  $\alpha$ , que controla o peso relativo entre as duas penalidades. O método estabiliza o processo de seleção de variáveis correlacionadas, permitindo que o modelo mantenha vários atributos simultaneamente (ZOU; HASTIE, 2005; FRIEDMAN *et al.*, 2010).

De uma perspectiva algorítmica, os três métodos podem ser representados como soluções para problemas de otimização convexa, nos quais o equilíbrio entre erro de ajuste e complexidade do modelo é mediado pelos parâmetros de regularização. A equação da regressão de Ridge é a seguinte:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \|\beta\|_2^2 \right\}$$

Equação 4 - Regressão Ridge

A regressão LASSO é formulada da seguinte maneira:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \|\beta\|_1 \right\}$$

Equação 5 - Regressão LASSO

E *Elastic Net*, por sua vez, é formulada como:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda [(1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1] \right\}$$

Equação 6 - Regressão *Elastic Net*

#### 4.9. Clusterização hierárquica

Para criar um modelo com parâmetros explicativos adequados para as variáveis de interesse, uma das aplicações realizadas no banco de dados imprimiu a verificação dos agrupamentos estatísticos/de habilidades. O agrupamento hierárquico foi um dos métodos utilizados. Este método baseia-se na proposta de que cada observação começa como um cluster separado e os agrupamentos são realizados um após o outro até que uma estrutura em árvore (dendrograma) seja criada (AUFFARTH, 2021; COHEN-ADDAD *et al.*, 2017).

Certas técnicas de ligação são empregadas para verificar a fusão entre *clusters*. Essas técnicas incluem as distâncias mais curtas e mais longas entre os pontos, a média das distâncias e, por fim, a redução da soma dos quadrados das diferenças dentro de todos os grupos (AUFFARTH, 2021; COHEN-ADDAD *et al.*, 2017).

#### 4.10. *Fuzzy C-Means*

A segunda técnica aplicada foi o algoritmo *Fuzzy C-Means*, abordagem de agrupamento que permite que cada instância pertença simultaneamente a mais de um *cluster*. Essa abordagem opera com uma matriz de pertinência, que expressa a intensidade de pertencimento a cada agrupamento (HASHEMI *et al.*, 2023).

O *Fuzzy C-Means* minimiza uma função que equilibra a distância entre os pontos e os centróides dos *clusters* ponderada pelos graus de pertinência e por um parâmetro que regula a difusão das associações (DANGETI, 2017). A flexibilidade do modelo é útil para contextos cujas fronteiras entre os grupos são difusas.

#### 4.11. *Random Forest*

A análise dos modelos propostos em relação à explicação das variáveis de interesse foi executada utilizando o método *Random Forest*. Este é um método de aprendizado de máquina supervisionado que escolhe repetidamente um subconjunto do espaço amostral para ajustar o modelo de um conjunto de árvores de classificação selecionando aleatoriamente um grupo de variáveis do espaço de variáveis (CONESA; HERNÁNDEZ, 2014).

A concepção do método se baseia na criação de múltiplas árvores de decisão independentes, construídas por meio da técnica de *bagging* (*bootstrap aggregation*), com o objetivo de aumentar a estabilidade e a precisão preditiva dos modelos (BREIMAN, 2001; HO, 1995). O modelo final é constituído por um conjunto de florestas, em que cada árvore contribui com um voto individual e a classe final atribuída é aquela que recebe o maior número de votos, também chamado consenso majoritário.

O processo de construção do *Random Forest* envolve a seleção aleatória de subconjuntos de observações com reposição a partir da base original, procedimento conhecido como *bootstrap sampling*. Para cada árvore, um subconjunto diferente de dados é amostrado, promovendo diversidade entre os classificadores individuais (BREIMAN, 2001). A cada divisão de nó durante a construção da árvore, uma quantidade limitada de variáveis independentes  $m$  (com  $m < M$ , sendo  $M$  o total de variáveis) é selecionada aleatoriamente para avaliação. A divisão é realizada com base na variável que apresenta a melhor separação para aquele nó específico. Essa aleatoriedade contribui para reduzir a correlação entre as árvores e, por consequência, o sobreajuste (*overfitting*) (CAI *et al.*, 2020).

Cada árvore é construída até sua profundidade máxima, ou até que um critério de parada seja atingido, como o número mínimo de amostras por folha. Após o treinamento, a classificação de uma nova observação ocorre a partir da votação de todas as árvores da floresta. No caso de regressão, a média dos valores previstos por cada árvore é calculada como valor final (GARCÍA-NIETO *et al.*, 2021).

O algoritmo oferece mecanismos intrínsecos para avaliação da importância das variáveis, com base na redução da impureza ou na diminuição da acurácia preditiva quando determinada variável é aleatorizada (BREIMAN, 2001; ZHANG, 2019). Essa funcionalidade foi relevante para esta pesquisa, pois permitiu identificar as variáveis de desempenho que mais

influenciavam na predição da variável de interesse (valor de transferência), após a *fuzzificação* da base original.

No contexto da presente tese, o *Random Forest* foi aplicado como uma segunda camada analítica após a *fuzzificação* dos dados, com o objetivo de classificar os jogadores segundo a principal variável dependente (valores de transferência) e mensurar a contribuição relativa de cada estatística de desempenho para essa predição. Apesar da complexidade dos dados e das correlações multivariadas, a flexibilidade do algoritmo permitiu executar iterações robustas e analisar a estrutura subjacente da base mesmo diante de desempenhos iniciais aquém dos parâmetros desejados.

#### **4.12. Descrição do procedimento**

O procedimento adotado neste estudo partiu das proposições iniciais relacionada ao comportamento dos dados e à performance esperada dos modelos de aprendizado de máquina aplicados na observação das variáveis de interesse. De maneira que o presente estudo adotou uma abordagem exploratória e quantitativa baseada em ensaios, visando obter modelos preditivos relacionados ao valor de transferência de atletas de futebol profissional. Essas proposições foram testadas por meio da aplicação de algoritmos de *machine learning*, utilizando-se conjuntos de dados originais e executando ajustes adequados.

Os primeiros ensaios foram executados utilizando dois bancos de dados distintos. O primeiro, de natureza estatística, foi oriundo da plataforma *Wyscout*, especializada na coleta e organização de dados de desempenho de jogadores. O segundo banco de dados foi extraído do jogo *Football Manager 2024*, que fornece avaliações de atributos técnicos, físicos e mentais de jogadores, derivadas de uma base consolidada de observadores internacionais.

Inicialmente, os algoritmos de regressão penalizada (LASSO, Ridge e *Elastic Net*) foram aplicados sobre um conjunto de dados que compreendia as estatísticas médias dos jogadores entre os anos de 2019 e 2023. Para esse propósito, os dados foram segmentados em dois períodos, um conjunto histórico (2019 – 2022) e o ano de 2023, com o objetivo de estimar o peso de cada período de estatísticas na formação das habilidades atribuídas no *Football Manager*.

A partir dos resultados das regressões, os dados da plataforma *Wyscout* foram calibrados utilizando transformações *fuzzy* previamente apresentadas nesta tese. Foram aplicadas as

seguintes técnicas da *Totally Fuzzy and Relative* alternativa (FILIPPONE; CHELI; D'AGOSTINO, 2001), o método *Integrated Fuzzy Relative* (BETTI *et al.*, 2020) e *Interval Type-2 Fuzzy Sets* (MENDEL; JOHN, 2002) com seus respectivos limites inferior, central e superior.

Cada uma dessas transformações deu origem a duas estruturas distintas. A primeira agregava por média aritmética as estatísticas *fuzzificadas* a partir da base estatística original, considerando o universo completo de atletas, independentemente da posição em campo. A segunda incorporava os pesos obtidos nas regressões penalizadas, multiplicando os valores *fuzzy* das estatísticas relevantes por seus respectivos coeficientes, excluindo os interceptos, de modo a compor uma pontuação agregada por habilidade. Essa pontuação foi posteriormente normalizada para o intervalo  $[0,1]$ , replicando-se o processo para as três técnicas de regressão (LASSO, Ridge e *Elastic Net*), resultando em diferentes versões da base de habilidades derivadas do *Football Manager*. Essa proposta resultou em 20 modelos distintos de pontuações *fuzzy* para os atletas mapeados.

A etapa subsequente consistiu na análise crítica dos resultados obtidos, especialmente dos parâmetros de desempenho como o Erro Quadrático Médio da Raiz (RMSE) e o Coeficiente de Determinação ( $R^2$ ), em relação às variáveis de interesse. O objetivo central dessa fase foi identificar os fatores com maior relevância para a variável de interesse – o valor de transferência do atleta – por meio da aplicação do algoritmo *Random Forest*, com análise da importância das variáveis.

Com o intuito de refinar os modelos, foi realizada uma análise de clusterização hierárquica, a fim de compreender os agrupamentos naturais das estatísticas entre os jogadores. A partir disso, observando algumas colunas da base de dados, identificou-se que certos fatores que não se agrupavam consistentemente às posições dos atletas, ou apresentavam reduzida variação entre valores máximos e mínimos, o que pode ter contribuído para a baixa capacidade explicativa observada. Um ensaio com um recorte aleatório da amostra indicou que o ajuste poderia encontrar resultados interessantes. Diante disso, optou-se por excluir alguns desses indicadores, tais como ações totais, ações totais bem-sucedidas,  $xG$ , carrinhos, carrinhos bem-sucedidos, cartões amarelos, cartões vermelhos, impedimentos (*offside*),  $xA$  e segundas assistências.

Considerando que a base de dados continha diversos indicadores, buscou-se uma orientação teórica para a interpretação das relações, como a proposta no trabalho de Markopoulo *et al.* (2024), que utilizou modelos de *machine learning* para prever a marcação de gols nas

principais ligas de futebol. Essa abordagem inspirou a transformação da coluna estatística referente aos gols marcados pelos atletas. De variável explicativa, essa coluna passou a ser interpretada como variável de interesse.

Nessa mesma lógica, outras duas colunas também foram transformadas em variáveis de interesse: a de minutos jogados e a de assistências. Com isso, em vez de utilizar esses dados na tentativa de explicar outras variáveis, optou-se por tentar explicar esses próprios fenômenos. A marcação de gols, objetivo principal das equipes de futebol em uma partida, as assistências, ações que precedem diretamente os gols e são executadas por outros jogadores, e os minutos jogados, que representam um indicador de participação constante nas partidas, portanto, tiveram suas colunas reposicionadas no foco analítico.

Dessa maneira, foi realizada a comparação entre os resultados obtidos a partir da base *fuzzificada* com 55 colunas, utilizada para a explicação dos fenômenos financeiros (valor de transferência, com e sem transferências a custo zero, e valor de mercado), e aqueles obtidos com a base ajustada para 42 colunas. Também foi executada uma nova regressão penalizada sobre a base de estatísticas do *Wyscout*, contendo 42 colunas estatísticas, em relação às 36 habilidades do *Football Manager*. Alguns dos resultados foram comparados entre as duas versões, sem que se verificasse aumento relevante no coeficiente de determinação ( $R^2$ ).

Portanto, uma vez que os resultados obtidos com a base ajustada de 42 colunas se mostraram semelhantes aos da primeira versão, apresentando apenas ganhos e perdas marginais, optou-se por apresentar apenas os resultados da versão inicial, com 55 colunas estatísticas explicando as relações financeiras e 52 colunas explicando as relações de performance (gols, assistências e minutos jogados). Essa decisão foi motivada pela complexidade envolvida no recálculo de todos os modelos a partir das novas regressões penalizadas, considerando que, para cada tipo de transformação *fuzzy*, há cerca de 12 *gigabytes* de dados atrelados ao processo de ponderação via coeficientes.

Um outro ajuste no modelo foi aplicado à transformação IFR para testes. Esse ajuste limitou a base de estatísticas a 30 colunas, considerando apenas ações bem-sucedidas no campo de jogo (*e.g.* Remates à baliza (chutes ao gol), duelos defensivos bem-sucedidos, passes certos, passes em profundidades certos entre outras). As habilidades oriundas do jogo *Football Manager* também foram restritas apenas às 14 habilidades técnicas. Neste modelo foi testado apenas a transformação *Elastic Net*. O baixo ganho no coeficiente de determinação também afetou a decisão de abandonar a explicação deste modelo.

Por fim, um ensaio encontrou um modelo com 39 fatores estatísticos com bons resultados para  $R^2$  de oito variáveis de interesse. Esse modelo, utilizando apenas as bases *fuzzificadas* agregou entre as variáveis de interesse, além de gols, minutos jogados e assistências, as faltas cometidas e faltas sofridas, as interceptações e as variáveis *expected goals (xG)* e *expected assists (xA)*. A análise de *clusters*, primeiro com a análise hierárquica e na sequência utilizando a técnica *Fuzzy C-means* encontrou melhor ajuste para explicar funções táticas dos jogadores que em relação à posição atribuída. Dado o volume de ensaios realizados, optou-se por detalhar na seção de análise alguns dos modelos que melhor se ajustaram na explicação dos fenômenos que foram estudados nesta tese.

#### **4.13. Análise**

Este tópico compreenderá as análises executadas para a elaboração do modelo proposto por esta tese. Serão indicados os resultados relacionados aos procedimentos descritos anteriormente. Houve diversas experimentações incidindo sobre o banco de dados e a execução de ajustes no modelo inicial, na intenção de indicar o que melhor alcançou parâmetros estatísticos de explicação para as variáveis de interesse.

##### **4.13.1. Regressões penalizadas entre a base de estatística reais e a base de dados do *Football Manager***

Nesta etapa, modelos de regressão penalizada (LASSO, Ridge e *Elastic Net*) foram empregados sobre os dados brutos oriundos das estatísticas do *Wyscout* para identificar pesos interpretáveis para cada variável, indicando sua influência nas habilidades dos jogadores, mensurada no jogo *Football Manager*.

As equipes de pesquisa e desenvolvimento em videogames de futebol, como EA FIFA, *Pro Evolution Soccer* e *Football Manager*, regularmente reúnem e examinam enormes conjuntos de dados para replicar com precisão as operações de futebol do mundo real (LI, 2021). Ao mostrar a capacidade da análise de processar dados intrincados de código aberto e aplicar algoritmos de aprendizado de máquina interpretáveis para dar sentido às decisões de negócios, o uso desses conjuntos de dados pode promover a pesquisa em disciplinas de negócios esportivos (LI, 2021).

O realismo do banco de dados do *Football Manager* tem revelado jogadores de futebol da vida real que se tornaram estrelas do futebol alguns anos depois que o *game* previu sua ascensão ao escalão internacional, embora uma parte desses jovens prodígios se tornem jogadores anônimos (NAKRANI, 2013). No entanto, essas instâncias de sucesso são suficientes para estabelecer a legitimidade do *Football Manager* fora da comunidade de jogos e começar a ganhar reconhecimento no mundo real do futebol (HOCQUET, 2016). O banco de dados do *Football Manager* levantou suspeitas em meados dos anos 2000 de que clubes de futebol reais contratavam jogadores jovens a partir dos dados do *game*, com clubes reconhecendo isso posteriormente (STUART, 2014).

Na intenção de unificar os bancos de dados, elaborou-se uma identificação única para cada atleta, denominada IDdob, composta pelo nome e pela data de nascimento (e.g. Ângelo Araos1997-01-06). Esse procedimento possibilitou a consolidação das informações dos jogadores, mesmo na ocorrência de alcunhas semelhantes, como Dudu ou João Victor, permitindo, assim, a realização dos cruzamentos entre as fontes.

Dentro dessa lógica, a obtenção das datas de nascimento e a consequente geração dos identificadores não foi possível para todos os atletas. Ainda que um tempo considerável tenha sido dedicado a essa etapa, a ausência desses dados implicou na exclusão de 7.632 linhas de informação, correspondentes a 3,71% da amostra. Além disso, uma falha no processo de extração do banco de dados da plataforma *Wyscout* resultou em lacunas em determinadas colunas para alguns atletas, ocasionando a eliminação de outras 217 linhas. Após esse processo de limpeza, restaram 197.469 linhas válidas, referentes a 1.263 atletas no banco oriundo do *Wyscout*.

Por sua vez, a base de dados extraída do jogo *Football Manager* foi tratada para obtenção do IDdob de cada jogador. A partir do cruzamento entre as duas fontes, foi possível identificar correspondência para 1.028 atletas, o que representa 83,17% da amostra de jogadores do *Wyscout*. Na sequência, utilizou-se o ambiente *RStudio* para a aplicação de regressões penalizadas.

As estatísticas dos atletas extraídas da plataforma *Wyscout* foram organizadas em dois blocos. O primeiro bloco agrupava os dados históricos, anteriores à temporada 2023. O segundo bloco concentrava apenas as estatísticas referentes à temporada 2023, a última temporada completa do futebol brasileiro no momento da coleta. A agregação das informações foi realizada por média aritmética.

Após a execução das regressões penalizadas, foram verificados os parâmetros de predição com o intuito de avaliar se as estatísticas reais obtidas por meio da plataforma *Wyscout* seriam suficientes para justificar as pontuações de habilidades atribuídas aos jogadores no videogame *Football Manager 2024*. As pontuações dos jogadores podem ser observadas na figura 18.



Figura 18 - Atributos do jogo *Football Manager*

Para a primeira etapa de modelagem preditiva, foram utilizadas regressões penalizadas com dados médios agregados de desempenho dos atletas. A variável dependente selecionada para demonstrar a comparação foi o atributo finalização atribuído aos jogadores no banco de dados do *Football Manager 2024*, enquanto as variáveis independentes corresponderam às estatísticas extraídas da plataforma *Wyscout*.

Os dados foram divididos em dois subconjuntos: 80% da amostra para treinamento do modelo e os 20% restantes para teste, conforme prática usual em procedimentos de validação cruzada. Foram estimados três modelos distintos: LASSO ( $\alpha = 1$ ), Ridge ( $\alpha = 0$ ) e *Elastic Net* ( $\alpha = 0,5$ ). Para cada modelo, utilizou-se o valor de penalização  $\lambda$  previamente calibrado por meio de validação cruzada com o objetivo de minimizar o erro de predição.

Os resultados obtidos no conjunto de teste estão sintetizados no quadro 15:

Modelo	Alpha	Best Lambda	Erro Médio Absoluto (MAE)	Erro Quadrático Médio (RMSE)
LASSO	1	0,043995	1,395	1,930
Ridge	0	1,497421	1,412	1,987

<i>Elastic Net</i>	0,5	0,058790	1,408	1,969
--------------------	-----	----------	-------	-------

Quadro 15 - Resultados obtidos nas regressões

O modelo LASSO apresentou o menor erro médio absoluto ( $MAE = 1,395$ ) e o menor erro quadrático médio ( $RMSE = 1,930$ ), indicando uma leve superioridade preditiva em relação aos demais modelos. Ainda que as diferenças entre os métodos tenham sido discretas, observa-se que a penalização L1 favoreceu um modelo mais parcimonioso, ao mesmo tempo em que manteve desempenho superior na capacidade preditiva sobre o conjunto de teste.

Os modelos Ridge e *Elastic Net*, apesar de apresentarem desempenho semelhante, não superaram o LASSO, o que sugere que a regularização via L1 foi suficiente para lidar com a possível colinearidade entre os preditores, sem a necessidade de incorporar a penalização L2. A seguir, as análises continuam com a exploração da relação entre estatísticas reais e demais atributos atribuídos no jogo, utilizando os coeficientes obtidos na regressão para fins de *fuzzificação* e etapas subsequentes da análise preditiva.

#### 4.13.2. Fuzzificação

O processo de *fuzzificação* foi executado para os cinco tipos de calibração previamente apresentados. Para essa etapa, alguns filtros foram aplicados com o intuito de garantir a coerência contextual dos dados utilizados. O primeiro filtro consistiu na restrição da amostra apenas aos jogos disputados em competições brasileiras. Tal decisão se justifica pelo fato de os dados terem sido extraídos de clubes atuantes no futebol nacional, sendo, portanto, mais adequado comparar atletas que disputam predominantemente os mesmos campeonatos ao longo da temporada.

A partir desse critério, a base de dados foi tratada para conter exclusivamente partidas referentes a torneios organizados em território brasileiro, o que resultou em uma redução para 143.488 linhas. No entanto, dois outros filtros foram empregados antes da aplicação das transformações *fuzzy*. O segundo tratamento visou restringir a análise apenas às competições que apresentassem volume minimamente relevante de partidas, garantindo robustez às comparações. Para isso, foi contabilizada a frequência de registros por competição no banco de dados.

Como exemplo, a Série A do Campeonato Brasileiro concentrava 48.266 registros, a Série B contava com 30.697, o Campeonato Paulista A1 possuía 11.060 linhas, entre outras. A análise da distribuição dessas frequências, por meio de quartis, indicou uma mediana de 958, ou seja,

metade das competições possuíam menos de 958 registros. Assim, definiu-se como critério de corte a exclusão das competições com volume de dados inferior à mediana. Com isso, certames como o Campeonato Brasileiro da Série D, o Campeonato Brasileiro de Aspirantes (Sub-23) e a Copa Verde foram removidos da análise. Ainda assim, o conjunto final de dados incluiu registros de 20 diferentes competições nacionais.

O terceiro e mais sensível filtro consistiu na exclusão dos dados referentes à posição de goleiro. Tal escolha foi motivada pela especificidade das ações associadas a essa função em campo, cuja lógica de atuação difere substancialmente das demais posições. Apesar do papel crescente do goleiro na construção ofensiva e defensiva das equipes no futebol contemporâneo, sua natureza única implicaria em um aumento significativo da complexidade da modelagem estatística descrita nesta tese. A partir desta exclusão as colunas de estatísticas foram restritas a 55 indicadores, como descritos no quadro 16.

Minutos jogados	Passes longos certos	Perdas no seu meio-campo	Cartões amarelos	$x_A$
Ações totais	Cruzamentos	Recuperações	Cartões vermelhos	Segundas assistências
Ações totais bem-sucedidos	Cruzamentos certos	Recuperações no meio-campo do adversário	Assistências para remate	Passes para terço final
Golos	Dribles	Duelos defensivos	Duelos ofensivos	Passes para terço final certos
Assistências	Dribles com sucesso	Duelos defensivos ganhos	Duelos ofensivos ganhos	Passes para a grande área

Remates	Duelos	Duelos de bola livre	Toques na área	Passes para a grande área precisos
Remates à baliza	Duelos ganhos	Duelos de bola livre ganhos	Foras de jogo	Passes recebidos
$xG$	Duelos aéreos	Carrinhos	Corridas seguidas	Passes para a frente
Passes	Duelos aéreos ganhos	Carrinhos bem-sucedidos	Faltas sofridas	Passes para a frente certos
Passes certos	Intercepções	Alívios	Passes em profundidade	Passes para trás
Passes longos	Perdas	Faltas	Passes em profundidade certos	Passes para trás certos

Quadro 16 - Estatísticas utilizadas na *fuzzificação*

Com base nesses critérios, a amostra final utilizada para as transformações *fuzzy* foi composta por 125.573 observações, correspondentes a jogadores de linha que atuaram em competições nacionais com volume representativo de registros. Esse filtro implicou na redução para 1.138 atletas. As variáveis de interesse também seguiram o processo de calibração para atribuição de valores *fuzzy* adequados a cada uma das transformações.

A idade dos atletas também foi adicionada ao modelo como um dos possíveis fatores explicativos. A partir do levantamento feito na plataforma *Transfermarkt* optou-se por adicionar o valor de mercado atribuído pelos usuários como uma das variáveis de interesse a ser explicada pela modelagem. Esse valor de mercado atribuído também foi transformado em pontuação *fuzzy* adequada a cada um dos modelos.

Os resultados das primeiras comparações entre os modelos, com foco na explicação da variável de interesse associada ao valor de transferência dos atletas e valor de mercado atribuído, serão apresentados na próxima seção.

#### 4.13.3. Algoritmo *Random Forest* aplicado aos modelos *fuzzificados* para explicação dos valores de transferência

Com o objetivo de avaliar a capacidade preditiva dos dados transformados por diferentes técnicas de *fuzzificação* no contexto da análise do valor de transferência dos jogadores de futebol, foram executadas cinco variações do algoritmo *Random Forest*. Essas análises utilizaram como variável dependente o valor das transferências declarado no banco de dados, incluindo as transferências realizadas a custo zero, o que insere viés na distribuição do desfecho, mas que condiz com a realidade do mercado esportivo.

Esta etapa encontrou desafio com a atualização da plataforma *Transfermarkt*, que implicou na falta de funcionalidade de um dos códigos do pacote *worldfootballR* dedicados a extrair o histórico de transferência de cada atleta. Desta maneira, os valores de transferência foram obtidos a partir da coleta executada para o capítulo 3, que analisou o fluxo de transferências entre diversas ligas no mundo. Assim, para a primeira etapa das análises a seguir, relacionado ao valor de transferência, há avaliação de 579 transações (50,87%) dos atletas filtrados.

Para cada uma das transformações *fuzzy* – TFRa, IFR, IT2FS *center*, IT2FS *lower* e IT2FS *upper* – os dados foram divididos em subconjuntos de treino (70%) e teste (30%), mantendo-se a reprodutibilidade das divisões com o uso do mesmo valor de semente aleatória (*set seed*). Em todas as análises, o número de árvores no conjunto (*ensemble*) foi fixado em 500 (*ntree* = 500), e foram mantidas constantes as variáveis preditoras, compreendendo o conjunto de atributos técnicos extraídos da base *Wyscout*.

#### 4.13.4. Resultados Obtidos

O quadro 17 resume os valores de erro quadrático médio (RMSE) e coeficiente de determinação ( $R^2$ ) obtidos em cada uma das execuções:

Modelo	RMSE	$R^2$
TFRa	0,2463	0,0323

IFR	0,1030	0,0359
IT2FS ( <i>center</i> )	0,1394	0,0005
IT2FS ( <i>lower</i> )	0,1418	0,0006
IT2FS ( <i>upper</i> )	0,1388	0,0005

Quadro 17 - Resultado para valores de transferência

Como pode ser observado, todos os modelos apresentaram desempenho modesto em termos preditivos, com valores de  $R^2$  consideravelmente baixos. A transformação baseada na abordagem IFR apresentou o melhor desempenho relativo ( $R^2 = 0,0359$ ), seguida da TFRa ( $R^2 = 0,0323$ ). As demais técnicas, notadamente as variações do IT2FS, registraram valores de  $R^2$  próximos de zero, indicando baixa capacidade explicativa para os valores de transferência a partir dos dados *fuzzy* não ponderados.

O erro quadrático médio (RMSE), embora mais sensível à escala dos dados, também demonstrou variação entre os modelos. Novamente, a IFR se destacou como a transformação com menor RMSE (0,1030), indicando um desempenho ligeiramente superior ao dos demais métodos no ajuste do modelo *Random Forest*.

Esses resultados implicaram na compreensão de que, apesar da aplicação de técnicas de transformação *fuzzy*, a capacidade de predição dos valores de transferência permaneceu limitada sem mecanismos adicionais de ponderação, como aqueles oriundos de regressões penalizadas. Isso sugere que, para além da representação *fuzzy* dos dados, era necessário incorporar critérios de relevância estatística na ponderação dos indicadores para que o modelo pudesse capturar as relações subjacentes entre o desempenho técnico dos atletas e seu valor de mercado.

#### **4.13.5. Análise dos modelos *fuzzificados* com filtro para transferências com valor positivo**

Conforme apresentado na seção anterior, a aplicação das técnicas de *fuzzificação* aos dados brutos possibilitou a análise da relação entre as estatísticas de desempenho dos atletas e os valores de suas transferências, ainda que com resultados preditivos modestos. No entanto, naquele primeiro conjunto de análises, foram incluídas também as transferências realizadas a custo zero. Isso implicou uma limitação, uma vez que esses casos podem representar não uma ausência de valor esportivo, mas sim estratégias de negociação específicas, término de contrato ou reestruturações financeiras dos clubes.

Com o intuito de refinar os modelos e torná-los mais sensíveis à variação real dos valores de mercado, foi conduzida uma nova etapa analítica, agora restringindo a amostra exclusivamente às transferências que envolveram movimentação financeira. Essa filtragem reduziu a amostra para 199 observações válidas, o que naturalmente impacta a estabilidade dos modelos e deve ser levado em conta na interpretação dos resultados.

Tal como na etapa anterior, foram testados cinco modelos distintos, cada um baseado em um método específico de transformação *fuzzy Totally Fuzzy and Relative* alternativa (TFRa), *Integrated Fuzzy and Relative* (IFR) e as três variações do *Interval Type-2 Fuzzy Sets* (IT2FS). Para todos os modelos, a estrutura do algoritmo *Random Forest* foi mantida, com 500 árvores e a mesma semente de aleatoriedade para garantir reprodutibilidade.

O quadro 18 sintetiza os resultados obtidos, expressos por meio do erro quadrático médio (RMSE) e do coeficiente de determinação ( $R^2$ ):

<b>Modelo</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
TFRa	0,2844	0,0961
IFR	0,2853	0,0886
IT2FS ( <i>center</i> )	0,1798	0,1824
IT2FS ( <i>lower</i> )	0,1793	0,1930
IT2FS ( <i>upper</i> )	0,1797	0,1853

Quadro 18 - Resultados para valores de transferência positivo

A análise dos novos resultados revela um padrão diferente em relação à primeira etapa. As variações da transformação IT2FS – especialmente os conjuntos *lower* e *upper* – passaram a registrar melhor desempenho, com valores de  $R^2$  próximos de 0,19. Embora esses coeficientes ainda representem uma capacidade explicativa limitada, apontam para um aumento da sensibilidade do modelo em captar as nuances das transferências monetárias, quando comparadas às transferências sem valor declarado.

Em contrapartida, os modelos baseados em TFRa e IFR apresentaram desempenho inferior nesta nova amostra, com valores de  $R^2$  abaixo de 0,10. Esse comportamento pode ser atribuído à natureza das transformações baseadas em distribuição relativa, que, embora úteis em grandes volumes de dados, podem perder precisão em amostras menores e mais concentradas.

O aumento do RMSE observado nos modelos TFRa e IFR também reforça o impacto da redução da base amostral, indicando que o erro absoluto em relação à variável de interesse se tornou mais acentuado. Já nos modelos IT2FS, o RMSE menor sugere que a transformação *fuzzy*, ao preservar a heterogeneidade dos dados com maior granularidade, contribuiu para estabilizar as previsões, mesmo em uma amostra reduzida. Essa rodada de experimentações aprofunda a compreensão das potencialidades e limitações dos diferentes métodos de transformação *fuzzy* no contexto de modelagem preditiva com base em desempenho esportivo.

#### 4.13.6. Análise dos modelos *fuzzificados* com a variável de interesse valor de mercado

Na sequência dos experimentos conduzidos para avaliar o desempenho dos modelos *fuzzificados* aplicados à predição de variáveis econômicas associadas aos atletas, esta terceira e última etapa buscou analisar o valor de mercado estimado pela comunidade de usuários da plataforma *Transfermarkt*. Esta variável, ao contrário da variável de transferência observada nas análises anteriores, é construída de maneira colaborativa e agrega percepções sobre potencial, regularidade e prestígio dos jogadores, apresentando-se como um indicador *proxy* de reputação no mercado futebolístico.

Nesta etapa da análise, as transformações *fuzzy* aplicadas foram mantidas conforme a lógica das análises anteriores *Totally Fuzzy and Relative* alternativa (TFRa), *Integrated Fuzzy and Relative* (IFR) e as três variantes do *Interval Type-2 Fuzzy Sets* (IT2FS – *center*, *lower* e *upper*). O algoritmo *Random Forest* foi novamente utilizado como método de predição, com as mesmas configurações paramétricas e controle de reprodutibilidade. A amostra utilizada, composta por 973 observações, representa a maior entre as três rodadas de ensaios realizadas nesta etapa da tese, proporcionando maior estabilidade aos modelos e, conseqüentemente, resultados mais robustos.

O quadro 19 sintetiza os resultados obtidos:

<b>Modelo</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
TFRa	0,2371	0,2816
IFR	0,2234	0,2776
IT2FS ( <i>center</i> )	0,0374	0,0792

IT2FS ( <i>lower</i> )	0,0367	0,0818
IT2FS ( <i>upper</i> )	0,0371	0,0810

Quadro 19 - Resultados para valor de mercado

Os resultados evidenciam um desempenho superior dos modelos baseados nas transformações TFRa e IFR quando comparados às demais alternativas *fuzzy* nesta aplicação específica. Com valores de  $R^2$  em torno de 0,28, esses modelos demonstraram maior capacidade de explicar a variabilidade do valor de mercado dos atletas a partir das estatísticas de desempenho observadas. Esse desempenho é relevante quando comparado às análises anteriores, nas quais as mesmas transformações apresentaram coeficientes de determinação inferiores.

A melhoria observada sugere que os métodos TFRa e IFR, por suas características de calibração relativa, tornam-se mais eficazes quando aplicados em contextos com volume mais amplo de dados – condição atendida nesta rodada, considerando as 973 observações. A modelagem *fuzzy* relativa depende da variação e da distribuição das variáveis na amostra, e seu desempenho pode ser comprometido em cenários com baixa densidade informacional, como observado nas transferências com valor financeiro, especialmente após filtragens mais restritivas.

Por outro lado, os modelos baseados em IT2FS apresentaram valores de  $R^2$  inferiores nesta análise, mantendo-se próximos a 0,08, apesar de seus RMSE serem menores. Isso indica que, embora as previsões numéricas tenham sido mais próximas em termos absolutos (erro médio), esses modelos foram menos eficazes na explicação da variância do valor de mercado. Este é um comportamento coerente com o que foi observado nas análises anteriores e que reforça o caráter mais descritivo e menos explicativo dessas transformações quando aplicadas isoladamente.

#### 4.13.7. Análise dos modelos *fuzzificados* com a variável de interesse gols

Como desdobramento das etapas anteriores e como forma de complementar as análises, buscou-se interpretar a capacidade explicativa dos mesmos modelos *fuzzificados* sobre a variável de interesse gols, previamente tratada como variável explicativa. A motivação para essa reinterpretação analítica já havia sido apresentada anteriormente, a partir do exemplo metodológico de Markopoulo *et al.* (2024), que propuseram o uso de algoritmos de *machine learning* para prever a marcação de gols em grandes ligas. A partir dessa inspiração, a variável

que representa o número de gols marcados por cada atleta passou a ser entendida como fenômeno a ser explicado, sendo, portanto, alvo direto da modelagem preditiva.

Os resultados obtidos com os modelos *Random Forest* sobre a variável gols revelaram um desempenho superior àquele observado nas análises que buscavam explicar os valores financeiros associados aos atletas. Dentre os cinco modelos *fuzzificados* avaliados, todos apresentaram coeficientes de determinação ( $R^2$ ) superiores a 0,75, com destaque para o modelo baseado no padrão IT2FS *center*, que alcançou um  $R^2$  de 0,7735. Os demais resultados mantiveram valores semelhantes, com TFRa (0,762), IFR (0,7502), IT2FS *lower* (0,7666) e IT2FS *upper* (0,7697), apontando para uma robusta capacidade de explicação do número de gols marcados com base nas variáveis disponíveis nas bases *fuzzificadas*.

O erro quadrático médio (RMSE) se manteve em níveis baixos, indicando que a magnitude das diferenças entre os valores observados e os valores previstos foi reduzida. O modelo baseado em IFR apresentou o menor RMSE entre os cinco (0,0035), seguido pelos modelos IT2FS *center* (0,0096), IT2FS *upper* (0,0097), IT2FS *lower* (0,0098) e, por fim, o TFRa (0,0211), que embora tenha apresentado o maior erro médio dentre os modelos, ainda mantém uma performance robusta em termos comparativos.

Esses achados sugerem que os modelos *fuzzificados* apresentam adequação para explicar fenômenos diretamente ligados ao desempenho em campo, como a marcação de gols. Esse resultado reforça a hipótese de que variáveis técnicas captadas na base *Wyscout* se conectam de maneira mais direta com aspectos do desempenho esportivo imediato do que com os valores atribuídos no mercado de transferências, os quais estão sujeitos a outras variáveis contextuais e econômicas.

Na sequência, a análise se concentrará nas variáveis assistências e minutos jogados, com o objetivo de verificar se a tendência de boa performance preditiva se mantém quando o foco da explicação recai sobre outros aspectos do desempenho no campo de jogo.

#### **4.13.8. Análise dos modelos *fuzzificados* com a variável de interesse assistências**

Dando continuidade à ampliação das variáveis de interesse, esta etapa do estudo concentrou-se na análise do fenômeno das assistências. Por representarem as contribuições diretas para a finalização bem-sucedida das jogadas ofensivas, essa estatística é compreendida como uma ação decisiva no contexto do futebol profissional. As assistências foram reposicionadas de

variável explicativa à variável explicada, de modo a avaliar em que medida os atributos técnicos presentes nas bases *fuzzificadas* são capazes de explicar esse desfecho coletivo.

Para tanto, os cinco modelos *fuzzificados* foram aplicados utilizando o algoritmo *Random Forest*. Os resultados revelam desempenho inferior em comparação ao observado na previsão de gols, ainda que alguns modelos tenham alcançado níveis moderados de explicação da variável de interesse.

Entre os modelos testados, o padrão IT2FS *upper* apresentou o melhor desempenho, com um coeficiente de determinação ( $R^2$ ) de 0,3297 e um erro quadrático médio (RMSE) de 0,0204. Resultados próximos foram encontrados no modelo IT2FS *lower* ( $R^2 = 0,3203$  e RMSE = 0,0205) e IT2FS *center* ( $R^2 = 0,2882$  e RMSE = 0,021), confirmando a ligeira superioridade do paradigma de conjuntos *fuzzy* intervalares (IT2FS) na explicação deste tipo específico de ação ofensiva.

Já os modelos baseados nas transformações IFR ( $R^2 = 0,2285$  e RMSE = 0,0034) e TFRa ( $R^2 = 0,2404$  e RMSE = 0,0316) apresentaram desempenhos razoáveis, mas inferiores ao conjunto IT2FS. Isso pode sugerir que abordagens que incorporam graus mais amplos de incerteza e variabilidade linguística (como ocorre nos modelos IT2FS) são mais adequadas para captar nuances relacionadas à criação de jogadas ofensivas coletivas.

Embora os níveis de  $R^2$  observados não sejam tão elevados quanto os relacionados à variável gols, os modelos mostraram-se suficientemente sensíveis para identificar certos padrões explicativos. Destaca-se que a ação de assistência depende simultaneamente do desempenho do assistente e do jogador que finaliza, o que pode introduzir um ruído adicional na previsão e justificar a redução na capacidade explicativa.

Assim, os resultados reforçam a validade do redesenho analítico que reposicionou assistências como variável dependente, ao mesmo tempo em que sugerem a continuidade da investigação sobre a influência de diferentes padrões de *fuzzificação* na modelagem de ações coletivas ofensivas. A próxima seção se voltará à variável minutos jogados, indicador de participação regular e consistência de um atleta ao longo das competições.

#### **4.13.9. Análise dos modelos *fuzzificados* com a variável de interesse minutos jogados**

Os minutos jogados foram incorporados ao escopo analítico como uma medida de participação e confiabilidade dos atletas ao longo da temporada, revelando-se como um indicador de

continuidade, condição física e regularidade no contexto competitivo. Assim como no caso das variáveis gols e assistências, esta análise buscou avaliar a capacidade explicativa dos modelos *fuzzificados* em sua forma original, isto é, ainda sem aplicação dos pesos obtidos pelas regressões penalizadas.

Todos os modelos apresentaram capacidade preditiva com coeficientes de determinação ( $R^2$ ) superiores a 0,84, indicando que a maior parte da variabilidade dos minutos jogados pôde ser explicada pelas demais variáveis técnicas incluídas nos modelos.

O modelo IFR alcançou o melhor desempenho geral, com um  $R^2$  de 0,8564 e RMSE de 0,0605. Em seguida, aparecem os modelos IT2FS *lower* ( $R^2 = 0,8494$  e RMSE = 0,0375) e TFRa ( $R^2 = 0,8488$  e RMSE = 0,0599), demonstrando resultados muito próximos. Os modelos IT2FS *center* e *upper* apresentaram desempenhos com  $R^2$  de 0,8454 e 0,8451, respectivamente, e RMSE em torno de 0,0379 em ambos os casos.

Esses dados sinalizam que a variável é mais estável e menos afetada por fatores aleatórios do jogo do que as demais variáveis analisadas. Essa estabilidade, associada à natureza cumulativa dos minutos ao longo da temporada, pode ter contribuído para a maior robustez dos modelos neste caso específico.

Além disso, a performance generalizada entre os modelos sugere que, ao contrário do que ocorreu em análises de variáveis mais sensíveis a contexto situacional (como valores de transferência ou assistências), a consistência da variável minutos jogados favoreceu o funcionamento eficiente das abordagens *fuzzy* mais simples. A comparação entre os modelos para explicação das variáveis de contexto financeiro pode ser vista no quadro 20.

<b>Transformação fuzzy</b>	<b>Transferência Geral (<math>R^2</math>)</b>	<b>Transferência com Valor (<math>R^2</math>)</b>	<b>Valor de Mercado (<math>R^2</math>)</b>
TFRa	0,0323	0,0961	0,2816
IFR	0,0359	0,0886	0,2776
IT2FS ( <i>center</i> )	0,0005	0,1824	0,0792
IT2FS ( <i>lower</i> )	0,0006	0,1930	0,0818
IT2FS ( <i>upper</i> )	0,0005	0,1853	0,0810

Quadro 20 - Resultados para variáveis financeiras

Com os resultados desta análise, encerra-se o ciclo de testes exploratórios com as bases *fuzzificadas* sem ponderações, tendo como variáveis de interesse valor de mercado, valores de transferência (com e sem custo zero), gols, assistências e minutos jogados. Em termos comparativos, os modelos demonstraram melhor desempenho explicativo para variáveis diretamente associadas ao desempenho em campo (como gols e minutos jogados), enquanto enfrentaram maior dificuldade para prever variáveis de natureza econômica, cuja determinação depende de fatores externos ao jogo, como contexto de mercado, duração de contrato, agentes e potencial de revenda. A comparação dos resultados da explicação das variáveis de desempenho pode ser observada no quadro 21.

<b>Transformação <i>fuzzy</i></b>	<b>Gols (R<sup>2</sup>)</b>	<b>Assistências (R<sup>2</sup>)</b>	<b>Minutos jogados (R<sup>2</sup>)</b>
TFRa	0,7620	0,2404	0,8488
IFR	0,7502	0,2285	0,8564
IT2FS ( <i>center</i> )	0,7735	0,2882	0,8454
IT2FS ( <i>lower</i> )	0,7666	0,3203	0,8494
IT2FS ( <i>upper</i> )	0,7697	0,3297	0,8451

Quadro 21 - Resultado para variáveis de performance

A partir desses achados, é possível reforçar a hipótese de que modelos de *fuzzificação* com base em calibrações relativas (como TFRa e IFR) são mais adequados para contextos em que há uma amostra suficientemente grande e variada, favorecendo a estruturação de pertinências *fuzzy* mais informativas. Já os modelos baseados em representações mais formais da incerteza (como IT2FS), embora úteis para capturar nuances semânticas, demonstraram maior limitação preditiva neste contexto específico.

As próximas etapas do trabalho exploram como a ponderação por regressões penalizadas pode refinar esse desempenho, especialmente nas variáveis cuja explicação se mostrou mais sensível ao ruído contextual. Os achados reforçam a necessidade de calibrar os modelos conforme o objetivo específico da análise e o perfil da base de dados utilizada. As próximas seções aprofundarão esse raciocínio, incorporando os pesos obtidos por meio das regressões penalizadas e analisando a performance dos modelos ponderados.

#### 4.13.10. Análise dos modelos *fuzzificados* com pesos de regressões penalizadas para *transfer fee* (TF)

Após a realização da primeira etapa das análises com dados *fuzzificados*, uma nova rodada de testes teve como objetivo avaliar o impacto da aplicação de ponderações derivadas das regressões penalizadas (*Elastic Net*, LASSO e Ridge) na modelagem dos dados. A lógica empregada nesta fase visou capturar a relevância das habilidades dos jogadores, de acordo com os pesos extraídos dos modelos de regressão, de modo a aprimorar a explicação da variável de interesse valor de transferência ("*transfer fee*"), incluindo na amostra tanto transferências com custos associados quanto aquelas a custo zero.

Inicialmente, as transformações foram ajustadas com as três diferentes penalizações. As bases de dados ponderadas foram divididas em conjuntos de treino (70%) e teste (30%) de maneira estratificada, proporcionando a aleatoriedade e a comparabilidade dos resultados. Manteve-se como modelo de *machine learning* o *Random Forest* configurado com 500 árvores ( $n = 500$ ).

Em comparação com a primeira rodada de análises, que utilizava apenas a transformação *fuzzy* sem ponderações pelos coeficientes das regressões, nota-se uma melhora nos valores de RMSE para praticamente todos os padrões de *fuzzificação*, com destaque para o padrão IFR, que apresentou os menores valores de erro absoluto e um ganho em termos de explicação da variável dependente (*transfer fee*).

No entanto, em termos de  $R^2$ , embora alguns modelos ponderados (especialmente IFR ponderado por *Elastic Net*) tenham apresentado desempenho superior aos modelos não ponderados, os valores absolutos de  $R^2$  ainda permanecem relativamente baixos. Tal fato sugere que, embora a ponderação tenha contribuído para uma melhora da capacidade preditiva, ainda existem limitações intrínsecas ao conjunto de variáveis disponíveis ou à natureza dos dados de transferência considerados. A comparação dos resultados pode ser observada no quadro 22.

Padrão de <i>fuzzificação</i>	Regressão Penalizada	RMSE	$R^2$
IFR	<i>Elastic Net</i>	0,0943	0,0984
IFR	LASSO	0,0960	0,0714
IFR	Ridge	0,0948	0,0899
IFR	(sem ponderação)	0,1030	0,0359

IT2FS ( <i>center</i> )	<i>Elastic Net</i>	0,1041	0,0061
IT2FS ( <i>center</i> )	LASSO	0,1012	0,0152
IT2FS ( <i>center</i> )	Ridge	0,0990	0,0267
IT2FS ( <i>center</i> )	(sem ponderação)	0,1394	0,0005
IT2FS ( <i>lower</i> )	<i>Elastic Net</i>	0,1039	0,0055
IT2FS ( <i>lower</i> )	LASSO	0,1018	0,0139
IT2FS ( <i>lower</i> )	Ridge	0,0989	0,0365
IT2FS ( <i>lower</i> )	(sem ponderação)	0,1418	0,0006
IT2FS ( <i>upper</i> )	<i>Elastic Net</i>	0,1042	0,0039
IT2FS ( <i>upper</i> )	LASSO	0,1002	0,0100
IT2FS ( <i>upper</i> )	Ridge	0,0989	0,0312
IT2FS ( <i>upper</i> )	(sem ponderação)	0,1388	0,0005
TFRa	<i>Elastic Net</i>	0,2337	0,0901
TFRa	LASSO	0,2359	0,0756
TFRa	Ridge	0,2356	0,0760
TFRa	(sem ponderação)	0,2463	0,0323

Quadro 22 - Comparação resultados para variável valor de transferência

Concluída esta primeira série de ensaios focados no valor de transferência dos atletas, a pesquisa avança para a aplicação dos mesmos modelos *fuzzificados* e ponderados pelas regressões penalizadas para a variável de interesse relacionada ao valor de mercado (*market value*), conforme estimado pelos usuários da plataforma *Transfermarkt*.

A escolha por esta variável se justifica tanto pelo volume mais elevado de observações disponíveis quanto pela expectativa de que o valor de mercado possa refletir de maneira mais ampla a percepção de desempenho dos atletas no cenário competitivo, potencialmente permitindo ganhos de explicação nos modelos a serem analisados.

#### 4.13.11. Modelos *fuzzificados* ponderados para explicação da variável valor de mercado (*market value*)

Com base nos resultados obtidos anteriormente na explicação da variável *transfer fee*, a presente etapa da pesquisa trata sobre a variável *market value*, estimada pela plataforma *Transfermarkt*. Essa variável, que representa o valor de mercado dos atletas segundo um processo coletivo de atribuição, apresenta maior volume de observações e, potencialmente, maior estabilidade em relação a flutuações pontuais de mercado, justificando sua seleção como objeto de análise complementar.

Mantendo a consistência metodológica, os modelos *fuzzy* foram ponderados pelos coeficientes extraídos das regressões penalizadas LASSO, Ridge e *Elastic Net*. A base de dados, composta por informações transformadas, foi submetida à técnica de *Random Forest*, com divisão padrão em 70% dos dados para treino e 30% para teste, e geração de 500 árvores por modelo.

Para uma melhor visualização da performance geral dos modelos *fuzzy* ponderados nesta etapa de análise, apresenta-se a seguir a tabela-síntese comparativa dos modelos aplicados à variável de valor de mercado. Como elemento de comparação, foram utilizados os parâmetros de resultados encontrados com os modelos não ponderados. Os resultados podem ser observados no quadro 23.

<b>Transformação <i>fuzzy</i></b>	<b>Ponderação</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
IFR	<i>Elastic Net</i>	0,2303	0,2645
IFR	LASSO	0,2300	0,2672
IFR	Ridge	0,2354	0,2287
IFR	(sem ponderação)	0,2234	0,2776
IT2FS <i>center</i>	<i>Elastic Net</i>	0,0508	0,1273
IT2FS <i>center</i>	LASSO	0,0509	0,1249
IT2FS <i>center</i>	Ridge	0,0537	0,0443
IT2FS <i>center</i>	(sem ponderação)	0,0374	0,0792
IT2FS <i>lower</i>	<i>Elastic Net</i>	0,0507	0,1330
IT2FS <i>lower</i>	LASSO	0,0512	0,1154
IT2FS <i>lower</i>	Ridge	0,0531	0,0558
IT2FS <i>lower</i>	(sem ponderação)	0,0367	0,0818

IT2FS <i>upper</i>	<i>Elastic Net</i>	0,0506	0,1354
IT2FS <i>upper</i>	LASSO	0,0509	0,1238
IT2FS <i>upper</i>	Ridge	0,0528	0,0641
IT2FS <i>upper</i>	(sem ponderação)	0,0371	0,0810
TFRa	<i>Elastic Net</i>	0,2416	0,2922
TFRa	LASSO	0,2425	0,2852
TFRa	Ridge	0,2448	0,2665
TFRa	(sem ponderação)	0,2371	0,2816

Quadro 23 - Comparação de resultados para a variável valor de mercado

Ao comparar os resultados com aqueles obtidos na etapa anterior, em que os modelos *fuzzy* foram utilizados sem ponderação, nota-se uma melhora nos índices de explicação ( $R^2$ ) para quase todos os conjuntos utilizando as ponderações de coeficiente oriundos das regressões *Elastic Net* e LASSO, exceção feita aos modelos ponderados aplicados à transformação IFR, que se mostraram piores nos índices de explicação.

A análise conduzida confirma a superioridade relativa dos modelos TFRa, especialmente quando ponderados por *Elastic Net* e *LASSO*, em cenários com maior volume de observações, como é o caso da base de dados de mercado utilizada aqui. Essa elevação na performance evidencia que a ponderação dos dados *fuzzy* com coeficientes derivados de regressões penalizadas contribui para maior acurácia e coerência nos modelos, sobretudo quando aplicadas a variáveis complexas como valor de mercado, que incorpora fatores técnicos, contextuais e subjetivos.

Com o encerramento da fase de avaliação dos modelos *fuzzy* ponderados aplicados à variável de valor de mercado, dá-se início à próxima etapa analítica dedicada à investigação dos modelos aplicados à variável de valor de transferência, considerando somente os casos com movimentação financeira (*only transfer fee* – OTF). A exclusão das transferências a custo zero visa eliminar ruídos estatísticos e permitir uma avaliação mais refinada das associações entre o desempenho técnico dos atletas e os valores efetivamente praticados no mercado de transferências. Essa abordagem deverá contribuir para uma compreensão mais precisa sobre os determinantes objetivos do valor de mercado transacional no futebol profissional brasileiro. Os resultados desta nova etapa serão discutidos na próxima seção.

#### 4.13.12. Modelos *fuzzificados* ponderados para a variável transferência com valores positivos (OTF)

Dando sequência à lógica analítica construída ao longo desta tese, esta etapa tem por objetivo avaliar a capacidade preditiva dos modelos *fuzzy* ponderados na explicação da variável de valor de transferência considerando exclusivamente os casos com movimentação financeira. Isto é, as transferências que envolveram pagamento de valores entre clubes que foi nomeado como a variável transferência com valores positivos (*only transfer fee* – OTF). Essa delimitação da amostra, ao excluir as transferências a custo zero, visa reduzir o ruído gerado por variações contratuais ou circunstâncias de mercado não explicadas pelas métricas técnico-funcionais dos atletas.

As variáveis foram ponderadas a partir dos coeficientes gerados por regressões penalizadas do tipo *Elastic Net*, LASSO e Ridge. Como nas etapas anteriores, o modelo de *Random Forest* foi executado com 500 árvores e os dados foram divididos em conjuntos de treino (70%) e teste (30%). Com essa análise, conclui-se a rodada de modelos aplicados à variável de interesse ligada às transferências com valores positivos. O quadro 24 apresenta um comparativo sintético entre os modelos testados para essa variável, possibilitando a leitura integradora dos achados.

<b>Transformação <i>fuzzy</i></b>	<b>Ponderação</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
IFR	<i>Elastic Net</i>	0,2783	0,0791
IFR	LASSO	0,2799	0,0674
IFR	Ridge	0,2765	0,0847
IFR	(sem ponderação)	0,2853	0,0886
IT2FS <i>center</i>	<i>Elastic Net</i>	0,1683	0,0040
IT2FS <i>center</i>	LASSO	0,1665	0,0052
IT2FS <i>center</i>	Ridge	0,1677	0,0128
IT2FS <i>center</i>	(sem ponderação)	0,1798	0,1824
IT2FS <i>lower</i>	<i>Elastic Net</i>	0,1726	0,0024
IT2FS <i>lower</i>	LASSO	0,1702	0,0068
IT2FS <i>lower</i>	Ridge	0,1728	0,0093

IT2FS <i>lower</i>	(sem ponderação)	0,1793	0,1930
IT2FS <i>upper</i>	<i>Elastic Net</i>	0,1659	0,0009
IT2FS <i>upper</i>	LASSO	0,1627	0,0060
IT2FS <i>upper</i>	Ridge	0,1657	0,0132
IT2FS <i>upper</i>	(sem ponderação)	0,1797	0,1853
TFRa	<i>Elastic Net</i>	0,2813	0,0613
TFRa	LASSO	0,2814	0,0576
TFRa	Ridge	0,2855	0,0415
TFRa	(sem ponderação)	0,2844	0,0961

Quadro 24 - Comparação de resultados para a variável valor de transferência positivo

A aplicação dos modelos *fuzzificados* ponderados à variável OTF revelou um cenário metodológico desafiador. Embora os dados tenham sido tratados com técnicas de ponderação baseadas em regressões penalizadas e as transformações *fuzzy* tenham permitido abordagens de representação mais adequadas ao domínio esportivo, o desempenho preditivo dos modelos permaneceu restrito.

A análise dos resultados indica que os modelos baseados na transformação IFR, especialmente quando ponderados com Ridge ( $R^2 = 0,0847$ ), obtiveram desempenho ligeiramente superior. Ainda assim, os valores de  $R^2$  obtidos foram, de forma geral, bastante modestos, sugerindo uma limitada capacidade explicativa dos modelos testados sobre a variação nos valores efetivamente pagos por atletas no mercado de transferências.

As transformações IT2FS, ainda que sofisticadas na representação da ambiguidade, apresentaram  $R^2$  extremamente baixos, com destaque negativo para os modelos *Elastic Net* e LASSO das versões *center* e *upper*, onde a variabilidade explicada foi praticamente nula. Isso reforça a hipótese de que a natureza mais subjetiva e distribuída destas representações pode não se alinhar adequadamente à lógica econômica de precificação direta no futebol, sobretudo em mercados tão assimétricos quanto o brasileiro. O resultado contrasta com os valores encontrados nos modelos em que não houve aplicação das penalizações, indicando que o modelo intervalar sem interferências pode ter captado melhor as nuances da variável de interesse neste caso.

A transformação TFRa, por sua vez, manteve um desempenho consistente em comparação com a etapa anterior sem ponderação, mas sem apresentar ganhos significativos. Isso sugere que, na ausência de outros fatores contextuais relevantes (como histórico de lesões, influência de agentes, tempo de contrato, entre outros), o uso isolado das estatísticas de desempenho técnico mesmo ponderadas ainda é insuficiente para captar a complexidade dos valores transacionais pagos.

Em relação ao processo de reanálise decorrente da baixa capacidade explicativa dos modelos sobre os valores de transferência, encontrou-se respaldo no trabalho de Markopoulo *et al.* (2024), que propõe a observação de resultados de performance dentro de campo. A aplicação dos cinco modelos *fuzzy*, sem ponderação, sobre a base de dados transformada, com a transposição das colunas de gols, assistências e minutos jogados de fatores explicativos para variáveis de interesse, apresentou bons coeficientes de determinação ( $R^2$ ) para gols e minutos jogados, mas revelou uma explicação limitada para o fenômeno das assistências.

Entretanto, as regressões penalizadas desenvolvidas nesta tese utilizaram justamente esses fatores como referência para encontrar coeficientes capazes de ponderar a importância das estatísticas sobre as habilidades, conceito oriundo da base de dados do jogo *Football Manager* 2024. Na análise realizada sobre um dos modelos que utiliza a ponderação, observou-se que o ganho na explicação das assistências mais do que dobrou (aumento de 104% na capacidade explicativa), ainda que tal resultado seja, em parte, influenciado pelo fato de se tratar de uma variável utilizada na geração dos próprios pesos. Apesar disso, a capacidade explicativa das variáveis gols e minutos jogados apresentou redução em algumas das transformações avaliadas. Motivados por isso, outros ensaios foram executados com variações na quantidade de fatores dos bancos de dados, e os resultados serão apresentados na próxima seção.

#### **4.14. Outros modelos *fuzzificados* testados e resultados obtidos**

Antes de apresentar os próximos resultados, é necessário indicar aos leitores que, apesar dos esforços empreendidos, os resultados referentes às variáveis financeiras não apresentaram melhorias significativas em nenhuma das novas versões dos modelos testados. Em contrapartida, os coeficientes de determinação ( $R^2$ ) para os indicadores de performance mantiveram valores adequados e, em alguns casos, demonstraram melhora na capacidade de explicação das variáveis de interesse. Esses efeitos corroboram o entendimento de que os valores de mercado e de transferência reúnem outros fatores explicativos não contemplados

pela base de dados utilizada neste trabalho, como, por exemplo, dimensões táticas e físicas da performance, histórico de lesões, tempo de contrato, entre outros elementos influenciadores.

Para alcançar esta conclusão ainda foram executados dois testes, adotando uma opção metodológica adicional em realizar as análises com apenas um tipo de transformação *fuzzy* – a *Integrated Fuzzy and Relative* (IFR) – de forma agregada, utilizando as colunas estatísticas específicas de cada teste, com e sem ponderação pelos coeficientes oriundos da regressão *Elastic Net*. A escolha pelo IFR se deve, em primeiro lugar, à preferência do autor desta tese por transformações *fuzzy* relativas, sendo esta apontada por seus autores como uma evolução da abordagem TFRa (BETTI *et al.*, 2023). Já a ponderação via *Elastic Net* foi escolhida por ser um modelo intermediário entre as abordagens LASSO e Ridge.

A primeira redução de fatores da base de dados oriunda do *Wyscout* foi realizada com a transformação das colunas gols, assistências e minutos jogados em variáveis de interesse, além da exclusão de outras 10 colunas com baixas dispersões nos dados *fuzzificados*. As colunas excluídas foram ações totais e ações totais bem-sucedidas,  $xG$ , carrinhos e carrinhos bem-sucedidos, cartões amarelos, cartões vermelhos, impedimentos (*offsides*),  $xA$  e segundas assistências. Com isso, uma nova regressão penalizada foi executada, obtendo-se novos coeficientes, os quais foram aplicados em uma base agregada com a pontuação *fuzzy* de cada jogador em relação às 36 habilidades modeladas a partir do jogo *Football Manager*.

Os resultados demonstraram que a configuração agregada e reduzida para 42 colunas do IFR, sem ponderação, apresentou melhorias na explicação de duas das três variáveis de interesse. Para minutos jogados, o modelo atingiu um coeficiente de determinação ( $R^2$ ) de 0,8680, superior tanto à versão completa (com 52 colunas) da IFR ( $R^2 = 0,8564$ ) quanto à versão ponderada por *Elastic Net* ( $R^2 = 0,7782$ ). Da mesma forma, para assistências, observou-se um ganho de explicação com  $R^2 = 0,4882$ , valor também superior à versão completa ( $R^2 = 0,2285$ ) e ao modelo ponderado ( $R^2 = 0,2011$ ). A variável gols, por sua vez, apresentou leve redução em relação à versão completa ( $R^2 = 0,7487$  frente a  $R^2 = 0,7502$ ), mas manteve desempenho superior ao modelo ponderado ( $R^2 = 0,7026$ ), evidenciando a robustez da abordagem mesmo após a exclusão de variáveis com baixa variabilidade estatística.

O desempenho inferior do modelo ponderado em relação à base reduzida sem ponderação em duas das três variáveis analisadas sugere que, embora o uso de penalizações via *Elastic Net* possa contribuir para o controle da multicolinearidade e a seleção de variáveis, sua aplicação direta sobre o conjunto de 36 habilidades pode comprometer a explicação de fenômenos cuja

complexidade contextual exige a preservação de determinados indicadores estatísticos. A escolha por uma regressão penalizada para calibrar os pesos do desempenho histórico, segmentando as estatísticas entre períodos anteriores a 2023 (*e.g.*, 2015–2022) e a temporada mais recente, se mostrou mais adequada para melhorar os resultados dos dados *fuzzificados* em relação às variáveis financeiras. Contudo, nenhum modelo ponderado superou as versões sem ponderação na explicação dos indicadores de performance em campo.

Esse mesmo padrão de desempenho foi observado em uma última regressão experimental adicional, restrita apenas às habilidades técnicas oriundas do jogo. Foram consideradas, conforme o português de Portugal utilizado no *game*, as habilidades finalização, cantos, cruzamentos, finta, primeiro toque, livres, cabeceamento, remates de longe, lançamento longo, marcação, passe, marcação de pênalti, desarme e técnica. A regressão foi executada com base em 30 colunas estatísticas do *Wyscout*, considerando apenas ações bem-sucedidas, como remates à baliza, dribles certos, entre outros indicadores. O desempenho desse modelo, no entanto, ficou aquém dos anteriores, tanto na explicação de gols, assistências e minutos jogados quanto nas variáveis financeiras.

Como último ensaio, foi realizado um ajuste nas colunas de fatores e adicionadas novas variáveis de interesse que já estavam presentes na versão original da base *fuzzificada*. Essa versão do modelo apresentou os melhores desempenhos nos coeficientes de determinação para gols e minutos jogados, além de ganhos relevantes na explicação de assistências. Entre as novas variáveis de interesse analisadas estão faltas cometidas, faltas sofridas, interceptações, além dos indicadores de gols esperados (*expected goals – xG*) e assistências esperadas (*expected assists – xA*), métricas que avaliam, respectivamente, a probabilidade de que uma finalização resulte em gol e a probabilidade de que um passe se transforme em assistência. Os 39 fatores estatísticos calibrados em *fuzzy* e aplicados à explicação das oito variáveis de interesse estão apresentados no quadro 25.

Fatores explicativos				Variáveis de interesse
Remates	Duelos	Duelos de bola livre	Passes para terço final	Minutos jogados
Remates a baliza	Duelos ganhos	Duelos de bola livre ganho	Passes para terço final certos	Golos

Passes	Duelos aéreos	Alívios	Passes para a grande área	Assistências
Passes certos	Duelos aéreos ganhos	Assistências para remate	Passes para a grande área precisos	Interceptações
Passes longos	Perdas	Duelos ofensivos	Passes recebidos	Faltas
Passes longos certos	Perdas meio-campo	Duelos ofensivos ganhos	Passes para a frente	Faltas sofridas
Cruzamentos	Recuperações	Toques na área	Passes para a frente certos	$xG$
Cruzamentos certos	Recuperações campo adversário	Corridas seguidas	Passes para trás	$xA$
Dribles	Duelos defensivos	Passes em profundidade	Passes para trás certos	
Dribles sucesso	Duelos defensivos ganhos	Passes em profundidade certos		

Quadro 25 - Fatores estatísticos e variáveis de interesse para modelo final

Uma vez que o resultado deste modelo, de 39 colunas estatísticas e oito variáveis de interesse, apresentou os melhores desempenhos para a transformação IFR optou-se por aplicar o algoritmo *Random Forest* nas demais transformações *fuzzy* e avaliar o melhor desempenho entre elas. Os resultados serão descritos a seguir, apresentando o modelo final desta tese.

#### 4.14.1. Modelo IFR para explicação de variáveis de performance

A primeira configuração avaliada, com 39 colunas estatísticas transformadas via abordagem *fuzzy* relativa (IFR), apresentou bons resultados na explicação das variáveis de interesse, corroborando a consistência da modelagem. Utilizando o algoritmo *Random Forest*, foram testadas oito variáveis de interesse diretamente relacionadas à performance esportiva, sendo elas gols, assistências, minutos jogados, faltas cometidas, faltas sofridas, interceptações,  $xG$  e  $xA$ .

Os resultados revelam que a modelagem *fuzzy* baseada no padrão IFR foi eficaz na explicação de  $xG$  ( $R^2 = 0,9283$ ), interceptações ( $R^2 = 0,9312$ ) e minutos jogados ( $R^2 = 0,8691$ ), indicando

boa capacidade do modelo para essas métricas. O desempenho da variável interceptações se vincula à leitura de jogo e posicionamento, evidencia a pertinência dos fatores estatísticos considerados, especialmente os ligados a ações defensivas e duelos. Por sua vez, as outras variáveis possuem ligação com o volume de participação do jogador e a incidência de ações ofensivas registradas na base estatística, fatores bem capturados pelas variáveis explicativas selecionadas.

A variável  $x_A$  também apresentou um desempenho elevado ( $R^2 = 0,8881$ ), sugerindo que a abordagem *fuzzy* relativa, mesmo sem ponderações adicionais, é capaz de capturar nuances associadas à criação de jogadas com potencial de assistência. Tais resultados indicam que os indicadores de qualidade do passe, ações ofensivas e movimentações em zonas decisivas do campo, quando tratados sob a lógica *fuzzy*, oferecem substrato analítico relevante para predição de assistências esperadas.

A variável correspondente às faltas sofridas ( $R^2 = 0,6817$ ) mostrou bom desempenho preditivo. Este resultado está em consonância com o entendimento de que os atletas mais participativos e com maior volume de ações ofensivas são frequentemente os mais visados e, portanto, vítimas de faltas. Já a variável faltas cometidas ( $R^2 = 0,5564$ ) apresentou coeficiente moderado de explicação, ainda assim relevante, considerando a natureza reativa desse comportamento e a sua possível relação com características não mensuradas pela base estatística, como posicionamento defensivo e controle emocional.

No caso de gols, o modelo IFR obteve um coeficiente de determinação de  $R^2 = 0,7509$ , consolidando-se como uma das variáveis mais bem explicadas dentro do conjunto testado. Esse resultado mantém a tendência observada nas análises anteriores de bom desempenho na predição de ações finalizadoras. Por sua vez, assistências foram explicadas com um coeficiente de  $R^2 = 0,4943$ , representando um avanço em relação às análises preliminares e confirmando que ajustes no conjunto de fatores explicativos elevaram a sensibilidade do modelo a esse tipo de contribuição ofensiva.

Em suma, a transformação *fuzzy* IFR com 39 colunas explicativas demonstrou robustez analítica e aplicabilidade. Os resultados mais relevantes foram observados em interceptações,  $x_G$ ,  $x_A$  e minutos jogados, sugerindo que a modelagem *fuzzy* foi eficaz na captação de métricas associadas à intensidade, regularidade e contribuição tática e técnica dos jogadores em campo. Os resultados dessa configuração serão utilizados como referência para comparação com as demais abordagens *fuzzy* nos próximos tópicos.

#### 4.14.2. Modelo IT2FS *center* para explicação de variáveis de performance

A configuração baseada na transformação *fuzzy Interval Type-2 Fuzzy Sets* (IT2FS) utilizando o centro dos intervalos como valor representativo apresentou bons desempenhos em praticamente todas as variáveis de interesse, consolidando-se como uma das alternativas consistente do conjunto de testes. A modelagem adotada manteve as 39 colunas estatísticas previamente ajustadas, e o algoritmo *Random Forest* foi novamente utilizado como motor preditivo.

A variável interceptações obteve o melhor desempenho explicativo entre as oito analisadas, com  $R^2 = 0,9136$ , confirmando que a estrutura *fuzzy IT2FS center* tem capacidade de capturar nuances de comportamentos defensivos relacionados à leitura tática e antecipação. Este resultado se alinha aos elevados valores de explicação obtidos para  $xG$  ( $R^2 = 0,8911$ ) e minutos jogados ( $R^2 = 0,8565$ ), os quais também figuram entre os melhores desempenhos desta transformação. Em conjunto, esses resultados evidenciam que a lógica *fuzzy* baseada no centro dos intervalos é eficaz para mapear atributos objetivos e de intensidade, muitas vezes bem quantificados por estatísticas agregadas ao longo da temporada.

No que se refere à variável gols, o modelo apresentou  $R^2 = 0,7895$ , um desempenho robusto e compatível com a capacidade do modelo em captar comportamentos de finalização e presença ofensiva em zonas decisivas. Já a explicação para assistências foi moderada ( $R^2 = 0,2868$ ), mas ainda superior às versões iniciais não ajustadas de outros modelos. Esse desempenho sugere que, embora a criação de jogadas esteja relacionada a uma variedade de estatísticas (passes para zonas decisivas, dribles, presença no terço final), o modelo IT2FS *center* capta parte relevante desse comportamento, ainda que com limitações no aspecto subjetivo da assistência.

As variáveis faltas cometidas ( $R^2 = 0,5071$ ) e faltas sofridas ( $R^2 = 0,6695$ ) apresentaram desempenho relevante. A primeira se relaciona com o estilo de jogo e a agressividade dos atletas, enquanto a segunda reflete a frequência com que os jogadores são alvos de desarmes ou ações defensivas adversárias. Ambos os indicadores foram bem absorvidos pela lógica *fuzzy* aplicada, mesmo com variações de contexto e posicionamento.

Por fim, a variável  $xA$  obteve um  $R^2 = 0,7989$ , valor elevado se comparado à explicação para as assistências efetivamente registradas. Este resultado destaca que a previsão de potencial de assistência ( $xA$ ) é menos dependente de fatores contextuais incontrolláveis – como a finalização correta por parte do companheiro de equipe –, e mais próxima das ações que antecedem a

chance real de gol, captadas por variáveis como passes em profundidade, cruzamentos e assistências para remate.

Em resumo, o padrão IT2FS *center* demonstrou consistência e estabilidade na explicação das variáveis de interesse, com destaque para intercepções,  $xG$ ,  $xA$ , minutos jogados e gols. A transformação *fuzzy* do tipo Interval Type-2, ao utilizar o centro como valor de referência, mostra-se eficaz para fenômenos objetivos e com dispersão estatística relevante, confirmando sua aplicabilidade em análises que demandam robustez preditiva com preservação de incertezas inerentes ao jogo.

#### 4.14.3. Modelo IT2FS *lower* para explicação de variáveis de performance

A aplicação da transformação IT2FS *lower*, que considera o limite inferior dos intervalos de pertinência, também apresentou desempenhos consistentes para a maioria das variáveis de interesse. Isso reforça o potencial explicativo dos dados *fuzzificados* mesmo em cenários de maior incerteza e com critérios de associação mais conservadores.

A variável intercepções novamente destacou-se como uma das mais bem explicadas pelo modelo, com  $R^2 = 0,9135$ , valor praticamente idêntico ao obtido com a transformação IT2FS *center*. Essa estabilidade entre os dois modelos indica que a estrutura estatística das intercepções, associada a comportamentos defensivos observáveis, é captada de maneira robusta independentemente da variação na abordagem de pertinência *fuzzy*.

Outros bons resultados foram observados nas variáveis  $xG$  ( $R^2 = 0,8879$ ) e  $xA$  ( $R^2 = 0,8000$ ), reiterando a capacidade explicativa da transformação para métricas avançadas relacionadas ao potencial ofensivo dos atletas. Tais métricas sintetizam eventos de jogo com base em probabilidade de sucesso, beneficiando-se diretamente do grau contínuo de associação fornecido pelas estruturas *fuzzy* do tipo IT2FS.

A variável gols obteve um bom desempenho ( $R^2 = 0,7933$ ), ligeiramente superior ao observado com a versão IT2FS *center* ( $R^2 = 0,7895$ ), mantendo a estabilidade do modelo na explicação de ações objetivas e mensuráveis. Já a variável assistências apresentou melhora em relação ao modelo anterior, com  $R^2 = 0,5166$ , alcançando o melhor resultado entre os modelos IT2FS até então testados. Esse dado sugere que, ao adotar uma abordagem mais cautelosa na associação das estatísticas, o modelo IT2FS *lower* capta melhor nuances da criação de jogadas, muitas vezes dependentes de características indiretas e contexto situacional.

A variável relacionada aos minutos jogados manteve-se com elevado desempenho ( $R^2 = 0,8580$ ), reforçando a constância do padrão *fuzzy* na explicação de participação efetiva em partidas. Em linha semelhante, as variáveis faltas cometidas ( $R^2 = 0,5094$ ) e faltas sofridas ( $R^2 = 0,6697$ ) também sustentaram níveis explicativos adequados, reafirmando a robustez da abordagem mesmo para variáveis associadas a estilo de jogo e características de contato físico.

Em síntese, o modelo baseado na transformação IT2FS *lower* revelou-se consistente para a maioria das variáveis analisadas, com destaque para o desempenho explicativo nas métricas interceptações,  $xG$ ,  $xA$ , gols e minutos jogados. Além disso, demonstrou leve superioridade na explicação das assistências em relação à versão *center*, o que pode indicar que a abordagem mais restritiva da pertinência *fuzzy* contribuiu para maior discriminação de ações ofensivas complexas, mesmo com elevada variabilidade entre atletas e partidas.

#### 4.14.4. Modelo IT2FS *upper* para explicação de variáveis de performance

A transformação IT2FS *upper*, que considera os limites superiores das funções de pertinência *fuzzy*, apresentou desempenho estável na explicação das variáveis de interesse, com variações discretas em relação às demais configurações do tipo IT2FS. Esta abordagem assume uma perspectiva mais inclusiva, ou seja, tende a atribuir maior grau de associação entre os elementos e os conjuntos *fuzzy*, favorecendo interpretações mais abrangentes das ações registradas em campo.

A variável interceptações manteve desempenho notavelmente elevado, com  $R^2 = 0,9143$ , confirmando a robustez da explicação já observada nas outras versões IT2FS. Esse resultado reforça a capacidade do modelo em capturar atributos defensivos objetivos e com baixa ambiguidade interpretativa, especialmente relevantes na avaliação de atletas com perfil de contenção.

Outro destaque foi a métrica *expected goals* ( $xG$ ), com  $R^2 = 0,8862$ , valor próximo dos obtidos com as demais transformações. A consistência dos resultados em  $xG$  evidencia que a modelagem *fuzzy* aplicada às estatísticas ofensivas dos atletas é assertiva na antecipação de potenciais finalizações com base no comportamento em campo. Em linha semelhante, a variável  $xA$  apresentou  $R^2 = 0,7957$ , também confirmando o desempenho competitivo do modelo para métricas probabilísticas relacionadas à criação de jogadas.

A variável gols foi bem explicada, com  $R^2 = 0,7883$ , resultado levemente inferior ao observado com o IT2FS *lower*. A explicação para assistências, por sua vez, alcançou  $R^2 = 0,5044$ , valor

que manteve o padrão de desempenho observado na versão *lower*, reiterando que esse fenômeno ofensivo depende de variáveis com menor previsibilidade direta ou influências contextuais difíceis de serem representadas por estatísticas agregadas.

Para minutos jogados, o modelo apresentou  $R^2 = 0,8586$ , consolidando outra vez o padrão de alto desempenho explicativo para a variável, presente em todas as transformações. Também se mantiveram satisfatórias as explicações das variáveis relacionadas a comportamento de jogo, como as faltas cometidas ( $R^2 = 0,5043$ ) e faltas sofridas ( $R^2 = 0,6696$ ), ambas com desempenho compatível com o observado nas demais abordagens *fuzzy* do tipo IT2FS.

Em suma, a transformação IT2FS *upper* manteve o bom comportamento das transformações intervalares na explicação de múltiplos indicadores de desempenho, em especial aqueles ligados à dimensão defensiva, tal como a métrica de interceptações. Além disso, manteve bons níveis explicativos para minutos jogados, gols,  $xG$  e  $xA$ . A leve tendência à ampliação dos graus de pertinência, característica dessa abordagem, não comprometeu a acurácia das previsões, sugerindo que o modelo tolera bem um maior espaço de ambiguidade nas associações entre estatísticas e variáveis-alvo, sobretudo quando essas envolvem comportamento coletivo ou construções táticas complexas.

#### **4.14.5. Modelo TFRa para explicação de variáveis de performance**

A transformação TFRa, considerada uma das primeiras alternativas *fuzzy* relativas aplicadas à modelagem de dados com variabilidade e subjetividade contextual, foi submetida ao mesmo conjunto de testes com 39 fatores estatísticos explicativos. Essa transformação, embora anterior à IFR, apresenta características úteis para análises que exigem um mapeamento relativo da distribuição dos dados, com preservação das diferenças entre atletas em suas respectivas métricas.

Os resultados obtidos com a TFRa demonstraram desempenho competitivo e, em alguns casos, superior às demais transformações *fuzzy* aplicadas. O principal destaque foi a variável *expected assists* ( $xA$ ), com  $R^2 = 0,9148$ , a melhor explicação estatística obtida entre todos os modelos testados para essa variável. Esse achado reforça o potencial da TFRa, e das transformações relativas, para capturar nuances associadas à criação de jogadas e ao envolvimento dos atletas na fase ofensiva.

A variável interceptações também apresentou elevado desempenho explicativo ( $R^2 = 0,9166$ ), mantendo o padrão observado nas demais transformações e reafirmando a estabilidade do

modelo em fenômenos defensivos bem definidos. Similarmente, a métrica  $xG$  apresentou  $R^2 = 0,8994$ , confirmando o bom ajuste do modelo à predição de situações de finalização com potencial para resultar em gol.

Em relação à variável minutos jogados, o valor de  $R^2 = 0,8741$  posiciona a TFRa entre as transformações com melhor desempenho nessa métrica, demonstrando boa sensibilidade à frequência de participação dos atletas ao longo das partidas. Já a variável gols atingiu  $R^2 = 0,7407$ , valor consistente com as demais transformações *fuzzy* e suficiente para indicar acurácia razoável na predição de eventos conclusivos. Para assistências, observou-se  $R^2 = 0,5507$ , o melhor desempenho em relação aos demais modelos testados, sugerindo que a TFRa pode capturar mais adequadamente as variáveis envolvidas nesse fenômeno ofensivo.

No que diz respeito às variáveis disciplinares e de contato, faltas cometidas alcançou  $R^2 = 0,5571$  e faltas sofridas atingiu  $R^2 = 0,6677$ , valores compatíveis com os melhores desempenhos observados nessas métricas. Esses resultados indicam que a TFRa é capaz de acomodar variações contextuais que caracterizam essas ocorrências dentro do jogo, muitas vezes influenciadas por estilo de jogo, posição e arbitragem.

De modo geral, a transformação TFRa se apresentou como uma alternativa sólida e eficaz, sobretudo para variáveis que envolvem a antecipação de jogadas ou a criação ofensiva ( $xA$ ,  $xG$ , assistências). Embora ligeiramente superada pela IFR em variáveis como minutos jogados ou interceptações, a TFRa demonstrou notável capacidade de adaptação ao modelo *Random Forest*, contribuindo de maneira relevante para a explicação estatística das oito variáveis de interesse.

#### **4.14.6. Síntese dos resultados do modelo para explicação de variáveis de performance**

A superioridade observada das transformações relativas – especialmente IFR e TFRa – na explicação das variáveis de interesse reforça a capacidade desse tipo de abordagem em capturar nuances contextuais presentes na avaliação de performance esportiva. Entre os modelos testados, a TFRa obteve os melhores resultados para quatro variáveis (assistências,  $xA$ , minutos jogados e faltas), enquanto a IFR apresentou melhor desempenho nas demais (gols,  $xG$ , interceptações e faltas sofridas). Tal padrão revela que, embora compartilhem uma base teórica comum ao evitarem categorizações fixas, as duas abordagens operam com sensibilidades distintas no processo de *fuzzificação*. A compilação dos resultados pode ser observada no quadro 26.

Variável de Interesse	IFR	IT2FS <i>center</i>	IT2FS <i>lower</i>	IT2FS <i>upper</i>	TFRa
<b>Faltas</b>	0,5564	0,5071	0,5094	0,5043	<b>0,5571</b>
<b>Faltas Sofridas</b>	<b>0,6817</b>	0,6695	0,6697	0,6696	0,6677
<b>Minutos Jogados</b>	0,8691	0,8565	0,8580	0,8586	<b>0,8741</b>
<b>Assistências</b>	0,4943	0,2868	0,5166	0,5044	<b>0,5507</b>
<b>Gols</b>	<b>0,7509</b>	0,7895	0,7933	0,7883	0,7407
<i>Expected Goals (xG)</i>	<b>0,9283</b>	0,8911	0,8879	0,8862	0,8994
<i>Expected Assists (xA)</i>	0,8881	0,7989	0,8000	0,7957	<b>0,9148</b>
<b>Intercepções</b>	<b>0,9312</b>	0,9136	0,9135	0,9143	0,9166

Quadro 26 - Compilação de resultados para variável de performance

Uma hipótese para a eficácia da IFR na explicação de variáveis como gols e *expected goals* (*xG*) está relacionada ao uso da curva de Lorenz como mecanismo de suavização na transformação relativa. Essa suavização não elimina a sensibilidade às posições relativas entre os jogadores, mas redistribui os valores de pertinência de forma mais progressiva e contínua. Essa propriedade da IFR pode favorecer a modelagem de variáveis de natureza mais concentrada e com maior assimetria estatística, como os gols e o *xG*, em que poucos jogadores acumulam a maior parte dos eventos. Ao aplicar essa suavização, a IFR conserva a diferenciação entre desempenhos altos e baixos, sem intensificar artificialmente os extremos, resultando em maior estabilidade explicativa.

Por outro lado, a TFRa, por não incorporar essa suavização adicional, tende a produzir funções de pertinência mais abruptas entre posições próximas no ranking relativo. Essa característica pode explicar seu melhor desempenho na modelagem de variáveis como assistências e *xA*, nas quais pequenas variações contextuais e interpretativas (como a qualidade do passe final ou o posicionamento do receptor) podem ser bem captadas por transformações com maior sensibilidade à posição ordinal pura. Da mesma forma, seu desempenho superior para minutos jogados e faltas pode refletir o fato de serem variáveis com maior dispersão relativa, nas quais a granularidade preservada pela TFRa se alinha melhor à variabilidade dos dados.

Essas observações permitem levantar uma hipótese adicional: a IFR se mostra mais eficaz na modelagem de variáveis em que a incerteza estatística (epistêmica) é predominante, como nos

desfechos objetivos (gols,  $xG$ ), enquanto a TFRa favorece a explicação de variáveis mais sujeitas à interpretação ou à ação tática (ontológica), como assistências e  $xA$ . Em ambos os casos, os resultados reforçam que a escolha da técnica de *fuzzificação* deve considerar não apenas a estrutura estatística da variável, mas também o tipo de incerteza subjacente ao fenômeno modelado.

Em síntese, a comparação entre as transformações revela que a suavização promovida pela IFR por meio da curva de Lorenz ampliou sua capacidade de capturar variações relevantes em eventos concentrados, enquanto a TFRa manteve vantagem em eventos distribuídos ou com alta dependência do contexto interpretativo. Esta distinção contribui para futuras aplicações metodológicas em domínios que exigem modelagem de desempenho sob condições complexas, como o futebol de alto rendimento.

Em relação à importância de cada um dos fatores na explicação das variáveis, optou-se por demonstrar apenas para a transformação IFR. O indicador observado foi o incremento no erro médio quadrático, ou %IncMSE, que é o resultado da alteração aleatória do valor de cada variável preditora, caso ela seja mais significativa. Isso aumenta o erro de previsão do modelo. Consequentemente, um valor mais alto denota a importância da variável (SONG *et al.*, 2022).

Para a variável relacionada às faltas sofridas, a importância de alguns fatores condiz com as ações da partida. Podem ser observadas as métricas de duelos ofensivos ganhos (38,9%), duelos ofensivos (16,6%), duelos ganhos (14,5%), dribles e dribles com sucesso (ambos em torno de 12-13%). Isso denota que jogadores mais envolvidos em ações ofensivas e dribles tendem a sofrer mais faltas. As estatísticas por ordem de importância podem ser observadas na figura 19.

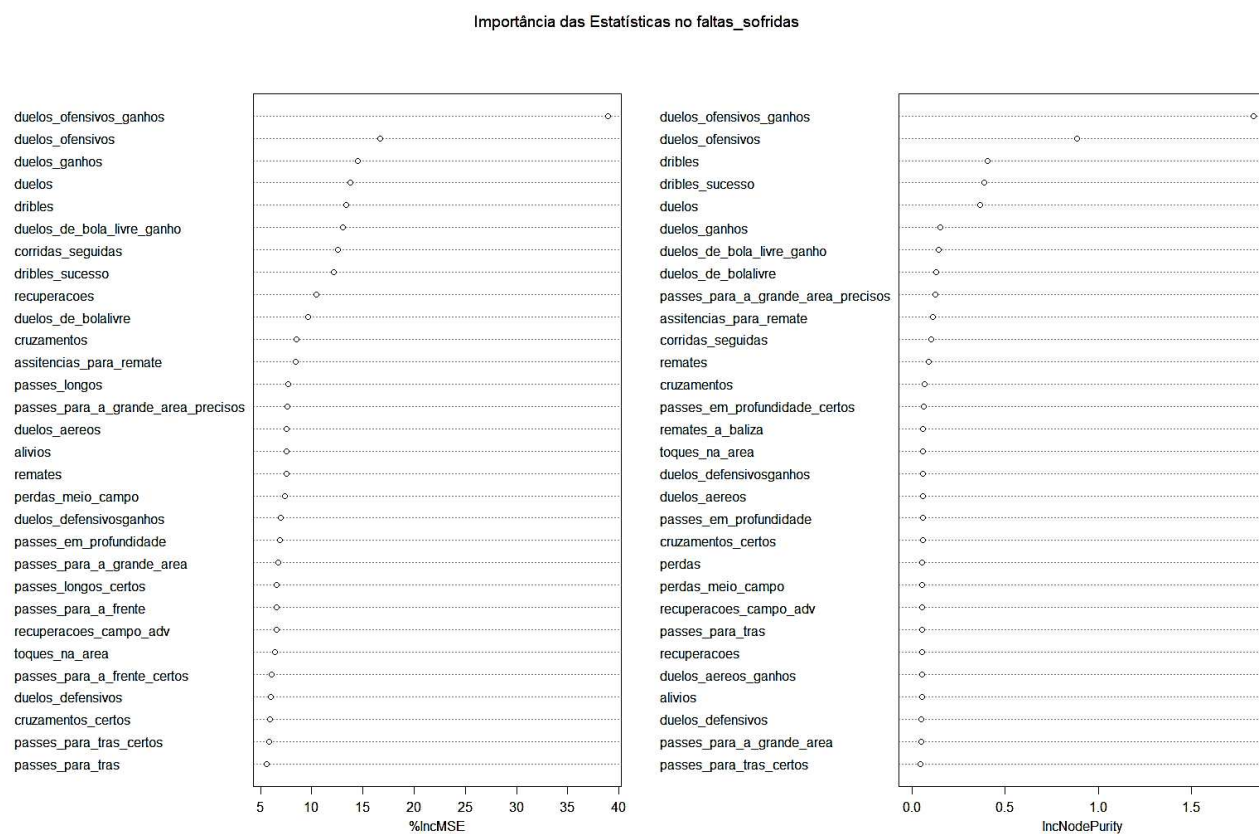


Figura 19 - Importância dos fatores para variável faltas sofridas

Para a variável gols, entre os fatores com métricas mais relevantes estiveram os chutes a gol (remates à baliza) (36,5%), toques na área (20,9%) e remates (15%). Estes são fatores diretamente relacionados ao ato de marcar gols e, portanto, a dominância dessas estatísticas reforça a efetividade do modelo em captar o comportamento ofensivo. A observação da importância das estatísticas por ordem pode ser vista na figura 20.

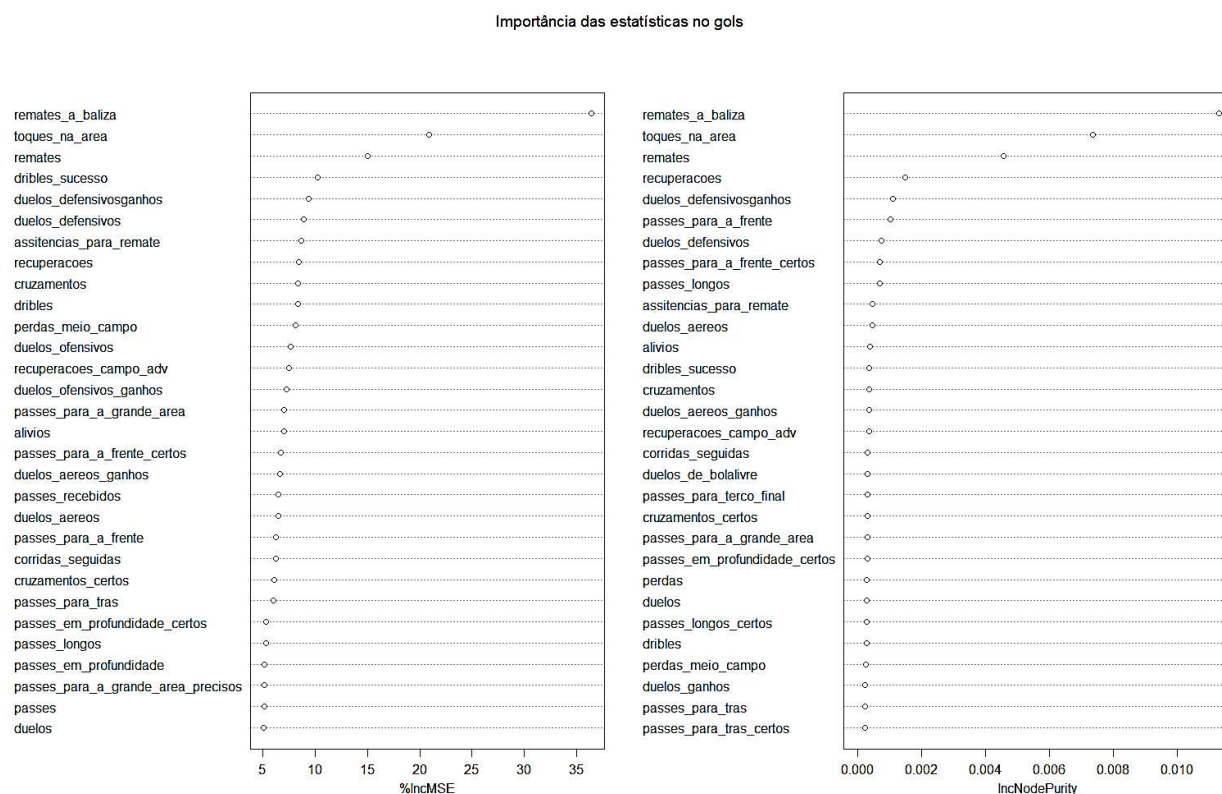


Figura 20 - Importância das estatísticas na variável gol

Para a variável minutos jogados, entre os fatores mais relevantes estiveram as perdas (22,3%), recuperações (21%) e perdas no meio-campo (20%). Esses indicadores sugerem que jogadores que participam ativamente das transições e disputas de posse tendem a permanecer mais tempo em campo, possivelmente por exercerem papel central nas dinâmicas de jogo. Além disso, estatísticas como passes (14,6%) e duelos (14,5%) complementam esse entendimento ao refletirem a intensidade e volume de participação durante as partidas. A figura 21 indica por ordem de importância os fatores relevantes para a variável.

Importância das Estatísticas nos minutos\_jogados

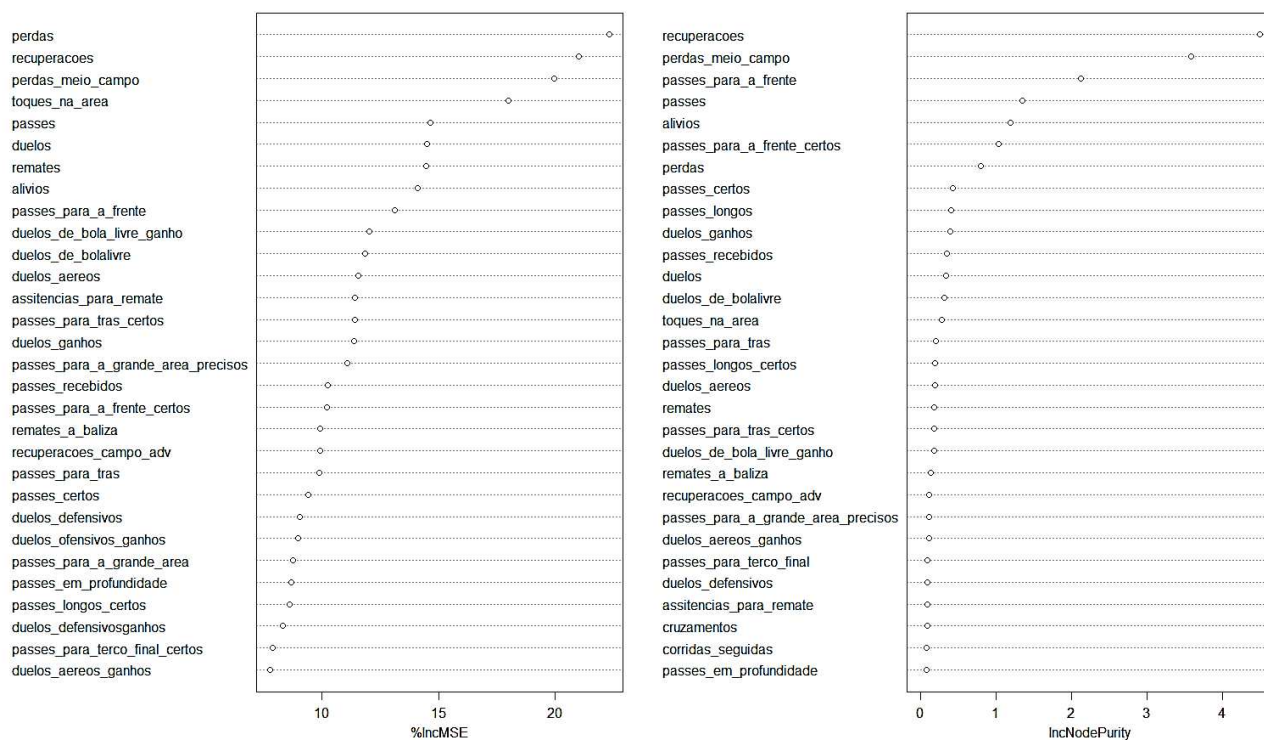


Figura 21 - Importância das estatísticas na variável minutos jogados

No caso das interceptações, dentre os fatores mais expressivos estavam recuperações (27,3%), alívios (21,5%) e perdas no meio-campo (17,1%). Isso indica o perfil de jogadores com forte contribuição defensiva, cuja atuação está associada ao rompimento de jogadas adversárias, antecipações e alívio da pressão. A relação entre perdas no setor central e interceptações também revela que atletas envolvidos em zonas críticas tendem a realizar mais ações de corte e recuperação. A ordem de importância dos fatores para explicação da variável relacionada às interceptações pode ser observada na figura 22.

Importância das estatísticas nas interceptações

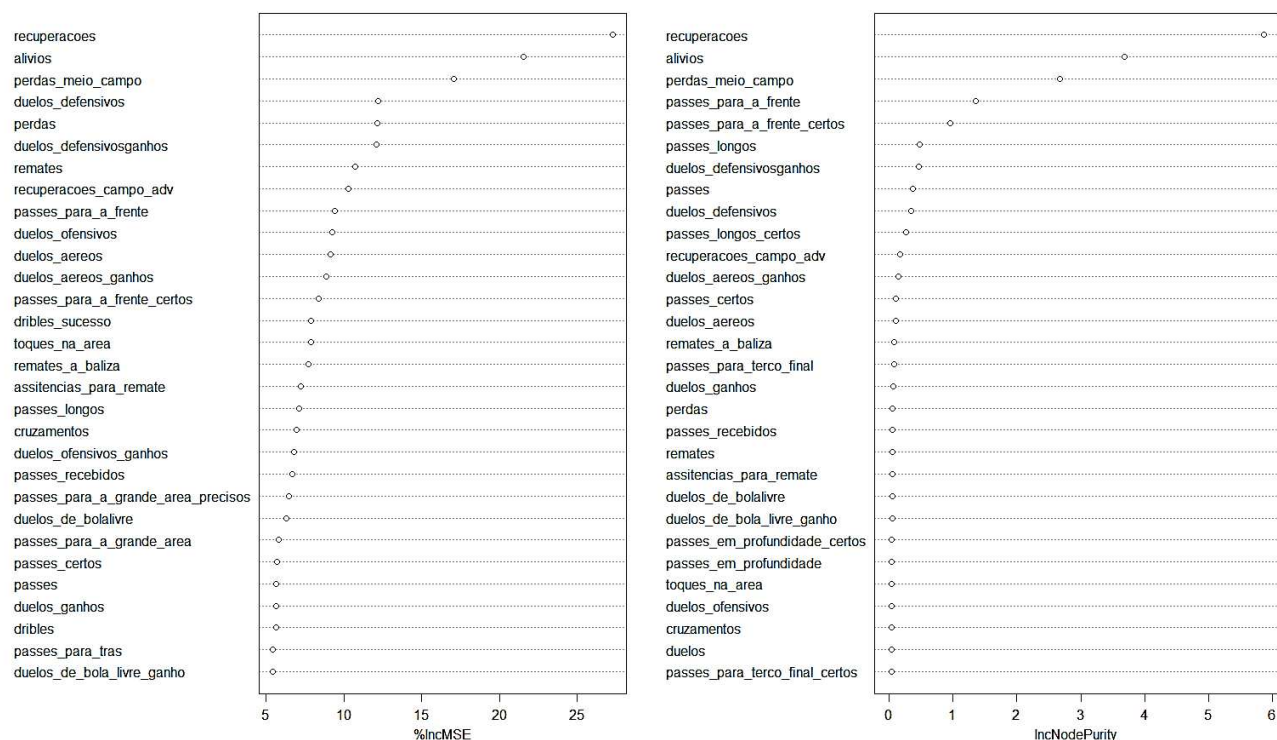


Figura 22 - Importância das estatísticas na variável interceptações

Para a variável *expected assists* ( $x_A$ ), alguns dos fatores mais importantes foram assistências para remate (44%), passes precisos para a grande área (14,7%) e cruzamentos certos (8%). Tais estatísticas são relacionadas à criação de chances de gol, e implicam que a métrica de  $x_A$  está explicada por ações ofensivas qualificadas e orientadas à finalização. O destaque da assistência para remate confirma sua centralidade na composição da expectativa de assistência. Os fatores podem ser observados em ordem de importância na figura 23.

Importância das estatísticas no xA

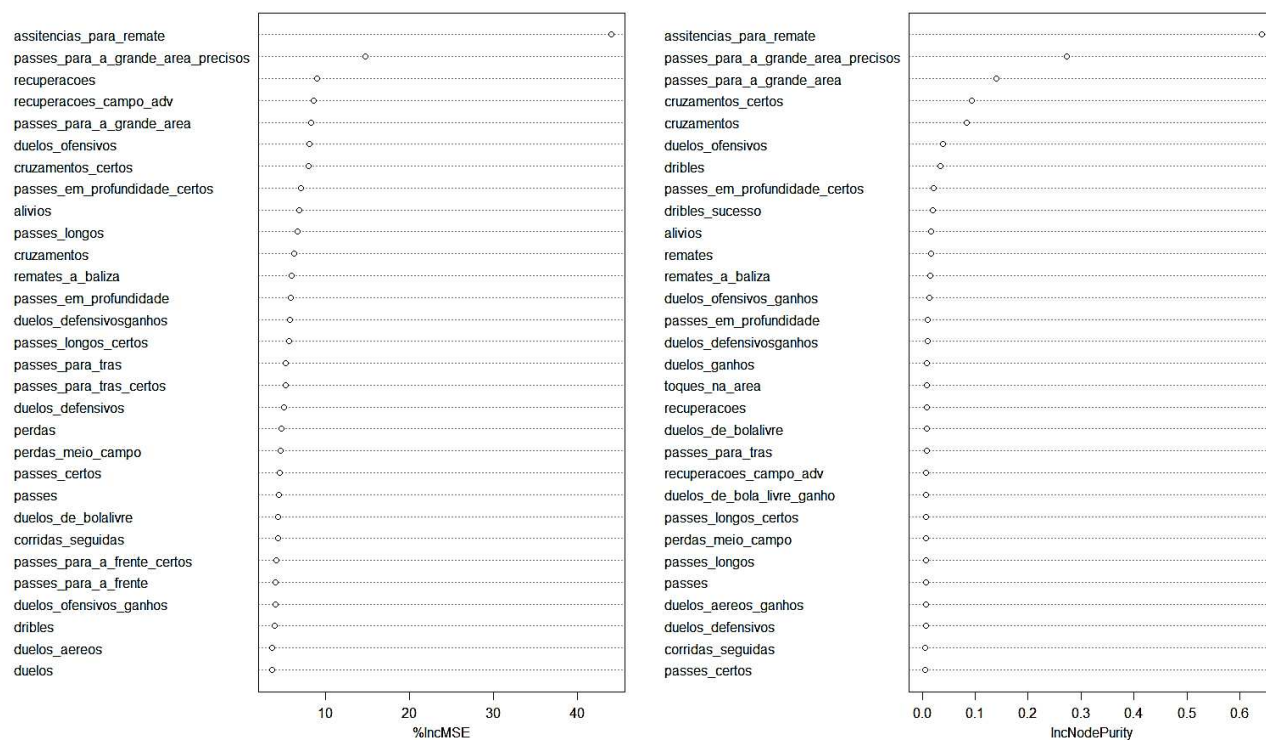


Figura 23 - Importância das estatísticas na variável  $x_A$

Na variável *expected goals* ( $xG$ ), alguns dos fatores com maior impacto foram remates à baliza (29,7%), remates totais (27,1%) e toques na área (25%). Esses três fatores refletem diretamente as ações que precedem tentativas de finalização e reforçam o entendimento de que a métrica  $xG$  é eficientemente modelada por comportamentos ofensivos que aproximam o jogador da meta adversária, tanto em volume quanto em qualidade de ações ofensivas. As estatísticas que explicam a variável  $xG$  podem ser observadas em ordem de importância na figura 24.

Importância das estatísticas no xG

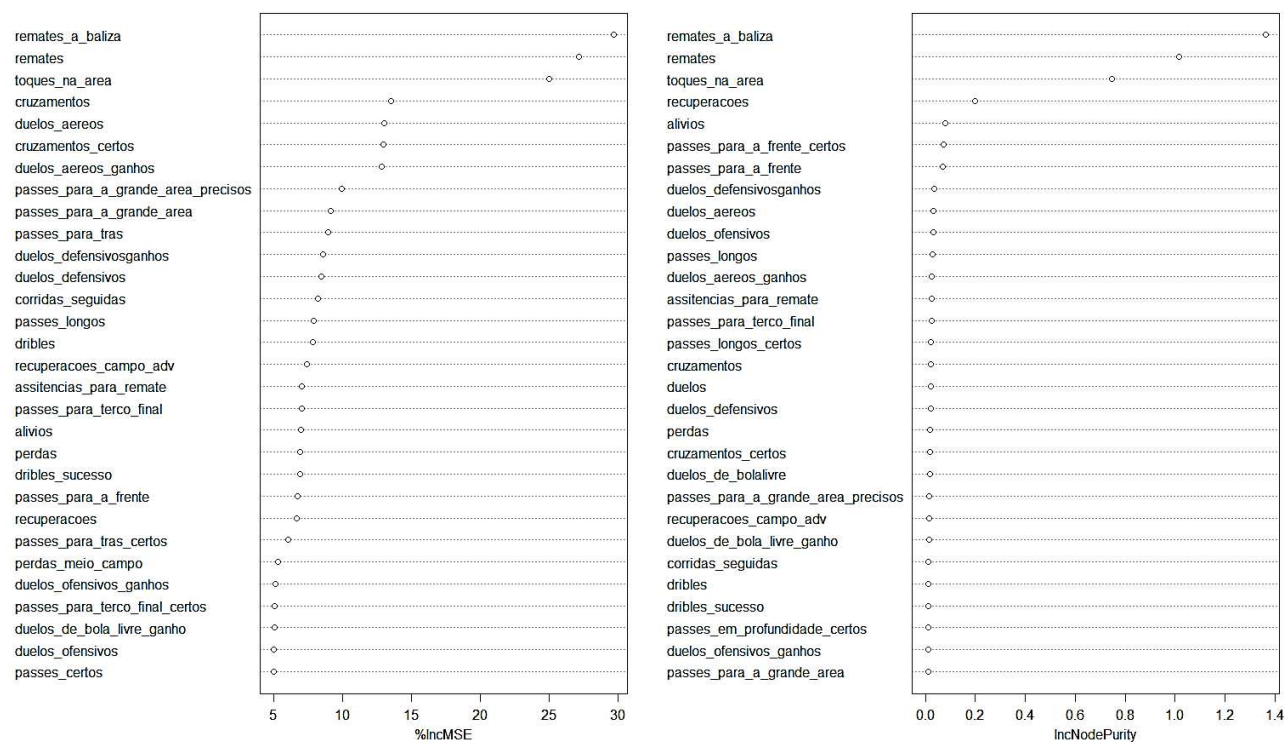


Figura 24 - Importância das estatísticas na variável xG

Com relação às assistências, os fatores mais relevantes foram assistências para remate (26,1%), passes precisos para a grande área (15,3%) e cruzamentos certos (11,6%). Essa combinação de variáveis ressalta a importância da qualidade do passe final e da capacidade de explorar zonas decisivas para gerar oportunidades, corroborando a convergência explicativa com os fatores de  $xA$ . Além disso, variáveis como dribles (11,3%) e passes em profundidade certos (10,4%) reforçam a ideia de ofensividade ativa por parte dos jogadores mais criativos. Os fatores mais importantes para explicação da variável assistências podem ser notados na figura 25.

Importância das estatísticas nas assistências

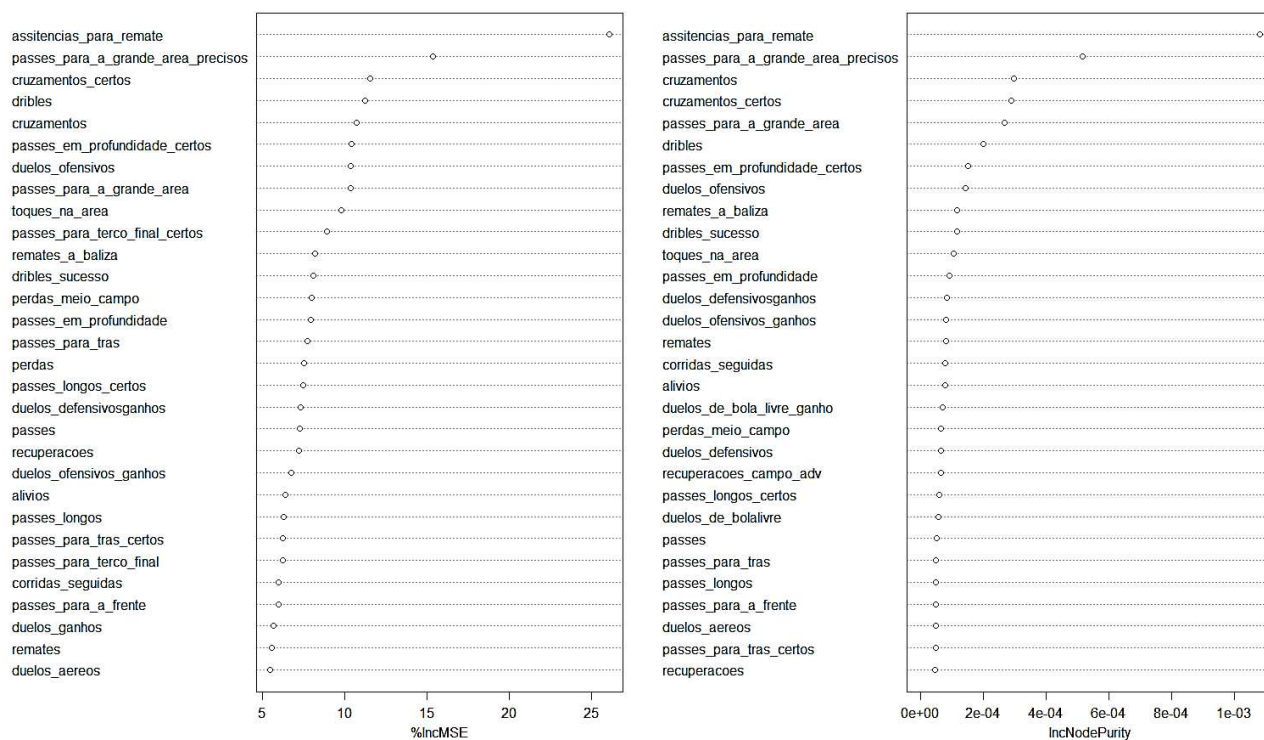


Figura 25 - Importância das estatísticas na variável assistências

Por fim, na variável faltas cometidas, entre os fatores mais impactantes estavam listados os duelos defensivos (32%), duelos de bola livre (25,5%) e duelos totais (20,4%). Esses dados sugerem que os jogadores mais envolvidos em disputas físicas e combativas tendem a cometer mais infrações, o que é coerente com o perfil de atletas com alta intensidade defensiva e constante exposição a confrontos individuais. Os fatores mais importantes para explicação das faltas cometidas podem ser observados na figura 26.

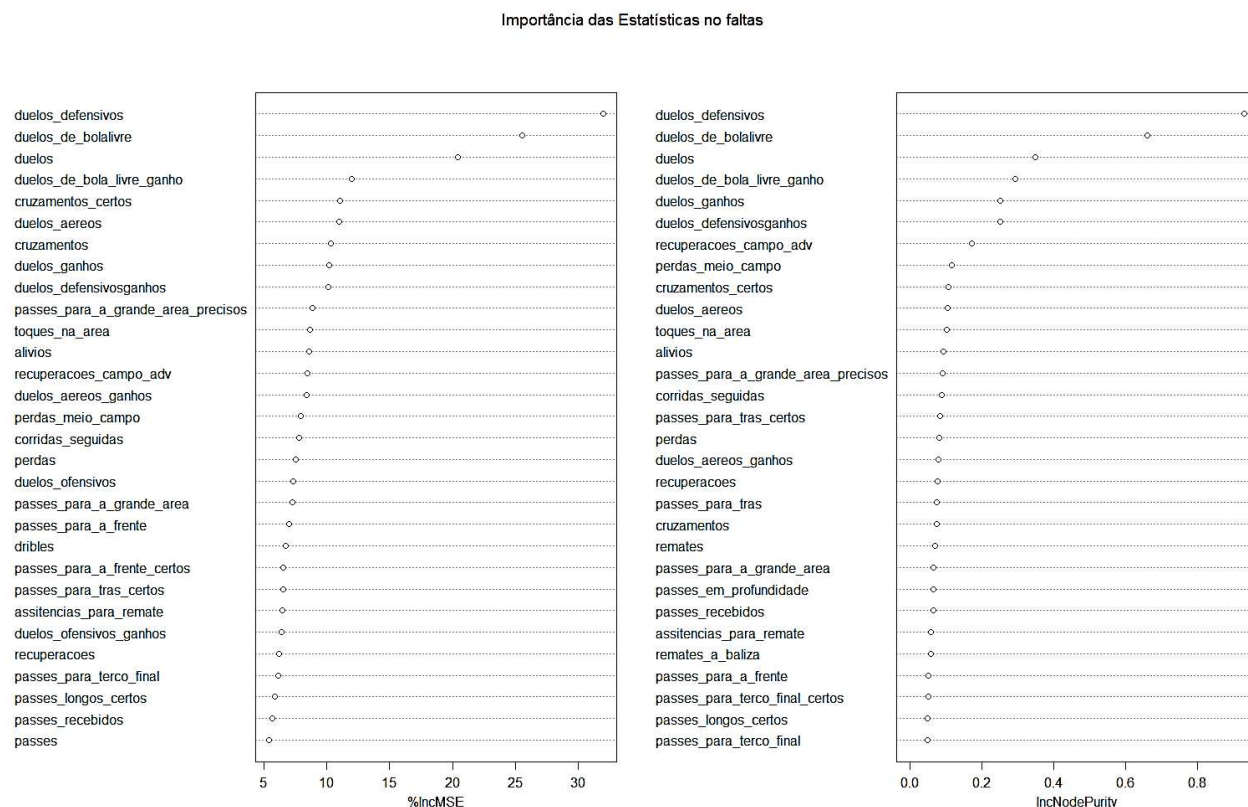


Figura 26 - Importância das estatísticas na variável faltas cometidas

Com base na análise das importâncias das variáveis, observa-se que o modelo *Random Forest*, alimentado pelas transformações *fuzzy* do tipo IFR, foi capaz de identificar relações coerentes entre os fatores estatísticos e as variáveis de interesse, evidenciando sua capacidade de interpretar padrões do desempenho em campo. Em cada variável analisada, os fatores com maior influência apresentaram aderência ao comportamento esperado dentro do jogo – isto é, o protagonismo de estatísticas ofensivas em métricas como gols, assistências e *expected goals*, e o destaque de ações defensivas e de transição em variáveis como interceptações, faltas e minutos jogados.

Essas relações apontam para a robustez explicativa da modelagem adotada e sustenta a relevância do uso da lógica *fuzzy* para representar nuances de desempenho em contextos como o futebol. Como etapa complementar, serão conduzidas análises de agrupamento com base em clusterização hierárquica e no algoritmo *fuzzy c-means*, que foram realizadas ainda na busca pelo melhor ajuste do modelo e aprofundaram a compreensão sobre os perfis de jogadores e suas similaridades a partir das variáveis transformadas.

#### 4.15. Análise de agrupamentos – Hierárquico e *Fuzzy C-Means*

Diante da ausência de progresso significativo na predição das variáveis financeiras, recorreu-se ao uso de algoritmos de aprendizado não supervisionado como instrumento complementar de compreensão da base de dados. Inicialmente, foi aplicada uma análise de agrupamento hierárquico sobre a base original do *Wyscout*, ainda sem transformação *fuzzy*. O método utilizado foi o *Ward.D2*, que minimiza o aumento na soma dos desvios quadrados em torno dos centroides dos grupos formados, utilizando a soma das diferenças quadradas entre os elementos e o centroide como critério de otimização. Tal abordagem tende a gerar agrupamentos mais compactos e com formas esféricas, favorecendo uma representação mais clara das similaridades entre os jogadores (QCBS R WORKSHOP SERIES, 2023).

A partir da indicação do número ideal de *clusters* sugerido por esta primeira etapa, passou-se à execução do algoritmo *fuzzy c-means* sobre a base de dados já transformada com lógica *fuzzy*. Essa aplicação visou compreender os comportamentos estatísticos de jogadores com graus distintos de pertencimento a diferentes agrupamentos, oferecendo uma representação mais flexível das características de performance. A análise foi conduzida com base na versão final do modelo da tese, que utiliza 39 fatores estatísticos para explicar oito variáveis de interesse.

Com o objetivo de aprofundar os resultados obtidos, os algoritmos de agrupamento foram aplicados novamente, agora sobre a base de dados *fuzzificada* e agregada por jogador, mantendo cada uma das colunas representando as estatísticas utilizadas na modelagem. Foi executado, primeiramente, o agrupamento hierárquico com o método *Ward.D2* para a sugestão do número de grupos. Em seguida, realizou-se a análise *fuzzy c-means*, adotando dois cenários: um com o número de agrupamentos sugerido pela análise hierárquica ( $k = 10$ ) e outro com sete agrupamentos ( $k = 7$ ), em alusão às sete posições registradas na base de dados (zagueiro central [CB], atacante de beirada [WF], meio-campista defensivo [DMF], meio-campista ofensivo [AMF], laterais [SB], atacantes [CF] e meio-campistas centrais [CMF]).

O coeficiente de silhueta, métrica oriunda da análise hierárquica que avalia a coesão e separação dos grupos formados, indicou que a solução com 10 *clusters* seria a mais adequada para o modelo. Curiosamente, o modelo com sete agrupamentos, que serviria como base comparativa posicional, foi um dos que apresentou menor pontuação nessa métrica. Essa divergência pode estar relacionada a características funcionais desempenhadas pelos jogadores em campo (como finalizadores, passadores, marcadores ou construtores), mais do que às suas posições táticas formais. As pontuações do escore de silhueta podem ser observados na figura 27.

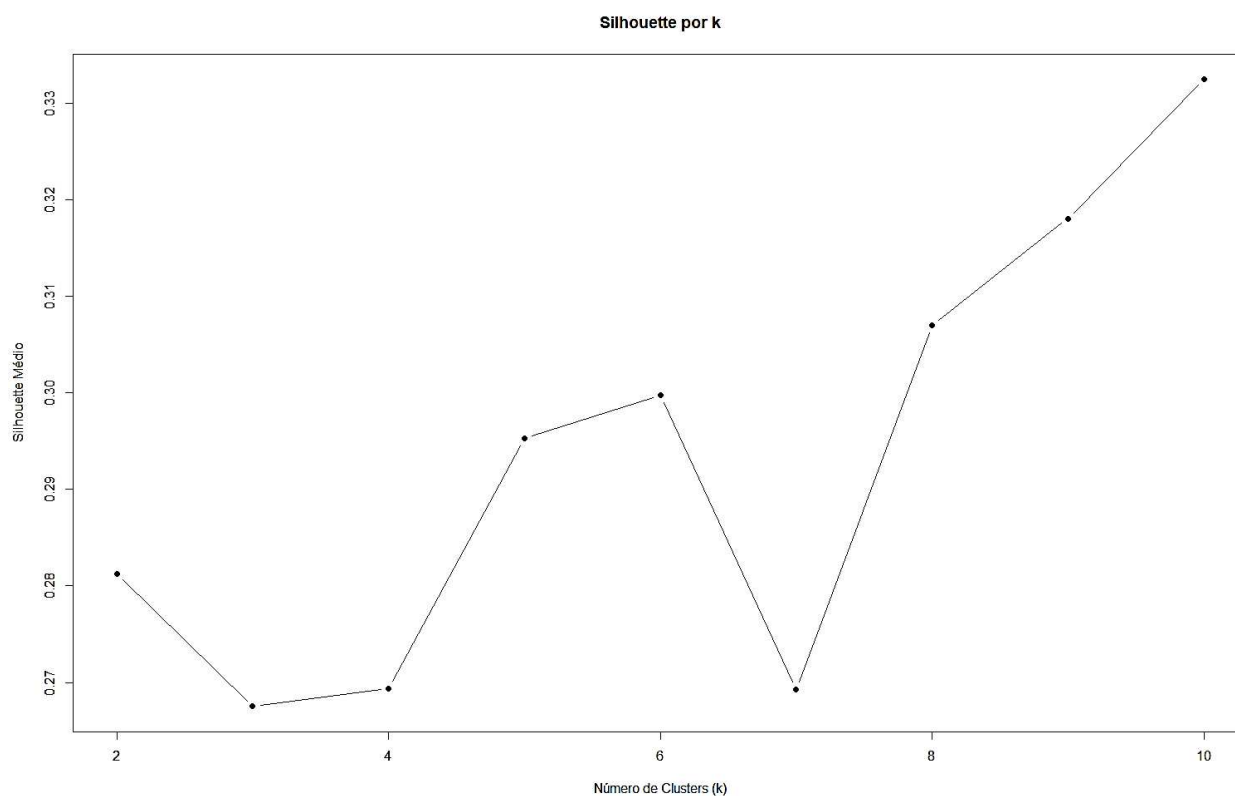


Figura 27 - Pontuação do escore de silhueta

Para a execução do algoritmo *fuzzy c-means*, foram utilizados os números de *clusters* previamente definidos, sendo ( $m = 2$ ) referente ao grau de *fuzzificação*, um controle da suavidade dos limites de pertencimento entre os *clusters*. A seguir, serão apresentados os resultados da análise para a solução com melhor pontuação no escore de silhueta ( $k = 10$ ).

A aplicação da análise de cluster *fuzzy c-means* com dez agrupamentos ( $k = 10$ ) resultou em uma segmentação mais refinada da amostra de jogadores, revelando comportamentos funcionais específicos para além das posições táticas registradas na plataforma de estatísticas. Essa abordagem permitiu interpretar os agrupamentos como representações de papéis desempenhados em campo com base nas combinações de atributos estatísticos dominantes.

Desta maneira, o Cluster 1 mostrou predomínio em passes para o terço final certos (91,6%) e passes para o terço final (64,3%), revelando um grupo de atletas com vocação para o avanço da bola e construção de jogo progressivo. Esses jogadores funcionam como atletas organizadores ou de apoio, com função de distribuir a bola para o momento ofensivo no campo de jogo.

Por sua vez o Cluster 2 apresentou maior membresia em variáveis como duelos (50,9%), perdas (44,8%), duelos de bola livre (31,9%), e passes para trás (24,8%), caracterizando um perfil de atletas de contenção e suporte, com alta participação nas transições e disputas. O Cluster 3 foi

composto por duelos aéreos ganhos (89,7%), duelos aéreos (34%), alívios (33,7%), e duelos de bola livre ganhos (21,3%), o que sugere com clareza o perfil de zagueiros centrais clássicos, exigidos nos combates defensivos e no jogo aéreo.

O Cluster 4 apresentou a maior associação com passes recebidos (79,1%), passes certos (87,4%), passes (92,7%) e passes para trás (73%), perfilando jogadores com alta rotatividade de bola, responsáveis por manter a posse e distribuir passes curtos, podendo ser zagueiros ou volantes recuados que participam da primeira fase de construção do momento ofensivo. O Cluster 5 destacou-se em passes para a frente certos (91,5%), passes para a frente (89,6%), e passes (58,5%), descrevendo atletas com papel de conexão vertical entre setores, como meias que cumprem o papel de ir de uma grande área à outra (*box-to-box*) ou mesmo laterais construtores com bom passe progressivo.

O Cluster 6 revelou concentração em dribles (75,2%), duelos ofensivos ganhos (89%), duelos ofensivos (52,5%), corridas seguidas, e ações ofensivas individuais, caracterizando um grupo com perfil de jogadores de rompimento das linhas defensivas, sejam eles pontas, meias ofensivos ou segundos atacantes. O Cluster 7 apresentou pertinência em remates (40%), assistências para remate (49%), passes para a grande área precisos (69%), toques na área, e cruzamentos, indicando um perfil funcional de jogadores de último terço, com vocação para criação e finalização de jogadas, como meias ofensivos centrais ou atacantes de beirada.

O Cluster 8 exibiu associação com passes em profundidade certos (58%), passes em profundidade (51%), e cruzamentos certos (83%), definindo jogadores com elevada capacidade de criar espaços e alimentar atacantes, provavelmente laterais ofensivos, meias criativos ou extremos com boa visão de jogo. O Cluster 9 se destacou em passes longos certos (86%) e passes longos (28%), reforçando o perfil de zagueiros armadores, laterais de transição longa ou meio-campistas defensivos com boa inversão de jogo. Por fim, o Cluster 10 apresentou o maior valor de pertinência para duelos defensivos ganhos (83,5%), duelos defensivos (69,1%), além de recuperações e perdas no meio-campo, consolidando o perfil de especialistas defensivos de marcação, como volantes de contenção ou zagueiros marcadores. Os agrupamentos podem ser observados no quadro 27.

<b>Estatísticas</b>	<b>Cluster_1</b>	<b>Cluster_2</b>	<b>Cluster_3</b>	<b>Cluster_4</b>	<b>Cluster_5</b>	<b>Cluster_6</b>	<b>Cluster_7</b>	<b>Cluster_8</b>	<b>Cluster_9</b>	<b>Cluster_10</b>
Duelos defensivos ganhos	0,032529 08	0,024442 18	0,017288 87	0,016193 7	0,018780 64	0,009317 69	0,011260 73	0,004995 87	0,030234 73	<b>0,8349564</b> <b>9</b>
Duelos defensivos	0,063049 13	0,055026 77	0,026358 33	0,036678 54	0,037808 28	0,019185 95	0,019184 1	0,008767 56	0,042513 85	<b>0,6914274</b> <b>9</b>
Duelos ganhos	0,087917 75	<b>0,315563</b> <b>67</b>	0,066521 95	0,063322 08	0,060059 21	0,058707 28	0,042213 79	0,017362 36	0,074604 78	0,2137271 3
Perdas meio-campo	0,117511 51	0,102387 64	0,136728 78	0,063119 18	0,077481 86	0,035487 42	0,047749 96	0,021024 2	<b>0,219828</b> <b>7</b>	0,1786807 4
Recuperações no campo adversário	0,125087 66	<b>0,164584</b> <b>68</b>	0,084870 95	0,049191 3	0,046198 17	0,094524 61	0,124981 6	0,036737 67	0,110141 93	0,1636814 4
Recuperações	0,123417 18	0,050809 82	0,059746 13	0,147248 01	<b>0,319382</b> <b>2</b>	0,022678 74	0,026603 93	0,015782 64	0,109560 59	0,1247707 6
Passes para trás certos	0,114161 75	<b>0,242191</b> <b>27</b>	0,062115 11	0,073694 58	0,054446 02	0,137480 92	0,088674 26	0,031707 74	0,077453 91	0,1180744 5
Passes para trás	0,110614 52	<b>0,248631</b> <b>47</b>	0,061199 22	0,073179 35	0,054069 74	0,142301 06	0,087420 32	0,031591 19	0,075300 93	0,1156922 1
Duelos de bola livre	0,076699 73	<b>0,319195</b> <b>3</b>	0,111460 79	0,041341 45	0,039224 07	0,091479 52	0,088797 56	0,027993 84	0,090175 4	0,1136323 4

Perdas	0,076578 32	<b>0,44896</b>	0,061118 31	0,052939 53	0,047899 99	0,075678 75	0,051270 52	0,018966 85	0,066766 52	0,0998212 1
Duelos	0,056938 07	<b>0,509643</b> <b>73</b>	0,049315 07	0,042617 34	0,036942 91	0,098467 04	0,044680 24	0,017733 08	0,047603 57	0,0960589 5
Duelos aéreos	0,075701	0,112688 73	<b>0,340481</b> <b>14</b>	0,057464 02	0,063544 8	0,049163 19	0,061293 75	0,032900 26	0,115987 74	0,0907753 9
Passes para a grande área	0,076036 68	0,140888 55	0,068285 32	0,039430 74	0,037787 68	0,170439 1	<b>0,248731</b> <b>63</b>	0,046022 75	0,083134 24	0,0892433
Passes longos	0,162473 29	0,045181 72	0,062344 83	0,104115 12	0,186616 79	0,022011 38	0,029974 71	0,017023 54	<b>0,284465</b> <b>76</b>	0,0857928 5
Alívios	0,067500 21	0,052075 39	<b>0,337365</b> <b>23</b>	0,040211 17	0,052939 92	0,032563 12	0,073259 84	0,045840 81	0,214597 69	0,0836466 2
Duelos de bola livre ganho	0,065069 97	0,080534 1	0,213566 28	0,032559 66	0,035258 26	0,061277 81	<b>0,220967</b> <b>78</b>	0,071383 19	0,136050 33	0,0833326 3
Cruzamentos	0,049339 05	0,085147 57	0,069763 05	0,028190 89	0,028611 3	0,125611 21	<b>0,416285</b> <b>82</b>	0,064323 05	0,067191 4	0,0655366 7
Toques na área	0,045809 82	0,089907 24	0,104053 03	0,028448 2	0,027953 4	0,132993 18	<b>0,364165</b> <b>77</b>	0,085300 37	0,066234 39	0,0551346 1
Duelos ofensivos	0,043306 78	0,130656 01	0,042038 28	0,031243 91	0,027493 57	<b>0,525553</b> <b>16</b>	0,076592 72	0,029570 5	0,040259 43	0,0532856 4

Remates	0,046743 71	0,080699 58	0,090218 59	0,026749 4	0,026222 53	0,121537 68	<b>0,400241</b> <b>48</b>	0,087859 31	0,066746 02	0,0529817 1
Corridas seguidas	0,039824 05	0,061983 18	0,066310 72	0,022467 61	0,022802 46	0,098101 29	<b>0,517971</b> <b>89</b>	0,061080 13	0,059585 23	0,0498734 4
Passes para terço final	<b>0,643281</b> <b>19</b>	0,025962 17	0,020067 8	0,087691 43	0,088411 68	0,011405 62	0,013145 82	0,006741 26	0,055549 31	0,0477437 2
Dribles sucesso	0,035360 56	0,070225 41	0,054408 18	0,020509 75	0,020270 25	0,213421 38	<b>0,436706</b> <b>83</b>	0,054857 92	0,047714 55	0,0465251 6
Passes em profundidade certos	0,036446 54	0,036050 01	0,062216 22	0,025438 94	0,027524 82	0,043047 08	0,096824 57	<b>0,581445</b> <b>4</b>	0,052325 2	0,0386812 3
Assistências para remate	0,032646 03	0,039768 41	0,059889 37	0,018352 96	0,019064 14	0,054906 38	<b>0,495680</b> <b>97</b>	0,192400 7	0,050674 23	0,0366168 1
Passes em profundidade	0,032386 55	0,032968 11	0,064403 39	0,018629 83	0,019990 13	0,039905 44	0,191120 72	<b>0,512062</b> <b>96</b>	0,054316 99	0,0342158 7
Passes para a grande área precisos	0,023656 21	0,030028 77	0,040280 09	0,012769 67	0,013233 11	0,041196 42	<b>0,698548</b> <b>41</b>	0,076096 8	0,036891 33	0,0272991 9
Dribles	0,019273 6	0,053193 61	0,021894 29	0,012386 02	0,011549 26	<b>0,752319</b> <b>28</b>	0,065871 96	0,016988 1	0,020895 65	0,0256282 3
Passes recebidos	0,049646 22	0,015151 68	0,010058 97	<b>0,791636</b> <b>82</b>	0,074686 86	0,006759 25	0,006812 72	0,003635 62	0,019797 52	0,0218143 4

Passes longos certos	0,030419 12	0,010883 16	0,023172 58	0,012453 68	0,016491 05	0,005882 36	0,010597 2	0,005519 56	<b>0,865455</b> <b>77</b>	0,0191255 1
Passes para terço final certos	<b>0,915858</b> <b>06</b>	0,007786 62	0,005826 94	0,015647 46	0,014278 3	0,003454 37	0,004194 36	0,001979 3	0,016678 06	0,0142965 2
Cruzamentos certos	0,011330 01	0,012151 28	0,022206 45	0,007103 32	0,007684 74	0,015315 14	0,059087 77	<b>0,834372</b> <b>95</b>	0,018003 64	0,0127447 1
Duelos aéreos ganhos	0,010701 85	0,011421 19	<b>0,897838</b> <b>61</b>	0,006544 02	0,007818 01	0,006626 67	0,013259 83	0,007701 46	0,025381 07	0,0127072 9
Duelos ofensivos ganhos	0,009052 68	0,028981 72	0,009494 18	0,005707 64	0,005178 51	<b>0,890726</b> <b>05</b>	0,023411 08	0,006536 33	0,009161 6	0,0117502 3
Passes certos	0,023789 07	0,007230 79	0,005560 74	<b>0,874298</b> <b>41</b>	0,058589 55	0,003341 24	0,003440 96	0,001960 4	0,010623 52	0,0111653 1
Passes para a frente	0,018493 17	0,005466 24	0,005291 56	0,046551 2	<b>0,895991</b> <b>7</b>	0,002603 29	0,002934 76	0,001747 28	0,010968 47	0,0099523 4
Passes para a frente certos	0,015256 26	0,004083 51	0,004022 72	0,039739 19	<b>0,915225</b> <b>67</b>	0,001946 63	0,002224 5	0,001308 86	0,008550 55	0,0076421
Remates a baliza	0,006446 44	0,007032 57	0,013168 51	0,004148 98	0,004430 06	0,008886 57	0,028767 79	<b>0,910068</b> <b>44</b>	0,009996 39	0,0070542 5
Passes	0,012961 84	0,003977 91	0,002918 41	<b>0,927475</b> <b>93</b>	0,036220 9	0,001785 08	0,001821 87	0,001021 5	0,005679 17	0,0061373 8

Quadro 27 - Agrupamento de estatísticas a partir de funções em campo

Essa análise reforça que o uso do algoritmo *fuzzy c-means* sobre a base em 39 fatores estatísticos foi eficaz para captar dinâmicas funcionais complexas no comportamento dos atletas. Ao transcender as classificações posicionais, os agrupamentos revelam funções táticas implícitas e auxiliam na compreensão das contribuições específicas dos jogadores para o coletivo. Na próxima seção serão discutidas as utilidades a partir da combinação entre os resultados desses agrupamentos e os resultados dos modelos explicativos, que poderá oferecer uma avaliação sobre a performance dos atletas, prestando suporte na seleção e otimização de elencos.

#### **4.16. Possíveis aplicações do modelo**

Na intenção de demonstrar uma aplicação do modelo proposto, utilizou-se um algoritmo no RStudio a partir do pacote *ggplot2* para gerar uma série histórica comparando a pontuação de dois atletas distintos em um conjunto de estatísticas selecionadas. Optou-se por destacar os dois meio-campistas ofensivos vencedores da Bola de Prata – prêmio anual organizado pela revista Placar e pela rede de TV ESPN Brasil. Em 2023, os vencedores foram o uruguaio Giorgian De Arrascaeta, atleta do Clube de Regatas do Flamengo, e o brasileiro Raphael Veiga, da Sociedade Esportiva Palmeiras.

Entre as estatísticas selecionadas para a comparação, foram escolhidas seis métricas com pontuação de membresia relevante no Cluster 6 da análise de agrupamentos com dez *clusters* ( $k = 10$ ). Essa escolha foi motivada pela associação deste grupo à função de meio-campista ofensivo, perfil de jogador tradicionalmente criativo, com bons recursos de passe e drible, características históricas valorizadas no futebol brasileiro.

As comparações evidenciaram que o jogador uruguaio parece apresentar maior consistência ao longo das temporadas, com suas habilidades mais adequadas à função analisada. As figuras 28 e 29 ilustram essa vantagem relativa.

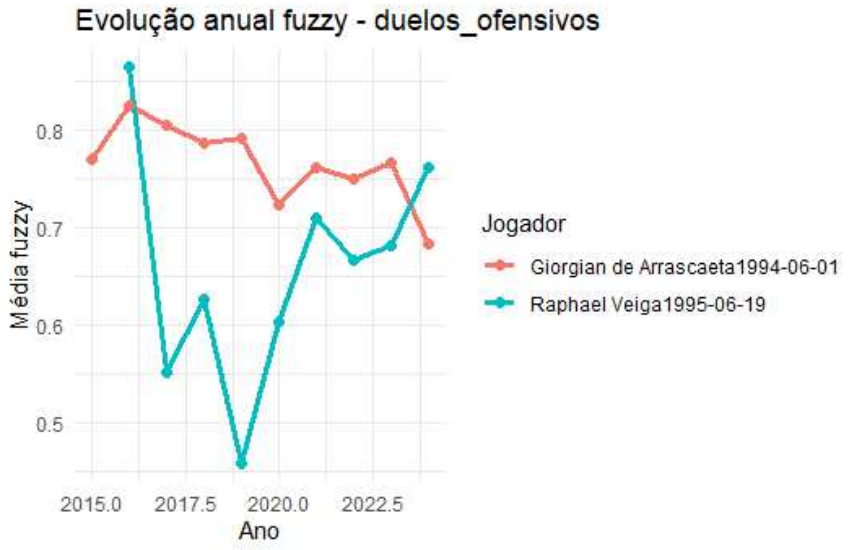
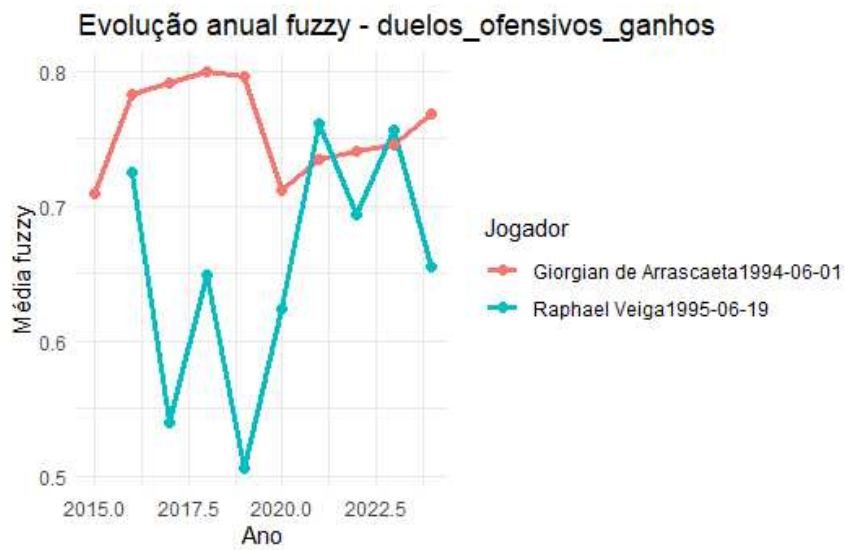
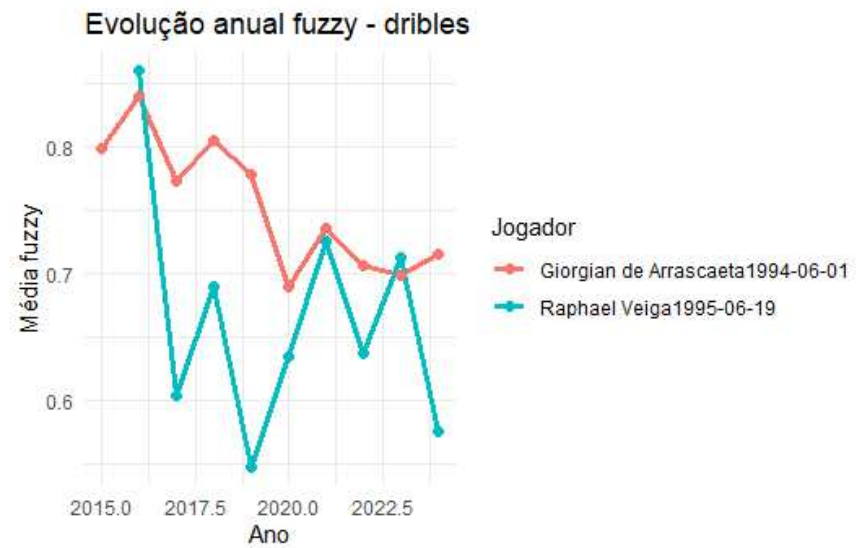
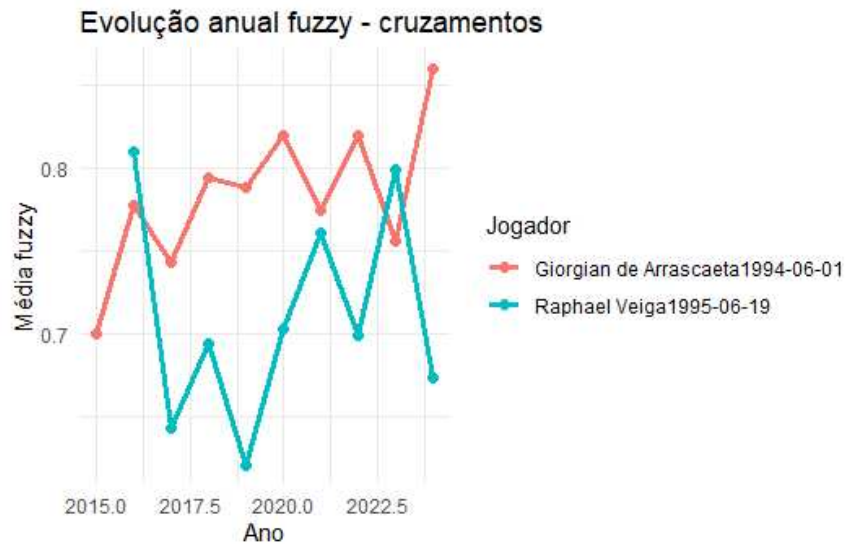


Figura 28 - Comparação entre Arrascaeta e Raphael Veiga

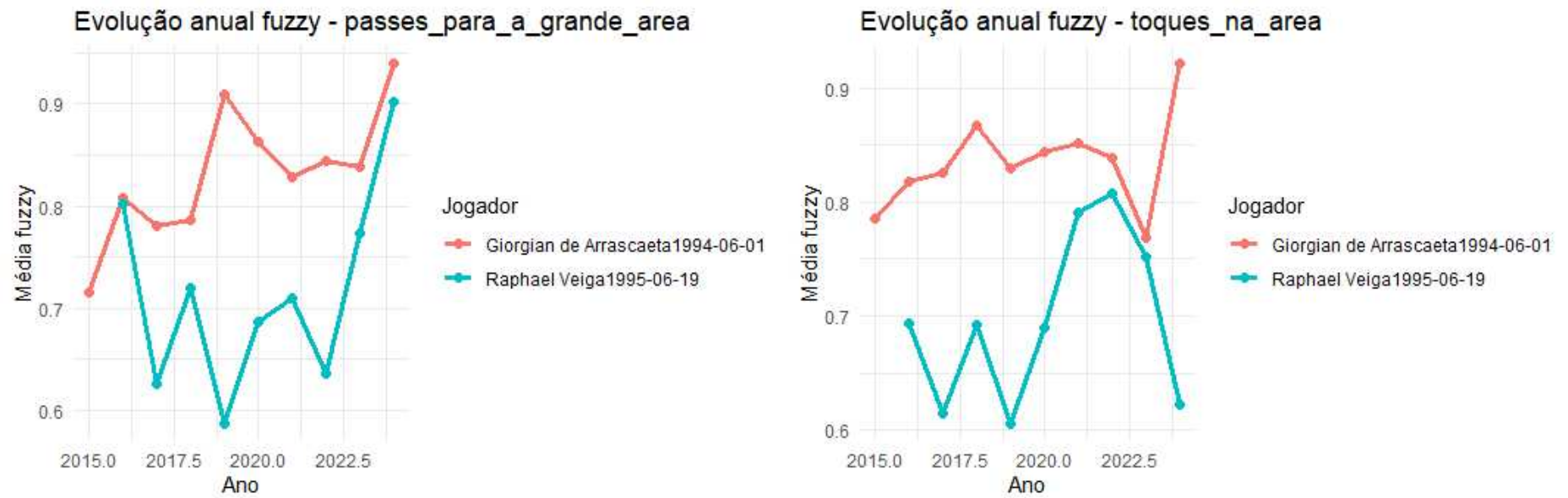


Figura 29 - Comparação entre Arrascaeta e Raphael Veiga

Em outro exemplo, sem comparação entre jogadores diferentes, utilizaram-se os gráficos da série histórica do meio-campista Jean Lucas, atleta do Esporte Clube Bahia, e uma das contratações mais valiosas no intervalo entre o final de 2023 e o início da temporada 2024, ao custo de 4,5 milhões de euros. Para Jean Lucas, foram selecionadas as estatísticas mais representativas do Cluster 4, grupo associado a uma função em campo relacionada à manutenção da posse e distribuição de passes curtos. A análise gráfica revelou um momento de crescimento em sua performance, sinalizando uma possível consolidação da boa forma do atleta em sua nova equipe. A figura 30 reúne os gráficos de função para Jean Lucas.

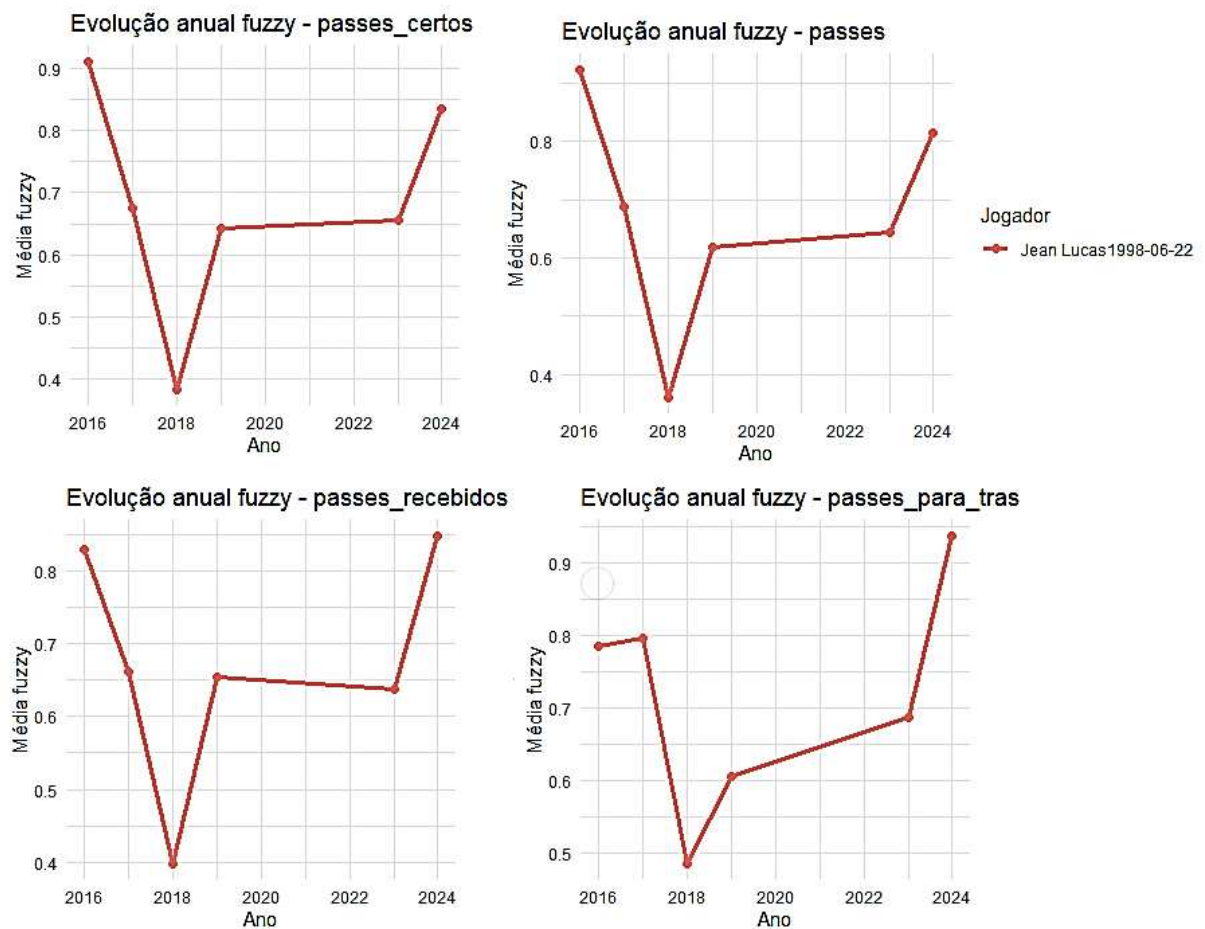


Figura 30 - Evolução dos indicadores para as estatísticas de Jean Lucas

O modelo apresentado demonstrou sua capacidade de oferecer interpretações funcionais sobre os dados, combinando métodos de *fuzzificação*, agrupamento e visualização histórica de desempenho de maneira robusta e replicável. A identificação de perfis por função, e não apenas por posição, sugere potencial para análises preditivas mais refinadas, com aplicações práticas em *scouting*, planejamento de contratações e monitoramento longitudinal de performance. A

seguir, serão discutidas as implicações analíticas e teóricas destes resultados à luz da literatura sobre desempenho, profissionalização e gestão estratégica no futebol.

#### 4.17. Discussão

Durante as análises desenvolvidas neste capítulo, buscou-se a construção e a aplicação de um instrumento de calibração capaz de comparar o desempenho técnico de atletas de futebol a partir de uma base estatística detalhada. Tal construção metodológica encontra respaldo na literatura da ciência do esporte, especialmente naqueles trabalhos que destacam a relevância da mensuração dos gestos técnicos como expressão do desempenho em campo (HUGHES *et al.*, 2012; MEYLAN *et al.*, 2010; PAOLI, 2007; SARMENTO *et al.*, 2018; ZHOU *et al.*, 2018). O modelo desenvolvido respeita essa perspectiva quantitativa e a expande ao aplicar transformações *fuzzy* capazes de suavizar a rigidez dos cortes lineares e permitir interpretações mais próximas da realidade fluida do jogo.

A orientação metodológica levou à aplicação de transformações *fuzzy* – tanto intervalares (IT2FS) quanto relativas (TFRa e IFR) – que permitiram comparar as métricas em diferentes níveis de abstração e granularidade. A análise revelou que as transformações relativas apresentaram melhor desempenho explicativo para a maioria das variáveis analisadas, com destaque para a IFR, cujos ajustes por curva de Lorenz permitiram incorporar suavizações que refletem as variações contextuais dos dados. Este é o ponto que se conecta diretamente aos fundamentos da abordagem ecológica dinâmica: o desempenho não é apenas um dado isolado, mas emerge das relações dinâmicas entre atleta e ambiente, onde contexto e momento histórico interferem na expressão da habilidade.

Inicialmente, a proposta de agrupamento dos dados se baseava em um estudo de Wakelam *et al.*, (2022) a partir de uma análise de componentes principais (PCA) como tentativa de reduzir a alta dimensionalidade dos dados extraídos da plataforma *Wyscout*. No entanto, de imediato, verificou-se que essa redução comprometeria as nuances informacionais presentes nas estatísticas brutas, resultando na perda de importantes diferenciações contextuais entre os atletas. Testaram-se regressões penalizadas para obter coeficientes de ponderação, na intenção de reduzir a base de estatísticas orientadas pelos atributos do jogo *Football Manager 2024*, buscando atribuir pesos distintos aos diferentes intervalos de tempo. A ideia inicial foi avaliar a influência da performance histórica do atleta frente à performance da temporada anterior.

Apesar de promissora, essa abordagem não foi suficientemente eficaz na previsão de variáveis financeiras, revelando limitações na capacidade dos modelos lineares penalizados em capturar a complexidade interacional e não linear da performance esportiva. Em resumo, os modelos ponderados pelas regressões obtiveram alguns resultados em relação às variáveis financeiras, entretanto a interação entre as duas propostas – a de utilizar dados *fuzzificados* ponderados por coeficientes lineares – não obteve resultado positivo.

A escolha de utilizar os dados das competições nacionais também se mostrou coerente com o princípio das circunstâncias semelhantes (FERNANDEZ-NAVARRO *et al.*, 2018; LIU *et al.*, 2015), ao reduzir a interferência de variáveis externas, como calendários e estilos de jogo distintos entre ligas. Ainda assim, a aplicação longitudinal em séries históricas – tanto para comparações entre atletas quanto para análise da evolução de um mesmo jogador ao longo do tempo – demonstrou a robustez do modelo em captar tendências contextuais. A pontuação atribuída pela membresia *fuzzy* variou de acordo com as alterações temporais de desempenho, sinalizando a adequação do instrumento para a leitura dinâmica da performance.

Os algoritmos de agrupamento hierárquico e *fuzzy c-means* corroboraram essa visão integrada da performance. Ao invés de delimitar os jogadores apenas por posições pré-definidas, os *clusters* sugeridos emergiram a partir de padrões funcionais, refletindo comportamentos específicos no campo de jogo (como finalizadores, armadores, construtores ou marcadores). Essa diferenciação é essencial, pois está mais alinhada com os conceitos de *affordances* e de auto-organização sob restrições, que sustentam que o comportamento é moldado de forma contínua pela interação entre as características individuais e as demandas situacionais (GIBSON, 2014; ARAÚJO *et al.*, 2019; WOODS; DAVIDS, 2023).

Embora o modelo tenha priorizado os gestos técnicos como principal dimensão de análise, ele não se opõe à complexidade multidimensional do desempenho esportivo. Pelo contrário, ao permitir a padronização e comparação entre diferentes tipos de estatísticas, cria-se uma base interpretativa passível de extensão para as dimensões tática e física. Isso reforça o princípio do *design* representativo, segundo o qual a avaliação de desempenho deve manter fidelidade ao ambiente de atuação do atleta (ARAÚJO *et al.*, 2019). A partir da calibração *fuzzy*, torna-se possível capturar com mais sensibilidade as microvariações que emergem do jogo, permitindo a incorporação de outras camadas de complexidade sem perder a coerência interpretativa. A abordagem proposta neste estudo aponta para uma reconfiguração metodológica na análise da performance esportiva, de uma lógica prescritiva e classificatória para uma orientação mais

flexível, contextual e centrada no indivíduo em relação ao seu ambiente. A seguir, serão discutidas as implicações teóricas, práticas e epistemológicas desta proposta, suas limitações e possibilidades de aplicação em outros contextos e categorias esportivas.

#### 4.17.1. Implicações Teóricas

A proposta metodológica apresentada nesta tese carrega determinadas implicações teóricas, especialmente quando situada no contexto da abordagem ecológica da performance esportiva e nas discussões contemporâneas sobre modelagem centrada na pessoa. Em primeiro plano, ao utilizar transformações *fuzzy* para calibrar estatísticas técnicas de jogadores de futebol, o estudo desloca o foco da análise de desempenho de uma lógica exclusivamente estatística e descritiva para uma abordagem relacional e dinâmica. Esse deslocamento é coerente com a perspectiva do realismo ecológico, que entende o comportamento não como produto de uma estrutura interna isolada, mas como um fenômeno emergente da interação contínua entre o atleta e seu ambiente (GIBSON, 2014; INGOLD, 2021).

Nesse sentido, o modelo *fuzzy* proposto não se limita a quantificar ações técnicas, mas oferece uma lente de análise que respeita as restrições situacionais, temporais e relacionais que moldam a performance. Isso reforça o entendimento de que o desempenho atlético não pode ser reduzido à repetição de padrões estáticos, mas sim interpretado como manifestação de um sistema auto-organizado sob restrições (ARAÚJO *et al.*, 2019; SARMENTO; ARAÚJO, 2021).

A tese contribui para o avanço do pensamento centrado na pessoa ao desenvolver um instrumento que reconhece a heterogeneidade entre atletas. A análise de agrupamentos demonstrou que os jogadores não necessariamente se organizam em perfis técnicos idênticos às suas posições formais, e sim por meio de comportamentos funcionais expressos em campo, um achado alinhado às abordagens idiográficas, que enfatizam a importância da trajetória individual e das variações entre sujeitos (HOFMANS; WILLE; SCHREURS, 2020; HOWARD; HOFFMAN, 2018). A lógica *fuzzy*, ao permitir graus de pertencimento a diferentes *clusters* e ao evitar classificações rígidas, reforça a natureza prototípica da performance e a ideia de que o jogador é sempre mais do que a soma de suas estatísticas.

Por outro lado, o instrumento de calibração também contribui para o debate epistemológico sobre mensuração na ciência do esporte. Frequentemente tensionada entre a necessidade de precisão e a complexidade dos contextos analisados, essa área encontra na proposta *fuzzy* uma alternativa que suaviza as dicotomias tradicionais (bom/ruim, eficiente/ineficiente,

certo/errado) e valoriza as zonas de transição que caracterizam o jogo de futebol em alto nível. Tal perspectiva se mostra sugestiva quando relacionada à noção de *affordances*, entendida como a abertura para possibilidades de ação que emergem das relações entre as capacidades do organismo e os recursos oferecidos pelo ambiente (CHEMERO, 2006; WOODS; DAVIDS, 2023).

Por fim, ao permitir comparações intertemporais, entre indivíduos e entre agrupamentos funcionais, o modelo apresentado contribui para a construção de sistemas mais representativos de monitoramento e avaliação da performance. Essa capacidade dialoga com o princípio de *design* representativo, propondo que o ambiente de avaliação reflita as propriedades do ambiente de atuação real (ARAÚJO *et al.*, 2019). Assim, os dados deixam de ser meramente registros históricos para se transformarem em guias interpretativos capazes de apoiar decisões de treinamento, observação técnica e recrutamento, respeitando o caráter ecológico, dinâmico e humano do jogo.

#### **4.17.2. Limitações e sugestões para pesquisas futuras**

Apesar das contribuições teóricas e metodológicas apresentadas neste trabalho, é necessário reconhecer as limitações e apontar caminhos para futuras investigações.

A primeira limitação refere-se à restrição da base de dados às competições nacionais. Ainda que essa escolha tenha se mostrado coerente com os pressupostos da dinâmica ecológica, que enfatizam a necessidade de contextualização do desempenho em ambientes específicos (FERNANDEZ-NAVARRO *et al.*, 2018; SARMENTO; ARAÚJO, 2021), ela compromete a generalização imediata dos achados para outras ligas, contextos competitivos ou culturas táticas distintas. Estudos futuros poderiam expandir a base para incluir dados intertemporais e interligas, permitindo a verificação da estabilidade e adaptabilidade do modelo *fuzzy* em diferentes ecossistemas de jogo.

Além disso, a ausência de variáveis táticas, físicas e psicossociais limita a multidimensionalidade proposta no próprio arcabouço teórico da tese. A decisão metodológica de utilizar exclusivamente estatísticas técnicas – extraídas da plataforma *Wyscout* – se deu por motivos de acessibilidade atrelados ao custo (em euros) da assinatura da plataforma, e compatibilidade com os métodos de modelagem *fuzzy*. Entretanto, isso não esgota a complexidade do desempenho atlético. Pesquisas futuras podem incorporar dados de GPS (para mapear cargas físicas), análises posicionais (para captar padrões táticos) ou até indicadores

psicométricos (como resiliência e motivação), ampliando o escopo explicativo do modelo e tornando-o ainda mais representativo da interação organismo-ambiente. Em alguns destes campos de pesquisa a lógica *fuzzy* já é aplicada na transformação dos dados.

Outra limitação observada foi a exclusão dos goleiros do modelo final, devido à estrutura de variáveis técnicas distintas aplicadas a essa posição, as quais não se alinharam com os fatores utilizados para a *fuzzificação* dos demais atletas. Futuras pesquisas podem desenvolver um módulo específico para goleiros, respeitando suas peculiaridades técnicas, táticas e cognitivas, como já sugerido por estudos que tratam do papel determinante dessa função na partida (DI SALVO *et al.*, 2008; LIU; GÓMEZ; LAGO-PEÑAS, 2015).

Do ponto de vista técnico, a sensibilidade dos modelos *fuzzy* a pequenas alterações na base de dados ou no número de agrupamentos escolhidos (como demonstrado na análise de *clusters*) indica a necessidade de mecanismos de validação cruzada mais robustos em estudos futuros. Investigações adicionais podem explorar outras técnicas de aprendizado não supervisionado, ou combinar o modelo *fuzzy* com redes neurais, sistemas adaptativos ou técnicas de *explainable AI*, avançando no campo da inteligência artificial aplicada ao esporte.

Por fim, sugere-se o uso exploratório do modelo em contextos formativos, como categorias de base ou processos de *scouting*. Avaliar a sensibilidade do modelo à identificação de trajetórias promissoras e comparar suas sugestões com as percepções de observadores técnicos pode oferecer contribuições valiosas para a prática profissional. Além disso, o modelo pode ser adaptado a esportes coletivos com dinâmicas semelhantes ao futebol, como o futsal, o rúgbi ou o handebol, o que representaria uma relevante ampliação do campo de aplicação.

Em síntese, embora esta pesquisa tenha avançado na construção de um instrumento representativo, flexível e coerente com as bases epistemológicas contemporâneas da ciência do esporte, ela também abre um leque de possibilidades para aprofundamentos, ajustes e inovações metodológicas, com vistas à complexa e multifacetada tarefa de compreender o desempenho humano em contextos de alta performance.

## 5. Conclusão

A conclusão desta tese não é, apenas, o trabalho de finalização e apontamentos de uma série de ensaios e análises realizados para este devido fim. Na realidade, este trabalho é tratado pelo seu autor como um *checkpoint* para as pesquisas desenvolvidas ao longo de 48 meses de trabalho no Centro de Pós-Graduação e Pesquisas em Administração da Universidade Federal de Minas Gerais. No entanto, ele ainda segue o propósito de oferecer uma contribuição original e aplicável à gestão estratégica do futebol profissional.

A proposta inicial era tratar um caso específico de fracasso organizacional, embora o ponto de inflexão surgido se deva ao fato de que o fenômeno que seria observado estava em pleno andamento. Portanto, era imprevisível lidar com um evento que estava em curso e suas eventuais dificuldades em coleta, tratamento, organização e análise de dados. Destarte, a proposta de contribuição à gestão esportiva tratou de migrar para uma ferramenta de suporte à tomada de decisão que lidasse com o desafio que envolve o processo de recrutamento e seleção de atletas em um ambiente de incertezas, pressão por resultados e hipercompetitivo.

A ferramenta já havia sido sugerida pelo autor desta tese em trabalhos anteriores, a saber, a monografia para o curso de especialização em futebol da Universidade Federal de Viçosa, que foi aplicada em outros dois trabalhos apresentados ao Congresso Brasileiro de Gestão do Esporte. As sugestões oriundas destas apresentações guiaram a evolução do modelo comparativo de performance de atletas de futebol. Isto é, este projeto apresentado como tese foi iniciado em 2018 e, embora os resultados das análises apresentadas nesta tese fortaleçam os principais argumentos da ferramenta, ainda há campo para evolução do modelo e sua eventual aplicação.

Ainda que o objeto de pesquisa se movesse da avaliação de uma estrutura organizacional do futebol no Brasil para o desenvolvimento de uma ferramenta de suporte à tomada de decisão, os trabalhos de pesquisa e orientação em paralelo à construção da tese ainda conseguiram abarcar o contexto da evolução institucional do futebol no Brasil. Algumas destas pesquisas se debruçaram nos estudos de casos derivados a partir da sanção da Lei 14.193/2021, também chamada de Lei da Sociedade Anônima do Futebol (SAF), no qual os clubes passaram a deter a opção de uma regulamentação específica para se transformarem em “clubes-empresa”.

A investigação central deste trabalho foi orientada pela premissa de que o futebol contemporâneo demanda intuição e experiência aliado à adoção de ferramentas analíticas que

transformam dados em informações úteis e acionáveis. A partir disso, propôs-se compreender como clubes de futebol podem utilizar algoritmos de aprendizado de máquina, lógica *fuzzy* e análise de redes sociais para estruturar estratégias de recrutamento de atletas. Essa proposta foi desdobrada em objetivos que englobaram as macroestruturas do mercado internacional de transferências, passando pelas dinâmicas organizacionais e alcançando à avaliação técnica individual de jogadores, com a intenção de integrar perspectivas complementares em um sistema de apoio à decisão.

O percurso teórico e empírico delineado ao longo do trabalho permitiu assumir uma abordagem interdisciplinar, sustentada nos campos da estratégia organizacional, dos estudos sobre desempenho esportivo e das metodologias computacionais aplicadas à decisão. A estrutura da tese refletiu esse compromisso metodológico ao articular capítulos independentes, mas interconectados, que avançaram progressivamente do mapeamento bibliométrico às análises estatísticas e proposições aplicadas.

Com esse arranjo, buscou-se respeitar tanto o rigor acadêmico quanto a utilidade prática da pesquisa, promovendo a dialética entre teoria e *práxis*. Essa estrutura permitiu manter a coerência com a alteração da proposta inicial, reconhecendo a necessidade de ir além da descrição de crises organizacionais pontuais para investigar, de maneira estrutural, as lógicas que sustentam as decisões estratégicas no futebol. Assim, serão apresentadas as principais conclusões de cada um dos capítulos empíricos, discutindo os achados à luz dos referenciais adotados e suas implicações teóricas, práticas e futuras.

O Capítulo 2 desta tese teve como propósito identificar o estado da arte sobre a aplicação de modelos estatísticos baseados em aprendizado de máquina, análise de performance esportiva e lógica *fuzzy*. Por meio de uma revisão bibliométrica estruturada, conduzida com o apoio do pacote *Bibliometrix* no ambiente RStudio, foram analisadas publicações indexadas nas bases *Web of Science* e *Scopus*, no intervalo de 2013 a 2024, permitindo a extração de tendências temáticas, abordagens metodológicas predominantes e lacunas teóricas ainda não exploradas.

Os resultados revelaram o crescimento na produção científica sobre a interseção entre ciência de dados e esporte, especialmente nos estudos voltados ao uso de algoritmos de *machine learning* em processos decisórios estratégicos. Ainda que a área seja relativamente recente, a análise demonstrou a consolidação de técnicas como regressões e modelos baseados em árvores como alternativas para a previsão de indicadores de performance.

Um dos principais achados desta etapa foi a constatação de que a lógica *fuzzy* ainda ocupa uma posição periférica no campo da análise esportiva, sendo pouco sistematizada e raramente integrada aos modelos interpretáveis de aprendizado de máquina. Essa observação reforça a originalidade do escopo proposto nesta tese e delinea a necessidade de desenvolver modelagens híbridas, que articulem o poder explicativo dos algoritmos preditivos com a flexibilidade interpretativa e gradual da lógica *fuzzy*.

Ademais, a revisão apontou oportunidades metodológicas futuras, como a possibilidade de calibrar funções de pertinência específicas por posição. Outra observação deu-se sobre a capacidade de adaptação a um modelo multidimensional, adotando variáveis táticas, biomecânicas, físicas e situacionais em modelos *fuzzy*, o que pode representar um próximo salto metodológico no campo da análise esportiva. Outra identificação foi da vantagem analítica da *fuzzificação* de bancos de dados, ao permitir a capacidade dos modelos de aprendizado de máquina em lidar com a ambiguidade e a variabilidade natural do comportamento humano no esporte.

Desta maneira, os achados deste capítulo forneceram o embasamento teórico e metodológico para o desenvolvimento do modelo *fuzzy* aplicado no Capítulo 4. Em termos acadêmicos, o capítulo posicionou a tese em uma vertente emergente de pesquisa, numa intercessão entre os campos da estratégia, ciência de dados e esportes de alto rendimento.

O Capítulo 3 desta tese teve como objetivo estruturar e interpretar os fluxos internacionais de transferência de jogadores por meio da Análise de Redes Sociais (ARS), abrangendo as dinâmicas interdependentes que envolvem o mercado global de atletas do futebol. A partir da modelagem de mais de 100 mil transações entre clubes de diferentes países, a análise revelou que o sistema de transferências não opera de maneira aleatória ou desorganizada, mas sim segundo padrões estruturais consolidados, ancorados em rotas específicas e em posições estratégicas assumidas por determinados países.

Os principais achados desta etapa implicaram na posição que países, como Bélgica e Portugal, assumem como nós centrais na rede global, desempenhando papéis diversos, porém sistemáticos, na intermediação ou valorização de ativos esportivos. Tanto Bélgica quanto Portugal demonstraram ser portas de entrada e redistribuição para os mercados mais valorizados da Europa, enquanto os Estados Unidos da América se posicionam como uma liga em transição estratégica, capaz de atrair e desenvolver talentos jovens. Já os Emirados Árabes Unidos atuam

como mercado de consolidação de carreira, atraindo jogadores em fase final de trajetória competitiva.

Tais posições derivam da articulação de capacidades dinâmicas organizacionais, como redes de relacionamento, processos ágeis de negociação, expertise regulatória e sistemas eficientes de prospecção. Esses elementos contribuem para a formação de vantagens competitivas sustentadas no mercado de transferências, segundo a literatura estratégica. Por meio da lente da estratégia como prática, é possível compreender como essas vantagens emergem de recursos formais e rotinas de clubes, dirigentes, empresários e intermediários, cujas decisões moldam o fluxo dos atletas.

Nesse contexto, o caso brasileiro emerge com ambivalência. O país ocupa uma posição de destaque como origem de atletas, reafirmando seu papel tradicional como celeiro de talentos, mas carece de protagonismo como intermediário. Sua baixa centralidade de intermediação e limitada presença em rotas estratégicas indicam uma subutilização de suas capacidades organizacionais e institucionais. Ou seja, exporta valor, mas capta pouco retorno ao longo da cadeia. Tal configuração reforça a necessidade de que os clubes brasileiros invistam em sistemas mais sofisticados de avaliação e planejamento estratégico, capazes de maximizar a captura de valor de seus ativos.

Além do mapeamento das rotas e das relações entre países, os resultados também oferecem evidências empíricas que sustentam a necessidade de integração entre o conhecimento estrutural das redes e ferramentas analíticas de avaliação de atletas. A compreensão das rotas, das posições na rede e dos intermediários frequentes precisa ser acompanhada da capacidade de estimar com precisão o valor potencial de um atleta, tanto em termos técnicos quanto de mercado.

Neste sentido, o capítulo fornece a base contextual e relacional sobre a qual se apoia a próxima etapa da tese. A proposta de desenvolver um modelo *fuzzy* de avaliação de performance emerge como resposta prática à constatação de que clubes precisam reduzir o grau de incerteza envolvido nas decisões de contratação. Se a posição em uma rede oferece oportunidades, é a qualidade da tomada de decisão interna que define se essas oportunidades serão ou não convertidas em vantagem competitiva.

A orientação para pesquisas futuras nesta área é o aprofundamento da análise de redes aplicadas aos contextos intrarregionais, isto é, recorte considerando apenas transferências dentro de um

mesmo continente. Outro contexto seria investigar a evolução temporal da rede sob o impacto de eventos externos, tais como novas regulações de contratos entre atletas e clubes. Por fim, o próprio autor desta tese tem como propósito investigar a articulação entre as redes de transferência e as múltiplas propriedades de clubes (*multi-club ownership*), fenômeno recente com capacidade de influenciar decisões de transferência e desenvolvimento de atletas.

Este capítulo contribuiu para consolidar o conhecimento estrutural sobre o mercado global de transferências, entregando resultados para a literatura de gestão esportiva e estudos organizacionais. Este trabalho também oferece embasamento à tomada de decisão nos clubes, sinalizando gargalos e caminhos estratégicos que podem ser explorados por meio de posicionamentos inteligentes na rede do futebol internacional.

Por fim, o último capítulo desta tese apresentou a construção, validação e aplicação de um modelo de avaliação de desempenho técnico de jogadores de futebol baseado em transformações *fuzzy*. Os principais achados revelaram que, ao incorporar lógicas de pertencimento graduais e estruturas de agrupamento funcional, o modelo permitiu representar a complexidade do jogo de maneira mais próxima à realidade prática e dinâmica do futebol. A comparação entre versões da modelagem (IT2FS, TFRa e IFR) demonstrou que a abordagem baseada na *Integrated Fuzzy and Relative* (IFR) apresentou os melhores resultados em termos de capacidade explicativa para variáveis objetivas como minutagem, gols e assistências, mesmo que ainda limitada quanto à previsão de variáveis financeiras.

Esses achados implicam duas contribuições teóricas do capítulo. A primeira está ancorada na epistemologia da abordagem ecológica dinâmica, que considera o desempenho como um fenômeno emergente da interação entre o jogador e o seu ambiente. A modelagem *fuzzy* alinha-se a esse princípio ao evitar classificações rígidas, permitindo capturar nuances e variações de contexto que podem ser negligenciadas em análises puramente estatísticas. A segunda contribuição reside no campo da modelagem centrada na pessoa: ao permitir que jogadores compartilhem características de múltiplos agrupamentos, a ferramenta valoriza a trajetória individual e reconhece a diversidade funcional dentro de um mesmo sistema tático.

Do ponto de vista metodológico, o capítulo inovou ao utilizar dados técnicos extraídos do *Wyscout*, ponderados e transformados por funções *fuzzy*, para construir um sistema comparativo de pontuação interpretável. O uso de agrupamentos *fuzzy c-means* e hierárquicos contribuiu para revelar estruturas latentes que não dependem apenas da posição nominal dos jogadores, mas sim de funções exercidas em campo, como finalizadores, construtores ou marcadores. Essa

categorização funcional tem potencial para ser aprofundada em estudos futuros e utilizada em contextos de *scouting* e análise técnica aplicada.

Os limites enfrentados nesta pesquisa oferecem potencial para investigações futuras. Uma primeira possibilidade refere-se à ampliação da base de dados, dado que o modelo foi testado exclusivamente com atletas que atuaram em competições nacionais. Estudos posteriores poderão incorporar competições internacionais, diferentes contextos táticos e variações culturais de jogo, permitindo avaliar a transferibilidade da modelagem *fuzzy* para ecossistemas mais diversos.

Além disso, a incorporação de variáveis de outras dimensões do jogo de futebol, tais como táticas, físicas e psicossociais, incluindo, mas não restrito, aos dados de GPS, métricas de posicionamento coletivo ou indicadores psicométricos, ampliaria o escopo da análise, contribuindo para a construção de modelos mais holísticos, capazes de representar a performance esportiva em sua complexidade multidimensional. Também se destaca a necessidade de desenvolver um módulo específico para goleiros, cuja exclusão neste estudo decorreu das particularidades técnicas da posição, que exigem critérios próprios de avaliação e *fuzzificação*.

De uma perspectiva técnica, o refinamento da modelagem *fuzzy* poderá incluir validações cruzadas distintas das que foram aplicadas neste estudo. Isto implica na possibilidade de utilização de outras bases estatísticas, aplicando, ou não, técnicas de regressão para redução dos fatores. Outros ensaios, como testes com diferentes critérios de agrupamento e a exploração de técnicas avançadas como *explainable AI*, redes neurais ou sistemas adaptativos, podem promover o desenvolvimento do campo da inteligência artificial aplicada ao esporte.

Outro caminho relevante para aplicação prática do modelo encontra-se nas categorias de base, onde seu uso experimental pode auxiliar na detecção precoce de padrões de talentos, comparando estes padrões com as avaliações subjetivas realizadas por olheiros e comissões técnicas. Finalmente, a lógica proposta nesta tese pode ser adaptada e testada em outras modalidades esportivas coletivas com características semelhantes ao futebol, como o futebol de salão, o handebol ou o rúgbi, contribuindo para consolidar um paradigma interpretativo e flexível de avaliação de performance em diferentes contextos esportivos.

Em síntese, o modelo *fuzzy* apresentado neste trabalho demonstrou ser uma ferramenta sensível às variações contextuais do jogo e teoricamente alinhada às correntes contemporâneas da

ciência do esporte. Sua virtude está na capacidade de equilibrar objetividade estatística com complexidade contextual e interpretabilidade prática, elementos para o apoio à tomada de decisão em ambientes competitivos de alta exigência. Concomitantemente, abre caminho para avanços metodológicos e aplicações práticas que poderão ampliar seu alcance e impacto no campo da gestão esportiva e da análise de desempenho.

## Referências Bibliográficas

- ADAM, Stavros P. *et al.* No Free Lunch Theorem: A Review. In: DEMETRIOU, IOANNIS C.; PARDALOS, PANOS M. (Org.). *Approximation and Optimization: Algorithms, Complexity and Applications*. Cham: Springer International Publishing, 2019. p. 57–82. Disponível em: <[https://doi.org/10.1007/978-3-030-12767-1\\_5](https://doi.org/10.1007/978-3-030-12767-1_5)>. Acesso em: 26 maio 2025.
- ADNER, Ron; HELFAT, Constance E. Corporate effects and dynamic managerial capabilities. *Strategic Management Journal*, v. 24, n. 10, p. 1011–1025, 2003.
- ALKIRE, Sabina *et al.* *Multidimensional Poverty Measurement and Analysis*. [S.l.]: Oxford University Press, 2015. Disponível em: <<https://doi.org/10.1093/acprof:oso/9780199689491.001.0001>>. (Oxford Scholarship Online).
- ALMEIDA, Luís Manuel de Oliveira Marques De. *Perfil psicológico de prestação, orientações cognitivas e negativismo do futebolista português*. 2010. 164 f. Dissertação de Mestrado – Universidade de Trás-os-Montes e Alto Douro, Vila Real, 2010.
- ALMULLA, Jassim; ALAM, Tanvir. Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League. *IEEE Access*, v. 8, p. 213695–213705, 2020.
- ALZUBI, Jafar; NAYYAR, Anand; KUMAR, Akshi. Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, v. 1142, p. 012012, nov. 2018.
- ANDERSON, Chris; SALLY, David. *The Numbers Game: Why Everything You Know About Football is Wrong*. [S.l.]: Penguin UK, 2013.
- ARAÚJO, Duarte *et al.* Ecological cognition: expert decision-making behaviour in sport. *International Review of Sport and Exercise Psychology*, v. 12, n. 1, p. 1–25, 1 jan. 2019.
- ARAÚJO, Duarte; DAVIDS, Keith. The (Sport) Performer-Environment System as the Base Unit in Explanations of Expert Performance. *Journal of Expertise*, v. 1, n. 3, 31 dez. 2018. Disponível em: <[https://www.journalofexpertise.org/articles/volume1\\_issue3/JoE\\_2018\\_1\\_3\\_AraujoDavids.html](https://www.journalofexpertise.org/articles/volume1_issue3/JoE_2018_1_3_AraujoDavids.html)>. Acesso em: 5 set. 2023.
- ARIA, Massimo; CUCCURULLO, Corrado. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, v. 11, n. 4, p. 959–975, 1 nov. 2017.

- AUFFARTH, Ben. *Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods*. [S.l.]: Packt Publishing, 2021.
- BABOOTA, Rahul; KAUR, Harleen. Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, v. 35, n. 2, p. 741–755, 1 abr. 2019.
- BARAJAS, Angel; RODRÍGUEZ, Plácido. Spanish Football Clubs Finances: Crisis and Player Salaries. *International Journal of Sport Finance*, v. 5, 14 mar. 2010.
- BARNETT-ITZHAKI, Zohar *et al.* Machine learning vs. classic statistics for the prediction of IVF outcomes. *Journal of Assisted Reproduction and Genetics*, v. 37, n. 10, p. 2405–2412, 1 out. 2020.
- BARNEY, Jay. Firm Resources and Sustained Competitive Advantage. *Journal of Management*, v. 17, n. 1, p. 99–120, 1 mar. 1991.
- BARROS, Laécio Carvalho De; ESMI, Estevão; *et al.* Cálculo para processos fuzzy interativos. In: VI CONGRESSO BRASILEIRO DE SISTEMAS FUZZY (VI CBSF), nov. 2021, São José do Rio Preto. *Anais...* São José do Rio Preto: [s.n.], nov. 2021.
- BARROS, Laécio Carvalho De; WASQUES, Vinícius Francisco; *et al.* Sistemas p-fuzzy e aplicações. In: VI CONGRESSO BRASILEIRO DE SISTEMAS FUZZY (VI CBSF), nov. 2021, São José do Rio Preto. *Anais...* São José do Rio Preto: [s.n.], nov. 2021.
- BARTLETT, Jonathan D. *et al.* Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *International Journal of Sports Physiology and Performance*, v. 12, n. 2, p. 230–234, 1 fev. 2017.
- BAUER, Daniel J.; SHANAHAN, Michael J. Modeling complex interactions: Person-centered and variable-centered approaches. *Modeling contextual effects in longitudinal studies*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2007. p. 255–283.
- BAZMARA, Mohammad. A Novel Fuzzy Approach for Determining Best Position of Soccer Players. *International Journal of Intelligent Systems and Applications*, v. 6, n. 9, p. 62–67, 2014.
- BEIDERBECK, Daniel *et al.* The impact of technology on the future of football – A global Delphi study. *Technological Forecasting and Social Change*, v. 187, p. 122186, 1 fev. 2023.

BENAVIDES ESPINOSA, María del Mar; MERIGÓ LINDAHL, José María. Organizational design as a learning enabler: A fuzzy-set approach. *Journal of Business Research*, Set-Theoretic research in business. v. 69, n. 4, p. 1340–1344, 1 abr. 2016.

BERGMAN, Lars R.; MAGNUSSON, David. A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, v. 9, n. 2, p. 291–319, jun. 1997.

BETTI, Gianni *et al.* Comparative measures of multidimensional deprivation in the European Union. *Empirical Economics*, v. 49, n. 3, p. 1071–1100, 1 nov. 2015.

BETTI, Gianni *et al.* Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach. In: LEMMI, ACHILLE; BETTI, GIANNI (Org.). . *Fuzzy Set Approach to Multidimensional Poverty Measurement*. Boston, MA: Springer US, 2006. p. 115–137. Disponível em: <[https://doi.org/10.1007/978-0-387-34251-1\\_7](https://doi.org/10.1007/978-0-387-34251-1_7)>. Acesso em: 4 abr. 2025.

BETTI, Gianni *et al.* The fuzzy approach to poverty measurement. *Research Handbook on Measuring Poverty and Deprivation*. [S.l.]: Edward Elgar Publishing, 2023. p. 489–500. Disponível em: <<https://www.elgaronline.com/edcollchap/book/9781800883451/book-part-9781800883451-61.xml>>. Acesso em: 28 set. 2024.

BETTI, Gianni *et al.* THE INTEGRATED FUZZY AND RELATIVE INDEX FOR POVERTY ANALYSIS: A REVIEW OF APPLICATIONS IN THE SOCIAL SCIENCES. *Studies of Applied Economics*, v. 38, n. 1, 4 fev. 2020. Disponível em: <<https://ojs.ual.es/ojs/index.php/eea/article/view/2985>>. Acesso em: 28 set. 2024.

BETTI, Gianni; GAGLIARDI, Francesca; VERMA, Vijay. Simplified Jackknife Variance Estimates for Fuzzy Measures of Multidimensional Poverty. *International Statistical Review*, v. 86, n. 1, p. 68–86, 2018.

BETTI, Gianni; VERMA, Vijay. Measuring the degree of poverty in a dynamic and comparative context: a multidimensional approach using fuzzy set theory. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE, 1999, Lahore. *Anais...* Lahore: [s.n.], 1999. p. 289–301.

BLUM, Avrim. Machine learning theory. 26. ed. Pittsburgh: Carnegie Melon Universit, School of Computer Science, 2007. .

BORGATTI, Stephen P.; EVERETT, Martin G.; JOHNSON, Jeffrey C. *Analyzing Social Networks*. London: SAGE, 2013.

- BORGES, Henrique do Nascimento; SANTOS, Ana Lúcia Padrão Dos. Panorama do sistema de captação de atletas dos clubes brasileiros de futebol. *RBFF - Revista Brasileira de Futsal e Futebol*, v. 13, n. 53, p. 355–366, 7 nov. 2021.
- BRACKE, Paul. Social networks and relational capital in library service assessment. *Performance Measurement and Metrics*, v. 17, n. 2, p. 134–141, 11 jul. 2016.
- BREIMAN, Leo. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 1 out. 2001.
- BRUNSWIK, Egon. *Perception and the Representative Design of Psychological Experiments*. [S.l.]: University of California Press, 1956.
- BUCK, Christoph; IFLAND, Sebastian. Toward an enduring football economy: a business model taxonomy for Europe’s professional football clubs. *European Sport Management Quarterly*, v. 23, n. 5, p. 1409–1429, 3 set. 2023.
- BUNKER, Rory P.; THABTAH, Fadi. A machine learning framework for sport result prediction. *Applied Computing and Informatics*, v. 15, n. 1, p. 27–33, 1 jan. 2019.
- BURAIMO, Babatunde; SIMMONS, Rob. Uncertainty of Outcome or Star Quality? Television Audience Demand for English Premier League Football. *International Journal of the Economics of Business*, v. 22, n. 3, p. 449–469, 2 set. 2015.
- BURGELMAN, Robert A. *et al.* Strategy processes and practices: Dialogues and intersections. *Strategic Management Journal*, v. 39, n. 3, p. 531–558, 2018.
- BURNAP, Pete *et al.* Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*, v. 95, p. 96–108, 1 jun. 2015.
- CAI, Jianchao *et al.* Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Applied Energy*, v. 262, p. 114566, 15 mar. 2020.
- CAMACHO, David *et al.* The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, v. 63, p. 88–120, 1 nov. 2020.
- CASARIN, Rodrigo Vicenzi *et al.* Modelo de jogo e processo de ensino no futebol: princípios globais e específicos. *Movimento*, v. 17, n. 3, p. 133–152, 2011.
- CHELI, Bruno *et al.* Evolution of the Fuzzy-Set Approach to Multi-Dimensional Poverty Measurement. *Analysis of Socio-Economic Conditions*. [S.l.]: Routledge, 2021. .

CHELI, Bruno; LEMMI, A. A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. v. 24, p. 115–134, 1995.

CHEMERO, Anthony. Information and direct perception: A new approach. *Advanced issues in cognitive science and semiotics*. [S.l: s.n.], 2006. p. 59–72.

CHEN, Qiuwen; MYNETT, Arthur E. Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling*, v. 162, n. 1, p. 55–67, 1 abr. 2003.

CHIU, Stephen L. Fuzzy Model Identification Based on Cluster Estimation. *J. Intell. Fuzzy Syst.*, v. 2, n. 3, p. 267–278, 1 maio 1994.

CINTRA, Marcos Evandro *et al.* Fuzzy Rule Base Generation through Genetic Algorithms and Bayesian Classifiers A Comparative Approach. In: SEVENTH INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS DESIGN AND APPLICATIONS (ISDA 2007), out. 2007, [S.l: s.n.], out. 2007. p. 315–322. Disponível em: <<https://ieeexplore.ieee.org/document/4389627>>. Acesso em: 4 abr. 2025.

COBO, M. J. *et al.* An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, v. 5, n. 1, p. 146–166, 1 jan. 2011.

COHEN-ADDAD, Vincent *et al.* Hierarchical Clustering: Objective Functions and Algorithms. *J. ACM*, v. 66, n. 4, p. 26:1-26:42, 5 jun. 2019.

COLYER, Steffi L. *et al.* A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Medicine - Open*, v. 4, n. 1, p. 24, 5 jun. 2018.

CONESA, Ana; HERNÁNDEZ, Rafael. Chapter 16 - Omics Data Integration in Systems Biology: Methods and Applications. In: GARCÍA-CAÑAS, VIRGINIA; CIFUENTES, ALEJANDRO; SIMÓ, CAROLINA (Org.). . *Comprehensive Analytical Chemistry. Applications of Advanced Omics Technologies: From Genes to Metabolites*. [S.l.]: Elsevier, 2014. v. 64. p. 441–459. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B97804444626509000166>>. Acesso em: 10 abr. 2025.

CORDÓN, Oscar. *Genetic Fuzzy Systems: Evolutionary Tuning And Learning Of Fuzzy Knowledge Bases*. Singapore: Wspc, 2002.

CORNELL UNIVERSITY. *The Art of Prediction and Responding. Networks Course*. [S.l.: s.n.]. Disponível em: <<https://blogs.cornell.edu/info2040/2022/09/19/the-art-of-prediction-and-responding/>>. Acesso em: 27 dez. 2022. , 19 set. 2022

COSTA, Israel Teoldo Da *et al.* Princípios Táticos do Jogo de Futebol: conceitos e aplicação. *Revista Motriz*, v. 15, n. 3, p. 657–668, 2009.

COSTA, Israel Teoldo Da; CARDOSO, Felipe; MACHADO, Guilherme. Tomada de decisão no futebol: da avaliação à aplicação em campo. [S.l.: s.n.], 2021. p. 187–205.

DANGETI, Pratap. *Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R*. [S.l.]: Packt Publishing, 2017.

DAVIDS, Keith *et al.* An ecological dynamics approach to skill acquisition: Implications for development of talent in sport. *Talent Development and Excellence*, v. 5, n. 1, p. 21–34, 2013.

DAVIDS, Keith *et al.* Understanding Environmental and Task Constraints on Talent Development: Analysis of micro-structure of practice and macro-structure of development histories. *Routledge Handbook of Talent Identification and Development in Sport*. [S.l.]: Routledge, 2017. .

DAY, Dave. Craft Coaching and the ‘Discerning Eye’ of the Coach. *International Journal of Sports Science & Coaching*, v. 6, n. 1, p. 179–195, 1 mar. 2011.

DELOITTE. *Football Money League 2023*. , Deloitte Football Money League. *Econômico-Financeiro*, n° 26th. [S.l.]: Deloitte, 2023. Disponível em: <<https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html>>. Acesso em: 5 out. 2023.

DEMROZI, Florenc *et al.* Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey. *IEEE Access*, v. 8, p. 210816–210836, 2020.

DETZEN, Dominic; LÖHLEIN, Lukas. Valuing soccer players: on the valuation dynamics of an online user community. *Accounting, Auditing & Accountability Journal*, v. ahead-of-print, n. ahead-of-print, 1 jan. 2023. Disponível em: <<https://doi.org/10.1108/AAAJ-05-2022-5824>>. Acesso em: 16 nov. 2023.

DI SALVO, Valter *et al.* Activity profile of elite goalkeepers during football match-play. *JOURNAL OF SPORTS MEDICINE AND PHYSICAL FITNESS*, v. 48, n. 4, p. 443–446, 2008.

DIMITRIOU, Loukas; TSEKERIS, Theodore; STATHOPOULOS, Antony. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. *Transportation Research Part C: Emerging Technologies*, v. 16, n. 5, p. 554–573, 1 out. 2008.

DIODATO, Virgil P.; GELLATLY, Peter. *Dictionary of Bibliometrics*. [S.l.]: Routledge, 2013. Disponível em: <<https://www.taylorfrancis.com/books/mono/10.4324/9780203714133/dictionary-bibliometrics-virgil-diodato-peter-gellatly>>. Acesso em: 6 jul. 2022.

DONTHU, Naveen *et al.* How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, v. 133, p. 285–296, 1 set. 2021.

DUSA, Adrian. *QCA with R: A Comprehensive Resource*. [S.l.]: Springer, 2018.

ECHCHAKOUI, Saïd. Why and how to merge Scopus and Web of Science during bibliometric analysis: the case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, v. 8, n. 3, p. 165–184, 1 set. 2020.

EISENHARDT, Kathleen M.; MARTIN, Jeffrey A. Dynamic capabilities: what are they? *Strategic Management Journal*, v. 21, n. 10–11, p. 1105–1121, 2000.

ELKANO, Mikel; BUSTINCE, Humberto; GALAR, Mikel. Do we still need fuzzy classifiers for Small Data in the Era of Big Data? In: 2019 IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS (FUZZ-IEEE), jun. 2019, [S.l.: s.n.], jun. 2019. p. 1–6. Disponível em: <[https://ieeexplore.ieee.org/abstract/document/8858943?casa\\_token=Wn-4QRtPjhwAAAAA:0MKY6ozZ4-bB4j1rpPW9jCqoSctg6tloX2K7z4ShBTt0DIYDNGnv23zhJttAoF8hjCWpt83JXp0](https://ieeexplore.ieee.org/abstract/document/8858943?casa_token=Wn-4QRtPjhwAAAAA:0MKY6ozZ4-bB4j1rpPW9jCqoSctg6tloX2K7z4ShBTt0DIYDNGnv23zhJttAoF8hjCWpt83JXp0)>. Acesso em: 24 out. 2023.

EMMENEGGER, Patrick. Job security regulations in Western democracies: A fuzzy set analysis. *European Journal of Political Research*, v. 50, n. 3, p. 336–364, 1 maio 2011.

ERMIDIS, Georgios *et al.* Technical demands across playing positions of the Asian Cup in male football. *International Journal of Performance Analysis in Sport*, v. 19, n. 4, p. 530–542, 4 jul. 2019.

EXAME. Clube com 200 torcedores ganha US\$ 14 mi com transferências. *Exame*, 19 mar. 2014. Disponível em: <<https://exame.com/negocios/clube-com-200-torcedores-ganha-us-14-mi-com-transferencias/>>. Acesso em: 30 mar. 2025.

FAINSHMIDT, Stav; SMITH, Adam; GULDIKEN, Orhun. Orchestrating the flow of human resources: Insights from Spanish soccer clubs. *Strategic Organization*, v. 15, n. 4, p. 441–460, 1 nov. 2017.

FERNANDEZ-NAVARRO, Javier *et al.* Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, v. 18, n. 3, p. 423–436, 4 maio 2018.

FERREIRA, João José M.; FERNANDES, Cristina I.; RATTEN, Vanessa. A co-citation bibliometric analysis of strategic management research. *Scientometrics*, v. 109, n. 1, p. 1–32, 1 out. 2016.

FIALHO, Joaquim. Pressupostos para a construção de uma sociologia das redes sociais. *Sociologia: Revista da Faculdade de Letras da Universidade do Porto*, v. 29, 1 dez. 2015. Disponível em: <<https://ojs.letras.up.pt/index.php/Sociologia/article/view/1297>>. Acesso em: 26 maio 2025.

FILIPPONE, Andrea; CHELI, Bruno; D'AGOSTINO, Antonella. Addressing the interpretation and the aggregation problems in totally fuzzy and relative poverty measures. *ISER Working Paper Series*, Institute for Social and Economic Research. n. 22, 2001.

FRIEDMAN, Jerome H.; HASTIE, Trevor; TIBSHIRANI, Rob. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, v. 33, p. 1–22, 2 fev. 2010.

FUERTES, Guillermo *et al.* Conceptual Framework for the Strategic Management: A Literature Review—Descriptive. *Journal of Engineering*, v. 2020, p. e6253013, 30 jan. 2020.

FURRER, Olivier; THOMAS, Howard; GOUSSEVSKAIA, Anna. The structure and evolution of the strategic management field: A content analysis of 26 years of strategic management research. *International Journal of Management Reviews*, v. 10, n. 1, p. 1–23, 2008.

GABAIX, Xavier *et al.* Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review*, v. 96, n. 4, p. 1043–1068, set. 2006.

GABRIEL, Allison S. *et al.* Fuzzy Profiles: Comparing and Contrasting Latent Profile Analysis and Fuzzy Set Qualitative Comparative Analysis for Person-Centered Research. *Organizational Research Methods*, v. 21, n. 4, p. 877–904, 1 out. 2018.

GARCÍA-NIETO, Paulino José; GARCÍA-GONZALO, Esperanza; PAREDES-SÁNCHEZ, José Pablo. Prediction of the critical temperature of a superconductor by using the

WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques. *Neural Computing and Applications*, v. 33, n. 24, p. 17131–17145, 1 dez. 2021.

GARGANTA, Júlio. *Atrás do palco, nas oficinas do futebol*. Centro de Pesquisa e Desenvolvimento Desportivo. [S.l.]: Comitê Olímpico de Portugal, 2004.

GARGANTA, Júlio. Trends of tactical performance analysis in team sports: bridging the gap between research, training and competition. *Revista Portuguesa de Ciências do desporto*, v. 9, n. 1, p. 81–89, 2009.

GASPARETTO, Thadeu; BARAJAS, Angel. Wage Dispersion and Team Performance: The Moderation Role of Club Size. *Journal of Sports Economics*, v. 23, n. 5, p. 548–566, 1 jun. 2022.

GIBSON, James J. *The Ecological Approach to Visual Perception: Classic Edition*. New York: Psychology Press, 2014.

GONZÁLEZ-ALBO, Borja; BORDONS, María. Articles vs. proceedings papers: Do they differ in research relevance and impact? A case study in the Library and Information Science field. *Journal of Informetrics*, v. 5, n. 3, p. 369–381, 1 jul. 2011.

GOSWAMI, R. *et al.* Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. *Brain Structure and Function*, v. 221, n. 4, p. 1911–1925, 1 maio 2016.

GRAFIETTI, Cesar. *Relatório Convocados 2024: Finanças, História e Mercado do Futebol Brasileiro*. Econômico-Financeiro. [S.l.]: Convocados e OutField Inc., 2024. Disponível em: <<https://static.poder360.com.br/2024/11/Relatorio-Convocados-2024-.pdf>>. Acesso em: 15 out. 2023.

GUAN, Shuo; WANG, Xiaochen. Optimization analysis of football match prediction model based on neural network. *Neural Computing and Applications*, v. 34, n. 4, p. 2525–2541, 1 fev. 2022.

GUERRAS-MARTÍN, Luis Ángel; MADHOK, Anoop; MONTORO-SÁNCHEZ, Ángeles. The evolution of strategic management research: Recent trends and current directions. *BRQ Business Research Quarterly*, v. 17, n. 2, p. 69–76, 1 abr. 2014.

GUIMARÃES, Murilo Balbino; OLIVEIRA, Augusto Moura De; PAOLI, Próspero Brum. *A prospecção do talento no futebol brasileiro:: diagnóstico estrutural e financeiro do processo de captação de atletas*. [S.l.]: Editora Appris, 2020.

HACKEL, Timo *et al.* *Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark*. . [S.l.]: arXiv. Disponível em: <<http://arxiv.org/abs/1704.03847>>. Acesso em: 26 maio 2025. , 12 abr. 2017

HASHEMI, Seyed Emadedin; GHOLIAN-JOUYBARI, Fatemeh; HAJIAGHAEI-KESHTALI, Mostafa. A fuzzy C-means algorithm for optimizing data clustering. *Expert Systems with Applications*, v. 227, p. 120377, 1 out. 2023.

HASSANNIAKALAGER, Arman *et al.* A conditional fuzzy inference approach in forecasting. *European Journal of Operational Research*, v. 283, n. 1, p. 196–216, 16 maio 2020.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome H. *The Elements of Statistical Learning*. 2. ed. New York: Springer, 2009. Disponível em: <<https://doi.org/10.1007/978-0-387-84858-7>>. (Springer Series in Statistics).

HAYAT, Khizar *et al.* New group-based generalized interval-valued q-rung orthopair fuzzy soft aggregation operators and their applications in sports decision-making problems. *Computational and Applied Mathematics*, v. 42, n. 1, p. 4, 8 dez. 2022.

HEALD, Graeme. Issues with Reliability of Fuzzy Logic. *International Journal of Trend in Scientific Research and Development*, v. Volume-2, n. Issue-6, 10 set. 2018. Disponível em: <<https://www.ijtsrd.com/engineering/computer-engineering/18573/issues-with-reliability-of-fuzzy-logic/graeme-heald>>. Acesso em: 4 jan. 2020.

HEFT, Harry. An Ecological Approach to Psychology. *Review of General Psychology*, v. 17, n. 2, p. 162–167, 1 jun. 2013.

HELFAT, Constance E.; MARTIN, Jeffrey A. Dynamic Managerial Capabilities: Review and Assessment of Managerial Impact on Strategic Change. *Journal of Management*, v. 41, n. 5, p. 1281–1312, 1 jul. 2015.

HEROLD, Mat *et al.* Machine learning in men’s professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, v. 14, n. 6, p. 798–817, 1 dez. 2019.

HEVNER, Alan R. *et al.* Design Science in Information Systems Research. *MIS Quarterly*, v. 28, n. 1, p. 75–105, 2004.

HEWA WELEGE, Nandun Madhusanka; PAN, Wei; KUMARASWAMY, Mohan. Social network analysis applications in sustainable construction and built environment management:

a review. *Built Environment Project and Asset Management*, v. 11, n. 4, p. 511–528, 1 jan. 2021.

HO, Tin Kam. Random decision forests. In: 3RD INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, ago. 1995, [S.l: s.n.], ago. 1995. p. 278–282 vol.1. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/598994>>. Acesso em: 10 abr. 2025.

HOCQUET, Alexandre. Football Manager: Mutual Shaping between Game, Sport, and Community. *Kinephanos*, 2016. Disponível em: <<https://www.kinephanos.ca/2016/football-manager/>>.

HOFMANS, Joeri; WILLE, Bart; SCHREURS, Bert. Person-centered methods in vocational research. *Journal of Vocational Behavior*, v. 118, p. 103398, 1 abr. 2020.

HOLZINGER, Andreas. From Machine Learning to Explainable AI. In: 2018 WORLD SYMPOSIUM ON DIGITAL INTELLIGENCE FOR SYSTEMS AND MACHINES (DISA), ago. 2018, [S.l: s.n.], ago. 2018. p. 55–66.

HOLZMAYER, Florian; SCHMIDT, Sascha L. Financial performance and corporate diversification strategies in professional football – evidence from the English Premier League. *Sport, Business and Management: An International Journal*, v. 10, n. 3, p. 291–315, 1 jan. 2020.

HOSKISSON, Robert E *et al.* Theory and research in strategic management: swings of a pendulum. *Journal of Management*, v. 25, n. 3, p. 417–456, 1 jan. 1999.

HOWARD, Matt C.; HOFFMAN, Michael E. Variable-Centered, Person-Centered, and Person-Specific Approaches: Where Theory Meets the Method. *Organizational Research Methods*, v. 21, n. 4, p. 846–876, 1 out. 2018.

HUGHES, Michael David *et al.* Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position. jun. 2012. Disponível em: <<http://rua.ua.es/dspace/handle/10045/23161>>. Acesso em: 17 dez. 2019.

HUGHES, Mike; PROBERT, Graham. A technical analysis of elite male soccer players by position and success. *Notational Analysis of Sport-VII*. Cardiff: UWIC, 2006. p. 76–91.

HUNG, Andrew J. Can machine-learning algorithms replace conventional statistics? *BJU International*, v. 123, n. 1, p. 1–1, 2019.

INGERSOLL, Keith; MALESKY, Edmund; SAIEGH, Sebastian M. Heterogeneity and team performance: Evaluating the effect of cultural diversity in the world's top soccer league. *Journal of Sports Analytics*, v. 3, n. 2, p. 67–92, 1 jan. 2017.

INGOLD, Tim. *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*. London: Routledge, 2021.

JAMIL, Mikael. Where do the best technical football players in the world come from? Analysing the association between technical proficiency and geographical origin in elite football. *Journal of Human Sport and Exercise*, 30 set. 2020. Disponível em: <[https://rua.ua.es/dspace/bitstream/10045/109475/1/JHSE\\_17-2\\_InPress\\_02.pdf](https://rua.ua.es/dspace/bitstream/10045/109475/1/JHSE_17-2_InPress_02.pdf)>. Acesso em: 18 set. 2023.

JANG, J.-S.R. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 23, n. 3, p. 665–685, maio 1993.

JANG, Jyh-Shing Roger; SUN, Chuen-Tsai; MIZUTANI, Eiji. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. 1st edition ed. Upper Saddle River, NJ: Pearson College Div, 1997.

JEDELHAUSER, Fabienne; FLEPP, Raphael; FRANCK, Egon. Overshadowed by Popularity: The Value of Second-Tier Stars in European Football. *Journal of Sports Economics*, v. 24, n. 8, p. 1026–1054, 1 dez. 2023.

KENWORTHY, Thomas P.; VERBEKE, Alain. The future of strategic management research: Assessing the quality of theory borrowing. *European Management Journal*, v. 33, n. 3, p. 179–190, 1 jun. 2015.

KERO, Chalchissa Amentie; BOGALE, Addisalem Tadesse. A Systematic Review of Resource-Based View and Dynamic Capabilities of Firms and Future Research Avenues. *International Journal of Sustainable Development and Planning*, v. 18, n. 10, p. 3137–3154, out. 2023.

KHALIFA, Azaddin Salem. The “strategy frame” and the four Es of strategy drivers. *Management Decision*, v. 46, n. 6, p. 894–917, 1 jan. 2008.

KIRSCH, Benedikt; SAUER, Tim; ZÜLCH, Henning. Decrypting the specifics of professional football club investments: Why? What? How? An integrative review. *Sport, Business and Management: An International Journal*, v. 14, n. 2, p. 136–168, 1 jan. 2024.

- KOCKELMAN, Paul. Information is the enclosure of meaning: Cybernetics, semiotics, and alternative theories of information. *Language & Communication*, v. 33, n. 2, p. 115–127, 1 abr. 2013.
- KONEFAŁ, Marek *et al.* Match outcome vs match status and frequency of selected technical activities of soccer players during UEFA Euro 2016. *International Journal of Performance Analysis in Sport*, v. 18, n. 4, p. 568–581, 4 jul. 2018.
- KUSMAKAR, Shitanshu *et al.* Machine Learning Enabled Team Performance Analysis in the Dynamical Environment of Soccer. *IEEE Access*, v. 8, p. 90266–90279, 2020.
- LANGEROUDI, Milad Keshtkar; YAMAGHANI, Mohammad Reza; KHODAPARAST, Siavash. FD-LSTM: A Fuzzy LSTM Model for Chaotic Time-Series Prediction. *IEEE Intelligent Systems*, v. 37, n. 4, p. 70–78, jul. 2022.
- LARKIN, Paul; O’CONNOR, Donna. Talent identification and recruitment in youth soccer: Recruiter’s perceptions of the key attributes for player recruitment. *PLOS ONE*, v. 12, n. 4, p. e0175716, 18 abr. 2017.
- LARKIN, Paul; REEVES, Matthew J. Junior-elite football: time to re-position talent identification? *Soccer & Society*, v. 19, n. 8, p. 1183–1192, 17 nov. 2018.
- LERNER, Richard M. Theories of human development: Contemporary perspectives. *Handbook of child psychology: Theoretical models of human development, Volume 1, 5th ed.* Hoboken, NJ, US: John Wiley & Sons Inc, 1998. p. 1–24.
- LI, Rongfeng *et al.* A new similarity function for Pythagorean fuzzy sets with application in football analysis. *AIMS Mathematics*, Cc\_license\_type: cc\_byPrimary\_atype: AIMS MathematicsSubject\_term: Research articleSubject\_term\_id: Research article, v. 9, n. 2, p. 4990–5014, 2024.
- LI, Yisheng. *When Moneyball Meets the Beautiful Game: A Predictive Analytics Approach to Exploring Key Drivers for Soccer Player Valuation*. 2021. Dissertação de Mestrado – Brock University, St. Catharines, 2021. Disponível em: <<https://dr.library.brocku.ca/handle/10464/15088>>. Acesso em: 10 nov. 2023.
- LIU, Hongyou; GOMEZ, Miguel-Ángel; *et al.* Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, v. 33, n. 12, p. 1205–1213, 21 jul. 2015.

LIU, Hongyou; YI, Qing; *et al.* Performance profiles of football teams in the UEFA Champions League considering situational efficiency. *International Journal of Performance Analysis in Sport*, v. 15, n. 1, p. 371–390, 1 mar. 2015.

LÓPEZ-VALENCIANO, Alejandro *et al.* A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms. *Medicine and Science in Sports and Exercise*, v. 50, n. 5, p. 915–927, maio 2018.

LÜDIN, Dennis; DONATH, Lars; ROMANN, Michael. Disagreement between talent scouts: Implications for improved talent assessment in youth football. *Journal of Sports Sciences*, v. 41, n. 8, p. 758–765, 18 abr. 2023.

MACDONALD, Robert D. *The struggle for the body, mind and soul of AFL Footballers [Book Review]*. . [S.l.]: Australian Society for Sports History Publications, maio 2011.

MACHADO, Guilherme; GONZÁLEZ-VÍLLORA, Sixto; TEOLDO, Israel. Selected soccer players are quicker and better decision-makers in elite Brazilian youth academies. *International Journal of Performance Analysis in Sport*, v. 23, n. 2, p. 65–82, 4 mar. 2023.

MAITLAND, Elizabeth; SAMMARTINO, André. Decision making and uncertainty: The role of heuristics and experience in assessing a politically hazardous environment. *Strategic Management Journal*, v. 36, n. 10, p. 1554–1578, 2015.

MAJEWSKI, Sebastian. Identification of Factors Determining Market Value of the Most Valuable Football Players. *Central European Management Journal*, v. 24, n. 3, p. 91–104, 2016.

MARCH, Salvatore T.; SMITH, Gerald F. Design and natural science research on information technology. *Decision Support Systems*, v. 15, n. 4, p. 251–266, 1 dez. 1995.

MARKOPOULOU, Christina; PAPAGEORGIOU, George; TJORTJIS, Christos. Diverse Machine Learning for Forecasting Goal-Scoring Likelihood in Elite Football Leagues. *MACHINE LEARNING AND KNOWLEDGE EXTRACTION*, Web of Science ID: WOS:001323712400001, v. 6, n. 3, p. 1762–1781, set. 2024.

MARQUES, Tânia; REIS, Nuno; GOMES, Jorge F. S. Responsible leadership research: A bibliometric review. *Brazilian Administration Review*, v. 15, n. 1, p. e170112, 13 mar. 2018.

MARTÍN, Gracia Rubio *et al.* The wisdom of crowd, real option and game theory decisions: can they be used by clubs to improve their investment in football players? *Managerial Finance*, v. 49, n. 6, p. 1036–1055, 27 dez. 2022.

MARTINS, Pedro Nuno Vieira Junqueira. *Análise aos processos de identificação, seleção, recrutamento e desenvolvimento de jovens Futebolistas. Relatório de Estágio Profissionalizante realizado no Departamento de Scouting da Futebol Clube do Porto–Futebol, SAD*. 2019. 127 f. Relatório de Estágio – Universidade do Porto, Porto, 2019.

MATESANZ, David *et al.* Transfer market activities and sportive performance in European first football leagues: A dynamic network approach. *PLOS ONE*, v. 13, n. 12, p. e0209362, 19 dez. 2018.

MELLAHI, Kamel; SMINIA, Harry. The frontiers of strategic management research: introduction to the special issue. *International Journal of Management Reviews*, v. 11, p. 1–7, 7 mar. 2009.

MENDEL, J.M.; JOHN, R.I.B. Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, v. 10, n. 2, p. 117–127, abr. 2002.

MENİZ, Büşra; ÖZKAN, E. Mehmet. Vaccine selection for COVID-19 by AHP and novel VIKOR hybrid approach with interval type-2 fuzzy sets. *Engineering Applications of Artificial Intelligence*, v. 119, p. 105812, 1 mar. 2023.

MEYLAN, César *et al.* Talent Identification in Soccer: The Role of Maturity Status on Physical, Physiological and Technical Characteristics. *International Journal of Sports Science & Coaching*, v. 5, n. 4, p. 571–592, dez. 2010.

MILOŠEVIĆ, Dušan; MILOŠEVIĆ, Mimica; SIMJANOVIĆ, Dušan. A Comparative Study of FAHP with Type-1 and Interval Type-2 Fuzzy Sets for ICT Implementation in Smart Cities. 2022, Cham. *Anais...* Cham: Springer International Publishing, 2022. p. 845–852.

MITROTASIOS, Michalis *et al.* The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. *International Journal of Performance Analysis in Sport*, v. 19, n. 3, p. 452–465, 4 maio 2019.

MIZUKOSHI, Marina T. *Estabilidade de Sistemas Dinâmicos Fuzzy*. 2004. 120 f. Tese de Doutorado – Universidade Estadual de Campinas, Campinas, 2004.

MJOLSNESS, Eric; DECOSTE, Dennis. Machine Learning for Science: State of the Art and Future Prospects. *Science*, v. 293, n. 5537, p. 2051–2055, 14 set. 2001.

MORIN, Alexandre J. S.; BUJACZ, Aleksandra; GAGNÉ, Marylène. Person-Centered Methodologies in the Organizational Sciences: Introduction to the Feature Topic. *Organizational Research Methods*, v. 21, n. 4, p. 803–813, 1 out. 2018.

MUGGLETON, Stephen. Alan Turing and the development of Artificial Intelligence. *AI Communications*, v. 27, n. 1, p. 3–10, 1 jan. 2014.

MUSTAFI, Zahid; BAYLE, Emmanuel; TERRIEN, Mickael. Non-Big five football clubs' strategies for generating transfer revenues: the case of Switzerland's Super League. *Soccer & Society*, v. 0, n. 0, p. 1–18, 2024.

NAKRANI, Sachin. Football Manager craze has matured into a full-grown national obsession. *The Guardian*, 21 ago. 2013. FootballDisponível em: <<https://www.theguardian.com/football/blog/2013/aug/21/football-manager-craze-lionel-messi>>. Acesso em: 28 maio 2025.

NAUCK, Detlef; KRUSE, Rudolf. Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine, Fuzzy Diagnosis*. v. 16, n. 2, p. 149–169, 1 jun. 1999.

NEALE, Walter C. The Peculiar Economics of Professional Sports\*. *The Quarterly Journal of Economics*, v. 78, n. 1, p. 1–14, 1 fev. 1964.

NEWELL, Karl M. Constraints on the development of coordination. *Motor development on children: Aspects of coordination and control*. [S.l: s.n.], 1986. .

NORMAN, Joel. Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, v. 25, n. 1, p. 73–96, fev. 2002.

NUNES, Felipe Alexandre de Souza Félix *et al.* A constituição da rivalidade polarizada entre torcedores de futebol: uma análise da rivalidade entre Atlético-MG e Cruzeiro. *ReMark - Revista Brasileira de Marketing*, v. 21, n. 5, p. 1677–1729, 2022.

NUNES, Felipe Alexandre de Souza Félix. *Comportamento de não consumo de produtos dos patrocinadores de Cruzeiro e Atlético-MG pelos torcedores do clube rival*. 2020a. 213 f. Dissertação de Mestrado – Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <<https://repositorio.ufmg.br/handle/1843/34231>>.

NUNES, Felipe Alexandre de Souza Félix. *Uma perspectiva configuracional sobre indicadores técnicos de desempenho do futebolista*. 2020b. 58 f. Monografia – Universidade Federal de Viçosa, Viçosa, 2020.

NUNES, Felipe Alexandre de Souza Félix; ANON, Iago Cambre; FREITAS, Jonathan Simões. O uso da lógica difusa na comparação de performance de atacantes do futebol no Brasil. In: CONGRESSO BRASILEIRO DE GESTÃO DO ESPORTE, 10 mar. 2021, Juiz de Fora. *Anais...* Juiz de Fora: [s.n.], 10 mar. 2021.

NUNES, Felipe Alexandre de Souza Félix; FLEURY, Fernando A.; GONÇALVES, Carlos Alberto. Negociações do mercado da bola: Identificando combinações de fatores necessários e suficientes para altos valores de transferência de atacantes. In: CONGRESSO BRASILEIRO DE GESTÃO DO ESPORTE, 25 nov. 2021, Novo Hamburgo. *Anais...* Novo Hamburgo: [s.n.], 25 nov. 2021.

NUNES, Felipe Alexandre de Souza Félix; GONÇALVES, Carlos Alberto; FREITAS, Jonathan Simões. A combinação de variáveis que leva a um melhor relacionamento entre torcedores e patrocinadores de times de futebol: uma análise de múltiplos métodos. *PODIUM Sport, Leisure and Tourism Review*, v. 13, n. 2, p. 277–299, 20 ago. 2024.

PAIXÃO, Paulo *et al.* How does match status affects the passing sequences of top-level European soccer teams? *International Journal of Performance Analysis in Sport*, v. 15, n. 1, p. 229–240, 1 mar. 2015.

PALAFIX, Damian. *Thinking poker through Game Theory*. 2016. 86 f. Dissertação de Mestrado – California State University, San Bernardino, 2016. Disponível em: <<https://scholarworks.lib.csusb.edu/etd/314>>.

PALMEIRA, Eduardo Silva. Inferência em sistemas fuzzy baseado em regras: variações e aplicabilidades. In: VI CONGRESSO BRASILEIRO DE SISTEMAS FUZZY (VI CBSF), nov. 2021, São José do Rio Preto. *Anais...* São José do Rio Preto: [s.n.], nov. 2021.

PAOLI, Próspero Brum. *Os estilos de futebol e os processos de seleção e detecção de talentos*. 2007. 187 f. Tese de Doutorado – Universidade Gama Filho, Rio de Janeiro, 2007.

PAPPALARDO, Luca *et al.* PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach. *ACM Transactions on Intelligent Systems and Technology*, v. 10, n. 5, p. 59:1-59:27, 12 set. 2019.

PAPPALARDO, Luca; CINTIA, Paolo. Quantifying the relation between performance and success in soccer. *Advances in Complex Systems*, v. 21, n. 03n04, p. 1750014, maio 2018.

PAREEK, Preksha; THAKKAR, Ankit. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Review*, v. 54, n. 3, p. 2259–2322, 1 mar. 2021.

PARIATH, Richard *et al.* Player Performance Prediction in Football Game. In: 2018 SECOND INTERNATIONAL CONFERENCE ON ELECTRONICS, COMMUNICATION AND AEROSPACE TECHNOLOGY (ICECA), mar. 2018, [S.l.: s.n.], mar. 2018. p. 1148–1153.

PASTORE, Luca. Third party ownership and multi-club ownership: where football is heading for. *Rivista di Diritto ed Economia dello Sport*, v. 14, n. 1, p. 23–58, 2018.

PATEL, Devansh; SHAH, Dhwanil; SHAH, Manan. The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. *Annals of Data Science*, v. 7, n. 1, p. 1–16, 1 mar. 2020.

PAULA, Ana Paula Paes De. Para além dos paradigmas nos Estudos Organizacionais: o Círculo das Matrizes Epistêmicas. *Cadernos EBAPE.BR*, p. 24 a 46–24 46, 7 mar. 2016.

PAYYAPPALLI, Vineet M.; ZHUANG, Jun. A data-driven integer programming model for soccer clubs' decision making on player transfers. *Environment Systems and Decisions*, v. 39, n. 4, p. 466–481, 1 dez. 2019.

PEPPARD, Joe; WARD, John. *The Strategic Management of Information Systems: Building a Digital Strategy*. [S.l.]: John Wiley & Sons, 2016.

PERARNAU, Martí. *Guardiola confidencial: Um ano dentro do Bayern de Munique acompanhando de perto o técnico que mudou o futebol para sempre*. [S.l.]: KasaFutebol Editora LTDA - Grande Área, 2015.

POLI, Raffaele; BESSON, Roger; RAVENEL, Loïc. Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies*, v. 10, n. 1, p. 4, jan. 2022.

POLI, Raffaele; RAVENEL, Loïc; BESSON, Roger. *Global study of football expatriates (2017-2023)*. , CIES Football Observatory Monthly Report., n° n°85. Neuchâtel, Switzerland: CIES Football Observatory, maio 2023. Disponível em: <<https://football-observatory.com/IMG/sites/mr/mr85/en/>>.

QCBS R WORKSHOP SERIES. *Workshop 9: Multivariate Analyses in R*. [S.l: s.n.], 2023. Disponível em: <<https://r.qcbs.ca/workshop09/book-en/index.html#expected-behaviour>>. Acesso em: 16 maio 2025.

RAGIN, Charles C. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. [S.l.]: University of Chicago Press, 2009.

RAJESWARI, A. M.; DEISY, C. Fuzzy logic based associative classifier for slow learners prediction. *Journal of Intelligent & Fuzzy Systems*, v. 36, n. 3, p. 2691–2704, 1 jan. 2019.

RATTEN, Vanessa. Introduction: Sport Technology and Innovation. In: RATTEN, VANESSA (Org.). *Sports Technology and Innovation: Assessing Cultural and Social Factors*. Cham: Springer International Publishing, 2019. p. 1–18. Disponível em: <[https://doi.org/10.1007/978-3-319-75046-0\\_1](https://doi.org/10.1007/978-3-319-75046-0_1)>. Acesso em: 8 nov. 2023.

RAVI, Daniele *et al.* Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In: 2016 IEEE 13TH INTERNATIONAL CONFERENCE ON WEARABLE AND IMPLANTABLE BODY SENSOR NETWORKS (BSN), jun. 2016, [S.l: s.n.], jun. 2016. p. 71–76. Disponível em: <<https://ieeexplore.ieee.org/document/7516235>>. Acesso em: 14 nov. 2023.

REEL, Parminder S. *et al.* Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances*, v. 49, p. 107739, 1 jul. 2021.

REIN, Robert; MEMMERT, Daniel. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, v. 5, n. 1, p. 1410, 24 ago. 2016.

REY-MARTÍ, Andrea *et al.* A comprehensive conceptual and bibliometric study of person-centered methodologies. *Quality & Quantity*, v. 56, n. 5, p. 3665–3683, 1 out. 2022.

ROBLEK, Vasja *et al.* Evolution of organisational agility: a bibliometric study. *Kybernetes*, v. 51, n. 13, p. 119–137, 1 jan. 2022.

RODRIGUES, Alexandre de Cássio *et al.* Efeitos da incorporação de julgamentos na avaliação da eficiência de clubes de futebol: uma abordagem por Data Envelopment Analysis. *Exacta*, v. 20, n. 1, p. 234–251, 21 jan. 2022.

ROMMERS, Nikki *et al.* A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine & Science in Sports & Exercise*, v. 52, n. 8, p. 1745–1751, ago. 2020.

- RØNNINGEN, Martin Holthe. *The genesis of data-driven decision-making in the world of soccer tactics : deciphering the potential of big data*. 2021. Master thesis – University of Agder, 2021. Disponível em: <<https://uia.brage.unit.no/uia-xmlui/handle/11250/2826027>>. Acesso em: 30 dez. 2022.
- ROSENBLUETH, Arturo; WIENER, Norbert. The Role of Models in Science. *Philosophy of Science*, v. 12, n. 4, p. 316–321, out. 1945.
- ROSSI, Alessio *et al.* Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*, v. 13, n. 7, p. e0201264, 25 jul. 2018.
- ROSSI, Alessio *et al.* Relationship between External and Internal Workloads in Elite Soccer Players: Comparison between Rate of Perceived Exertion and Training Load. *Applied Sciences*, v. 9, n. 23, p. 5174, jan. 2019.
- ROTTENBERG, Simon. The Baseball Players' Labor Market. *Journal of Political Economy*, v. 64, n. 3, p. 242–258, jun. 1956.
- RUBIO MARTÍN, Gracia *et al.* Measuring football clubs' human capital: analytical and dynamic models based on footballers' life cycles. *Journal of Intellectual Capital*, v. 23, n. 5, p. 1107–1137, 1 jan. 2022.
- RUMELT, Richard P. Good Strategy/Bad Strategy: The Difference and Why It Matters. *Strategic Direction*, v. 28, n. 8, 1 jan. 2012. Disponível em: <<https://doi.org/10.1108/sd.2012.05628haa.002>>. Acesso em: 27 fev. 2024.
- RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 4<sup>a</sup> ed. [S.l.]: Pearson, 2021. (Pearson series in artificial intelligence).
- RUSU, Adrian; STOICA, Doru; BURNS, Edward. Analyzing Soccer Goalkeeper Performance Using a Metaphor-Based Visualization. In: 2011 15TH INTERNATIONAL CONFERENCE ON INFORMATION VISUALISATION, jul. 2011, [S.l.: s.n.], jul. 2011. p. 194–199.
- ŞAHİN, Mehmet; EROL, Rızzvan. Prediction of Attendance Demand in European Football Games: Comparison of ANFIS, Fuzzy Logic, and ANN. *Computational Intelligence and Neuroscience*, v. 2018, n. 1, p. 5714872, 2018.
- SALAŁUBUN, Wojciech *et al.* A Fuzzy Inference System for Players Evaluation in Multi-Player Sports: The Football Study Case. *Symmetry*, v. 12, n. 12, p. 2029, dez. 2020.
- SALGADO-BARANDELA, J.; BARAJAS, A.; SÁNCHEZ-FERNÁNDEZ, P. Impacto económico del deporte: tema de interés creciente en la literatura científica / Economic Impact

- of Sport: Topic of Growing Interest for the Scientific Literature. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte*, v. 68, n. 2017, 2017.
- SAMUEL, Arthur L. Machine learning. *The Technology Review*, v. 62, n. 1, p. 42–45, 1959.
- SÁNCHEZ, Luis Carlos; BARAJAS, Angel; SANCHEZ-FERNANDEZ, Patricio. Fans in the ownership of Big Five leagues: lessons for better football governance. *Soccer & Society*, v. 22, n. 4, p. 355–371, 19 maio 2021.
- SANTOS, José Manuel Sánchez; GARCÍA, Pablo Castellanos. A Bibliometric Analysis of Sport Economics Research. *International Journal of Sport Finance*, v. 6, n. 3, p. 222–244, 2011.
- SANTOS, Pedro Miguel Ribeiro. *O modus operandi de um Departamento de Scouting de Futebol Estágio Profissionalizante realizado na Futebol Clube do Porto - Futebol, SAD*. 2012. 265 f. Relatório de Estágio – Universidade do Porto, Porto, 2012.
- SARMENTO, Hugo *et al.* Match analysis in football: a systematic review. *Journal of Sports Sciences*, v. 32, n. 20, p. 1831–1843, 14 dez. 2014.
- SARMENTO, Hugo *et al.* Talent Identification and Development in Male Football: A Systematic Review. *Sports Medicine*, v. 48, n. 4, p. 907–931, 1 abr. 2018.
- SARMENTO, Hugo; ARAÚJO, Duarte. Readiness for career affordances in high-level football: Two case studies in Portugal. *High Ability Studies*, v. 32, n. 1, p. 89–103, 2 jan. 2021.
- SAUNDERS, Rob; BUCKMAN, Joshua E. J.; PILLING, Stephen. Latent variable mixture modelling and individual treatment prediction. *Behaviour Research and Therapy*, v. 124, p. 103505, 1 jan. 2020.
- SCHNEIDER, Carsten Q.; ROHLFING, Ingo. Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research. *Sociological Methods & Research*, v. 42, n. 4, p. 559–597, 1 nov. 2013.
- SETIONO, Rudy; LIU, Huan. Understanding neural networks via rule extraction. IJCAI'95, 20 ago. 1995, San Francisco, CA, USA. *Anais...* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 20 ago. 1995. p. 480–485. . Acesso em: 3 abr. 2025.
- SHERWIN, Ian; CAMPBELL, Mark J.; MACINTYRE, Tadhg Eoghan. Talent development of high performance coaches in team sports in Ireland. *European Journal of Sport Science*, v. 17, n. 3, p. 271–278, 16 mar. 2017.
- SIMON, Herbert Alexander. *Sciences of the Artificial*. Cambridge, Mass: Mit Press, 1996.

SKUPIN, André. Discrete and continuous conceptualizations of science: Implications for knowledge domain visualization. *Journal of Informetrics, Science of Science: Conceptualizations and Models of Science*. v. 3, n. 3, p. 233–245, 1 jul. 2009.

SONG, Wenzhu *et al.* Using random forest algorithm for glomerular and tubular injury diagnosis. *Frontiers in Medicine*, v. 9, 28 jul. 2022. Disponível em: <<https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.911737/full>>. Acesso em: 15 maio 2025.

STUART, Keith. Why clubs are using Football Manager as a real-life scouting tool. *The Guardian*, 12 ago. 2014. GamesDisponível em: <<https://www.theguardian.com/technology/2014/aug/12/why-clubs-football-manager-scouting-tool>>. Acesso em: 28 maio 2025.

SUMPTER, David. *Luck, skill and randomness. Soccermaths*. [S.l.: s.n.]. Disponível em: <<https://soccermaths.medium.com/luck-skill-and-randomness-aaa8d80836ea>>. Acesso em: 27 dez. 2022. , 8 jan. 2017

SUMPTER, David. *Soccermaths: Mathematical Adventures in the Beautiful Game*. Reprint ed. [S.l.]: Bloomsbury Publishing, 2016.

SZYMANSKI, Stefan. Entry into exit: insolvency in English professional football. *Scottish Journal of Political Economy*, v. 64, n. 4, p. 419–444, set. 2017.

TANG, Xiaoan *et al.* Derivation of personalized numerical scales from distribution linguistic preference relations: an expected consistency-based goal programming approach. *Neural Computing and Applications*, v. 31, n. 12, p. 8769–8786, 1 dez. 2019.

TAVANA, Madjid *et al.* A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, v. 16, n. 1, p. 97–110, 1 fev. 2013.

TAVARES, Fernando Flores; BETTI, Gianni. Gender Differences in Multidimensional Poverty in Brazil: A Fuzzy Approach. *Social Indicators Research*, 19 fev. 2024. Disponível em: <<https://doi.org/10.1007/s11205-024-03312-z>>. Acesso em: 28 abr. 2024.

TEECE, David J. A capability theory of the firm: an economics and (Strategic) management perspective. *New Zealand Economic Papers*, v. 53, n. 1, p. 1–43, 2 jan. 2019.

TEECE, David J.; PISANO, Gary; SHUEN, Amy. Dynamic capabilities and strategic management. *Strategic Management Journal*, v. 18, n. 7, p. 509–533, 1997.

- TIBSHIRANI, Robert. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 58, n. 1, p. 267–288, 1996.
- TRAVASSOS, Bruno *et al.* Performance analysis in team sports: Advances from an Ecological Dynamics approach. *International Journal of Performance Analysis in Sport*, v. 13, n. 1, p. 83–95, abr. 2013.
- TURING, Alan Mathison. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, v. LIX, n. 236, p. 433–460, 1 out. 1950.
- UPHAM, S. Phineas; SMALL, Henry. Emerging research fronts in science and technology: patterns of new knowledge development. *Scientometrics*, v. 83, n. 1, p. 15–38, 1 abr. 2010.
- VAN ECK, Nees Jan; WALTMAN, Ludo. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, v. 84, n. 2, p. 523–538, 1 ago. 2010.
- VAUGHAN, James *et al.* Developing Creativity to Enhance Human Potential in Sport: A Wicked Transdisciplinary Challenge. *Frontiers in Psychology*, v. 10, 2019. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02090>>. Acesso em: 10 out. 2023.
- VERMA, Vijay; BETTI, Gianni; GAGLIARDI, Francesca. Fuzzy Measures of Longitudinal Poverty in a Comparative Perspective. *Social Indicators Research*, v. 130, n. 2, p. 435–454, 1 jan. 2017.
- WAGNER, Fabio; PREUSS, Holger; KÖNECKE, Thomas. A Central Element of Europe’s Football Ecosystem: Competitive Intensity in the “Big Five”. *Sustainability*, v. 13, n. 6, p. 3097, jan. 2021.
- WAKELAM, Edward; STEUBER, Volker; WAKELAM, James. The collection, analysis and exploitation of footballer attributes: A systematic review. *Journal of Sports Analytics*, v. 8, n. 1, p. 31–67, 1 jan. 2022.
- WALDRON, Mark; WORSFOLD, Paul. Differences in the Game Specific Skills of Elite and Sub-Elite Youth Football Players: Implications for Talent Identification. *International Journal of Performance Analysis in Sport*, v. 10, n. 1, p. 9–24, 1 abr. 2010.
- WAND, Tobias. Analysis of the Football Transfer Market Network. *Journal of Statistical Physics*, v. 187, n. 3, p. 27, 19 abr. 2022.
- WARING, Jonathan; LINDVALL, Charlotta; UMETON, Renato. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial Intelligence in Medicine*, v. 104, p. 101822, 1 abr. 2020.

WASSERMAN, Stanley; FAUST, Katherine. *Social Network Analysis: Methods and Applications*. [S.l.]: Cambridge University Press, 1994.

WIEMEYER, Josef. Who should play in which position in soccer? Empirical evidence and unconventional modelling. *International Journal of Performance Analysis in Sport*, v. 3, n. 1, p. 1–18, 1 abr. 2003.

WILLIAMS, A. M.; WARD, P. Searching for the Holy Grail: Can there ever be such a thing as a “Grand Unified Theory of sports performance”? *Human Movement Science*, v. 56, n. Pt A, p. 181–183, 2017.

WILLIAMS, Jason J. An investigation into operational definitions used within performance analysis. In: WORKSHOP IN PERFORMANCE ANALYSIS, jan. 2009, Lincoln. *Anais...* Lincoln: [s.n.], jan. 2009.

WOO, Sang Eun *et al.* Putting the “Person” in the Center: Review and Synthesis of Person-Centered Approaches and Methods in Organizational Science. *Organizational Research Methods*, v. 21, n. 4, p. 814–845, 1 out. 2018.

WOODS, Carl T.; DAVIDS, Keith. Sport scientists in-becoming: from fulfilling one’s potential to finding our way along. *Sport, Education and Society*, p. 1–15, 5 jan. 2023.

WUNDERLICH, Fabian; SECK, Alessandro; MEMMERT, Daniel. The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*, v. 39, n. 20, p. 2322–2337, 18 out. 2021.

YAMABE, Tetsuo; NAKAJIMA, Tatsuo. Playful training with augmented reality games: case studies towards reality-oriented system design. *Multimedia Tools and Applications*, v. 62, n. 1, p. 259–286, 1 jan. 2013.

YANG, Luyao; AMIN, Osama; SHIHADA, Basem. Intelligent Wearable Systems: Opportunities and Challenges in Health and Sports. *ACM Comput. Surv.*, v. 56, n. 7, p. 190:1-190:42, 9 abr. 2024.

YAZDIPOUR, Rassoul; CONSTAND, Richard. Predicting Firm Failure: A Behavioral Finance Perspective. *The Journal of Entrepreneurial Finance*, v. 14, n. 3, p. 90–104, 1 dez. 2010.

YI, Qing *et al.* Differences in Technical Performance of Players From ‘The Big Five’ European Football Leagues in the UEFA Champions League. *Frontiers in Psychology*, v. 10, 2019.

Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02738>>. Acesso em: 13 set. 2023.

YI, Qing *et al.* Technical demands of different playing positions in the UEFA Champions League. *International Journal of Performance Analysis in Sport*, v. 18, n. 6, p. 926–937, 2 nov. 2018.

ZADEH, L. A. Fuzzy sets. *Information and Control*, v. 8, n. 3, p. 338–353, 1 jun. 1965.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, v. 8, n. 3, p. 199–249, 1 jan. 1975.

ZAMBOM-FERRARESI, Fabíola; RIOS, Vicente; LERA-LÓPEZ, Fernando. Determinants of sport performance in European football: What can we learn from the data? *Decision Support Systems*, v. 114, p. 18–28, 1 out. 2018.

ZHANG, Pin. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. *Applied Soft Computing*, v. 85, p. 105859, 1 dez. 2019.

ZHANG, Zixuan *et al.* Deep learning-enabled triboelectric smart socks for IoT-based gait analysis and VR applications. *npj Flexible Electronics*, v. 4, n. 1, p. 1–12, 26 out. 2020.

ZHOU, Changjing *et al.* Chinese soccer association super league, 2012–2017: key performance indicators in balance games. *International Journal of Performance Analysis in Sport*, v. 18, n. 4, p. 645–656, 4 jul. 2018.

ZHOU, Yuankai *et al.* Triboelectric nanogenerator based self-powered sensor for artificial intelligence. *Nano Energy*, v. 84, p. 105887, 1 jun. 2021.

ZHOU, Zhi-Hua; WU, Jianxin; TANG, Wei. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, v. 137, n. 1, p. 239–263, 1 maio 2002.

ZHU, Minglu *et al.* Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications. *Science Advances*, v. 6, n. 19, p. eaaz8693, 8 maio 2020.

ZOU, Hui; HASTIE, Trevor. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, v. 67, n. 2, p. 301–320, 1 abr. 2005.