Universidade do Minho
Escola de Economia, Gestão e Ciência Política

João Pedro Gomes Gonçalves

Tomada de decisão na contratação de jogadores de futebol utilizando multicritério

julho de 2026

Tomada de decisão na contratação de jogadores de futebol utilizando multicritério

João Pedro Gomes Gonçalves

UMinho | 2026

Universidade do Minho
Escola de Economia, Gestão e Ciência Política

João Pedro Gomes Gonçalves

Tomada de decisão na contratação de jogadores de futebol utilizando multicritério

Dissertação de Mestrado
Mestrado em Gestão
Especialização em Gestão Geral

Trabalho efetuado sob a orientação do
Professor Doutor José António Almeida Crispim

julho de 2026

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Esta é uma dissertação que pode ser utilizada por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites no que concerne aos direitos de autor e direitos conexos.

Assim, a presente investigação pode ser utilizada nos termos previstos da licença abaixo indicada.

Caso o utilizados necessite de permissão para poder fazer uso desta investigação em condições não previstas do licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-SemDerivações
CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

AGRADECIMENTOS

Antes de encerrar esta etapa, não poderia de deixar de agradecer a todos aqueles que de diferentes formas contribuíram para a concretização desta dissertação de gestão.

Ao meu orientador, Professor Doutor José Crispim, agradeço por toda a orientação, disponibilidade e acompanhamento ao longo de todo este processo. A sua exigência, aliada ao apoio constante, foi fundamental para a realização desta investigação.

Aos meus amigos, pelo apoio e incentivo ao longo deste percurso e em especial ao Ruben, pela disponibilidade permanente para ajudar e a esclarecer as muitas dúvidas, mesmo nos momentos mais exigentes.

À minha família, pelo apoio incondicional e pela confiança que sempre depositaram em mim, em particular à minha mãe e à minha irmã por estarem sempre presentes e por serem um pilar ao longo deste percurso.

À EY, pelo apoio demonstrado durante a realização desta dissertação que sempre mostraram compreensão e flexibilidade, que permitiu conciliar as exigências profissionais com o desenvolvimento desta investigação.

Por fim, a mim próprio pela resiliência e dedicação ao longo desta etapa, mesmo perante os desafios encontrados e momentos de maior exigência.

A todos, o meu sincero obrigado.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio, nem a qualquer forma indevida ou falsificação de informações ou resultados em nenhuma etapa na realização desta investigação.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

DECLARAÇÃO DE USO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri a qualquer forma de utilização indevida de ferramentas de Inteligência Artificial, nem à falsificação ou omissão de informações, fontes, dados, resultados ou processos em nenhuma das etapas na realização desta investigação.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

RESUMO EXECUTIVO

A presente investigação tem como objetivo desenvolver um modelo de apoio à tomada de decisão, em gestão desportiva num contexto para a contratação de jogadores de futebol profissional, com base numa abordagem multicritério. Neste contexto, é proposta uma abordagem integrada que combina métodos estatísticos tradicionais com técnicas de Machine Learning, utilizando uma base de dados bastante extensa de dados do FIFA 23. Num contexto caracterizado por uma elevada incerteza e grande impacto financeiro, pretende-se estruturar a avaliação dos jogadores de forma mais objetiva e mais fundamentada, considerando as especificidades para cada posição em campo.

Numa primeira fase, foi realizada uma análise exploratória e estatística, incluindo modelos de regressão linear múltipla com o intuito de identificar relações entre os critérios de desempenho dos jogadores para cada posição. Posteriormente, foi aplicada uma abordagem de Machine Learning utilizando técnicas de Random Forest e Redes Neuronais Artificiais de forma a estimar o potencial dos jogadores e o seu valor de mercado, permitindo capturar relações não lineares e aumentar a capacidade preditiva dos modelos.

De forma a garantir um maior rigor metodológico, os critérios analisados foram previamente selecionados com base na sua relevância média por posição, o que permitiu que apenas as variáveis consideradas significativas fossem incluídas, de modo a não acrescentar ruído na modelação. Esta abordagem permitiu focar a análise na importância relativa dos critérios dentro de cada grupo posicional aqui estudado, Guarda Redes, Defesas, Médios e Avançados.

Adicionalmente, utilizou-se técnicas de clustering para identificar diferentes padrões com base nos seus critérios, possibilitando uma segmentação complementar. De forma a aumentar a interpretação dos resultados obtidos pelas técnicas de RF e RNA, foi utilizada a metodologia de SHAP, permitindo avaliar o contributo individual dos diferentes critérios e validar a consistência dos resultados obtidos.

Ainda sobre estes dois modelos, observou-se uma elevada consistência nos resultados estimados, reforçando a robustez dos resultados obtidos. A integração de SHAP permitiu não só validar os resultados, mas também visualizar e identificar padrões menos evidentes, demonstrando a importância de uma abordagem multicritério na avaliação de jogadores.

Do ponto de vista prático, o modelo desenvolvido constitui uma ferramenta de apoio à decisão no contexto de avaliação dos jogadores, contribuindo para uma avaliação mais estruturada e fundamentada. Do ponto de vista académico, esta investigação destaca-se pela integração de técnicas

de Machine Learning, constituindo um contributo relevante para a literatura na área da gestão desportiva.

Por fim, pode-se concluir que o desempenho dos jogadores de futebol não depende de um único fator, mas sim da interação de diversos critérios, sendo fundamental a adoção de abordagens multicritério para obter uma decisão mais fundamentada na decisão.

Palavras-chave: Tomada de decisão, Contratação de Jogadores, Machine Learning, Análise Multicritério, Futebol Profissional.

ABSTRACT

The objective of this research is to develop a decision-support model for sports management in the context of signing professional soccer players, based on a multi-criteria strategy. In this context, an integrated approach that combines traditional statistical methods with machine learning techniques is proposed, utilizing a comprehensive database obtained from FIFA 23. In an environment described by high uncertainty where every decision has a significant financial impact, the goal is to structure player evaluations in a more objective and well-founded manner, taking into account the specific characteristics of each position on the field.

In the first phase, an exploratory and statistical analysis was conducted – including multiple linear regression models – with the aim of identifying relationships among player performance criteria for each position. Subsequently, a machine learning approach was applied using Random Forest and Artificial Neural Network techniques to estimate players' potential and market value, enabling the capture of nonlinear relationships and enhancing the models' predictive power.

To ensure greater methodological rigor, the analysed criteria were first selected based on their average relevance by position, which allowed only the variables deemed significant to be included, so as not to introduce noise into the modelling. This procedure let the research focus on the relative importance of the criteria within each positional group studied here: Goalkeepers, Defenders, Midfielders, and Forwards.

Additionally, clustering techniques were used to identify different patterns based on these criteria, enabling further segmentation. To enhance the interpretation of the results obtained through RF and RNA techniques, the SHAP methodology was employed, allowing for the evaluation of the individual contribution of the different criteria and the validation of the consistency of the results obtained.

Regarding these two models, a high degree of consistency was observed in the estimated results, reinforcing the robustness of the findings. The integration of SHAP allowed not only for the validation of the results but also enabled the visualization and identification of less obvious patterns, demonstrating the importance of a multi-criteria approach in player evaluation.

From a practical standpoint, the developed model serves as a decision-support tool in the context of player evaluation, contributing to a more structured and well-founded assessment. From an academic

standpoint, this research stands out for its integration of machine learning techniques, constituting a significant contribution to the literature in the field of sports management.

Finally, it can be concluded that the performance of soccer players does not depend on a single factor, but rather on the interaction of various indicators, making the adoption of multi-criteria approaches essential for reaching a more well-founded decision.

Keywords: Decision Making, Player Recruitment, Machine Learning, Multi-Criteria Decision Analysis, Professional Football

ÍNDICE

AGRADECIMENTOS ..... *i*
DECLARAÇÃO DE INTEGRIDADE ..... *ii*
DECLARAÇÃO DE USO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL ..... *ii*
RESUMO EXECUTIVO ..... *iii*
ÍNDICE DE FIGURAS ..... *ix*
ÍNDICE DE TABELAS ..... *x*
LISTA DE ABREVIATURAS ..... *xi*
1. INTRODUÇÃO *1*
1.1 Tópico/Problema 1
1.2 Conceitos principais 2
1.3 Objetivo do estudo 3
1.4 Pertinência do estudo 3
2. REVISÃO DA LITERATURA *5*
3. METODOLOGIA *10*
3.1 Enquadramento metodológico 10
3.2 Caracterização da Base de Dados 11
3.3 Construção da árvore de critérios 12
3.4 Preparação dos dados 15
3.5 Análise Empírica de Regressão Múltipla 16
3.6 Análise com Clustering e Machine Learning 19
3.6.1 Agrupamento dos Jogadores (Método Clustering) 22
3.6.2 Estimativa do potencial do jogador através de Random Forest 25
3.6.3 Estimativa do potencial do jogador através de Redes Neuronais Artificiais 27
3.6.4 Estimativa do Valor de Mercado através de Redes Neuronais Artificiais 29
3.6.5 Avaliação da Qualidade do Modelo 31
3.6.6 Interpretação do Overhall dos modelos com SHAP 33
3.6.7 Avaliação da Qualidade da Previsão do Valor de Mercado 38
4 DISCUSSÃO E RESULTADOS *40*
4.1 Importância dos critérios na avaliação dos jogadores 40
4.2 Comparação com os dados recolhidos 40
4.3 Identificação de perfis de jogadores através de clustering 45
4.4 Estimativa do potencial dos jogadores através de ML 46
4.5 Estimativa do valor de mercado dos jogadores 46
4.6 Implicações para a tomada de decisão na contratação de jogadores 47
5 CONCLUSÕES, LIMITAÇÕES E INVESTIGAÇÃO FUTURA *48*
5.1 Conclusões 48
5.2 Limitações do Estudo 49
5.3 Investigação Futura 50

6 REFERÊNCIAS.....52
7 APÊNDICES.....55

ÍNDICE DE FIGURAS

Figura 1 - Árvore de Critérios 14
Figura 2 - Pipeline do Processo do Machine Learning 21
Figura 3 - Visualização dos Clusters através de PCA 23
Figura 4 - Clusters por Posição Projetados em PCA 24
Figura 5 – Três Clusters de Desempenho dos Jogadores em PCA 25
Figura 6 - SHAP Guarda Redes Random Forest 33
Figura 7 - SHAP Guarda Redes Redes Neuronais Artificiais 34
Figura 8 - SHAP Defesas Random Forest 34
Figura 9 - SHAP Defesas Redes Neuronais Artificiais 35
Figura 10 - SHAP Médios Random Forest 35
Figura 11 - SHAP Médios Redes Neuronais Artificiais 36
Figura 12 - SHAP Avançados Random Forest 36
Figura 13 - SHAP Avançados Redes Neuronais Artificiais 37

ÍNDICE DE TABELAS

Tabela 1 – Critérios de desempenho por categoria 17
Tabela 2 – Estimativas de valor de mercado por regressão 18
Tabela 3 – Classificação do Potencial dos Jogadores com RF 27
Tabela 4 - Classificação do Potencial dos Jogadores com RNA 28
Tabela 5 - Estimativas de valor de mercado por RNA 30
Tabela 6 – Valores das Métricas de Clustering, RF e RNA 32
Tabela 7 - Métricas de avaliação do modelo (MAE e RMSE) 38
Tabela 8 - Comparação entre critérios estimados e critérios identificados pelos profissionais (GR) 41
Tabela 9 - Comparação entre critérios estimados e critérios identificados pelos profissionais (DEF) 42
Tabela 10 - Comparação entre critérios estimados e critérios identificados pelos profissionais (MC) 43
Tabela 11 - Comparação entre critérios estimados e critérios identificados pelos profissionais (AVA) 44

LISTA DE ABREVIATURAS

AHP: Analytic Hierarchy Process

ANP: Analytic Network Process

DEA: Data Envelopment Analysis

ELECTRE: ELimination Et Choix Traduisant la REalité

MAE: Mean Absolute Error

MCDA: Multi Criteria Decision Analysis

ML: Machine Learning

PCA: Principal Component Analysis

PROMETHEE: Preference Ranking Organization METhod for Enrichment Evaluation

ReLU: Rectified Linear Unit

RF: Random Forest

RMSE: Root Mean Squared Error

RNA: Redes Neuronais Artificiais

SHAP: SHapley Additive exPlanations

TOPSIS: Technique for Order of Preference by Similarity to Ideal Solution

SWARA: Step-Wise Weight Assessment Ratio Analysis

VIKOR: VIse Kriterijumska Optimizacija I Kompromisno Resenje

1. INTRODUÇÃO

1.1 Tópico/Problema

Nos últimos anos, o mercado europeu tem se deparado com valores de transferências que influenciam diretamente a situação financeira do clube, com valores que ultrapassam dezenas ou centenas de milhões de euros. Este crescimento no valor das transferências reforça a importância de haver decisões estratégicas na contratação de jogadores de futebol.

Nesse contexto, a contratação de jogadores representa uma das decisões mais estratégicas para os clubes, tendo estes um impacto direto no seu desempenho desportivo e na sua sustentabilidade financeira. Num contexto em que os valores de transferências atingem montantes com valores cada vez mais elevados, a escolha errada de um jogador pode levar a perdas significativas, tanto a nível económico como desportivo.

A decisão da contratação caracteriza-se por um grande grau de complexidade, que envolve múltiplos critérios, tais como critérios técnicos, psicológicos, mentais e físicos, mas também fatores externos como a liga onde estes atuam, o clube de origem ou a visibilidade mediática do jogador (Mourao, 2016). Para além disto, trata-se de um processo envolto em incerteza, uma vez que o desempenho futuro do jogador não pode ser previsto com total precisão devido a haver variáveis que não podemos mensurar.

Segundo Di Minin et al. (2014), clubes como o Udinese, alteraram a sua estratégia, passando a adotar uma abordagem mais racional e sustentada na contratação dos seus ativos, de modo a garantir estabilidade financeira e valorizar os seus ativos. Contudo, este processo de decisão costuma ser influenciado pela intuição por parte de treinadores e dirigentes do clube, o que pode comprometer a eficácia (Collins et al., 2016; Giske et al., 2013). A literatura mais recente, tem sugerido que existe uma necessidade de recorrer a metodologias científicas, de modo a minimizar o risco e assim aumentar a probabilidade de sucesso na escolha dos seus jogadores (Çaloğlu Büyükselçuk & Badem, 2024). Neste sentido, esta investigação propõe criar um modelo de apoio multicritério (MCDA), podendo contribuir para uma gestão desportiva mais eficaz e fundamentada do processo de contratação destes ativos.

1.2 Conceitos principais

A **tomada de decisão** na gestão desportiva constitui um processo bastante complexo, que envolve abordagens racionais e de intuição. De acordo com Giske et al. (2013), existe uma dinâmica entre os estilos racionais e intuitivos, sendo que o último tende a ter um comportamento mais predominante nos mais altos níveis de futebol. Segundo Collins et al. (2016), este observou que os treinadores utilizam primeiramente o estilo intuitivo, sendo posteriormente utilizado processos deliberativos, de modo a aumentar a probabilidade de sucesso nas suas escolhas.

Investigações mais recentes, têm evidenciado a importância do uso de metodologias cientificas para a tomada de decisão, nomeadamente modelos de apoio multicritério, onde estes fornecem uma estrutura mais racional, mais fundamentada para estas decisões, reduzindo a subjetividade associada à intuição, proporcionando decisões mais consistentes e fundamentadas (Çaloğlu Büyükselçuk & Badem, 2024).

O **valor do jogador** de futebol representa uma estimativa de quanto um clube estaria disposto a pagar pela sua contratação, refletindo o seu desempenho atual (valor presente) e o seu potencial que pode vir a atingir no futuro. De acordo com Kirschstein & Liebscher (2019), quando estudou os jogadores da Bundesliga (Campeonato Alemão), confirmou a relevância de diferentes critérios, tais como, o desempenho (golos, assistências, etc.), características pessoais (idade, altura, posição de jogo) e também fatores contextuais (liga, clube). Um exemplo referido pelo autor, destacou que a idade é um dos critérios que valoriza o valor de um jogador, ou que desvaloriza jogadores já em fase final de carreira.

Jogadores na linha da frente, principalmente avançados, tendem a ser valorizados pelos seus golos e assistências oferecidas, enquanto os jogadores mais defensivos são mais valorizados pelas suas ações defensivas. Serna Rodríguez et al. (2019), evidencia que os critérios variam de posição para posição, sendo que o valor do jogador resulta de múltiplos fatores, sendo eles técnicos, físicos e contextuais.

Por fim, é importante destacar que o valor de mercado de um jogador não depende exclusivamente das suas características individuais. Mourao (2016) analisou a forma como os clubes das principais ligas europeias dirigem as transferências e concluiu que o valor do jogador está também ligado à eficiência da equipa ao longo do seu tempo, ou seja, o valor do jogador não depende apenas das suas características, mas também a forma como o clube gere os seus jogadores, de modo a atingir bons resultados desportivos.

1.3 Objetivo do estudo

Esta investigação tem como objetivo desenvolver um modelo multicritério de apoio à decisão que integra métodos estatísticos e de Machine Learning com vista à melhoria da eficácia no processo de contratações de jogadores de futebol.

Este modelo tem como finalidade apoiar o processo de tomada de decisão dos gestores desportivos na hora de contratar jogadores para o seu clube, permitindo uma avaliação mais estruturada, objetiva e fundamentada, com base em múltiplos critérios relevantes para o contexto do futebol.

Adicionalmente, pretende-se que este modelo seja capaz de identificar os seus valores de mercado e o potencial que um jogador pode ter, contribuindo para uma maior eficácia das decisões de contratação e a diminuição do risco associado.

1.4 Pertinência do estudo

Com a crescente relevância económica e mediática do futebol, os clubes enfrentam cada vez mais desafios e mais exigentes na gestão dos seus recursos humanos, nomeadamente a contratação dos mesmos. O aumento significativo que tem acontecido nos últimos anos reforça a necessidade de decisões com mais estratégia, uma vez que o erro de avaliação na contratação destes jogadores pode impactar negativamente o seu nível financeiro e desportivo. O uso de modelos de tomada de decisão, serve para contribuir/auxiliar para contratações mais eficazes, minimizando assim a probabilidade de contratações falhadas e assim minimizar esses riscos.

A pertinência deste estudo assenta em três pilares fundamentais, sendo que o primeiro, o mais importante, segue a importância estratégica nas contratações dos principais ativos do clube, de modo a levar ao sucesso desportivo e financeiro. Em segundo, podemos expor a lacuna existente da literatura do uso de MCDA na gestão desportiva. Por último, podemos contribuir na prática com um modelo de apoio estruturado para as futuras decisões de contratações de jogadores de futebol, potenciando assim a eficácia dos investimentos destes ativos.

Adicionalmente, este estudo contribui para a literatura ao integrar abordagens multicritério com técnicas de ML, aplicadas no contexto do futebol, uma área ainda pouco explorada, especialmente no contexto português.

Esta dissertação encontra-se estruturada na seguinte forma. No capítulo 2, é apresentada a revisão da literatura com os temas mais relevantes para esta investigação. No capítulo 3 descreve-se a metodologia adotada. No capítulo 4 são apresentada e discutidos os resultados obtidos. No capítulo 5 são apresentadas as principais conclusões, limitações que esta investigação encontrou e sugestões para uma investigação futura.

2. REVISÃO DA LITERATURA

A presente revisão de literatura tem como objetivo estruturar o enquadramento teórico da investigação presente na literatura focando-se em quatro dimensões.

Em primeiro lugar, é analisada a contratação dos jogadores como um problema na gestão desportiva, caracterizado por elevados níveis de incerteza e altos impactos financeiros. Em segundo lugar, é explorado o processo de tomada de decisão, nomeadamente abordagens intuitivas e racionais. Em terceiro lugar exploraram-se como métodos multicritério aplicados à seleção de jogadores enquanto ferramentas de apoio à decisão em contextos complexos. Por fim, apresenta-se as limitações destes modelos, bem como o potencial contributo de abordagens em Machine Learning para ultrapassar algumas restrições existentes.

Neste contexto, a literatura evidencia que a escolha de contratar e renovar jogadores de futebol, é um dos principais desafios da gestão desportiva, o que exige um equilíbrio entre o desempenho desportivo e a sustentabilidade financeira do clube. Nos últimos anos, tem-se observado um crescente número de estudos que abordam as melhores práticas de gestão, onde são explorados diferentes métodos de apoio à tomada de decisão, de modo a obter uma maior probabilidade de sucesso na contratação do jogador. Um exemplo relevante é o caso do clube italiano Udinese, um exemplo de sucesso na identificação e valorização dos seus ativos (Di Minin et al., 2014), onde o clube tem adotado nos últimos tempos um modelo de investimento sustentado em jogadores com enorme potencial, formá-los e depois garantir uma grande venda, garantindo a sua estabilidade financeira. Contudo, existe um trade off nesta estratégia, em que para garantir a sustentabilidade financeira, abdica no seu desempenho desportivo (Di Minin et al., 2014; Sánchez et al., 2020).

Estudos sobre a tomada de decisão no futebol evidenciam uma complexa dinâmica entre abordagens analíticas e intuitivas. De acordo com Giske et al. (2013), nos níveis mais elevados no futebol, a intuição desempenha o papel central, sendo fortemente influenciada pela experiência dos treinadores. Tanto a criatividade como a intuição, contribui para a criação de opções, sendo que a intuitiva costuma ser a mais assertiva (Klatt et al., 2019).

Porém, os treinadores, em certas situações optam por decisões mais conservadoras, possivelmente para reduzir a incerteza. Em diferentes modalidades, Collins et al. (2016) questionou diversos treinadores, que referiram que utilizam como principal, o estilo intuitivo e posteriormente o estilo deliberativo. A sua experiência enquanto profissional no desporto, é considerado fundamental para o

desenvolvimento da "habilidade de tomada de decisão intuitiva" (Collins et al., 2016; Giske et al., 2013).

Estes estudos sugerem que, embora a intuição seja o principal fator para a tomada de decisão, esta é frequentemente complementada por outros processos para obter uma maior probabilidade de sucesso nas suas escolhas.

Perante esta complexidade, diversos autores têm defendido a utilização de metodologias científicas, de modo a minimizar a probabilidade de insucesso na contratação de jogadores. Os modelos multicritério têm vindo a ganhar algum destaque, que permite integrar múltiplos critérios na avaliação de jogadores. Métodos como SWARA e VIKOR, são utilizados para encontrar soluções de equilíbrio na seleção de jogadores. Um estudo recente, demonstrou que a utilização deste método, demostrou uma eficácia de 60% em relação ao onze escolhido pelo treinador (Çaloğlu Büyükselçuk & Badem, 2024).

Outro método também bastante estudado é o AHP, onde este ordena os atributos, segundo aqueles que os autores acham mais importantes para aquelas diferentes posições dos jogadores, de modo a permitir uma avaliação mais estruturada. Modelos matemáticos, em combinação com o método AHP, têm sido desenvolvidos para melhorar a composição das equipas, utilizando diversos critérios que mais beneficiem o clube (Ozceylan, 2016). Esta prevalência do AHP e das suas variantes é reforçada por uma análise bibliométrica recente sobre a aplicação de métodos multicritério no futebol que com base em 26 artigos indexados na Web of Science, entre o período de 2000 a 2025, comprova que o TOPSIS e o AHP são as técnicas mais frequentes utilizadas neste domínio, refletindo uma preferência por abordagens híbridas e baseadas em contexto de incerteza. Os mesmos autores identificam a seleção de jogadores, a avaliação de desempenho e o planeamento estratégico como os problemas de decisão mais recorrentes na literatura, onde se verifica um crescente número de publicações a partir de 2022 (Belhouchet et al., 2026).

Segundo Belhouchet et al. (2026), o trabalho de Qader et al. (2017) sobre uma metodologia de seleção de jogadores de futebol baseada em múltiplos critérios consta como a publicação mais citada nesta área, tendo estabelecido fundações metodológicas que continuam a moldar a investigação mais recente.

Apesar de todas estas vantagens, os modelos multicritério, também apresentam as suas limitações e desafios. Embora forneça uma ajuda para a avaliação da escolha de jogadores com os diversos critérios mais desejados e os seus respetivos dados estatísticos, a complexidade de todos os fatores, seja diretos ou indiretos, dificulta a criação de um modelo 100% correto (Nasiri et al., 2019).

Adicionalmente, limitações financeiras existentes num clube também são uma condicionante a ser exposta e que condiciona o clube (Nasiri et al., 2019).

Neste contexto, investigações mais recentes, sugerem que a adoção de técnicas de Machine Learning neste método de seleção de modelos de multicritério, pode aprimorar este método e torná-lo mais preciso, melhorando a capacidade preditiva e reduzindo a sua subjetividade (Ati et al., 2024). Numa revisão sistemática mais alargada sobre esta interseção, baseada na análise de 66 estudos publicados entre o período de 2018 a 2023, Ati et al. (2024) classificam os métodos multicritério existentes em três categorias principais, o primeiro métodos de agregação completa como o AHP e o ANP, o segundo métodos de outranking como o ELECTRE e o PROMETHEE e o terceiro métodos baseados num nível de referência ou objetivo como o TOPSIS e a DEA. Estes confirmam ainda que os critérios de avaliação tendem a ser estruturados por posição em campo, Guarda Redes, Defesas, Médios e Avançados, corroborando a abordagem presente nesta investigação e identificam a integração de MCDM e algoritmos de Machine Learning para prever critérios e subcritérios de decisão como uma das principais lacunas a colmatar na literatura (Ati et al., 2024). Esta aproximação entre estas duas técnicas, tem fundamento teórico. Doumpos & Zopounidis (2011), numa revisão sobre a desagregação de preferências em MCDA e a aprendizagem estatística, demonstram que ambos partilham o objetivo de identificar modelos de decisão a partir de exemplos, distinguindo-se sobretudo quanto à interpretabilidade em que os modelos MCDA privilegiam a compreensibilidade para o decisor, enquanto as técnicas como as redes neuronais artificiais ou os modelos de ensemble tendem a funcionar como "caixas-negras". Ainda assim, estes documentam diversas aplicações bem sucedidas com a utilização de redes neuronais artificiais na aprendizagem de modelos multicritério a partir de exemplos de decisão, sem as restrições de independência habitualmente exigidas pelos modelos de função de valor tradicionais. Também destacam ainda outras pontes entre os dois campos, nomeadamente a utilização de redes neuronais artificiais no contexto do próprio método AHP para aproximar as classificações de preferência em situações de julgamentos imprecisos do decisor e o recurso a métodos de kernel e a máquinas de vetores de suporte (SVM) para representar funções de valor e relações de preferência de forma generalizada, o que reforça a pertinência de combinar as abordagens multicritério tradicionais como o AHP e o VIKOR com técnicas de Machine Learning.

Este compromisso entre poder preditivo e a interpretabilidade é também discutida por Hecksteden et al. (2025) no contexto específico do futebol, que alertam para o facto de os algoritmos de Machine Learning que possuem uma elevada dimensionalidade, podem não fornecer necessariamente as

previsões mais precisas ou úteis, demonstrando que modelos deliberadamente disseminados e interpretáveis podem igualar ou até superar o desempenho preditivo de modelos denominados por "caixa-negra" como o "gradient boosting", fenómeno em que designam de "less-can-be-more". Hecksteden et al (2025) colocam esta discussão no âmbito mais vasto da chamada "revolução Moneyball", alertando para o risco de uma confiança excessiva nos dados em detrimento do julgamento humano, mas também para o risco inverso em que é desvalorizado aspetos do desempenho para os quais não se possui boas métricas, um fenómeno que designam de "falácia de Billy Beane". Apresentam ainda uma limitação particularmente no futebol profissional em que o facto de o número de jogadores ser por natureza limitado, o que entra em conflito com a natureza que é "faminta por dados" dos algoritmos complexos, em que necessitam de um número considerável de instâncias de treino para generalizarem de forma fiável, pelo que defendem que a colaboração entre o conhecimento dos especialistas desportivos e as técnicas computacionais é fundamental para obter modelos de apoio à decisão robustos e adaptáveis a novos contextos.

É exatamente para colmatar esta lacuna de interpretabilidade que se tem recorrido à metodologia SHAP, desenvolvida por Lundberg e Lee (2017) com base na teoria dos jogos cooperativos. O SHAP decompõe a previsão de qualquer modelo por mais complexo que seja, na contribuição individual de cada variável, satisfazendo as três propriedades desejáveis sendo elas a precisão local, ausência de atribuição a variáveis irrelevantes e a sua consistência tudo isto conferindo-lhe uma vantagem teórica sobre métodos alternativos como o LIME, cujas explicações assentam em modelos substitutos locais e tendem a ser mais instáveis perante pequenas perturbações dos dados (Patel et al., 2025). Esta robustez tem justificado a sua adoção crescente em domínios de elevado risco e forte exigência regulatória, como departamentos como o da saúde e as finanças, nos quais a possibilidade de auditar e justificar as decisões de um modelo é tão relevante quanto a sua exatidão preditiva. Porém, Patel et al (2025) identificam limitações importantes ao nível da complexidade computacional em modelos de aprendizagem profunda e dados de elevada dimensionalidade, bem como desafios no uso para utilizadores não especializados.

Para complementar, técnicas de clustering têm cada vez mais demonstrado um enorme potencial para extrair padrões a partir de dados, sem depender de uma estrutura de pesos definida como acontece nos métodos multicritério clássicos. Nikić et al. (2024) aplicaram o método k-means em que este encontra o número ótimo de clusters utilizando um conjunto de dados de 98 equipas das cinco principais ligas europeias de futebol a época 2021-2022, em que foi considerado variáveis como golos, remates, cartões, percentagem de posse de bola, precisão de passe e duelos aéreos ganhos.

Os resultados obtidos puderam identificar um cluster onde surge equipas que geralmente atuam nas maiores competições em que os autores designam de “Super League”, composto por clubes como o Manchester City, o Bayern de Munique, o Real Madrid ou o Paris Saint-Germain em que caracterizam por um estilo de jogo ofensivo, um elevado número de remates e uma valorização consistente da posse de bola, demonstrando o valor da segmentação não supervisionada na identificação de padrões estratégicos relevantes para a gestão desportiva.

Contudo, devido a esta adoção conjunta de MCDM e Machine Learning ainda ser bastante recente, tendência também identificada por Belhouchet et al. (2026) através de um cluster temático emergente que agrega a inteligência artificial à seleção de jogadores. A eficácia destes métodos combinados ainda não está totalmente comprovada, pelo que existe uma lacuna na literatura quanto à aplicação integrada destas abordagens no contexto do futebol profissional pelo que apontam como prioridades para a investigação futura a comunicação sistemática da sensibilidade e da robustez dos modelos face à escolha dos pesos e dos métodos, o enriquecimento dos critérios de avaliação com dados contextuais de rastreamento e de zonas táticas e a validação da reprodutibilidade destes modelos através de aplicações prospetivas ou de estudos de caso integrados em clubes (Belhouchet et al., 2026).

Face ao exposto, esta dissertação pretende contribuir com esta integração destes 2 métodos combinados ao explorar a contratação de jogadores como um problema de gestão desportiva, de forma a desenvolver um modelo mais robusto e adaptado à realidade do futebol profissional.

De forma a sistematizar os contributos principais da literatura analisada, foi elaborada uma tabela de extração de dados com os estudos considerados nesta investigação. Devido à sua extensão, a tabela completa encontra-se no Apêndice A.

3. METODOLOGIA

3.1 Enquadramento metodológico

A presente investigação percorre um desenho metodológico estruturado com o objetivo de criar um modelo de apoio à decisão na contratação de jogadores de futebol. Este modelo pretende combinar diferentes abordagens analíticas, preditivas e exploratórias.

O processo deste estudo é organizado em várias fases sequenciais. No início é definido o problema de decisão associado à contratação e valorização dos jogadores, considerado o contexto existente de elevada incerteza e o impacto significativo financeiro. Depois, procede-se à identificação e estruturação dos critérios relevantes com base na literatura e na disponibilidade existente dos dados. Posteriormente, realiza-se a recolha e preparação da base de dados, garantindo a sua conformidade para a análise.

Em seguida, são aplicadas diferentes técnicas analíticas e preditivas, nomeadamente a regressão linear múltipla e modelos de ML respetivamente, onde estas têm como objetivo compreender relações entre as variáveis, segmentar perfis de jogadores e estimar o seu potencial futuro e valor de mercado, assim, este desenho metodológico permite uma abordagem estruturada, replicável e orientada para a prática.

Neste contexto, o propósito desta fase foi entender como os diferentes tipos de variáveis se encontram representados na base de dados utilizada e assim identificar eventuais lacunas entre o que é valorizado na literatura académica e o que é efetivamente mensurável nos dados existentes.

A metodologia adotada nesta dissertação foi elaborada com o objetivo de desenvolver um modelo analítico robusto e aplicável na realidade da gestão desportiva, de modo a possibilitar apoiar decisões como a contratação/renovação dos jogadores. Num mercado altamente competitivo, com jogadores cada vez mais a serem transferidos por valores que ascendem a vários milhões de euros, é essencial criar um modelo robusto para minimizar decisões incorretas. Assim esta fase inicial procura demonstrar o valor que este modelo poderá ter aplicabilidade prática, não se limitando unicamente ao plano teórico, mas procurar construir uma ferramenta com utilidade real para o processo da tomada de decisão. É importante referir que a criação deste modelo não garante previsões exatas, uma vez que atua num contexto de incerteza e sujeito a variáveis intangíveis que não conseguimos incluir no modelo.

Nesta etapa, o objetivo ainda não é definir quais os critérios essenciais que um jogador deve ter para certa posição, mas sim analisar e testar diferentes abordagens para construir uma base metodológica mais robusta possível. Consequentemente irá ser possível antecipadamente compreender e comparar técnicas analíticas, estimar valores, identificar quais as variáveis poderão ter uma maior influência e perceber quais os elementos serão utilizados para inserir no modelo final. Ao realizar esta base metodológica antecipadamente, quando se realizar o processo da recolha de dados será possível incorporar os dados num modelo já criado e tecnicamente validado para oferecer resultados consistentes.

Neste enquadramento, o problema central desta investigação prende-se na tomada de decisão na contratação e renovação dos jogadores de futebol, num contexto de elevada incerteza, múltiplos critérios e forte impacto desportivo e financeiro. Os clubes enfrentam restrições orçamentais e pressão por desempenho imediato, levando a que este seja um processo de decisão altamente complexo e dependente de múltiplas dimensões de análise.

3.2 Caracterização da Base de Dados

A base de dados utilizada nesta investigação é referente ao ano de 2023 do jogo FIFA (atualmente EA Sports FC) que é desenvolvida pela empresa EA Sports, sendo uma das fontes mais reconhecidas de informação quantitativa sobre jogadores de futebol profissional. Esta base de dados, por ser de um jogo de simulação, pode ser considerada, pois é construída por um processo semiestruturado de recolha, análise e posteriormente validada pela equipa de "data editors" da empresa.

> "With the help of EA's 300 data editors, all that feedback from those 9,000 data reviewers is collated into 300 different data fields and 35 specific attribute categories, eventually yielding each player's overall rating." (*FIFA 17 Player Ratings Blend Stats, Subjectivity*, 2016)

A sua construção resulta de um processo que envolve a participação de milhares de especialistas de futebol, cerca de 9000 da área de futebol (treinadores, ex-treinadores, olheiros, jornalistas, etc.). Estes submetem o seu feedback sobre o desempenho do jogador, avaliando posteriormente os critérios definidos na BD. Estas observações, posteriormente recolhidas são analisadas e padronizadas por cerca de 300 editores da empresa, integrando a informação e consolidando num único formato (*FIFA 17 Player Ratings Blend Stats, Subjectivity*, 2016; FourFourTwo Staff, 2021). As avaliações dos

jogadores são expressas numa escala de 1 a 100 , abrangendo diversas dimensões de desempenho, que permite avaliar diretamente a análise de múltiplos atributos técnicos, físicos, mentais e psicológicos (FourFourTwo Staff, 2021). Esta padronização constitui uma enorme vantagem para a aplicação de métodos estatísticos e técnicas de ML, facilitando a identificação de padrões e de relações entre as variáveis.

Esta base de dados é composta por um total de 18 539 jogadores abrangendo um conjunto vasto e diversificado de jogadores e clubes. É composto em termos geográficos pelas principais ligas europeias, Premier League (Inglaterra), La Liga (Espanha), Bundesliga (Alemanha), Série A (Itália), Ligue 1 (França) e Liga Portuguesa (Portugal). Também estão incluídas outras ligas relevantes a nível mundial, desde ligas sul americanas, Brasil e Argentina, norte americanas (MLS), asiáticas e médio oriente.

Para cada jogador, a base de dados disponibiliza um conjunto alargado de variáveis, desde os 35 critérios que serão utilizados para construir o modelo, mas também outras variáveis complementares, sendo elas nacionalidade, pé preferencial, posição em campo do jogador, contrato e o clube onde atuam. Esta riqueza e diversidade de informação reforçam o potencial analítico desta BD, possibilitando uma caracterização multidimensional dos jogadores.

Apesar destas vantagens, é importante reconhecer as suas limitações, desde as suas classificações basearem-se na sua perceção/intuição, não resultando de um processo fundamentado através de métricas do seu desempenho físico ou estatístico. Contudo, a sua grande amplitude, padronização e coerência temporal tornam esta base de dados uma fonte útil, sobretudo quando o seu objetivo é comparar perfis de jogadores ou identificar padrões de desempenho.

3.3 Construção da árvore de critérios

Com o objetivo de sistematizar e simplificar apresentando graficamente os critérios identificados, desenvolveu-se uma árvore de critérios (Figura 1), sendo constituída por:

- Dados pessoais e contratuais: variáveis de identificação (nome, idade, nacionalidade, altura, peso, pé preferido, etc.) e variáveis contratuais (valor de mercado, salário, cláusula de rescisão e duração do contrato);

- Características técnicas e táticas: subdivididas por posição em campo (guarda-redes e jogadores da frente) englobando métricas de desempenho;

- Outros critérios segundo a literatura: variáveis de natureza comportamental e psicológica, como comunicação, criatividade, trabalho em equipa, liderança, resiliência, motivação, atitude e linguagem corporal, etc.

É importante salientar que na representação da árvore, os critérios com preenchimento com a cor vermelha, correspondem a critérios identificados na literatura dos quais não existem dados na base de dados da FIFA. Desta maneira, permitiu mostrar de uma forma imediata as lacunas existentes entre a informação teórica e a informação empírica disponível, onde é de destacar que os aspetos comportamentais e psicológicos são difíceis de quantificar.

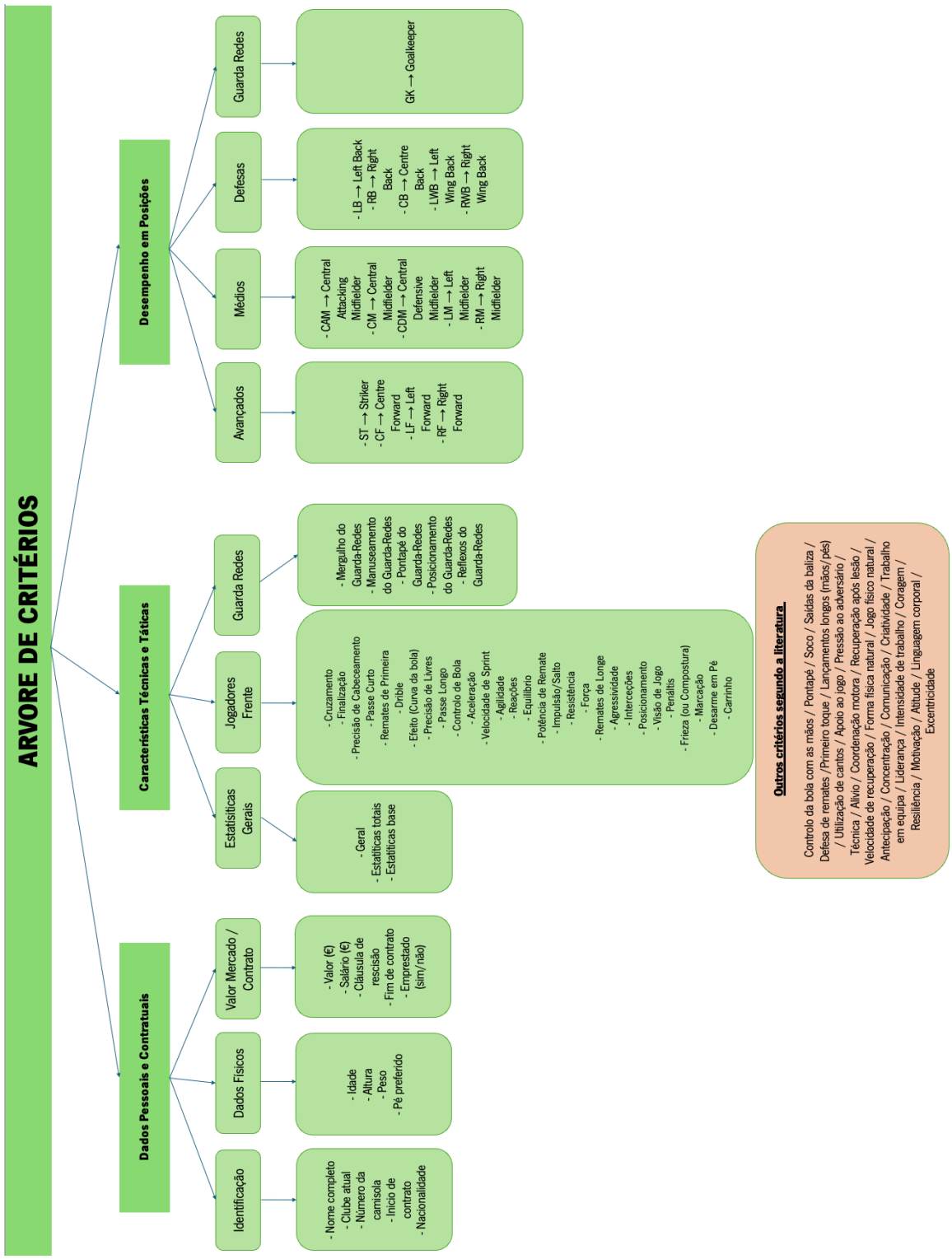


Figura 1 - Árvore de Critérios

3.4 Preparação dos dados

O conjunto de dados utilizado foi antecipadamente realizada a normalização dos dados, de modo a obtermos uma base de dados limpa e organizada. Caso existam valores em falta nas variáveis, os jogadores foram excluídos garantindo assim consistência na aplicação destas técnicas.

Nestas técnicas, os jogadores foram também previamente agrupados em 4 grupos, Guarda Redes, Defesas, Médios e Avançados. A agregação dos jogadores nestas quatro posições, foi realizada com o objetivo de garantir comparabilidade estatística e coerência analítica. Esta segmentação permite reduzir a heterogeneidade, uma vez que cada uma destas quatro posições apresenta exigências técnicas distintas. Deste modo, conseguimos obter uma dimensão amostral suficiente dentro de cada grupo, permitindo a aplicação robusta das técnicas estatísticas e de ML.

Esta separação permite comparações coerentes e assegurar que todas as análises reflitam as exigências para cada um destes grupos. A análise separada por grupos posicionais, permite captar de forma mais precisa o impacto dos diferentes critérios no seu desempenho e valorização, evitando distorções que poderiam acontecer caso não se tivesse realizado a desagregação das posições dos jogadores.

Para além dos procedimentos anteriormente descritos, foi realizada uma preparação adicional com o objetivo de garantir a adequação correta das diferentes técnicas utilizadas.

Referente às variáveis utilizadas, foram consideradas de forma consistente nas diversas técnicas, os atributos técnicos, físicos, mentais e psicológicos, bem como a idade. Estas variáveis foram utilizadas nos modelos de regressão e de ML, permitindo comparabilidade nos presentes métodos.

No que concerne aos outliers, foi realizada uma análise exploratória das variáveis, verificando-se valores extremos, mais concretamente quando analisado o valor de mercado dos jogadores no modelo de regressão. A transformação logarítmica contribuiu para reduzir este impacto, não tendo sido efetuada nenhuma remoção destes jogadores de forma a preservar a representatividade.

Relativamente à escala das variáveis, a normalização foi particularmente relevante nos modelos de Redes Neuronais Artificiais, garantindo estabilidade no processo de aprendizagem. No caso de Random Forest, não foi necessária a normalização adicional, visto que este modelo não é sensível à escala das variáveis.

3.5 Análise Empírica de Regressão Múltipla

Com o objetivo de compreender de forma quantitativa a influência das diferentes características no valor de mercado dos jogadores, foram estimados modelos de regressão linear múltipla simples onde a variável que quisemos explicar (variável dependente) é o valor de mercado do jogador e que as variáveis explicativas (variáveis independentes) são os diversos critérios, sendo estes o conjunto de atributos técnicos, físicos, psicológicos e mentais presentes na base de dados. Apesar de o propósito desta dissertação não ser a determinação do valor de mercado do jogador, esta análise possibilita identificar a relevância estatística dos diversos critérios de desempenho aqui estudados, contribuindo para a construção de um modelo fundamentado de decisão.

De forma a garantir uma maior precisão analítica, os jogadores foram previamente segmentados em quatro grupos posicionais como mencionado anteriormente Para isso utilizou-se a variável "Best_Position" da base de dados, garantindo que cada jogador pertencia a um único grupo, sendo divido em 4 posições, Guarda Redes (GR), Defesas (DF), Meio Campo (MC) e Avançados (AVA). Esta segmentação permite capturar as especificidades de cada posição, assegurando que os modelos refletem as diferentes necessidades associadas a cada perfil do jogador.

A seleção das variáveis explicativas teve como base a relevância média dos critérios por posição, tendo sido definido limiares mínimos de inclusão. Para as posições de Defesas, Médios e Avançados considerou-se critérios com valores médios ou superiores a 60, enquanto que para Guarda Redes foi utilizado um limiar mínimo de 40 devido à menor diversidade de critérios específicos para esta posição. A escolha destes valores mínimos baseou-se numa análise exploratória prévia da distribuição de critérios por posição, onde se verificou que acima destes valores médios havia um equilíbrio entre a inclusão de um número suficiente de variáveis e a sua relevância estatística, evitando simultaneamente a sobrecarga dos modelos com variáveis pouco informativas que fossem contribuir unicamente com ruído. Deste modo, permitiu que os GR tivessem uma inclusão de no mínimo 10 critérios e que os DEF, MC, e AVA tivessem uma inclusão de no mínimo 15 critérios.

Dado o número elevado de variáveis semelhantes, foi analisada a possível presença de multicolinearidade entre os critérios. Para o efeito, foi examinada a matriz de correlação entre as variáveis independentes, o que permitiu identificar variáveis com elevada correlação. Sempre que se verificaram correlações elevadas entre as variáveis, foi considerada a sua relevância e estatística, evitando a inclusão redundante de critérios com informação semelhante. Esta abordagem permitiu reduzir potenciais problemas de instabilidade nos coeficientes estimados, de forma a garantir uma

maior robustez nos resultados do modelo.

Antes de se realizar as regressões das 4 posições definidas, os critérios de desempenho foram organizados em 4 categorias, “Técnicos”, “Físicos”, “Mentais” e “Psicológicos”. Esta separação irá facilitar a interpretação e compreensão dos resultados permitindo analisar não apenas o impacto individual de cada critério, mas também perceber se certos grupos de competências contribuem mais para o valor de mercado do jogador.

Estes 4 grupos seguiram com base em 2 estudos (Çaloğlu Büyükselçuk & Badem, 2024; Ozceylan, 2016), onde abordaram os critérios nestas categorias. Assim, a estrutura foi considerada adequada para garantir a comparabilidade com investigações anteriores e garantir uma análise sistematizada e coerente com a literatura.

A Tabela 1 mostra a distribuição dos critérios pelas quatro categorias referenciadas, servindo de base para a interpretação dos modelos de regressão apresentados na secção seguinte.

TÉCNICO	FÍSICO	MENTAIS	PSICOLÓGICO
Cruzamentos	Aceleração	Reações	Calma / Serenidade
Finalização	Velocidade de Sprint	Interceções	Agressividade
Precisão de Cabeceamento	Agilidade	Posicionamento	
Passe Curto	Equilíbrio	Visão de Jogo	
Remate de Primeira	Impulsão		
Drible	Resistência		
Efeito	Força		
Precisão de Livres			
Passe Longo			
Controlo de Bola			
Potência de Remate			
Remates de Longa Distância			
Penáltis			
Marcação			
Desarme em Pé			
Carrinho			
Mergulho do Guarda-Redes			
Manuseamento do Guarda-Redes			
Pontapé do Guarda-Redes			
Posicionamento do Guarda-Redes			
Reflexos do Guarda-Redes			

Tabela 1 – Critérios de desempenho por categoria

A variável dependente que queremos explicar, o valor de mercado do jogador em euros, foi transformada em logaritmo com o objetivo de reduzir assimetrias e estabilizar a variância dos resíduos, ou seja, tornar constante a dispersão dos erros do modelo. Por sua vez, as variáveis independentes, à exceção da idade do jogador, estão numa escala de 1 a 100 referentes aos atributos técnicos, físicos, mentais e psicológicos. Estas variáveis foram incluídas tanto na sua forma linear como quadrática, $(x_i x_i^2)$ permitindo capturar relações não lineares, pois critérios como idade ou aceleração tendem a ter um efeito crescente até certo ponto e depois decrescente. Os modelos foram estimados separadamente para cada grupo posicional. Assim, permitiu identificar quais os critérios com maior impacto para cada posição. Os resultados obtidos possibilitam identificar os critérios mais relevantes, mas também a interpretação do seu efeito marginal no valor de mercado dos jogadores.

Para ilustrar a aplicabilidade do modelo, selecionou-se aleatoriamente cinco jogadores de diferentes posições para estimar os seus valores referentes ao ano de 2023. Para cada jogador, foi multiplicado os valores das regressões com os critérios, mais os valores com efeito quadrático e a constante do modelo correspondente, obtendo assim o valor em logaritmo. Com este valor, aplicou-se a exponenciação para converter o resultado em euros como mostra na Tabela 2. Por exemplo, o jogador Jude Bellingham (Posição: Médio, Idade: 19, Critérios: Valores altos), a soma ponderada dos efeitos das variáveis acrescida da sua constante, resulta num log (ValorMercado) ≈ 18,214, o equivalente a aproximado 74,4 milhões de euros como mostra a equação em baixo. Este processo permite demonstrar de forma clara como cada critério contribui para um valor estimado do jogador.

Jogador	Posição	Valor (€) aproximado
Paulo Oliveira	Defesa	2 090 576 €
Haaland	Avançado	24 424 322 €
Pedro Porro	Defesa	2 676 445 €
Ter Stegen	Guarda-Redes	5 389 698 €
Jude Bellingham	Médio	81 327 925 €

Tabela 2 – Estimativas de valor de mercado por regressão

$$\log(\text{Valor}) = 12,90 + (0,124 \times 19) + \cdots + (0,0209 \times 56 + (-0,000112 \times 56^2)) = 18,214$$

$$\text{Valor (€)} = e^{18,214} \approx 81,3 \text{ milhões de euros}$$

Contudo, estes valores estimados não refletem o seu valor total, pois fatores externos não estão refletidos diretamente nestas variáveis. De acordo com Mourao (2016), o valor do jogador de futebol não depende apenas do seu desempenho, mas também a forma como o clube a que pertence é administrado e os seus resultados desportivos ao longo do tempo. Mourao (2016) refere que a liga onde atua, a reputação do clube, o seu historial como troféus conquistados e a visibilidade mediática leva a um papel determinante na valorização do mercado. Assim, jogadores de equipas com maior sucesso desportivo e histórico tendem a apresentar valores de mercado mais elevados.

Na sequência desta análise, o jogador Erling Haaland que atua no Manchester City na Premier League, apresenta um valor de mercado real superior face ao estimado pelo modelo, evidenciando uma subestimação por parte da regressão, revelando algumas limitações na capacidade do modelo em captar plenamente os fatores que influenciam o valor de mercado.

Neste contexto, os coeficientes estimados para cada variável, bem como a sua respetiva significância estatística, estão apresentados no Apêndice B, de modo a permitir uma análise mais detalhada do impacto individual de cada critério na determinação do valor do mercado.

3.6 Análise com Clustering e Machine Learning

Esta secção combina métodos exploratórios e preditivos de forma sequencial. Primeiramente, os dados para análise foram preparados e normalizados. Foram utilizadas técnicas não supervisionadas para identificar padrões e perfis de jogadores e técnicas supervisionadas em que estas estimaram o seu overall máximo e potencial valor de mercado.

Para aprofundar a análise dos critérios dos jogadores e identificar padrões de desempenho, utilizou-se técnicas avançadas de Machine Learning. Estas permitiram agrupar jogadores com perfis idênticos e estimar o seu potencial máximo que poderiam atingir. A Figura 2 ilustra o pipeline do processo de ML. Tanto o Random Forest e as Redes Neuronais Artificiais são modelos do ML. O RF integra um conjunto de métodos baseados em árvores de decisão (Trees) enquanto as RNA pertencem à área de Deep Learning que conseguem capturar relações não lineares complexas.

A utilização destas técnicas permitem identificar padrões complexos entre as variáveis que as regressões lineares simples não conseguem captar. Todas estas técnicas permitem potenciar uma maior robustez para este modelo, no entanto, é importante relembrar que embora essa maior robustez dos resultados, eles devem ser encarados como estimativas sujeitas a incerteza, não garantindo previsões perfeitas, pois no futebol existe variáveis que não podem ser mensuradas, pelo que as conclusões aqui determinadas resultam numa natureza probabilística.

Foram incluídos diversos gráficos de comunicação para facilitar a leitura dos resultados obtidos, desde mapas PCAs, gráficos de barras e tabelas para transmitir de forma intuitiva os padrões obtidos.

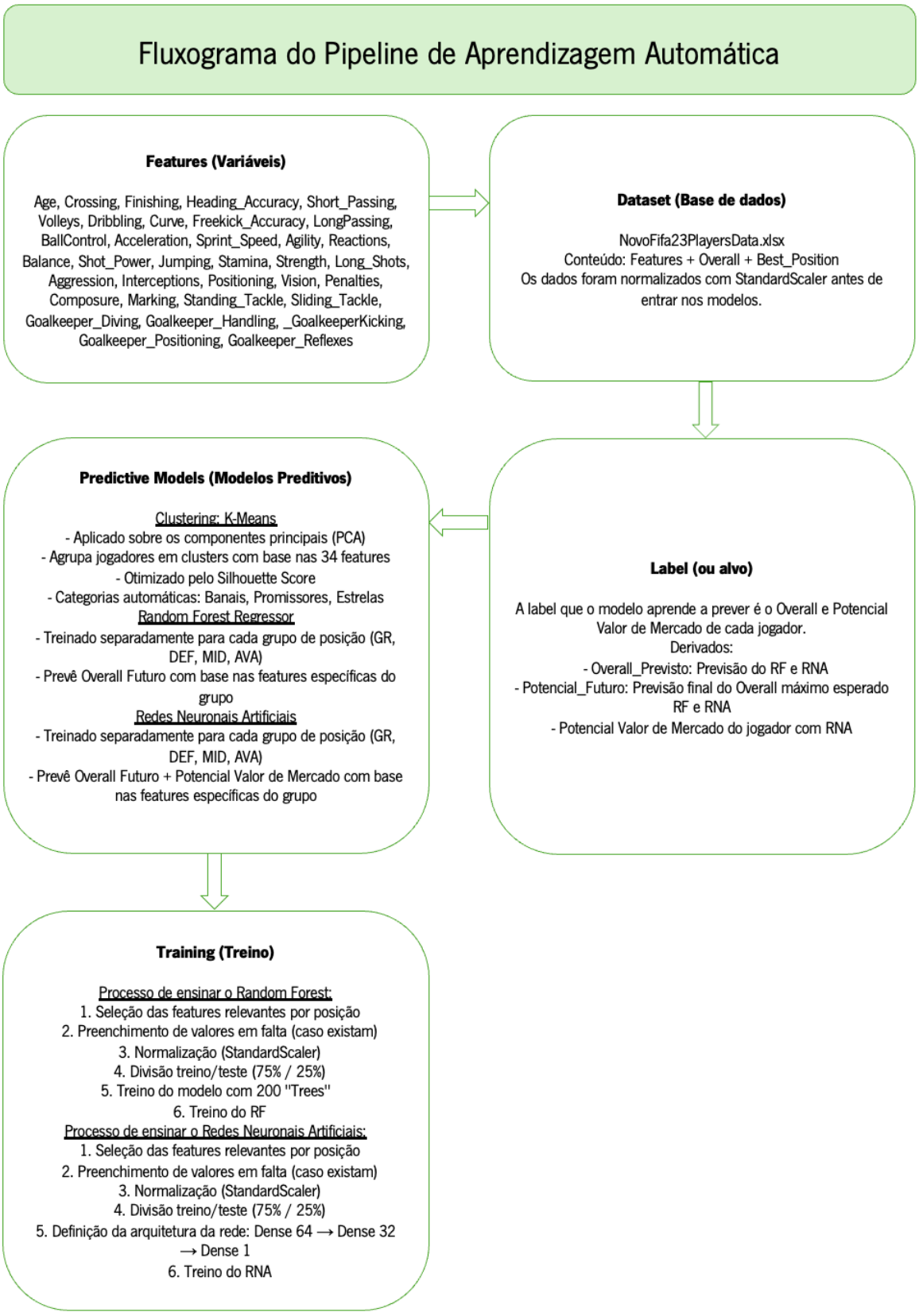


Figura 2 - Pipeline do Processo do Machine Learning

3.6.1 Agrupamento dos Jogadores (Método Clustering)

Para identificar padrões e segmentação de jogadores, aplicou-se o método clustering que constitui uma técnica não supervisionada, que tem o objetivo de identificar grupos de dados, agrupando com base na sua semelhança. Desta forma, cada cluster obtido pode ser interpretado como um perfil típico de jogador, caracterizado por um conjunto específico de critérios dominantes. Utilizando a técnica de "K-Means Clustering", é importante referir que esta depende da inicialização dos centroides. Primeiramente, foi testado qual o número ótimo de clusters, estimando a sua qualidade através da métrica "Silhouette Score", que indica o quão bem cada jogador se encaixa no cluster atribuído face aos outros clusters. Este número de clusters é selecionado automaticamente com base na métrica referenciada anteriormente.

Foi utilizada esta técnica, por ser simples, eficiente computacionalmente e facilmente interpretável (Kanungo et al., 2000), adequada para agrupar jogadores com perfis semelhantes em função dos critérios analisados. Embora exista outros algoritmos de clustering, o K-Means permite escolher o número de clusters pretendidos e apresenta resultados consistentes.

Para simplificar a visualização, utilizou-se a técnica de PCA, onde esta reduz a dimensão dos conjuntos de dados mantendo a maior parte da variabilidade o que permite observar de forma clara os agrupamentos de clusters como ilustra a Figura 3. Assim, a projeção dos dados em duas dimensões facilita a interpretação dos resultados.

Neste contexto, foi obtida uma solução inicial, onde foram identificados dois clusters automáticos. Estes dois clusters, como evidencia a Figura 3, foi possível observar uma separação clara entre os Guarda Redes e as restantes posições. Este resultado mostra a forte especificidade dos critérios associados à posição dos Guarda Redes.

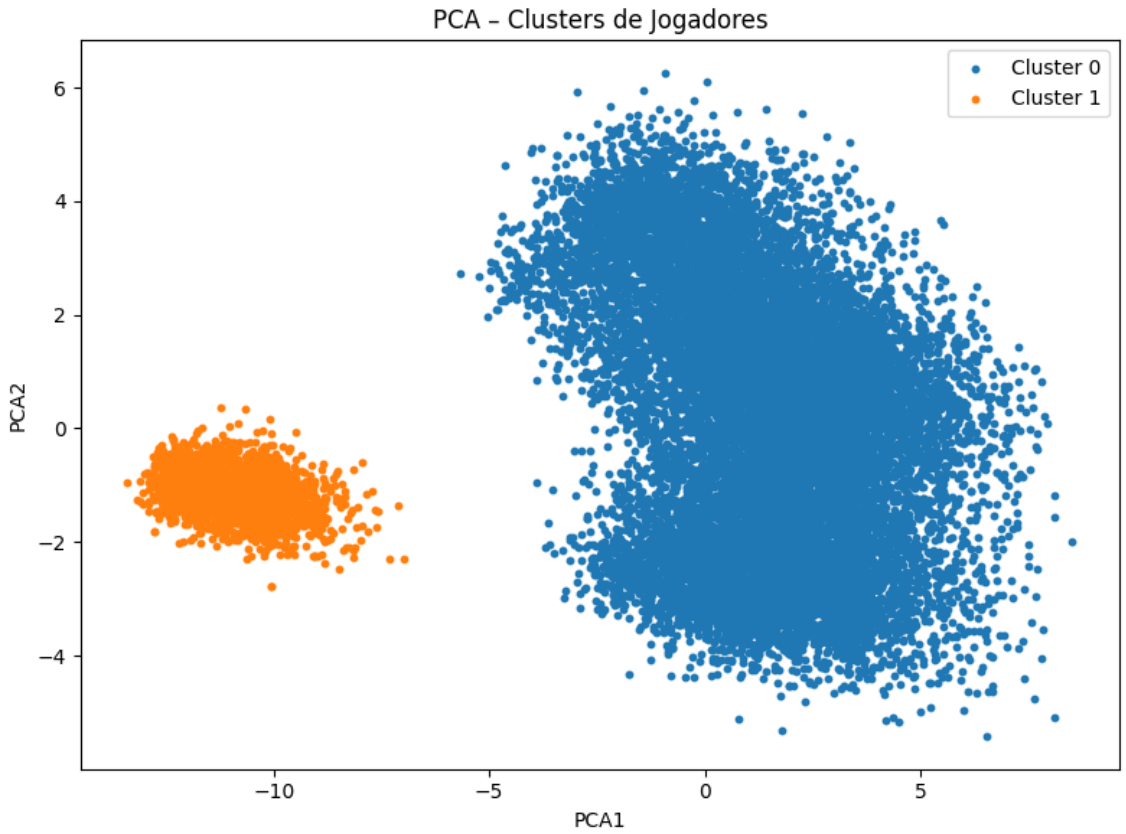


Figura 3 - Visualização dos Clusters através de PCA

Posteriormente, com o objetivo de aumentar o nível de detalhe para as quatro posições que queremos aqui analisar, o modelo foi refinado, sendo testada uma solução com quatro clusters, correspondentes aos Guarda Redes, Defesas, Médios e Avançados. Esta abordagem de separar os clusters, impõe uma segmentação mais interpretável.

A figura 4 mostra a segmentação com quatro clusters definidos com base nos critérios utilizados nas regressões. Isto permitiu uma visualização mais detalhada das diferenças entre estas 4 posições alinhada com as especificidades funcionais de cada grupo.

De forma geral, olhando para os "jogadores da frente", verifica-se uma maior proximidade entre os clusters, ainda que seja possível identificar padrões de agrupamentos distintos associados a cada posição.

Os Defesas, identificados com a cor laranja, os Médios com a cor verde e os Avançados com a cor vermelha, apresentam alguma sobreposição entre si, refletindo a existência de critérios comuns entre

posições. No caso dos Médios, é possível observar uma maior dispersão, o que sugere uma maior heterogeneidade desta posição.

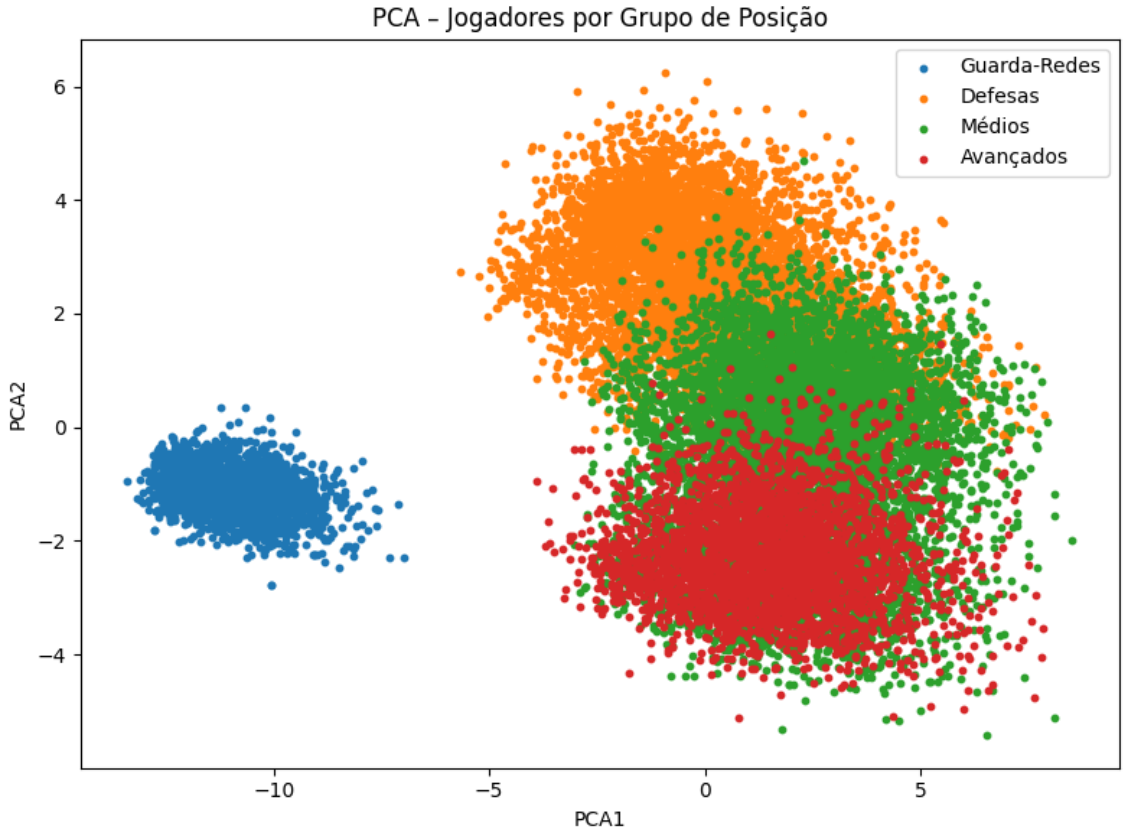


Figura 4 - Clusters por Posição Projetados em PCA

Complementarmente à segmentação em posições, foi desenvolvida uma abordagem diferente de clustering, onde esta tem como objetivo identificar diferenças ao nível da qualidade dos jogadores, independente da posição aonde estes se inserem. Esta segmentação procura identificar perfis de desempenho, permitindo uma leitura mais orientada para a avaliação de talento, assim procedeu-se à identificação de três grupos distintos, procurando identificar jogadores "Banais, "Promissores" e "Estrelas". Após a obtenção destes três clusters, através do algoritmo de K-Means, a sua identificação foi realizada com base na análise conjunta de perfis médios de cada grupo, considerando os seus critérios de desempenho, bem como o seu overhall e a idade para captar os ciclos de vida do jogador, sejam eles crescentes e decrescentes. Desta forma, o cluster que apresenta níveis médios de desempenho menores foi interpretado como jogadores "Banais", o cluster com valores intermédios e

características de desenvolvimento foi classificado como jogadores “Promissores” e por fim o cluster que concentra jogadores com níveis superiores de desempenho global foi identificado como jogadores “Estrelas” como mostra a Figura 5.

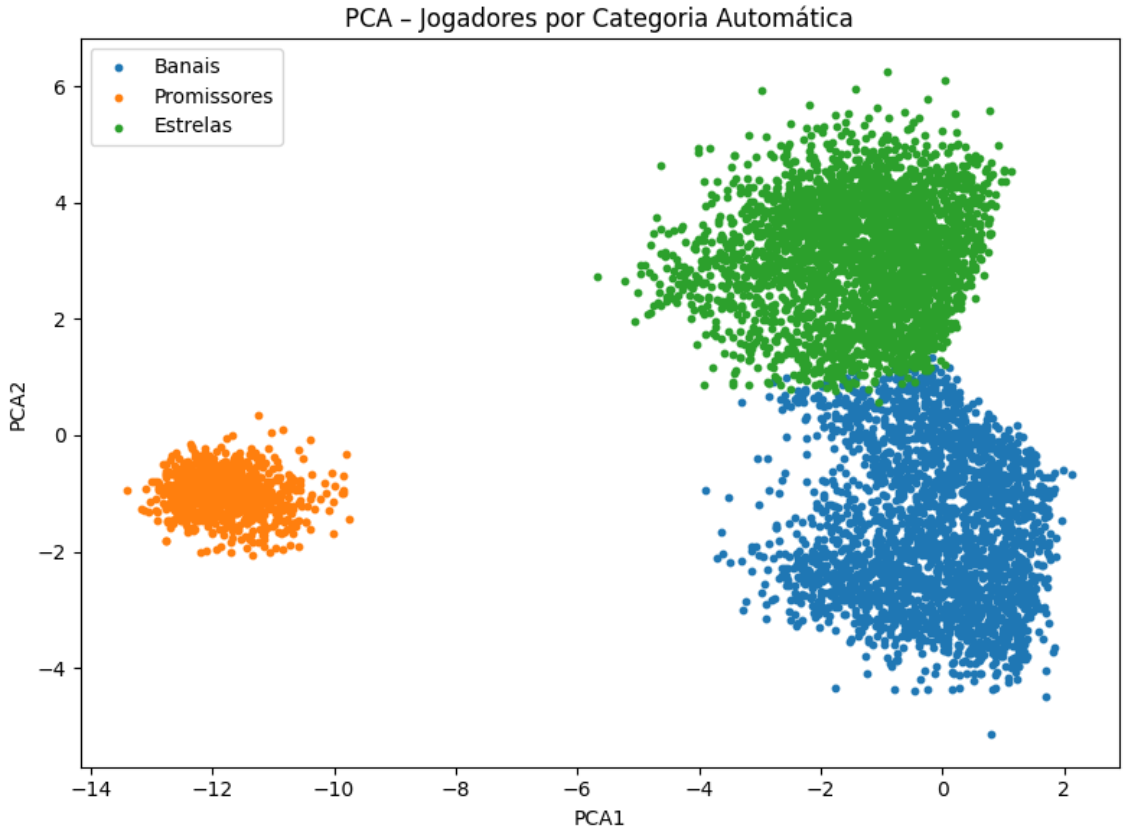


Figura 5 – Três Clusters de Desempenho dos Jogadores em PCA

3.6.2 Estimativa do potencial do jogador através de Random Forest

Para estimar o potencial que o jogador poderá atingir, ou seja, o seu máximo "Overall” (Potencial_Futuro), utilizou-se o modelo de RF. Esta é uma técnica automática de aprendizagem que combina múltiplas “Trees” para calcular valores de forma robusta, proposto por Breiman (2001). Este modelo foi escolhido devido a possuir uma grande robustez, à sua capacidade de trabalhar com relações não lineares entre as variáveis e a fácil leitura através da “Feature Importance” O RF ao combinar múltiplas “Trees” reduz a variância das previsões e diminui o risco de overfitting (Breiman, 2001). Em comparação com outros modelos existentes, este modelo apresenta uma boa performance

em bases de dados grandes, com muitas variáveis correlacionadas o que nos permite obter estimativas mais confiáveis sobre a nossa variável Overall.

O RF seleciona as variáveis automaticamente, conforme estas sejam mais relevantes para a previsão, através do cálculo da importância de cada característica (Feature Importance), em que esta mede quanto a variável contribui para diminuir o erro nos nós das "Trees". Nem todas as variáveis possuem o mesmo peso, algumas o modelo considera que têm pouca influência o que leva a que estas quase não tenham impacto nas previsões. Assim, esta abordagem permite que o modelo utilize as variáveis mais importantes, o que aumenta a robustez destas estimativas.

Antes de proceder ao treino do RF, os dados foram divididos em treino e teste, 75% para treino e 25% para teste. Estas percentagens permitem avaliar adequadamente a generalização do modelo, mantendo ao mesmo tempo um volume suficiente de dados para a aprendizagem (Breiman, 2001).

De forma a mitigar problemas de overfitting e underfitting, foram definidos limites para a profundidade de "Trees" garantindo uma capacidade de aprendizagem e generalização. Os hiper parâmetros do modelo foram definidos com base em testes preliminares, nos quais observou-se que o aumento de árvores acima de 200 resultava em ganhos marginais reduzidos, sendo assim adotado este número de "Trees". Este valor encontra também suporte na literatura, a qual sugere que a partir de determinado limiares, normalmente entre 100 a 300, os ganhos de performance tornam-se marginais, não justificando aumentos adicionais na complexidade dos modelos (Oshiro et al., 2012). Com esta opção, alienada com limites de profundidade e tamanho mínimo das folhas, ajuda a reduzir a probabilidade de overfitting, em que aprende demasiado bem os dados de treino incluindo o ruído. Exemplo de explicação de um autor: "Um aluno decora as respostas do teste passado. No exame final, sai perguntas diferentes, obteve mau resultado." e de underfitting, em que o modelo não aprendeu o suficiente sobre os dados, o autor dá outro exemplo, "Um aluno não estuda quase nada. Nem o teste passado tem (não aprende e não generaliza)" (Pinto, 2025).

Assim, o modelo foi treinado utilizando os critérios utilizados anteriormente. Foi dado uma maior preponderância à idade de forma a refletir um maior potencial em jogadores mais jovens em comparação a jogadores mais velhos. Desta forma permitiu identificar jogadores com "Potencial Alto", "Potencial Médio" e "Potencial Baixo" como evidencia a Tabela 3, uma amostra dos resultados obtidos. Esta classificação foi escolhida de modo a garantir uma interpretação simples e compreensiva, evitando categorias excessivas. Outros valores acima destes iriam aumentar a complexidade sem acrescentar algum valor interpretativo.

Full_Name	Age	Group_Position	Overall	Overall_Preditivo	Potencial_Ajustado	Potencial_Futuro	Potencial_Interpretação
Kylian Mbappé	23	AVA	91	90,415	3,415	93,83	Potencial médio
Erling Haaland	21	AVA	88	87,39	5,39	92,78	Potencial alto
Gianluigi Donnarumma	23	GR	88	87,46	3,46	90,92	Potencial médio
Mohamed Salah	30	AVA	90	89,665	0,665	90,33	Potencial baixo
Jude Bellingham	19	MC	84	83,61	6,61	90,22	Potencial alto
Pedro González López	19	MC	85	84,085	6,085	90,17	Potencial alto
Frenkie de Jong	25	MC	87	87,54	2,54	90,08	Potencial médio
Alisson Ramses Becker	29	GR	89	88,97	0,97	89,94	Potencial baixo
Harry Kane	28	AVA	89	88,925	0,925	89,85	Potencial baixo
Robert Lewandowski	33	AVA	91	90,285	-0,465	89,82	Potencial baixo
Kevin De Bruyne	31	MC	91	90,005	-0,245	89,76	Potencial baixo
C. Ronaldo dos Santos Aveiro	37	AVA	90	89,735	-0,265	89,47	Potencial baixo
Nicolò Barella	25	MC	86	86,73	2,73	89,46	Potencial médio
Toni Kroos	32	MC	88	88,45	0,95	89,4	Potencial baixo

Legenda: < 2 = Potencial Baixo; >= 2 e < 5 Potencial Médio; >= Potencial Alto

Tabela 3 – Classificação do Potencial dos Jogadores com RF

3.6.3 Estimativa do potencial do jogador através de Redes Neuronais Artificiais

Para aumentar a robustez e complementar e validar os resultados obtidos com o RF, utilizou-se também as RNA para estimar o potencial que um jogador poderá atingir. Este modelo é inspirado no funcionamento biológico do cérebro (Santana et al., 2025), em que é capaz de identificar relações complexas e não lineares entres as variáveis, sendo uteis em contextos onde existem interações entre múltiplas variáveis Todo o processo neste modelo foi realizado como no anterior, desde os 4 grupos, à sua limpeza e normalização através do StandardScaler para assegurar estabilidade no processo de aprendizagem.

A arquitetura utilizada consistiu em 2 camadas ocultas, a primeira com 64 neurónios e a segunda com 32 neurónios, ambas com a função de ativação ReLU que introduz a não linearidade, ou seja, corta tudo o que é negativo (Santana et al., 2025). Esta configuração foi definida com base em testes preliminares, procurando obter um equilíbrio entre a capacidade de aprendizagem e complexidade do modelo.

A divisão dos dados foi realizada da mesma forma como o RF, 75% para treino e 25% para teste para garantir comparabilidade entre os métodos. Para obter o "Overall_Previsto", foi utilizado a mesma lógica de ajustar a idade para atribuir um maior potencial aos jogadores mais jovens. As classificações foram mantidas a mesma escala, mantendo a coerência com o enquadramento metodológico anterior. Para controlo de overfitting, foi utilizada a técnica de Early Stopping, que interrompe o treino quando o erro de validação deixa de melhorar. Esta abordagem é bastante útil, pois permite evitar que o modelo se ajuste excessivamente aos dados do treino.

Os resultados obtidos apresentam uma similaridade aos do RF, reforçando a robustez do valor a estimar onde apresentam estimativas globalmente consistentes. Existem algumas diferenças, onde este modelo dá um potencial a diferentes jogadores como mostra a Tabela 4, mas em geral mantêm-se alinhado com os valores dados pelo modelo do RF, não se verificando alterações muito relevantes. Podemos assumir estas diferenças existentes devido a este modelo ter uma maior sensibilidade a combinações não lineares das variáveis que leva a ajustes distintos e à inclusão de outros jogadores face ao RF.

Full_Name	Age	Group_Position	Overall	Overall_Preditivo	Potencial_Ajustado	Potencial_Futuro	Potencial_Interpretação
Kylian Mbappé	23	AVA	91	91,231	4,231	95,463	Potencial médio
Pedro González López	19	MC	85	86,188	8,188	94,375	Potencial alto
Erling Haaland	21	AVA	88	88,038	6,038	94,075	Potencial alto
Frenkie de Jong	25	MC	87	88,854	3,854	92,708	Potencial médio
Lionel Messi	35	MC	91	91,450	0,450	91,900	Potencial baixo
João Félix Sequeira	22	MC	84	85,429	6,429	91,857	Potencial alto
Phil Foden	22	MC	85	85,833	5,833	91,666	Potencial alto
C. Ronaldo dos Santos Aveiro	37	AVA	90	90,639	0,639	91,278	Potencial baixo
Gianluigi Donnarumma	23	GR	88	87,536	3,536	91,072	Potencial médio
Dušan Vlahović	22	AVA	84	84,982	5,982	90,965	Potencial alto
Kevin De Bruyne	31	MC	91	90,373	0,123	90,496	Potencial baixo
Florian Wirtz	19	MC	82	82,745	7,745	90,490	Potencial alto
Christopher Nkunku	24	MC	86	86,662	3,662	90,323	Potencial médio
Robert Lewandowski	33	AVA	91	90,498	-0,252	90,246	Potencial baixo

Legenda: < 2 = Potencial Baixo; >= 2 e < 5 Potencial Médio; >= Potencial Alto

Tabela 4 - Classificação do Potencial dos Jogadores com RNA

3.6.4 Estimativa do Valor de Mercado através de Redes Neuronais Artificiais

Para além da previsão que um jogador pode atingir no seu Overall, foi também realizado a estimação do seu potencial económico, através das RNA. Foi escolhida esta técnica pois esta consegue captar relações complexas e não lineares o que não acontece no caso das regressões como verificado na secção 3.5, pois as regressões captam apenas as variáveis escolhidas, o que leva a que não consiga estimar um valor potencial de mercado. As RNA conseguem estimar padrões associados a fatores não diretamente medidos, tais como fatores intangíveis e voláteis, como a visibilidade mediática, prestígio do clube e liga onde se insere que influencia o seu valor de mercado (Mourao, 2016). Por este motivo, o modelo de RF não foi realizado, pois não consegue captar tão bem estas relações complexas e não lineares em que tende a simplificar.

O processo metodológico seguiu a mesma lógica dos processos anteriores na previsão do Overall, desde a construção dos 4 grupos, ao preenchimento de valores em falta, à normalização das variáveis com StandardScaler e a mesma divisão em treino e teste, com 75% e 25% respetivamente.

A variável do mercado de jogador, como nas regressões foi transformada em logaritmo para esta estabilizar a distribuição e reduzir o efeito dos valores extremos. A arquitetura utilizada também é igual à do modelo anterior de RNA, 64 e 32 neurónios em ReLU. Utilizou-se sempre os mesmos processos metodológicos para mantermos a coerência, de modo a termos comparabilidade dos resultados.

Após o treino, as estimativas destas previsões do potencial valor de mercado foram convertidas através de “expm1”, obtendo assim um valor em euros do jogador. A Tabela 5 mostra os potenciais valores de mercado. Para efeitos de comparação serão utilizados os mesmos jogadores previamente selecionados nas regressões realizadas na secção 3.5.

Jogador	Equipa	Posição	Valor (€)
Paulo Oliveira	SC Braga	DEF	4 245 892,22 €
Erling Haaland	Manchester City	AVA	169 087 783,61 €
Pedro Porro	Sporting CP	DEF	29 501 090,49 €
Ter Stegen	FC Barcelona	GR	87 975 000,00 €
Jude Bellingham	Borussia Dortmund	MC	109 500 000,00 €

Tabela 5 - Estimativas de valor de mercado por RNA

De forma a avaliar a qualidade dos valores estimado, recorreu-se à plataforma Transfermarket, amplamente conhecida por ser uma BD bastante completa, que inclui diversas métricas, com especial destaque para os valores de mercado. Assim observou-se a comparação entre os valores estimados versus os valores apresentados pela plataforma, assegurando que os dados considerados correspondiam ao mesmo período temporal da BD utilizada nesta investigação.

De acordo com o Transfermarket, o Paulo Oliveira estava avaliado em 3 500 000€, verificando-se uma diferença inferior de 750 000€. Por sua vez, Erling Haaland estava avaliado em 170 000 000€, registando uma diferença inferior a 1 000 000€. Relativamente a Pedro Porro, este estava avaliado em 25 000 000€, diferença de cerca de 4 500 000€. Já Ter Stegen, encontrava-se avaliado em 30 000 000€, evidenciando uma discrepância já considerada, sendo quase um terço do valor estimado. Por fim, Jude Bellingham encontrava-se avaliado em 110 000 000€, diferença de 500 000€.

Perante estas diferenças, verifica-se que os valores estimados apresentam de forma geral uma elevada proximidade face aos valores da plataforma Transfermarket, evidenciando uma boa capacidade preditiva. Ainda assim, verificou-se um valor nesta amostra que evidenciou uma discrepância, sugerindo a existência de fatores não captados pelo modelo que poderão influenciar o valor do mercado.

Em síntese, a metodologia adotada neste estudo integra diferentes abordagens. A regressão linear múltipla permite identificar os atributos com maior associação ao valor de mercado dos jogadores, contribuindo para uma compreensão explicativa dos critérios aqui analisados. O clustering é utilizado

para segmentar os jogadores em perfis homogéneos permitindo uma análise exploratória da estrutura dos dados utilizados. Os modelos de Random Forest e Redes Neuronais Artificiais são aplicados para estimar o potencial futuro e o valor do mercado do jogador, captando relações lineares e complexas. Estas metodologias adotadas, geraram estes outputs, os quais servem para alimentar um modelo multicritério de apoio à decisão, permitindo a integração de diferentes resultados para apoiar a decisão à contratação de jogadores de futebol.

3.6.5 Avaliação da Qualidade do Modelo

A qualidade deste modelo foi avaliada considerando as técnicas de Clustering, RF e RNA.

Para o Clustering, utilizou-se a métrica Silhouette Score, onde esta indica se o jogador encaixa-se no cluster atribuído em comparação com os outros clusters, sendo que valores perto de 1 refletem melhores agrupamentos (Shahapure & Nicholas, 2020). Posteriormente, usou-se PCA, uma técnica de redução de dimensionalidade que permite visualizar em 2D os clusters.

Para o RF, a avaliação da qualidade foi realizada através das seguintes métricas:

- R^2: Mede a proporção da variabilidade do Overall
- $RMSE$: Penaliza os erros maiores, dando uma medida sobre a precisão das previsões
- MAE: Média absoluta das diferenças dos valores previstos e observados o que fornece uma interpretação direta

A validação foi realizada com a divisão treino/teste (75%/25%) com o treino separado pelos 4 grupos (GR, DEF, MC e AVA) para obter uma maior precisão dentro destes grupos.

Para as RNA, a avaliação da qualidade foi realizada através das seguintes métricas:

- MAE: Média absoluta das diferenças dos valores previstos e observados o que fornece uma interpretação direta
- MSE: Função de custo minimizada durante a realização do treino

A validação foi realizada com a divisão treino/teste (75%/25%) com o treino separado pelos 4 grupos (GR, DEF, MC e AVA) para obter uma maior precisão dentro destes grupos. Para evitar overfitting, usou-se a técnica de "Early Stopping".

A Tabela 6 apresenta os valores das métricas obtidos deste modelo.

Modelo	Posição	R^2	RMSE	MAE	Silhouette
RF	GR	**0.993**	0.622	0.445	–
RF	DEF	**0.970**	1.139	0.859	–
RF	MC	**0.947**	1.569	1.199	–
RF	AVA	**0.976**	1.093	0.841	–
RNA	GR	–	–	**0.32**	–
RNA	DEF	–	–	**0.67**	–
RNA	MC	–	–	**1.11**	–
RNA	AVA	–	–	**0.53**	–
Clustering	Todas	–	–	–	**0.56 (k=2)**

Tabela 6 – Valores das Métricas de Clustering, RF e RNA

Olhando em termos quantitativos, o Clustering demonstra um desempenho sólido, sendo que foi identificado um número ótimo de clusters de 2, com um "Silhouette Score" de 0,56. Nesta métrica, o maior valor obtido, até um máximo de 1, diz que é o número ótimo de clusters (Shahapure & Nicholas, 2020).

Em relação ao RF os resultados apresentam uma elevada capacidade preditiva. Olhando para os valores dados do R^2 apresentam um alto valor o que indica que o modelo explica quase toda a variabilidade do Overall. Os erros médios (RMSE e MAE) mantiveram-se baixos, o que evidencia previsões estáveis.

No que diz respeito às RNA os resultados também evidenciaram uma extrema fiabilidade. O erro absoluto (MAE) variou entre 0,32 e 1,11. Estes valores mostram que o modelo consegue captar relações complexas e estimar previsões ajustadas. A função de custo foi minimizada durante o treino, contudo não é apresentada na Tabela 6, sendo unicamente utilizada para ajustar os pesos.

Com todas estas métricas é possível comprovar que os clusters identificados e as previsões dos Overall dos jogadores e o seu potencial de mercado apresentam um certo nível de robustez nos resultados obtidos, contudo, é necessário mais uma vez salientar que estes resultados são estimativas probabilísticas sujeitas à incerteza.

3.6.6 Interpretação do Overhall dos modelos com SHAP

De forma a aumentar a interpretabiliade dos modelos de ML utilizou-se a metodologia SHAP, para verificar visualmente quais os critérios que contribuem mais para a variável estimada overhall. Esta metodologia permite quantificar o contributo individual de cada critério para as previsões do modelo em que baseia-se na teoria dos jogos, atribuindo a cada critério um valor que representa o seu impacto marginal na estimação.

A análise SHAP seguiu os mesmos processos utilizados anteriormente, onde os critérios foram previamente selecionado com base na relevância para cada posição. Assim, nas figuras apresentadas, estão incluídos todos os critérios que são considerados relevantes para cada posição evidenciando depois a importância relativa dentro de cada grupo posicional, tanto para Random Forest como para Redes Neuronais Artificiais. Os resultados são apresentados sob a forma de gráficos SHAP, um para cada grupo posicional e para cada técnica utilizada, sendo que a cor indicativa do valor do critério e a sua posição no eixo dos Y corresponde ao impacto na previsão do modelo. As figuras 6 a 13 ilustram os resultados obtidos.

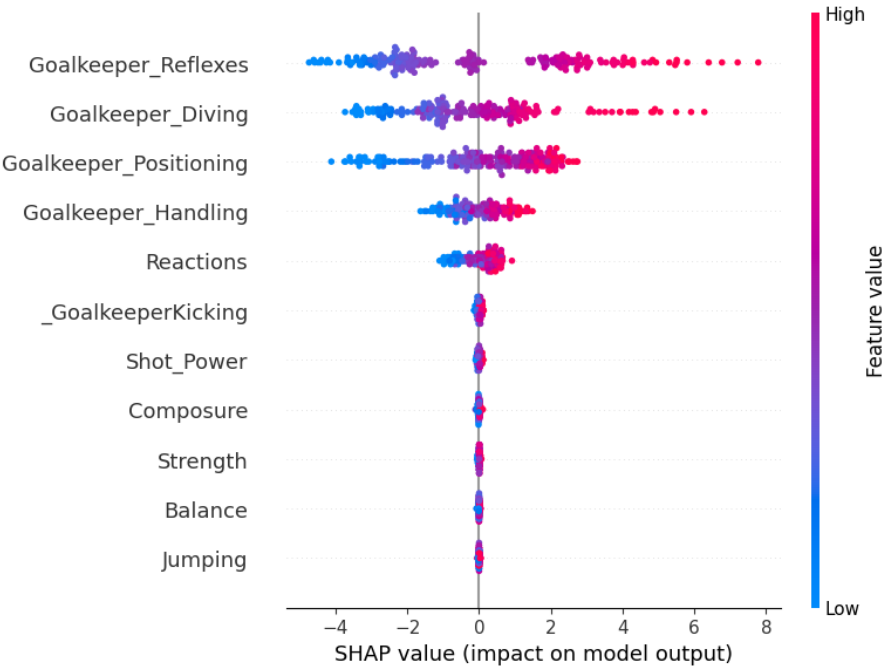


Figura 6 - SHAP Guarda Redes Random Forest



Figura 7 - SHAP Guarda Redes Redes Neuronais Artificiais

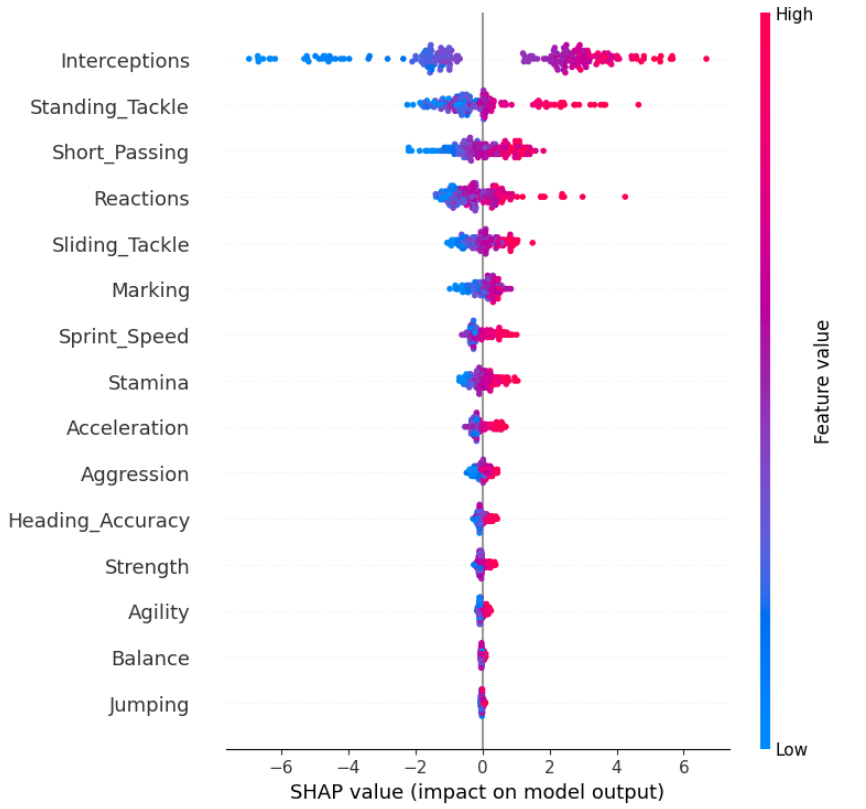


Figura 8 - SHAP Defesas Random Forest

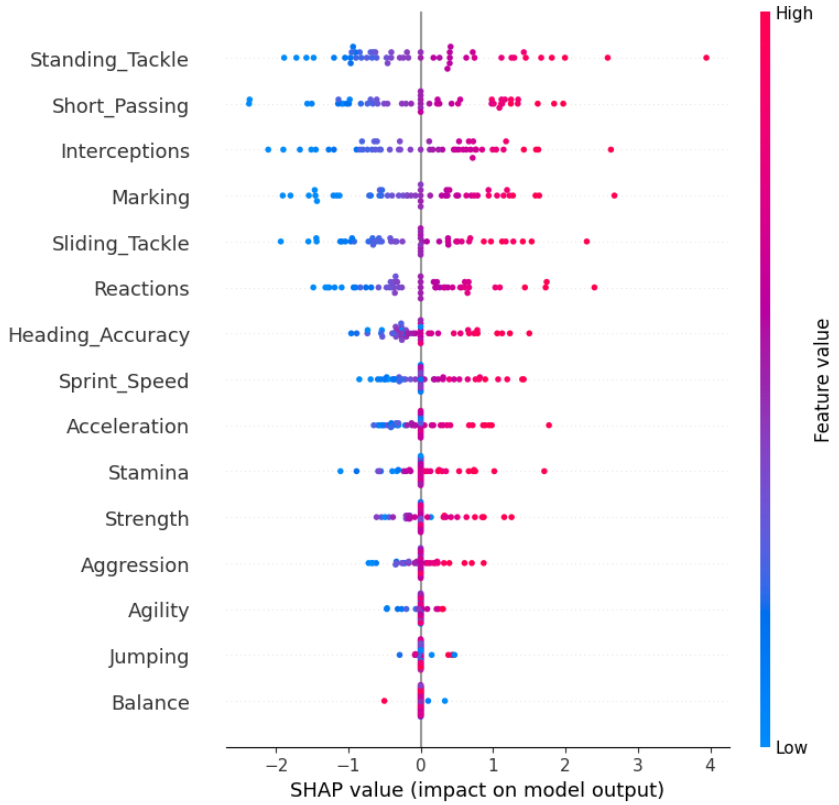


Figura 9 - SHAP Defesas Redes Neuronais Artificiais

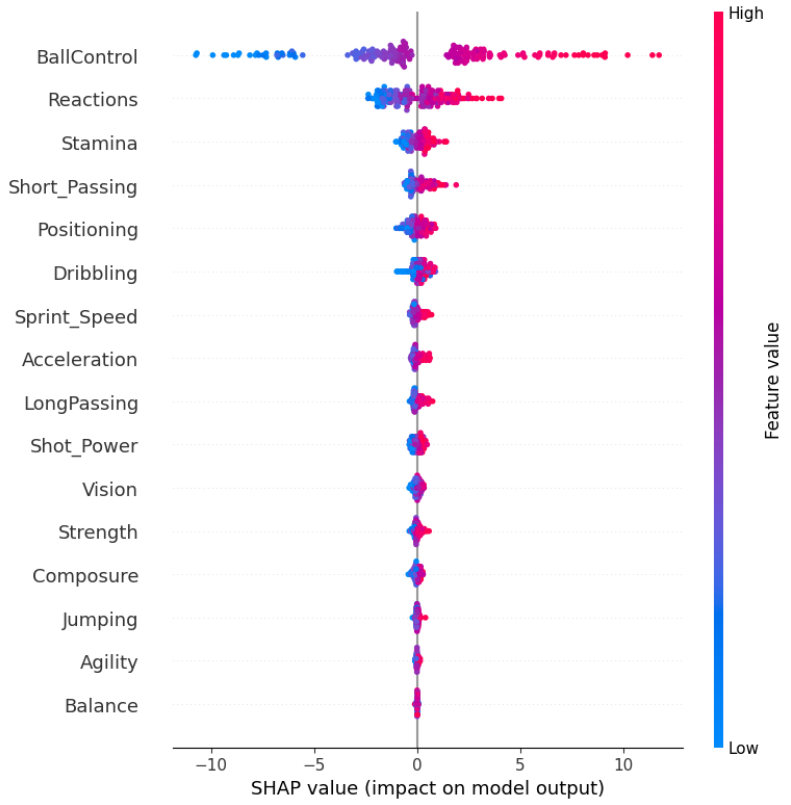


Figura 10 - SHAP Médios Random Forest

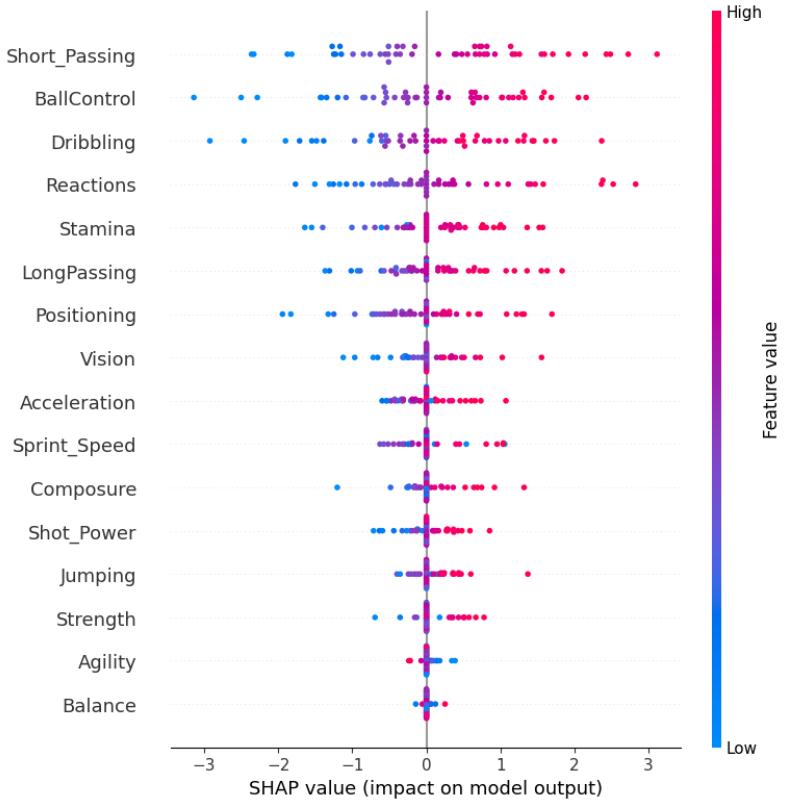


Figura 11 - SHAP Médios Redes Neuronais Artificiais

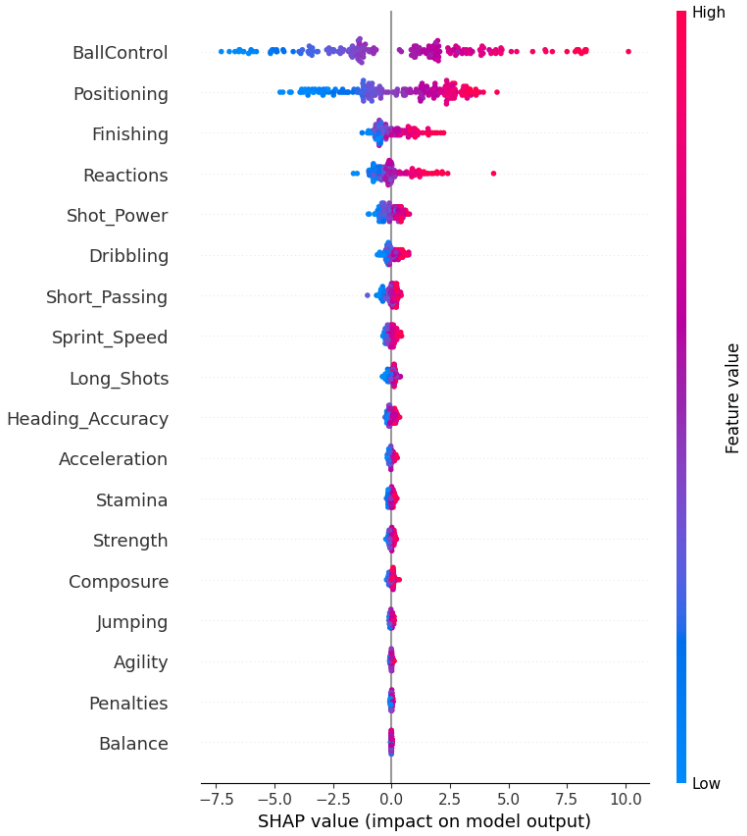


Figura 12 - SHAP Avançados Random Forest

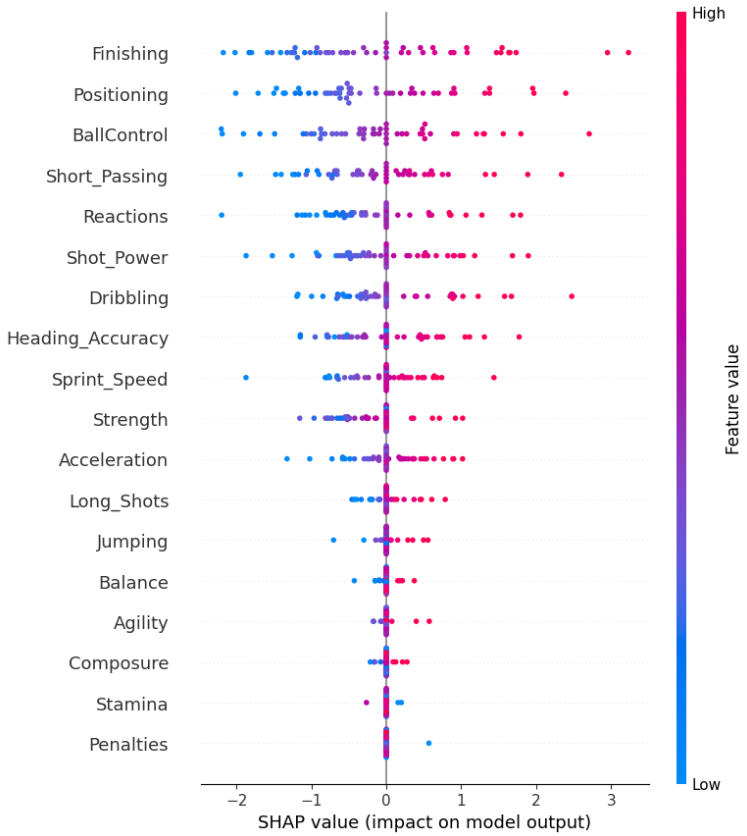


Figura 13 - SHAP Avançados Redes Neuronais Artificiais

A análise através da metodologia SHAP mostra padrões consistentes entre ambos os modelos, permitindo identificar os critérios com maior contributo dentro dos critérios já anteriormente selecionados para as respetivas posições

Conforme evidencia nas Figura 6 e Figura 7 relativas ao grupo de Guarda Redes, para além dos critérios que são naturalmente específicos para estes, destaca-se o critério Reação. Já referente ao grupo de Defesas, representado pela Figura 8 e Figura 9, destacam-se as competências mais defensivas, desde Interceção, Desarme em Pé, Marcação, Passe Curto, Desarme em Carrinho e Reação. Por outro lado, no grupo dos Médios, vemos como já mencionado anteriormente uma maior heterogeneidade nos critérios como mostra a Figura 10 e Figura 11, destacando critérios como Controlo de Bola, Passe Curto, Drible, Passe Longo, Resistência, Posicionamento, Visão e Aceleração, evidenciando uma diversidade face às outras posições. Por fim, no grupo dos Avançados temos critérios mais ofensivos como mostra a Figura 12 e Figura 13, destacando critérios como Finalização, Posicionamento, Controlo de Bola, Reação, Precisão de Cabeceamento e Velocidade em Sprint.

De forma geral, os resultados obtidos apresentam uma elevada consistência, reforçando a sua robustez das conclusões e evidenciando que diferentes abordagens de ML levam a identificação dos principais critérios para os respetivos grupos posicionais.

3.6.7 Avaliação da Qualidade da Previsão do Valor de Mercado

Para verificarmos a qualidade do nosso modelo de RNA, onde este estima o valor de mercado económico dos jogadores, recorremos a métricas já anteriormente utilizadas, sendo elas as MAE e RMSE.

Para realizar estas métricas, utilizamos a diferença dos nossos valores estimados com os valores da base de dados, de modo a verificar se o nosso modelo tem uma aproximação das previsões observadas na base de dados aqui estudada. O MAE, permite uma interpretação direta do erro médio das previsões, somando a diferença de todos os jogadores e posteriormente a sua média, indicando deste modo a sua capacidade preditiva global como indica a Tabela 7. O RMSE, penaliza de forma mais acentuada os erros de maior magnitude, ou seja, multiplica ao quadrado as suas diferenças para identificar a existência de previsões significativamente mais afastadas dos valores da base de dados como mostra a Tabela 7.

Valor de mercado	
MAE	RMSE
1 003 724,56 €	3 411 379,78 €

Tabela 7 - Métricas de avaliação do modelo (MAE e RMSE)

Conseguimos assim apresentar, de certa forma um desempenho global satisfatório, face aos valores praticados das transferências atuais, contudo existem situações específicas em que a previsão se afasta de forma mais acentuada do valor observado. Estes valores, poderão estar relacionados a jogadores com valores de mercado extremos, tais como jovens promessas ou jogadores altamente valorizados, cujos preços podem estar associados a fatores subjetivos e contextuais como já discutido anteriormente.

De forma geral, os resultados apresentados na Tabela 7, demonstram que o modelo de RNA apresenta uma capacidade adequada para estimar o valor de mercado dos jogadores, constituindo

uma ferramenta relevante de apoio para a tomada de decisão. Ainda assim, é sempre importante referir que todas as previsões são de contexto probabilístico, estando sempre sujeitas à incerteza, especialmente no mercado de jogadores de futebol com enorme volatilidade, com jogadores altamente valorizados ou jovens promessas, cujo valor é influenciado por fatores contextuais e subjetivos em que o modelo não consegue captar.

4 DISCUSSÃO E RESULTADOS

4.1 Importância dos critérios na avaliação dos jogadores

Os resultados obtidos através das análises realizadas permitiu identificar um conjunto de critérios que evidenciasse uma maior relevância para a investigação presente. A análise de regressão múltipla revelou que as variáveis associadas à idade e atributos técnicos e físicos evidenciam uma forte relação com o valor de mercado, verificando-se também a presença de efeitos não lineares através da inclusão dos termos quadráticos. Em particular, a idade revelou um comportamento típico do período de vida do jogador. Os coeficientes positivos associados à idade e os coeficientes quadráticos negativos indicam que o valor de mercado tende a aumentar no início da carreira profissional até atingir um ponto máximo e depois diminuindo até chegar ao fim da sua carreira enquanto jogador. Este resultado coincide com o que diz a literatura, que identifica a idade como um dos principais determinantes do valor de mercado (Kirschstein & Liebscher, 2019).

Na mesma análise, foi possível identificar que diferentes critérios apresentam níveis de relevância distintos consoante a posição do jogador. Por exemplo, critérios associados ao drible, controlo de bola e remate mostram uma maior relevância para jogadores que jogam nas posições de médios e avançados, enquanto critérios associados a interceções ou desarme em pé mostram uma maior importância para jogadores nas posições de defesas.

Assim, os resultados desta investigação sugerem de que a avaliação dos jogadores deve ser realizada através de uma abordagem multicritério, uma vez que diferentes dimensões de desempenho contribuem de forma distinta para diferentes posições do campo.

4.2 Comparação com os dados recolhidos

Para validar esta análise e complementá-la, foram recolhidos dados empíricos através de um questionário dirigido a profissionais ligados ao futebol, sendo estes elementos do clube do Sporting Clube de Braga. O objetivo desta recolha de dados consistiu em identificar quais os critérios considerados por parte de profissionais no processo de avaliação e contratação dos jogadores. Foram obtidas 7 respostas válidas, provenientes de elementos da equipa de scouting, todos eles envolvidos em processos de análise e avaliação de jogadores. Os resultados foram bastantes satisfatórios,

apresentando uma elevada consistência com os resultados anteriormente estimados. Em geral, os critérios estimados foram coincidentes com aqueles que os profissionais identificaram para as quatro posições aqui estudadas. No caso dos Guarda Redes, entre os 11 critérios mais relevantes identificados pelo modelo estimado, verificou-se a correspondência de 9 critérios com os apontados pelos profissionais como mostra a Tabela 8.

Guarda Redes = 11	
Estimação	Profissionais
Mergulho Guarda Redes	Mergulho Guarda Redes
Manuseamento Guarda Redes	Manuseamento Guarda Redes
Pontapé Guarda Redes	Pontapé Guarda Redes
Posicionamento Guarda Redes	Posicionamento Guarda Redes
Reflexos Guarda Redes	Reflexos Guarda Redes
Equilibrio	Equilibrio
Reações	Reações
Calma	Calma
Impulsão	Impulsão
Força	Posicionamento
Potencia de Remate	Agilidade

*A verde: critérios coincidentes; a amarelo: critérios divergentes.

Tabela 8 - Comparação entre critérios estimados e critérios identificados pelos profissionais (GR)

No caso dos Defesas, entre os 15 critérios mais relevantes identificados pelo modelo estimado, verificou-se a correspondência de 13 critérios com os apontados pelos profissionais como mostra a Tabela 9.

Defesas = 15	
Estimação	Profissionais
Intercepções	Intercepções
Força	Força
Desarme em pé	Desarme em pé
Carrinho	Carrinho
Precisão Cabeceamento	Precisão Cabeceamento
Marcação	Marcação
Agressividade	Agressividade
Equilíbrio	Equilíbrio
Impulsão	Impulsão
Passe Curto	Passe Curto
Agilidade	Agilidade
Resistência	Resistência
Reações	Reações
Aceleração	Calma
Velocidade Sprint	Posicionamento

*A verde: critérios coincidentes; a amarelo: critérios divergentes.

Tabela 9 - Comparação entre critérios estimados e critérios identificados pelos profissionais (DEF)

Em relação ao Médios, entre os 16 critérios mais relevantes identificados pelo modelo estimado, verificou-se a correspondência de 15 critérios com os apontados pelos profissionais como mostra a Tabela 10. Importa referir que, das respostas dos profissionais, a tradução para pontuações quantificáveis, resultou em seis critérios com a mesma pontuação pelo que foram incluídos na tabela para não haver interpretações enviesadas.

Médios = 16	
Estimação	Profissionais
Posicionamento	Posicionamento
Visão de Jogo	Visão de Jogo
Passe Curto	Passe Curto
Passe Longo	Passe Longo
Controlo de Bola	Controlo de Bola
Resistência	Resistência
Agilidade	Agilidade
Drible	Drible
Velocidade Sprint	Velocidade Sprint
Equilibrio	Equilibrio
Força	Força
Calma	Calma
Reações	Reações
Acelaração	Acelaração **
Potência de Remate	Potência de Remate **
Impulsão	Cruzamento
	Finalização**
	Precisão de Livres**
	Efeito **
	Precisão de Cabeceamento **
	Remate de Primeira **
	Remate de Longa Distância **

* A verde: critérios coincidentes; a amarelo: critérios divergentes.
** Critérios com a mesma pontuação foram igualmente incluídos na tabela.

Tabela 10 - Comparação entre critérios estimados e critérios identificados pelos profissionais (MC)

Por fim, nos Avançados, entre os 18 critérios mais relevantes identificados pelo modelo estimado, verificou-se a correspondência de 15 critérios com os apontados pelos profissionais como mostra a Tabela 11. À semelhança do que ocorreu nos médios, também aqui quatro critérios apresentaram a mesma pontuação na tradução das respostas, pelo que foram incluídos na tabela.

Alinhamento Inferior

Avançados = 18	
Estimação	Profissionais
Acelaração	Acelaração
Velocidade Sprint	Velocidade Sprint
Agilidade	Agilidade
Impulsão	Impulsão
Resistência	Resistência
Força	Força
Finalização	Finalização
Precisão de Cabeacemento	Precisão de Cabeacemento
Potência de Remate	Potência de Remate
Remate de Longa Distância	Remate de Longa Distância
Pénaltis	Pénaltis
Drible	Drible
Posicionamento	Posicionamento **
Controlo de Bola	Controlo de Bola
Equilíbrio	Equilíbrio **
Passe Curto	Remate de Primeira
Reações	Efeito
Calma	Agressividade **
	Equilíbrio **

*A verde: critérios coincidentes; a amarelo: critérios divergentes.
** Critérios com a mesma pontuação foram igualmente incluídos na tabela.

Tabela 11 - Comparação entre critérios estimados e critérios identificados pelos profissionais (AVA)

Assim, a comparação entre os resultados do modelo estimado e os resultados obtidos do questionário junto dos profissionais reforça a validade do modelo proposto, demonstrando que os padrões identificados nos dados quantitativos estão alinhados com a experiência dos profissionais do futebol.

4.3 Identificação de perfis de jogadores através de clustering

A aplicação desta técnica permitiu identificar diferentes perfis de jogadores com base nos critérios já aqui referidos, evidenciando a existência de grupos distintos de jogadores. Numa primeira análise, verificou-se se a separação clara entre os Guarda Redes e "jogadores de campo" o que evidencia que os atributos associados a estas posições apresentam características bastante distintas, confirmando a elevada especificidade para esta posição face às outras posições de campo. Assim, mesmo sem impor um número de clusters previamente, o algoritmo foi capaz de identificar uma divisão natural presente, reforçando a validade da abordagem utilizada.

Posteriormente, ao refinar a análise para as quatro posições de campo que queríamos analisar, foi possível identificar padrões de desempenho específicos para cada um destes grupos. Destaca-se particularmente no caso dos Médios, em que estes evidenciam uma maior dispersão dentro do seu grupo. Este resultado comprova uma elevada heterogeneidade refletindo a diversidade de papeis que os médios podem assumir, seja mais defensivos, organizadores ou ofensivos. Esta evidência tem implicações importantes, indicando que a avaliação de jogadores nesta posição deve ser mais contextualizada e menos padronizada.

Para além da dimensão posicional em que se criou a identificação de perfis de qualidade, foi acrescentada uma camada adicional para a análise, que permitiu distinguir jogadores com base no seu nível de desempenho e potencial de evolução. Esta segmentação é particularmente relevante do ponto de vista de gestão, pois pode permitir apoiar decisões estratégicas como recrutamento, desenvolvimento e retenção de talento. Assim, jogadores classificados como "Promissores", assumem uma especial importância para clubes que pretendam jogadores com potencial de valorização futura. Por outro lado, jogadores classificados como "Estrelas", refletem jogadores onde é pretendido um contributo imediato para o seu desempenho desportivo.

Em conjunto, estes resultados demonstram que a utilização de técnicas de clustering permite identificar padrões nos dados, mas também gerar outros insights relevantes para a tomada de decisão. Esta segmentação contribui para uma melhor compreensão do perfil dos jogadores, alinhando-se com a necessidade crescente de abordagens analíticas no futebol.

4.4 Estimativa do potencial dos jogadores através de ML

Os modelos de ML utilizados nesta investigação, mais concretamente, Random Forest e Redes Neuronais Artificiais, permitiu estimar o potencial que os jogadores podem atingir em termos de overhall. Os resultados mostram uma elevada capacidade preditiva dos modelos, evidenciada pelos baixos valores calculados das métricas de erro e também pelos elevados níveis de explicação da variabilidade dos dados. Em particular, o modelo de RF revelou-se ser o mais eficaz na identificação da importância relativa dos diferentes critérios.

Por sua vez, as RNA demonstraram uma elevada capacidade de captar relações não lineares complexas entre os critérios, uma característica distintiva desta técnica, que lhe permite modelar padrões complexos, produzindo estimativas do potencial parecidas às obtidas pelo RF. Assim, evidencia a consistência entre estes dois modelos, o que reforça a robustez dos resultados obtidos.

A utilização da metodologia SHAP possibilitou aprofundar a interpretação dos modelos de ML. Os resultados evidenciaram padrões alinhados, tanto como a literatura, tanto como os aqui estimados pelas técnicas aqui utilizadas. A consistência entre os resultados obtidos tanto por RF e RNA vem reforçar ainda mais a robustez das conclusões.

Assim podemos afirmar que a utilização destas técnicas de ML, pode constituir uma ferramenta valiosa para usar no processo de tomada de decisão, permitindo identificar jogadores com elevado potencial de desenvolvimento.

4.5 Estimativa do valor de mercado dos jogadores

A estimativa do valor do mercado foi realizado utilizando ML, neste caso unicamente as RNA, onde já explicado anteriormente, devido a este método conseguir captar relações não lineares, visto que existem diversas variáveis difíceis de mensurar e que dependem de diversos contextos em que o jogador está inserido (Mourao, 2016). Enquanto as regressões permitiram identificar relações estatísticas entre variáveis específicas e o valor de mercado, as RNA conseguem captar relações mais complexas entre os diferentes critérios da BD.

Os resultados das RNA, demonstram que o modelo apresenta uma capacidade satisfatória para aproximar os valores observados na BD aqui utilizada. Contudo verificou-se algumas diferenças entre

os valores estimados e os valores observados em certos jogadores. Também se verificou os dados de mercado junto de outra BD externa relevante, concretamente a plataforma Transfermarket. Verificou-se junto desta que os valores estimados encontram-se bastante satisfatórios, com apenas um jogador a mostrar uma grande diferença de valor face ao estimado.

Estas diferenças podem ser explicadas pela existência de fatores externos, dos quais os nossos critérios não conseguem captar, desde a visibilidade mediática que o jogador tem, o prestígio do clube ou a competitividade da liga aonde este se insere (Mourao, 2016).

Embora o modelo apresente uma boa capacidade de previsão, os resultados devem ser interpretados sempre como estimativas probabilísticas e não como valores precisos.

4.6 Implicações para a tomada de decisão na contratação de jogadores

Os resultados desta investigação mostram que a integração de métodos quantitativos e de perceções dos profissionais da área, podem contribuir para uma abordagem mais estruturada e racional no processo complexo que é a contratação dos jogadores de futebol.

A combinação de técnicas econométricas, de clustering, RF e RNA, permitiram identificar padrões de desempenho, estimar o seu potencial e avaliar o valor de mercado de forma mais sistemática. Complementando com a experiência prática de profissionais do futebol, com os dados recolhidos dos elementos do SC Braga, estes modelos podem fornecer uma base para apoiar decisões na contratação dos jogadores.

Desta forma, os modelos aqui desenvolvidos nesta investigação podem ser utilizados para apoiar o scouting e a gestão desportiva, de modo a reduzir o risco da contratação e melhorar a eficiência do investimento nos jogadores.

5 CONCLUSÕES, LIMITAÇÕES E INVESTIGAÇÃO FUTURA

5.1 Conclusões

O presente estudo teve como seu principal objetivo analisar de que forma modelos de apoio à decisão utilizando multicritério podem contribuir para melhorar o processo de tomada de decisão na contratação de jogadores de futebol. Num contexto caracterizado por elevada incerteza, forte pressão competitiva e investimento financeiros cada vez mais elevados, torna-se fundamental desenvolver ferramentas analíticas capazes de apoiar decisões estratégicas dos clubes, reduzindo o risco associado às contratações.

Para atingir este objetivo, foi desenvolvido um modelo que combina diferentes abordagens quantitativas, nomeadamente regressão linear múltipla e técnicas de Machine Learning, sendo este com modelos não supervisionados (Clustering) e modelos supervisionados (Random Forest e Redes Neuronais Artificiais). Esta abordagem permitiu não só analisar os principais critérios associados ao desempenho, como estimar o potencial futuro e segmentar perfis de forma sistemática.

Os resultados obtidos mostram que o valor de mercado dos jogadores é influenciado por um conjunto de diferentes critérios, sendo eles técnicos, físicos, mentais e psicológicos, sendo que a sua importância varia consoante a posição aonde este atua em campo. A análise econométrica revelou que critérios como idade, critérios técnicos e critérios físicos apresentam uma relação estatisticamente significativa com o valor de mercado, confirmando os resultados encontrados na literatura. Adicionalmente, verificou-se a existência de relações não lineares entre algumas variáveis e o valor de mercado, particularmente na idade do jogador, cujo impacto do valor de mercado tende a aumentar na fase inicial da sua carreira e começa a diminuir na sua fase final. Este resultado evidencia a importância de considerar o ciclo de "vida" do jogador no processo de avaliação na contratação do jogador de futebol.

A aplicação de técnicas de clustering permitiu identificar diferentes conjuntos de jogadores com base nos seus critérios. A análise evidenciou uma clara distinção entre os Guarda Redes face às restantes posições aqui estudadas, bem como posteriormente, quando se dividiu pelas quatro posições. Em particular, verificou-se que os médios apresentam uma maior diversidade de critérios, evidenciando a variedade de funções táticas em campo associadas a esta posição.

Referente aos modelos de ML, tanto o Random Forest como as Redes Neuronais Artificias demonstraram uma elevada capacidade de prever uma estimativa do potencial dos jogadores. A consistência obtida nos dois modelos reforça a robustez das estimativas apresentadas, o que sugere que estas técnicas podem constituir ferramentas úteis no processo de avaliação dos jogadores.

Adicionalmente e de modo a validar os dados obtidos através dos modelos aqui executados, comparamos com os dados recolhidos junto dos profissionais do SC Braga, que revelaram uma elevada consistência entre os critérios identificados pelos modelos. Esta convergência entre evidência estatística e perceção prática vem reforçar a validade do modelo desenvolvido.

Concluindo, podemos afirmar que os resultados deste estudo demonstram que a utilização de metodologias quantitativas pode contribuir significativamente para tornar o processo de tomada de decisão na contratação de jogadores mais estruturado e fundamentado. Embora estes modelos não substituam a experiência e a intuição dos profissionais do futebol, poderá funcionar como um complemento analítico, permitindo reduzir a incerteza associada a decisões envoltas em grande impacto financeiro e desportivo.

5.2 Limitações do Estudo

Apesar das contribuições apresentadas, este estudo apresenta algumas limitações que devem ser consideradas na interpretação dos resultados obtidos.

A primeira limitação prende-se pela base de dados utilizada. Apesar de os dados serem referentes a um jogo líder de simulação de futebol, este baseia-se em avaliação subjetivas realizadas por especialistas e observadores de futebol. Assim, embora exista um esforço significativo de validação por parte da empresa deste jogo, as classificações podem não refletir totalmente o valor verdadeiro para os critérios identificados na base de dados, os quais são aqui utilizados para criar estes modelos e que pode levar a enviesamentos nos dados e consequentemente afetar a precisão dos modelos preditivos.

A segunda limitação prende-se pela ausência de alguns critérios considerados importantes pela literatura sobre a avaliação dos jogadores. Em particular, fatores psicológicos e comportamentais, como liderança, capacidade de adaptação ou visibilidade mediática são difíceis de mensurar e não se encontram na base de dados utilizada. Como resultado desta limitação, estes elementos não

puderam ser incorporados diretamente nos modelos aqui desenvolvidos. O valor de mercado dos jogadores, também é influenciado por diversos fatores externos que não são captados pelos critérios utilizados neste estudo. Elementos como reputação do clube, a competitividade da liga, o histórico de conquistas do clube onde este se insere podem ter um impacto bastante significativo na valorização do jogador no mercado de transferências.

A terceira limitação diz respeito à natureza probabilística dos modelos de ML aqui utilizados. Embora estes modelos apresentem uma elevada capacidade preditiva conforme verificado na avaliação da qualidade dos mesmos, as suas estimativas são puramente previsões determinísticas, não devendo ser interpretadas como valores exatos. Num contexto como o futebol, onde o desempenho dos jogadores pode ser influenciado por tantos fatores imprevisíveis, como lesões, adaptações táticas ou contexto competitivo, qualquer previsão dever ser encarada com um certo grau de incerteza.

Por fim, a quarta limitação deve-se ao número de respostas obtidas no questionário a profissionais do futebol, que embora relevante para complementar a análise, pode não ser demasiado representativo devido ao número limitado de respostas obtidas.

5.3 Investigação Futura

Foram identificadas limitações existentes no presente estudo, bem com diversas oportunidades para futuras investigações nesta área.

Em primeiro lugar, investigações futuras poderão utilizar BD mais detalhadas provenientes de plataformas de análise do desempenho desportivo que o jogador obtém ao jogar nos jogos oficiais, como dados de tracking ou estatísticas avançadas do jogo. Assim, estas fontes de informação irão permitir captar de forma mais precisa o comportamento dos jogadores e contribuir para melhorar a precisão dos modelos analíticos.

Em segundo lugar, seria relevante adicionar variáveis difíceis de mensurar, relacionadas com fatores contextuais e comportamentais, como liderança, inteligência tática, capacidade de adaptação a diferentes ideias de jogo, bem como outras variáveis qualitativas relevantes. A incorporação destas variáveis permitiria construir um modelo mais robusto, mais completo e representativo da realidade que é o futebol profissional.

Outra linha de investigação futura, consistiria na aplicação de métodos de análise multicritério já

existentes na literatura, como AHP, SWARA ou VIKOR. A combinação destas abordagens poderia contribuir para desenvolver modelos híbridos que conciliem a capacidade explicativa dos métodos multicritério com o poder preditivo das técnicas de ML, proporcionando estimativas mais robustas.

Por fim, futuras investigações poderão explorar a aplicação prática destes modelos em contexto real, testando a sua utilidade no apoio de tomada de decisão. A validação empírica destes modelos em contextos reais permitirá avaliar de forma mais precisa o seu potencial contributo para a melhoria dos processos de tomada de decisão no futebol.

6 REFERÊNCIAS

Ati, A., Bouchet, P., & Ben Jeddou, R. (2024). Using multi-criteria decision-making and machine learning for football player selection and performance prediction: A systematic review. *Data Science and Management*, *7*(2), 79–88. https://doi.org/10.1016/j.dsm.2023.11.001

Belhouchet, H., Dergaa, I., Zoudji, B., Ceylan, H. İ., Stefanica, V., Khatrouch, I., Imane, G., Acoumambo, D., Bragazzi, N. L., & Azaiez, F. (2026). Multi-criteria decision-making in soccer: A bibliometric analysis. *Humanities and Social Sciences Communications*, *13*(1), 595. https://doi.org/10.1057/s41599-026-06968-9

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Çaloğlu Büyükselçuk, E., & Badem, E. (2024). Player selection in football by integrated SWARA-VIKOR methods under fuzzy environment. *Heliyon*, *10*(12), e33087. https://doi.org/10.1016/j.heliyon.2024.e33087

Collins, D., Collins, L., & Carson, H. J. (2016). "If It Feels Right, Do It": Intuitive Decision Making in a Sample of High-Level Sport Coaches. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00504

Di Minin, A., Frattini, F., Bianchi, M., Bortoluzzi, G., & Piccaluga, A. (2014). Udinese Calcio soccer club as a talents factory: Strategic agility, diverging objectives, and resource constraints. *European Management Journal*, *32*(2), 319–336. https://doi.org/10.1016/j.emj.2013.04.001

Doumpos, M., & Zopounidis, C. (2011). Preference disaggregation and statistical learning for multicriteria decision support: A review. *European Journal of Operational Research*, *209*(3), 203–214. https://doi.org/10.1016/j.ejor.2010.05.029

FIFA 17 player ratings blend stats, subjectivity. (2016, setembro 26). ESPN.Com. https://global.espn.com/football/story/_/id/37456940/fifa-17-player-ratings-system-blends-advanced-stats-subjective-scouting

FourFourTwo Staff. (2021, outubro 7). *FIFA 22: Player ratings explained*. FourFourTwo. https://www.fourfourtwo.com/features/fifa-22-player-ratings-explained-chemistry-fut-ultimate-team-pace-passing-speed-card

Giske, R., Benestad, B., Haraldstad, K., & Høigaard, R. (2013). Decision-Making Styles among Norwegian Soccer Coaches: An Analysis of Decision-Making Style in Relation to Elite and Non-Elite Coaching and Level of Playing History. *International Journal of Sports Science & Coaching*, *8*(4), 689–701. https://doi.org/10.1260/1747-9541.8.4.689

Hecksteden, A., Kempe, M., & Berger, J. (2025). Perspectives on data analytics for gaining a competitive advantage in football: Harnessing data for decision support. *Science and Medicine in Football*, *10*(2), 249–257. https://doi.org/10.1080/24733938.2025.2517056

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C., Silverman, R., & Wu, A. Y.

(2000). The Analysis of a Simple k-Means Clustering Algorithm. *2000*.

Kirschstein, T., & Liebscher, S. (2019). Assessing the market values of soccer players – a robust analysis of data from German 1. And 2. Bundesliga. *Journal of Applied Statistics*, *46*(7), 1336–1349. https://doi.org/10.1080/02664763.2018.1540689

Klatt, S., Noël, B., Musculus, L., Werner, K., Laborde, S., Lopes, M. C., Greco, P. J., Memmert, D., & Raab, M. (2019). Creative and Intuitive Decision-Making Processes: A Comparison of Brazilian and German Soccer Coaches and Players. *Research Quarterly for Exercise and Sport*, *90*(4), 651–665. https://doi.org/10.1080/02701367.2019.1642994

Mourao, P. R. (2016). Soccer transfers, team efficiency and the sports cycle in the most valued European soccer leagues – have European soccer teams been efficient in trading players? *Applied Economics*, *48*(56), 5513–5524. https://doi.org/10.1080/00036846.2016.1178851

Nasiri, M. M., Ranjbar, M., Tavana, M., Santos Arteaga, F. J., & Yazdanparast, R. (2019). A novel hybrid method for selecting soccer players during the transfer season. *Expert Systems*, *36*(1), e12342. https://doi.org/10.1111/exsy.12342

Nikić, A., Topalović, A., & Bach, M. P. (2024). From Data to Decision: Machine Learning in Football Team Management. *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, 1059–1064. https://doi.org/10.1109/MIPRO60963.2024.10569835

Oshiro, T., Perez, P., & Baranauskas, J. (2012). How Many Trees in a Random Forest? Em *Lecture notes in computer science* (Vol. 7376). https://doi.org/10.1007/978-3-642-31537-4_13

Ozceylan, E. O. (2016). A MATHEMATICAL MODEL USING AHP PRIORITIES FOR SOCCER PLAYER SELECTION: A CASE STUDY. *South African Journal of Industrial Engineering*, *27*(2). https://doi.org/10.7166/27-2-1265

Patel, P. Z., Vekariya, D. V., & Sutaria, D. K. (2025). ILLUMINATING THE BLACK BOX: A DEEP DIVE INTO SHAP FOR EXPLAINABLE AI. *LEX LOCALIS*, *23*.

Pinto, P. (2025, dezembro 6). Inteligência Artificial: Sabe o que é Overfitting vs. Underfitting? *Pplware*. https://pplware.sapo.pt/inteligencia-artificial/inteligencia-artificial-sabe-o-que-e-overfitting-vs-underfitting/

Sánchez, L. C., Barajas, Á., & Sanchez-Fernandez, P. (2020). Profits may lead teams to lose matches, but scoring goals does not lead to profit. *European Research on Management and Business Economics*, *26*(1), 26–32. https://doi.org/10.1016/j.iedeen.2019.12.007

Santana, E. H. D. S., Melo, F. R. D., & Reais, E. F. D. (2025). Implementação de uma rede neural artificial no processo de recomendação de cultura baseada em dados edafoclimáticos. *OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA*, *23*(1), e8542. https://doi.org/10.55905/oelv23n1-033

Serna Rodríguez, M., Ramírez Hassan, A., & Coad, A. (2019). Uncovering Value Drivers of High Performance Soccer Players. *Journal of Sports Economics*, *20*(6), 819–849. https://doi.org/10.1177/1527002518808344

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748. https://doi.org/10.1109/DSAA49011.2020.00096

7 APÊNDICES

Apêndice A

A1 - Number, DOI, Autor, Título do artigo, Ano

1	10.1016/j.dsm.2023.11.001	Ati et al	Using multi-criteria decision-making and machine learning for football player selection and performance prediction: A systematic review. *Data Science and Management*	2024
2	10.1016/j.heliyon.2024.e33087	Çaloğlu Büyükselçuk & Badem	Player selection in football by integrated SWARA-VIKOR methods under fuzzy environment. *Heliyon*	2024
3	10.3389/fpsyg.2016.00504	Collins et al	"If It Feels Right, Do It": Intuitive Decision Making in a Sample of High-Level Sport Coaches. *Frontiers in Psychology*	2016
4	10.1016/j.emj.2013.04.001	Di Minin et al	Udinese Calcio soccer club as a talents factory: Strategic agility, diverging objectives, and resource constraints. *European Management Journal*	2014
5	10.1260/1747-9541.8.4.689	Giske et al	Decision-Making Styles among Norwegian Soccer Coaches: An Analysis of Decision-Making Style in Relation to Elite and Non-Elite Coaching and Level of Playing History. *International Journal of Sports Science & Coaching*,	2013
6	10.1080/02701367.2019.1642994	Klatt et al	Creative and Intuitive Decision-Making Processes: A Comparison of Brazilian and German Soccer Coaches and Players. *Research Quarterly for Exercise and Sport*,	2019

7	10.1016/j.entcom.2024.100913	Liu & Liu	The application of artificial intelligence technology in the tactical training of football players. *Entertainment Computing*
8	10.1080/00036846.2016.1178851	Mourao	Soccer transfers, team efficiency and the sports cycle in the most valued European soccer leagues – have European soccer teams been efficient in trading players? *Applied Economics*
9	10.1111/exsy.12342	Nasiri et al	A novel hybrid method for selecting soccer players during the transfer season. *Expert Systems*,
10	10.7166/27-2-1265	Ozceylan	A MATHEMATICAL MODEL USING AHP PRIORITIES FOR SOCCER PLAYER SELECTION: A CASE STUDY. *South African Journal of Industrial Engineering*
11	10.1016/S0167-9236(02)00083-0	Ramos et al	Soccer strategies that live in the B2B world of negotiation and decision-making. *Decision Support Systems*,
12	10.1016/j.iedeen.2019.12.007	Sánchez et al	Profits may lead teams to lose matches, but scoring goals does not lead to profit. *European Research on Management and Business Economics*
13	10.1016/j.dss.2018.08.006	Zambom-Ferraresi et al	Determinants of sport performance in European football: What can we learn from the data? *Decision Support Systems*,

14	10.1080/02664763.2018.1540689	Kirschstein & Liebscher	Assessing the market values of soccer players – a robust analysis of data from German 1. and 2. Bundesliga	2019
15	10.1177/1527002518808344	Serna Rodríguez et al	Uncovering Value Drivers of High Performance Soccer Players	2019
16	10.2202/1559-0410.1238	Tunaru & Viney	Valuations of Soccer Players from Statistical Performance Data	2010
17	10.1016/j.ejor.2010.05.029	Doumpos & Zopounidis	Preference disaggregation and statistical learning for multicriteria decision support: A review	2010
18	0.1080/24733938.2025.2517056	Anne Hecksteden, Matthias Kempe & Julian Berger	Perspectives on data analytics for gaining a competitive advantage in football: harnessing data for decision support	2025
19	10.1109/MIPRO60963.2024.10569835	Anđelko Nikić, Amir Topalović & Mirjana Pejić Bach	From Data to Decision: Machine Learning in Football Team Management	2024
20	Revista Lex Localis, Vol. 23, No. S6, pp. 6493–6500	Piyush Zaverbhai Patel, Vipul Vekariya & Kruti Sutaria	Illuminating the black box: A deep dive into shap for explainable AI	2025
21	10.1057/s41599-026-06968-9	Hatem Belhouchet, Ismail Dergaa, Bachir Zoudji, Halil İbrahim Ceylan, Valentina Stefanica, Ikram Khatrouch, Ghalmi Imane, Dimitri Acoumambo, Nicola Luigi Bragazzi & Fairouz Azaiez	Multi-criteria decision-making in soccer: a bibliometric analysis	2026

A2 - Objetivos do artigo, Metodologia, Método de recolha de dados, Métodos para tratamento de análise de dados, "Setting Context".

• Revisão sistemática literatura (SLR) • Utilização do machine learning • Machine Learning para seleção dos melhores jogadores	Revisão sistemática da literatura	5 bases de dados científicas: IEEE, ScienceDirect, Springer, Taylor & Francis, Wiley	Análise de 66 artigos relevantes através de filtragem por palavras-chave, resumos, conteúdos e categorização por perguntas de investigação.	Decisão multicritério e machine learning para apoiar técnicos, gestores e recrutadores
Avaliar e selecionar objetivamente os jogadores mais adequados para a Seleção Nacional da Turquia utilizando métodos de decisão multicritério (SWARA e VIKOR) com base em números fuzzy neutrosóficos.	Aplicação integrada dos métodos SWARA e VIKOR sob ambiente fuzzy (números neutrosóficos).	Dados retirados da base de dados do site fminside (Jogo de Simulação), com pontuação de atributos dos jogadores (1–20), avaliados por 3 treinadores e 2 dirigentes.	Conversão dos dados em valores crisp (valores exatos, contrario de fuzzy); agregação dos juízos dos decisores; cálculo de pesos com SWARA; classificação de alternativas com VIKOR.	Contextualiza o declínio da Seleção Turca após 2008 e a necessidade de critérios objetivos para recuperar a estabilidade e sucesso.
Explorar o uso da tomada de decisão intuitiva por treinadores de desporto de alto nível, incluindo frequência, origem e interação com estilos deliberativos.	Qualitativa (entrevistas semi-estruturadas) Análise temática	Entrevistas com 18 treinadores de rugby e desportos de aventura no Reino Unido.	Transcrição literal, codificação indutiva, agrupamento em temas	Ambientes de alta exigência: rugby profissional e desportos de aventura, ambos com riscos e pressão elevados.
• Identificar jogadores promissores para obter no futuro receita para equilibrar as contas do clube. • Identificação de jogadores, não só talento técnico mas também estilo de vida • Demonstrar que a agilidade estratégica é crucial	Abordagem exploratória Longitudinal	Entrevistas semiestruturadas com dirigentes e treinadores da Udinese, jornalistas desportivos, documentação interna e análise de artigos publicados nos últimos 20 anos.	Análise qualitativa baseada na triangulação de fontes primárias e secundárias; com apoio do modelo teórico de agilidade estratégica	• Jogadores desconhecidos para valorizá-los para fazer encaixe financeiro • Trade-off entre rendimento desportivo e rendimento económico
Examinar os estilos de tomada de decisão dos treinadores noruegueses de futebol.	Estudo quantitativo	Questionário (General Decision-Making Style - GDMS)	Análise estatística (ANOVA, estatísticas descritivas, testes t, etc.)	Contexto do futebol norueguês, níveis elite e juvenil
Comparar os estilos de tomada de decisão criativa e intuitiva entre treinadores e jogadores de futebol do Brasil e da Alemanha.	Estudo qualitativo	Entrevistas semi-estruturadas com treinadores e jogadores de futebol	Análise de conteúdo qualitativo com categorização temática	Contexto desportivo e cultural do futebol no Brasil e na Alemanha

Propor um módulo tático baseado em IA que otimize o desempenho dos jogadores de futebol, via treino estratégico personalizado.	Desenvolvimento de modelo com redes de aprendizagem concatenadas, com aprendizagem recorrente e ANN	Dados de treino da base Football Talent Evaluation with Machine Learning do Kaggle (Site com a programação e explicação).	Algoritmos de IA: ANN, funções de ativação, aprendizagem recorrente, regressões baseadas em métricas e análise segmentada de desempenho.	Treinamento tático em futebol com apoio de IA; foco em personalização de treino por jogador com base em dados históricos e feedback contínuo.
Avaliar se os clubes europeus foram eficientes na venda de jogadores entre 2007 e 2013 e identificar determinantes para tal eficiência.	Análise de Fronteira Estocástica (SFA) com modelos de efeitos ineficientes; estimação por máxima verossimilhança (estima parametros de um modelo probabilístico).	Dados extraídos do site www.transfermarkt.com para clubes das ligas de Espanha, Inglaterra, Itália, Alemanha, França e Portugal.	Modelagem econométrica (SFA), regressão com painel de dados, estimação simultânea de sistemas com variáveis como títulos, saldo de transferências, pontos na liga, etc.	Transferências de jogadores no futebol europeu profissional; avaliação de 183 clubes entre 2007–2013.
Propor um modelo híbrido para selecionar jogadores de futebol durante a temporada de transferências, maximizando desempenho e minimizando custos.	Multicritério: Fuzzy ANP (analytic network process) + PROMETHEE II + Programação matemática multiobjetivo + DEA.	Questionário com especialistas e dados de um clube real (Irão).	Fuzzy ANP para pesar critérios, PROMETHEE II para pontuar jogadores, modelo de programação inteira biobjetivo resolvido com método ε-constraint, e DEA para selecionar solução eficiente.	Seleção de jogadores em clubes de futebol com restrições financeiras e operacionais (exemplo prático de um clube iraniano).
Determinar o 11 ideal através de um modelo de apoio de decisão com AHP e programação linear binária	Estudo quantitativo com modelagem matemática multicritério.	Dados reais de desempenho dos jogadores (qualitativos e quantitativos), extraídos do jogo Football Manager 2015; julgamentos de especialistas para pesos do AHP via entrevistas presenciais.	Aplicação do método AHP para priorização de critérios por posição, seguido de um modelo de programação linear binária (0-1) para selecionar a equipa ideal segundo múltiplos critérios e restrições operacionais (idade média, custo, número de estrangeiros, etc.).	Escolher o 11 com os criterios e testam com diferentes ajustes, tais como diminuição do orçamento e idade média da equipa
Desenvolver um modelo analógico entre estratégias de jogo no futebol e processos de decisão e negociação em ambientes B2B, com foco na resolução de conflitos e apoio à decisão.	Investigação teórica e desenvolvimento de simuladores em Java baseados em modelos de Q-learning e teoria dos jogos.	Simulação computacional	Q-learning (aprendizagem por reforço), matrizes de decisão	Negociação e tomada de decisão no contexto do e-business, comparando com dinâmicas do futebol
Analisar se há uma relação entre rentabilidade financeira e desempenho desportivo nos clubes de futebol europeus, e se essa relação é positiva ou negativa.	Regressão de equações simultâneas com método de três etapas (3SLS); comparado com OLS.	Dados financeiros e desportivos de 20 clubes europeus da "Deloitte Money League" de 2010 a 2017.	Regressão OLS, três estágios mínimos quadrados (3SLS), testes de Hausman e Hansen-Sargan para endogeneidade e instrumentos.	Clubes de elite do futebol europeu, com foco em como objetivos distintos dos donos (lucro vs. sucesso desportivo) afetam performance.
Analisar os determinantes de performance desportiva nas “Big Five” ligas europeias de futebol entre 2012/13 e 2014/15.	Modelagem estatística com Bayesian Model Averaging (BMA -> Lida com a incerteza da escolha) e análise de importância relativa (LMG, Genizi, CAR).	Dados da empresa **OptaPro**, que fornece estatísticas detalhadas de jogos.	Regressão linear com BMA (MC3 sampling), decomposição de variância com métricas LMG, Genizi e CAR.	Futebol europeu, ligas principais (“Big Five”), dados de três temporadas (2012/13 a 2014/15), 5532 jogos.

Identificar e quantificar os determinantes dos valores de mercado dos jogadores de futebol das duas principais ligas alemãs, com especial atenção a métodos robustos de estimação.	Quantitativa, com aplicação de regressões econométricas robustas.	Dados de jogadores das ligas 1. Bundesliga e 2. Bundesliga (desempenho, características pessoais, contratuais e do clube).	Regressão quantílica, regressão robusta, efeitos fixos por clube e posição, análise de sensibilidade e controlo de outliers.	Mercado de transferências da Alemanha, com foco nos jogadores profissionais das duas divisões principais.
Identificar os principais fatores que influenciam o valor de mercado de jogadores de futebol de alto desempenho, considerando diferenças entre posições.	Quantitativa, baseada em análise estatística multivariada.	Bases de dados estatísticas de jogadores de futebol profissionais com elevado desempenho em diversas ligas europeias.	Regressão multivariada para identificar variáveis explicativas significativas do valor de mercado por posição.	Mercado de futebol europeu, com foco em jogadores de alto desempenho e variáveis que influenciam a sua valorização.
Demonstrar que o valor de um clube de futebol como o Manchester United inclui opções reais e que estas explicam melhor o seu valor de mercado do que modelos tradicionais como DCF.	Estudo de caso com aplicação do modelo de opções reais ao contexto da valorização de clubes de futebol.	Dados financeiros públicos do Manchester United (ações, contas, histórico de desempenho).	Aplicação de modelos de opções reais (Black-Scholes e binomial); análise de sensibilidade sobre parâmetros como volatilidade, tempo, taxa livre de risco.	O artigo parte da observação de que o valor de mercado do Man. United excede em muito o valor contabilístico ou DCF, sugerindo que investidores valorizam as oportunidades futuras do clube.
Explorar a relação entre a desagregação de preferências (MCDA) e a aprendizagem estatística/ML, identificando semelhanças, diferenças e potencial de integração.	Revisão narrativa/conceptual	N/A	Comparação teórica/matemática dos dois paradigmas	Interseção crescente entre MCDA e Machine Learning, motivada pelo interesse mútuo dos dois campos.
Comentário crítico ao artigo de Olthof & Davis (2025), discutindo nuances sobre o valor da analítica de dados/ML para obter vantagem competitiva no futebol, com foco na tarefa de previsão como núcleo da tomada de decisão.	Artigo de opinião/comentário	Reanálise de dados de um estudo anterior (Hecksteden et al., 2023) em jogadores de futebol masculinos.	Aplicação de dois algoritmos de ML interpretável comparados com o modelo original "caixa-negra".	Apoio à decisão baseado em dados no futebol de elite. Discussão do fenómeno "Moneyball" e da tensão entre modelos complexos e interpretáveis.
Explorar como o Machine Learning (clustering) identifica padrões e estratégias entre as equipas europeias de futebol de topo, revelando as características associadas ao sucesso desportivo.	Estudo de caso quantitativo com clustering (K-means)	Base de dados secundária, descarregada do Kaggle, com dados de 98 equipas das 5 principais ligas europeias, época 2021-2022.	Clustering K-means, com número ótimo de clusters. Avaliação por Percentage Split. Duas análises: uma por torneio/liga (6 clusters) e outra por equipa (5 clusters).	Analítica desportiva e tomada de decisão na gestão de equipas de futebol europeias, comparando estilos de jogo entre ligas.
Explorar os fundamentos teóricos e as aplicações do SHAP em IA explicável, comparar as suas vantagens/desafios face a outras técnicas e identificar lacunas de investigação.	Revisão sistemática de literatura	Pesquisa em bases académicas (Scopus, Web of Science, IEEE Xplore, Google Scholar) e fontes complementares (blogs/relatórios técnicos de indústria); pesquisa restrita a 2017-2025	Análise qualitativa/síntese narrativa da literatura selecionada, organizada por domínios de aplicação (saúde, finanças, PLN, comportamento do consumidor).	Necessidade crescente de transparência e interpretabilidade em modelos de IA complexos ("caixa-negra"), sobretudo em domínios de alto risco (saúde, finanças, justiça).
Mapear a evolução da investigação sobre MCDM no futebol, tendências de publicação, autores/instituições influentes, redes de colaboração e clusters temáticos emergentes.	Análise bibliométrica, com abordagem descritiva	Pesquisa na Web of Science Core Collection (2000-2025), com estratégia de termos de tópico e de título (ex.: "football"/"soccer" + "MCDM"/"MCDA" + "AHP"/"TOPSIS").	Biblioshiny/Bibliometrix (indicadores descritivos e mapeamento científico) e VOSviewer (redes de co-autoria, co-citação e co-ocorrência de palavras-chave)	Crescente aplicação de métodos multicritério (MCDM/MCDA) na gestão do futebol (seleção de jogadores, avaliação de desempenho, planeamento estratégico), sem mapeamento abrangente do campo até então.

A3 - Termos de pesquisa, Palavras chave dos autores, Critérios eleitos, País de investigação, Resultados.

Soccer and Decision support system	Multi-criteria decision-making Machine learning Football player selection Managerial- nancial and sporting performance	• Competência técnicas • Competências táticas • Experiência de jogo profissional • Número médio de golos por jogo • Capacidade de coordenação com a equipa • Caracteristicas morais, comportamentais e prestigio social	Internacional (Razão: SRL)	Identificação das abordagens MCDM mais utilizadas (AHP, TOPSIS, PROMETHEE, etc.) Benefícios da combinação com machine learning para seleção e previsão de desempenho
Multi criteria decision analytics (MCDA)	Decision making Football player Neutrosophic numbers SWARA e VIKOR	Para guarda-redes: 21 critérios; Para jogadores de campo: 26 critérios (ex. agilidade, reflexos, drible, passe, marcação).	Turquia	Foram identificados critérios-chave por posição e posteriormente ordenados os melhores jogadores para cada função.
Limitations using multi criteria in soccer	Adventure sports Coaching practice Expertise Macro cognition Professional judgment and decision making Rugby	N/A	Reino Unido	A tomada de decisão intuitiva é amplamente usada por treinadores experientes; está relacionada com a prática reflexiva e surge com a experiência. Muitos combinam estilos intuitivo e deliberativo, usando a intuição especialmente em contextos de pressão.
Soccer and Decision support system	Strategic agility Business model innovation Talent management Intangible assets Case study Italy Sport management	N/A	Itália	• Ultimos 25 anos transformou-se numa fábrica de talentos • Sucesso no modelo de transformação de jovens promessas, não deteriorando o desempenho desportivo
Limitations using multi criteria in soccer	Association Football Coach Decision-Making Style Expertise Heuristics	Treinadores norueguese de futebol (elite e juvenil), com diferentes níveis de qualificação	Noruega	Predominio do estilo racional. Diferenças mínimas entre treinadores de elite e juvenis. Formação está associada a estilos mais racionais.

Limitations using multi criteria in soccer	Choice Cognitive process Evaluation Soccer	Treinadores e jogadores profissionais de alto nível de ambos os países	Alemanha e Brasil	Alemães tendem a estilos mais analíticos e estruturados; brasileiros preferem decisões criativas e intuitivas. As diferenças refletem dimensões culturais (e.g., evasão da incerteza, distância ao poder, orientação temporal).
Soccer and Decision support system	Football players Performance improvement Recurrent learning Tactical training	Precisão de passe, postura defensiva, consciência tática, capacidade de decisão, resposta a pressão, número e qualidade das sessões de treino.	N/A	A IA (PFSTM) melhorou o desempenho em até 95% de precisão, RMSE de 0.15. Identificou e corrigiu fraquezas específicas (lagging features) em tomada de decisão e precisão.
-	Soccer Transfers Efficiency	Títulos acumulados, saldo de transferências, número médio de espetadores, presença em competições europeias, idade do clube, percentagem de jogadores nacionais.	Europa BIG Five + Portugal	Clubes com mais títulos, maior histórico, e presença em competições europeias são mais eficientes em transferências. Clubes ingleses e franceses mostraram maior eficiência; italianos, menor. Média de eficiência: 0.772.
Soccer and Decision support system	integer programming, multicriteria decision making, player selection, sports management.	Critérios técnicos, físicos e mentais; subcritérios como passe, controle de bola, criatividade, velocidade, lesões, etc.	Irão	O modelo indicou que 15 jogadores deviam ser adquiridos (14 comprados e 1 emprestado) e 11 deviam sair. Solução otimizada equilibrando performance e rentabilidade financeira.
Soccer and Decision support system	-	20 criterios para: • Guarda redes • Defesas centrais • Defesas laterais • Médios centrais • Alas Avançados centrais	Turquia	• Modelo que encontra o melhor 11 onde utiliza AHP e depois programação linear • No entanto não garante a escolha perfeita
Soccer strategies	Soccer Agents Conflit negociation e-Business B2B	N/A	México	O modelo proposto permite simular negociações B2B de forma distribuída e adaptativa, utilizando estratégias inspiradas no futebol, com bons resultados na representação e gestão de conflitos dinâmicos.
Soccer strategies	Sports finance Firm objectives Firm performance Profitability	Inclui variáveis como retorno sobre ativos (ROA), UEFA coefficient, concentração acionista, alavancagem financeira, ativos intangíveis, payroll, etc.	Europa	A rentabilidade tem influência negativa na performance desportiva; o sucesso desportivo não afeta a rentabilidade. Concentração acionista também tem efeito negativo em ambas.
Soccer strategies	Performance analysis Sport management Soccer leagues Sport success Bayesian model averaging Relative importance analysis	Inclusão de 24 variáveis explicativas de ataque e defesa; seleção baseada em revisão de literatura e dados disponíveis.	Europa BIG Five	Determinantes robustos: assistências, remates sofridos, defesas do guarda-redes, precisão de passe e remates enquadrados. A precisão é mais importante que a quantidade de ações.

Soccer player value	Multivariate analysis Soccer data Marketvalue Soccer player performance Robust regression	Idade, altura, posição, nacionalidade, experiência internacional, desempenho (golos, assistências, minutos), duração de contrato, categoria da liga, desempenho do clube.	Alemanha	A idade tem um efeito não linear; desempenho e posição são altamente significativos; o clube e a liga onde o jogador atua também influenciam fortemente o valor de mercado.
Soccer player value	BMA IVBMA MC3 Soccer Value drive	Golos, assistências, passes, desarmes, idade, altura, posição, nacionalidade, clube, liga, visibilidade internacional.	Europa BIG Five	Fatores de valorização variam por posição; golos e assistências para jogadores ofensivos, ações defensivas para defesas; variáveis físicas e contextuais também são relevantes.
Soccer player value	English Premier League OPTA performance index Team performance Financial valuation	Valor de ativos tangíveis e intangíveis, volatilidade do valor do clube, taxa de crescimento, risco do mercado, oportunidades futuras (participações europeias, transferências, marketing).	Reino Unido	Mostra que uma grande parte do valor do Manchester United advém de opções reais (ex: sucesso futuro, marketing global), e que métodos tradicionais subestimam esse valor.
Interpretable machine learning football	Multiple criteria analysis Disaggregation analysis Preference learning Data mining	N/A	Grécia	MCDA e ML partilham a mesma base matemática, mas diferem em interpretabilidade, dimensionalidade dos dados e validação. Revê aplicações nos dois sentidos (ML para MCDA e MCDA para ML) e conclui que a integração ainda tem potencial por explorar.
Interpretable machine learning football	Artificial intelligence Machine learning Forecasting Decision making Decision support	Histórico de lesões, performance de resistência, composição corporal, carga de treino aguda (7 dias), carga do dia atual posição em campo, velocidade de sprint (30m).	Austria	Os modelos interpretáveis e esparsos (risk score: AUC=0,64; GAM: AUC=0,61) igualaram ou superaram ligeiramente o modelo "caixa-negra" original (gradient boosting, AUC=0,62), confirmando o efeito "less-can-be-more" e reforçando a importância de combinar conhecimento de domínio com ML.
Interpretable machine learning football	Football; Technology Knowledge discovery Cluster analysis Data mining	Golos, remates por jogo, cartões amarelos, cartões vermelhos, posse de bola (%), passe (%), duelos aéreos ganhos, rating.	Croácia	O cluster de equipas de elite caracteriza-se por jogo ofensivo, elevado número de remates, maior posse de bola e maior precisão de passe.
SHAP	Explainable AI SHAP Interpretability Model transparency Black box models Feature attribution	N/A	India	O SHAP destaca-se por explicações localmente e globalmente consistentes, superando o LIME em fiabilidade e reprodutibilidade. Identifica como principais desafios a complexidade computacional, a usabilidade para utilizadores não especialistas e a falta de frameworks de avaliação centrados no utilizador.
Multi-criteria Decision-making in football	N/A	N/A	Multinacional	TOPSIS e fuzzy AHP são as técnicas mais usadas, seleção de jogadores é o problema de decisão mais recorrente, artigo mais citado é Qader et al. (2017)

Apêndice B

VARIABLES	(1) Modelo 1: Guarda Redes	(2) Modelo 2: Defesas	(3) Modelo 3: Médios	(4) Modelo 4: Avançados
Age	0.135***	0.129***	0.124***	0.0311***
	(0.00986)	(0.00822)	(0.00923)	(0.00948)
Age2	-0.00523***	-0.00445***	-0.00412***	-0.00242***
	(0.000176)	(0.000154)	(0.000176)	(0.000179)
Heading_Accuracy		-0.0258***		-0.0470***
		(0.00380)		(0.00339)
Heading_Accuracy2		0.000280***		0.000453***
		(3.26e-05)		(2.95e-05)
Short_Passing		0.00633*	-0.0436***	-0.0106
		(0.00339)	(0.0118)	(0.00683)
Short_Passing2		0.000137***	0.000594***	0.000231***
		(3.07e-05)	(9.08e-05)	(5.87e-05)
Acceleration		-0.0306***	-0.0419***	0.00740
		(0.00343)	(0.00506)	(0.00511)
Acceleration2		0.000292***	0.000373***	-9.29e-06
		(2.81e-05)	(3.80e-05)	(3.81e-05)
Sprint_Speed		-0.00434	-0.0199***	0.00938*
		(0.00316)	(0.00443)	(0.00496)
Sprint_Speed2		0.000106***	0.000194***	4.44e-06
		(2.53e-05)	(3.39e-05)	(3.64e-05)
Agility		-0.0144***	-0.00672	-0.0108**
		(0.00270)	(0.00502)	(0.00433)
Agility2		0.000134***	4.23e-05	8.45e-05**
		(2.35e-05)	(3.64e-05)	(3.30e-05)
Reactions	0.00886	-0.0374***	-0.0439***	-0.00787
	(0.00616)	(0.00687)	(0.00602)	(0.00703)
Reactions2	9.42e-05*	0.000494***	0.000577***	0.000187***
	(5.58e-05)	(5.73e-05)	(5.06e-05)	(5.86e-05)
Balance	0.00204	-0.0102***	0.00847**	-0.00670**
	(0.00380)	(0.00252)	(0.00431)	(0.00310)
Balance2	-2.22e-05	8.70e-05***	-7.17e-05**	5.38e-05**
	(4.35e-05)	(2.12e-05)	(3.12e-05)	(2.45e-05)
Jumping	0.00154	0.00745***	-0.00927***	-0.00610**
	(0.00393)	(0.00282)	(0.00258)	(0.00274)
Jumping2	-2.23e-05	-5.69e-05***	7.17e-05***	4.94e-05**
	(3.55e-05)	(2.10e-05)	(2.14e-05)	(2.12e-05)
Stamina		-0.00532**	0.00890***	0.00987***
		(0.00266)	(0.00278)	(0.00298)
Stamina2		0.000107***	-3.96e-06	-7.65e-05***
		(2.04e-05)	(2.11e-05)	(2.41e-05)
Strength	0.00917**	-0.0282***	-0.0234***	-0.0192***
	(0.00401)	(0.00313)	(0.00265)	(0.00280)
Strength2	-8.21e-05**	0.000237***	0.000233***	0.000202***
	(3.50e-05)	(2.32e-05)	(2.30e-05)	(2.16e-05)
Aggression		-0.00391		
		(0.00337)		
Aggression2		6.13e-05**		
		(2.61e-05)		
Interceptions		0.0328***		
		(0.00991)		

Interceptions2		-8.52e-05		
		(7.88e-05)		
Marking		0.0181**		
		(0.00868)		
Marking2		-1.87e-05		
		(6.91e-05)		
Standing_Tackle		-0.0274*		
		(0.0140)		
Standing_Tackle2		0.000462***		
		(0.000107)		
Sliding_Tackle		0.000587		
		(0.0126)		
Sliding_Tackle2		0.000197**		
		(9.87e-05)		
Shot_Power	-0.120		0.00909**	0.0104
	(0.161)		(0.00436)	(0.00756)
Shot_Power2	0.00111		-2.38e-05	3.78e-06
	(0.00171)		(3.52e-05)	(5.69e-05)
Composure	-0.00160		-0.00103	-0.0153**
	(0.00301)		(0.00541)	(0.00631)
Composure2	1.44e-05		3.56e-05	0.000137***
	(3.57e-05)		(4.50e-05)	(5.22e-05)
Goalkeeper_Diving	-0.00189			
	(0.0179)			
Goalkeeper_Diving2	0.000293**			
	(0.000141)			
Goalkeeper_Handling	-0.0247			
	(0.0166)			
Goalkeeper_Handling2	0.000529***			
	(0.000133)			
_GoalkeeperKicking	0.109			
	(0.120)			
_GoalkeeperKicking2	-0.000716			
	(0.000966)			
Goalkeeper_Positioning	-0.00907			
	(0.0129)			
Goalkeeper_Positioning2	0.000381***			
	(0.000104)			
Goalkeeper_Reflexes	-0.0246			
	(0.0161)			
Goalkeeper_Reflexes2	0.000516***			
	(0.000126)			
Dribbling			-0.0453***	-0.0712***
			(0.00853)	(0.00894)
Dribbling2			0.000522***	0.000690***
			(6.70e-05)	(7.01e-05)
LongPassing			-0.0542***	
			(0.00659)	
LongPassing2			0.000505***	
			(5.54e-05)	
BallControl			0.106***	0.00715
			(0.0122)	(0.0122)
BallControl2			-0.000498***	0.000161*
			(9.50e-05)	(9.66e-05)
Positioning			-0.0437***	0.0371***
			(0.00436)	(0.00971)
Positioning2			0.000414***	-0.000112

			(3.88e-05)	(7.67e-05)
Vision			0.0209***	
			(0.00764)	
Vision2			-0.000112*	
			(6.23e-05)	
Finishing				0.00638
				(0.0114)
Finishing2				0.000207**
				(8.61e-05)
Long_Shots				0.0117*
				(0.00686)
Long_Shots2				-3.89e-05
				(5.68e-05)
Penalties				-0.00417
				(0.00558)
Penalties2				4.27e-05
				(4.44e-05)
Constant	8.268***	11.25***	12.90***	11.44***
	(0.385)	(0.262)	(0.303)	(0.306)
Observations	2,042	6,237	7,031	3,125
R-squared	0.968	0.952	0.933	0.968