



UFOP – UNIVERSIDADE FEDERAL DE OURO PRETO
DECOM – DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
BCC502 – METODOLOGIA CIENTÍFICA EM CIÊNCIA DA
COMPUTAÇÃO



GoldFit Soccer: Sistema Inteligente De Identificação de Talentos de Futebol

por Elisa Alves Veloso

Professores orientadores: Francisco Zacaron, Emerson Filipino e
Rodrigo Cesar Pedrosa Silva

Setembro de 2022

REVISÃO BIBLIOGRÁFICA

RESUMO:

O presente texto refere-se a uma pesquisa em andamento que tem como objetivo comparar algoritmos de aprendizado de máquina na previsão do sucesso futuro de jovens futebolistas. À questão orientadora busca-se responder se é possível utilizar tais técnicas na identificação objetiva de futebolistas de sucesso, tendo em vista a conjugação de indicadores de performance e o conhecimento de treinadores. Assim, o enfoque da pesquisa é qualitativo, ao procurar-se conhecer mais sobre a aplicação estatística das variáveis coletadas dos atletas em algoritmos de inteligência artificial e os empecilhos que circundam essa análise. O itinerário da pesquisa abordará o uso da Ciência de Dados na identificação de talentos do esporte (Pino-Ortega, 2021), Os problemas metodológicos na identificação de talentos do futebol (BERGKAMP, 2019), a aplicação do aprendizado de máquina no futebol (Rico-González, 2022) e, por fim, a manipulação dos dados desbalanceados de um banco de dados (Harsurinder, 2020). Destarte, a literatura estudada é a base para a comparação dos modelos de aprendizado de máquina,

INTRODUÇÃO:

Comum e historicamente, a identificação de talentos do futebol sempre foi feita de modo subjetivo e baseada na experiência e preferências individuais dos treinadores e olheiros, também chamados de *Scouters* ou Analistas de Mercado. Acerca dessa pessoalidade nos resultados, observa-se que a maior assertividade para tais ocorre ao alinhar-se esses indicadores de potencial esportivo com os métodos da Ciência da Dados. Assim, pesquisadores têm utilizado a modelagem estatística para estimar potencial de atletas. Nessa perspectiva, indicadores objetivos de performance e conhecimento dos treinadores são combinados e aplicados em técnicas de aprendizado de máquina a fim de otimizar o processo de identificação de atletas com desempenho de elite no futuro.

No contexto prático, a avaliação das cargas de trabalhos de treinos e competições é necessária para desenhar a prescrição do exercício (desenho da tarefa e análise de desempenho), bem como para identificação de talentos. Para isso, a identificação de tais variáveis que fornecem informações técnicas mais relevantes sobre os atletas são importantes para checar a explicação da performance de cada algoritmo. Ao mesmo tempo, como cada time e em cada tipo de esporte existem diferentes comportamentos dos atletas e em diferentes características da equipe, espera-se que as análises das variáveis sejam diversas em cada contexto (Pino-Ortega, 2021).

Devido ao desenvolvimento da tecnologia, uma grande quantidade de dados tornou-se uma ferramenta largamente utilizada para treinamentos e competições esportivas, sendo usadas na tomada de decisões. Atualmente, o desenvolvimento de algoritmos capazes de aprender conforme maiores quantidades de informações são coletadas e armazenadas tornou-se conhecido como Aprendizado de Máquina (AM) e é um ponto importante no futebol. Praticamente, a aplicação de AM no futebol tem performado utilizando-se uma ampla variedade de algoritmos preditivos, onde os mais considerados são as árvores de decisão (Rico-González, 2022).

Dentre os desafios de lidar com a distribuição dos dados, existe o desbalanceamento de dados. Nesse, ocorre de classes do banco de dados terem pesos de desproporcionalidade que comprometem o aprendizado do modelo naquele conjunto de informações. Conforme o desbalanceamento restringe a performance e a acurácia dos classificadores, vários métodos e técnicas são propostas para superar os efeitos negativos da desproporcionalidade de dados. Alguns pré-processamentos, abordagens com os algoritmos e técnicas de

reamostragem são exemplos a serem explorados para quitar os prejuízos das diferenças entre as classes.

REFERÊNCIAS:

TAN, J. How to deal with imbalanced data in Python. Disponível em:

<<https://towardsdatascience.com/how-to-deal-with-imbalanced-data-in-python-f9b71aba53eb>>.

Jantan, H. Razak, A. Human Talent Prediction in HRM using C4.5 Classification Algorithm. ResearchGate, [s.d.].

Beal, R., Norman, T., & Ramchurn, S. (2019). Artificial intelligence for team sports: A survey. *The Knowledge Engineering Review*, 34, E28.
doi:10.1017/S0269888919000225

Pino-Ortega, J.; Rojas-Valverde, D.; Gómez-Carmona, C.D.; Rico-González, M. Training Design, Performance Analysis, and Talent Identification—A Systematic Review about the Most Relevant Variables through the Principal Component Analysis in Soccer, Basketball, and Rugby. *Int. J. Environ. Res. Public Health* 2021, 18, 2642. <https://doi.org/10.3390/ijerph18052642>

BERGKAMP, T. L. G. et al. Methodological Issues in Soccer Talent Identification Research. *Sports Medicine*, v. 49, n. 9, p. 1317–1335, 3 jun. 2019.

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., and Baca, A. (2023). Machine learning application in soccer: A systematic review. *Biology of Sport*, 40(1), pp.249-263. <https://doi.org/10.5114/biolSport.2023.112970>

Harsurinder Kaur, Husanbir Singh Pannu, and Avleen Kaur Malhi. 2019. A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. *ACM Comput. Surv.* 52, 4, Article 79 (July 2020), 36 pages. <https://doi.org/10.1145/3343440>