GoldFit Soccer: Sistema Inteligente De Identificação de Talentos de Futebol

Autor: Elisa Alves Veloso
Orientadores: Francisco Zacaron Werneck, Emerson Filipino Coelho, Rodrigo César Pedrosa Silva

27 de outubro de 2022

Resumo

Tendo em vista a natureza do futebol, os modelos de previsão estatística tornaram-se um atual desafio para a tomada de decisão baseada em envidências científicas. O presente trabalho de pesquisa propõe uma comparação entre os diversos modelos de aprendizado de máquina no propósito de identificar, objetivamente, os potenciais atletas de elite no âmbito do futebol, ao conjugar indicadores de performance e o conhecimento de treinadores. A ideia é comparar métricas estatísticas a partir da aplicação dos citados algoritmos.

Abstract

Given the nature of football, statistical prediction models have become a current challenge for decision making based on scientific evidence. This paper proposes a comparison among the different machine learning models in order to objectively identify potential elite athletes in the field of soccer, by combining performance indicators and the knowledge of coaches. The idea is to compare statistical metrics from the application of the aforementioned algorithms.

1 Introdução

Na área esportiva, a identificação de talentos é classificada como um complexo processo com o fim de detectar as distintas potenciais características em atletas para alcançar sucesso na carreira, isto é, a contratação profissional do esportista. Tendo isso em vista, qualidades de condicionamento e habilidades motoras são medidas utilizadas nessas previsão.

Comum e historicamente, a identificação de talentos do futebol sempre foi feita de modo subjetivo e baseada nas experiências 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 Cência 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

das análises. 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). Isso posto, a partir do fato de cada variável ter uma respectiva importância para cada olheiro, as avaliações de desempenho e potencial mostra-se relativa e subjetiva.

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).

O estudo dos fatores de desempenho no esporte têm sido abordado levando em consideração variáveis antropométricas, fisiológicas, psicológicas, biomecânicas e de aptidão física e o uso de estatísticas descritivas e modelos lineares, sendo que a análise estatística não consegue descrever a relação entre variáveis dependentes e independentes. (Maszczyk et al., 2011). Assim, as variáveis mencionadas são de extrema relação à performance, desencolvimento e melhora dos jovens atletas e técnicas de AM são capazes de determinar fatores de performance dentre um grupo de atletas, conjugando as características morfológicas e os exercícios realizados (FERNÁNDEZ, E.2022).

En se tratando da previsão de talentos humanos, o classificador C4.5 tem um grande potencial em previsões. Nesse sentido, a administrção de potenciais funcionários, a identificação desses talentos têm sido um desafio, bem como o do âmbito esportivo citado acima. Para solucionar tal problema, a implementação do classificador de árvores de decisão mencionado é tal que pode produzir tanto a árvore de decisão quanto o conjunto de regras, além de construir as árvores com o propósito de melhorar a acurácia das previsões (Jantan, 2014). As regras de classificação geradas podem ser usadas para predizer a performance de um empregado em caso de promoção ou não. Tendo apresentado o algoritmo, esses estudos ainda não foram aplicados para o âmbito do esporte de modo específico. Assim, o êxito em outra área produz a potencialidade de sucesso no estudo em questão.

Em linhas gerais, considerando que o futebol depende de diferentes dimensões, tais como técnica, tática e condicional, as análises de dados multivariados deveriam ser realizadas no âmbito de destacar as variáveis mais importantes na identificação de talentos, dentre outros fatores. (Pino-Ortega, 2021). Assim, é primordial que todas as variáveis destacadas sejam abrangidas nos modelos de previsão.

Considerando os pontos apresentados, estuda-se se é possível identificar talentos do futebol de modo objetivo, conjugando indicadores de performance e conhecimento dos treinadores. Assim, apresenta-se alguns modelos com alto desempenho em previsões esportivas, sendo eles (i) Árvores de Decisão (AD), (ii) Gradient Boosting (GB), (iii) Gaussian Process (GP), (iv) K Vizinhos (KNN), (v) Regressão Logística (LR), (vi) Redes Neurais Artificiais (RNA), (vii) Floresta aleatória (FA), (viii) Máquinas de Vetores de Suporte (SVM) e (ix) Extreme Gradient Boosting (XGB). A extração de estatísticas de desempenho dos algoritmos utilizando Bootstrap é promissora nas estimativas. A divisão da amostra em treinamento (90%) e teste (10%) também é de grande importância. Para cada algoritmo, a execução dos experimentos e o cálculo da Acurácia (Acu), Sensibilidade (Sen) e Especificidade (Esp) indicarão os resultados das comparações da pesquisa.

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

compromentem o aprendizado do modelo naquele conjunto de informações. Conforme o desbalanceamento restringe a performance e a acurácia dos classificadores, várius 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.

Destarte, inicia-se a pesquisa de comparação de algoritmos de aprendizado de máquina, a fim de testar se é possível utilizar algoritmos de aprendizado de máquina na identificação objetiva de futebolistas de sucesso.

2 Referências

TAN, J. How to deal with imbalanced data in Python. Disponível em: <https://towardsdatascience.co to-deal-with-imbalanced-data-in-python-f9b71 aba53eb>.

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

FERNÁNDEZ, E. et al. Original Prediction Of Sports Talent In Young Throwers Using Machine Learning. [s.l: s.n.]. .

Maszczyk, A., Zając, A., y Ryguła, I. (2011). A neural network model approach to athlete selection. Sports Engineering , 13 (2), 83-93

Maszczyk, A., Gołaś,A., Pietraszewski, P., Roczniok, R., Zając, A., y Stanula, Procedia A. (2014). Application of neural and regression models in sports results prediction. Social Behavior Science , 117 , 482487.