

RANQUEANDO JOGADORES DE FUTEBOL: CRIAÇÃO DE UM ÍNDICE DE PERFORMANCE BASEADO EM APRENDIZADO DE MÁQUINA

José Américo Pereira Antunes^{a,b}

^a *Banco Central do Brasil*

^b *COPPEAD Instituto de Pós-Graduação e Pesquisa em Administração, Universidade
Federal do Rio de Janeiro (UFRJ)*

RESUMO

Este estudo desenvolve um modelo de classificação que utiliza algoritmos de aprendizado não-supervisionado e supervisionado para ranquear jogadores de futebol de acordo com seu desempenho e posição em campo. O modelo trabalha em duas fases: treinamento e classificação. A fase de treinamento envolve as 5 principais ligas europeias, enquanto a fase de classificação inclui as demais ligas de interesse, a critério do usuário. O modelo automatiza a fase inicial do processo de scouting, produzindo um *shortlist* de potenciais alvos a ser aprofundado pelo clube. Em complemento ao modelo de classificação objeto deste estudo, foi desenvolvida e disponibilizada publicamente ferramenta em Python que automatiza as consultas. Por meio da ferramenta, é possível acessar um conjunto de sete aplicativos, incluindo o modelo de classificação de jogadores.

1. PROBLEMA

A tecnologia pode alavancar o desempenho dos clubes e permitir que façam mais com menos. Brentford FC é um bom exemplo. Um clube de futebol intensamente orientado por dados, o Brentford FC alcançou a promoção à Premier League em 2021. O sucesso do clube chamou atenção para esta nova abordagem de gestão do futebol e colocou em dúvida a forma convencional de fazê-lo.

A tecnologia permite, entre outros, que os clubes sejam mais eficientes na descoberta de talentos. Há mais de uma centena de ligas profissionais e dezenas de milhares de jogadores de futebol em atividade. Nesse oceano de possibilidades, faz-se necessária uma etapa preliminar de mineração e identificação de talentos para posterior aprofundamento. Nesse contexto, proponho a seguinte pergunta de pesquisa:

Como desenvolver um índice de performance que permita ranquear jogadores segundo o desempenho?

2. OBJETIVO GERAL

Este estudo desenvolve um modelo de classificação destinado a ranquear jogadores de futebol (Pappalardo *et al*, 2019) e orientar a prospecção de talentos para o clube. O experimento usa dados do Wyscout nas ligas da UE para treinar um algoritmo de aprendizado de máquina e construir um modelo de classificação, que é então aplicado a um conjunto de ligas internacionais para ranquear jogadores.

3. METODOLOGIA

Foram utilizados dois tipos de aprendizado de máquina: na fase de treinamento, os jogadores das cinco principais ligas europeias recebem ratings atribuídos por meio de aprendizado não supervisionado, enquanto na fase de classificação os jogadores recebem ratings atribuídos por meio de aprendizado supervisionado.

3.1 Posições e perfis/funções correspondentes em campo

Como os jogadores são classificados de acordo com suas posições em campo, o primeiro passo é definir tais posições, conforme figura 1.


LWF 17/18/19	CF 20/21/22	RWF 17/18/19
LAMF	AMF 15/16	RAMF
LW	LCMF RCMF	RW
17/18/19	12/13/14  12/13/14	17/18/19
LWB	LDMF DMF RDMF 09/10/2011	RWB
LB	LCB CB RCB 6/7/8	RB
3/4/5	GK 1/2	3/4/5

Fig. 1 – Posições mais jogadas pelo jogador, de acordo com a terminologia da Wyscout.

Nota: GK significa goleiro; RB/LB, lateral direito/esquerdo; RWB/LWB, ala direito/esquerdo; CB, zagueiro; RCB/LCB, zagueiro direito/esquerdo; DMF, primeiro-volante; RDMF/LDMF, primeiro-volante direito/esquerdo; RW/LW, extremo direito/esquerdo; RAMF/LAMF, meio-campo avançado direito/esquerdo; RCMF/LCMF, segundo-volante direito/esquerdo; AMF, meia; RWF/LWF, extremo avançado direito/esquerdo; CF, centroavante.

As nove posições dão origem a vinte e dois perfis ou funções em campo, conforme tabela 1.

Tabela 1 - Posições e Perfis/Funções correspondentes

Posições	Perfis/Funções em campo
1/2 – Goleiro	1 – Goleiro Clássico 2 – Goleiro Líbero
3/5 – Lateral/Ala	3 – Lateral Defensivo 4 – Lateral Ofensivo 5 – Lateral Equilibrado
6/8 – Zagueiro	6 – Zagueiro Defensivo 7 – Zagueiro Construtor 8 – Zagueiro Equilibrado
9/11 – Primeiro Volante	9 – Primeiro Volante Defensivo 10 – Primeiro Volante Construtor 11 – Primeiro Volante Equilibrado
12/14 – Segundo Volante	12 – Segundo Volante Box to Box 13 – Segundo Volante Organizador 14 – Segundo Volante Equilibrado
15/16 – Meia	15 – Meia Organizador 16 – Meia Atacante
17/19 – Extremos	17 – Extremos Organizador 18 – Extremo Tático 19 – Extremo Agudo
20/22 – Atacante	20 – Atacante Referência 21 – Atacante Móvel 22 – Segundo Atacante

3.2 Matriz de features, Base de Treinamento e Base de Classificação

Os passos seguintes no desenvolvimento do modelo envolvem:

- i) determinação das *features* que devem ser mapeadas para cada um dos vinte e dois perfis/funções derivados das oito posições. Para tal, recorri a dados baixados do Wyscout (Wyscout, 2022).
- ii) definição da base de treinamento, a qual foi constituída a partir dos jogadores das cinco principais ligas europeias (Inglaterra, Alemanha, Espanha, Itália e França) com pelo menos 900 minutos jogados, em seis temporadas completas 2018/19, 2019/20, 2020/21, 2021/22, 2022/23 e 2023/24.
- iii) rotulagem da base de treinamento: utilização de algoritmo de aprendizado não-supervisionado por meio da técnica de redução de dimensionalidade para determinação do Índice de Performance dos jogadores.
- iv) definição da base de classificação, a partir das ligas de interesse do usuário. Para esse estudo, coletei dados de cerca de 30 mil jogadores, com mais de 900 minutos jogados, de 61 ligas, em 5 janelas de tempo distintas (temporadas fechadas de 2019/20, 2020/21, 2021/22, 2022/23, 2023/24; ou as equivalentes 2020, 2021, 2022, 2023 e 2024 – entre maio/2023 e maio/2024).
- v) aplicação do modelo de classificação e ranqueamento de jogadores na liga em que atuam, realizado por meio de aprendizado de máquina supervisionado. O rating final é um número contínuo, entre 0 e 1. O produto dessa análise será o ranqueamento liga a liga de todos os jogadores com mais de 900 minutos nas janelas de análise.

4. RESULTADOS

O modelo permite a classificação de todos os jogadores em atividade em sua liga. A partir dessas informações, os analistas podem rapidamente elaborar *shortlists* de prospectos nas diversas posições/funções. Dada a limitação do resumo, segue exemplo com os 5 melhores meia-atacantes do Brasileirão/23, com mais de 900 minutos em campo até a rodada 24.

Tabela 2 – 5 Meias Atacantes mais bem ranqueados na Liga Brasileira na temporada 2023

Atleta	Equipe	Jogos	Minutos	Rating
F. Cristaldo	Grêmio	31	2115	1,000
Cauly	Bahia	31	2746	0,840
G. De Arrascaeta	Flamengo	28	2072	0,836
Renato Augusto	Fluminense	26	1493	0,781
Lucas Lima	Santos	28	1838	0,707

Em complemento ao modelo de classificação objeto deste estudo, desenvolvi uma ferramenta em Python que automatiza as consultas e a disponibilizei na plataforma de compartilhamento Streamlit e que pode ser acessada por meio do seguinte *link*:

<https://jamerico1898-softplayerz-soft-playerz-vx07dg.streamlit.app/>

A ferramenta permite acessar um conjunto de sete aplicativos, incluindo o Ranking de Jogadores, produto deste estudo.

5. CONCLUSÃO

Este estudo desenvolve um modelo de classificação, que utiliza aprendizado de máquina supervisionado e não supervisionado para classificar jogadores de futebol de acordo com seu desempenho e posição em campo. Visa discriminar bons e maus jogadores por meio de análise quantitativa e pode ser usado como o primeiro passo na descoberta de talentos e prospecção de jogadores. Existem limites importantes a serem considerados nesta análise, uma vez que é baseada exclusivamente em dados notacionais e KPIs fornecidos pelo Wyscout e não inclui recursos importantes na análise dos jogadores, como histórico de lesões, habilidades emocionais, comprometimento profissional e ajustes para diferenças entre ligas.

REFERÊNCIAS

Pappalardo, L., Cintia, P., Rossi, A. *et al.* (2019). A public data set of spatio-temporal match events in soccer competitions. *Sci Data* **6**, 236. <https://doi.org/10.1038/s41597-019-0247-7>

Wyscout (2022). https://dataglossary.wyscout.com/player_metrics/ acessado em 10 de Abril de 2022