

Using ML to find a key Player Replacement

By Ruizhang Chen, Yuezhang Chen



The Problem

- Player Injuries
- Retirement
- Transfers & Trades



Dataset

The screenshot shows the homepage of FBREF, a football statistics website. At the top, there's a navigation bar with links for Sports Reference®, Baseball, Football (college), Basketball (college), Hockey, Soccer, Blog, Stathead®, and Immaculate Footy®. Below the navigation is a search bar with placeholder text "Enter Person, Team, Section, etc" and a "Search" button. A message below the search bar says "The most viewed football player and club in your state/city is here! [Check it out.](#)". The main content area is divided into sections: "Football Players" (with a grid of player faces), "Football Squads" (Men's and Women's squads with logos), and "STATHEAD EXPERTS POWERED BY FBREF" (with a purple background and a "Start your FREE trial" button). On the left, there's a sidebar with a "Search our database of over 234,000 players" section, a "Play Immaculate Footy" grid, and a "Play World Footy or English Footy (NEW)" link.

↑ **2024-2025
Season**

↑ **Top 5 Euro
leagues**

↑ **2500+ Players**

Feature Selection

Organic features include:

- Goals
- Assists
- Passes
- Age
- Shots
- Shots on goal



Feature Engineering

Important engineered features may include:

- Goals per 90 minutes
- Assists per 90 minutes
- xG: Expected goals
- xA: Expected Assists



Model: Random Forest

- Label:
 - Targeted player → Yes(1)
 - Other player → No(0)
- Training:
 - Learns non-linear rules
 - Each tree captures a different aspect of the player profile
 - Identify key playing-style attributes

onG	0.100000
Succ	0.085714
Att_stats_possession	0.068445
PPA	0.066662
TO	0.057143
TB	0.054292
FK	0.029784
Att 3rd_stats_possession	0.028571
GCA	0.028571
Sh	0.024984



Output

- Top 5 Players similar to the targeted player
- Comparison on key data(Goals, Assists, Shots)

	Player	Squad	Age	Pos	RF_Score	Gls	Ast
2304	Mohamed Salah	Liverpool	32.0	FW	0.05	29	18
1960	Michael Olise	Bayern Munich	22.0	FW,MF	0.04	12	15
1356	Joshua Kimmich	Bayern Munich	29.0	MF	0.02	3	7
1691	Kylian Mbappé	Real Madrid	25.0	FW	0.02	31	3
757	Jeremy Doku	Manchester City	22.0	FW,MF	0.02	3	6



Validation Process

- Cloning: Add a clone of target player into current dataset
- Run RF model on “new” dataset
- Compare the result between two dataset.

2792	Lamine Yamal	Barcelona	0.70
2854	Clone_Lamine Yamal	Barcelona	0.57
2304	Mohamed Salah	Liverpool	0.05
1960	Michael Olise	Bayern Munich	0.04
1691	Kylian Mbappé	Real Madrid	0.02
1356	Joshua Kimmich	Bayern Munich	0.02
757	Jeremy Doku	Manchester City	0.02



Future Work

Ideas to improve:

- Incorporate market value and contract info
- Use time-series trends for performance prediction
- Add team tactical fit metrics (Possession vs Counter)
- culture background



Reference:

<https://www.kaggle.com/code/mm6891/looking-for-a-replacement-for-my-retire-midfielder/notebook>

<https://www.kaggle.com/datasets/vivovinco/20222023-football-player-stats>

