

The Problem Area

Being passionate about football, I wanted to see how I could bring ML into that area. In particular, I wanted to see how ML could be utilized to maximize a teams productivity in the transfer market. So rather than focusing my attention on games and their outcomes like most models, I plan on including an analysis on individual players.

The User

With all the focus on money in club football these days, there is no doubt that no club, big or small, can afford to be wasteful in the transfer market. So whether the user is a big club like Manchester united with big budgets, they need to ensure that each purchase brings an impact corresponding to the price tag; or if the user is a smaller club like Brentford who have a much smaller budget and need to make sure none of that budget is being spent on players that wont fit in.

The Big Idea

Following the expression 'if it ain't broke, don't fix it', I plan to use ML to solve the problem of replacing players in a successful pre-existing team. For example, a team like Liverpool has a problem of critical players nearing the ends of their contract. So instead of trying to change the system in order to fit around a player like Harvey Elliot at right wing, this algorithm would find the closest player to Mohamed Salah in Europe's top 5 leagues by comparing their stats and career trajectory. Another application would be solving Man City's injury crisis with Rodri being out for the season.

One final, though extreme and finicky application, would be for a club like Manchester united, where the players aren't working well for the system so replacing pre-existing players wont make things any better, would be to create an artificial player that fits a profile they would like, then finding a real player that fits that profile.

The Impact

This would have a big impact on the way that football transfers are done. For example, large, recently successful clubs would be able to have a sense of continuity as players come and go, whilst big teams that are struggling would be able to use this model to keep similar profile for the players that work, whilst using other algorithms to find new profiles to complement the players that fit the system.

The Data

Detailed statistics about players dating back 3-5 years as well as potentially needing transfer market data.

Ex: <https://fbref.com/en/comps/Big5/2023-2024/stats/players/2023-2024-Big-5-European-Leagues-Stats>

From this source:

- We will gather the data on each of the players from each of the top 5 leagues as well as the champions league and the Europa league.
- Over general statistics, possession stats, and defensive actions categories.
- Then join them all on player name and figure out how to deal with the fact of multiple clubs duplicates.
- Over the previous 5 seasons (this way we can also study the evolution of players).

<https://www.transfermarkt.com/>

- To get the market value of the players.

I am considering on whether or not to include goalkeepers in the model.

The Alternative

Also related to football, it would be interesting to come up with a ML generated live commentary for football matches that can be tweaked to suit the preferences of the viewer. For example in a Liverpool vs Arsenal game, a Liverpool fan watching the game would like the commentator to be biased towards Liverpool whereas an arsenal fan would want an Arsenal bias. Or some fans might prefer stats based commentary where the commentator spews facts regarding the teams and their previous fixtures, whereas some fans would like story based commentary that is more focused on describing the game.