Evaluating The Market Value of English Premier League Players for New Season Investments by Arsenal F.C.

Amir Khoeilar

Abstract

Arsenal Football Club, one of the well known soccer teams in the world is planning to purchase couple of players in order to strengthen their squad for the upcoming season of 2021-22. As of their decision and approach, I have decided to gather information regarding all the players in the English Premier league with their attributes and stats from the past 2 seasons (2019-20,2020-21) from the FBREF-Soccer Website and also the market value of the players from Transfer Market Website. I have web scrapped both websites and performed a linear regression approach in order to provide the club with the real value of their targets in comparison to their hyped marked value.

Design

Arsenal Football Club is has been suffering in the league for the last couple of years. Not scoring enough goals, not being able to create chances, receiving more than usual goals and squad depth has been the main reason of them dropping in ranking in league season after season. But since their are still considered one of the begets club in the world, they have lots of connections in order to attract their targeted players to the club for the upcoming season of 2021-22. They have put aside a certain budget of for their summer investments and need to address and buy a player for the positions they are lacking depth in. One of their preferred criteria is purchasing British players, but yet again they come with higher prices due to Britain tax and that may take away too much from their budget. By knowing the real price of the players before inquiring for their targeted players, Arsenal can optimize their approach and have stronger negotiations in their purchases from other clubs.

Data

Due to Brexit rules, Arsenal prefers to purchase British players . Also, purchasing a player who has been playing in the premier league with most likely hit the ground running and will need less time to adapt.

I have scrapped and gather information of all the plates who have played in premier league in the season of 2019-20 and 2020-21 from the FBREF-Soccer Website. All players are compared and put in a table in accordance with their stats for each season. All players have been organized with features by 'Player', 'Nation', 'Pos', 'Squad', 'Age', 'MP', 'Starts', 'Min','90s', 'Gls', 'Ast','Gls', 'Ast','Height', ‘Foot’. I have matched all the players with their ‘Market Value’ from Transfer Market Website in order to collect a reliable dataset to perform linear regression and evaluate the mathematical value of the players in accordance to their performances.

After scrapping from both websites, I performed some data cleaning in order to get rid of duplicates and some null values present in the data frame. I started performing a baseline linear regression model on my clean data frame and received R^2 of .456 with obvious presence of multicollinearity between features and the target ('Market Value') , and also feature themselves.

In order to reduce multicollinearity in my data frame, I used the methods VIF, cross validation and Feature engineering to find multicollinearity , eliminate it and also raise the value of R^2 to .601 by introducing new features to each player in the table.Finally, I used LASS method from regularization in order to eliminate the final features that were causing multicollinearity and finally reaching the R^2 = 0.6450347772759658 and MAE = 7856451.605622319.With having the optimized situation, we could now calculate the real market value of each player in the clubs interest, choose and eliminate the best candidates and put Arsenal in strong negotiation position in their approaches.

Algorithms

Feature Engineering

1. Scrapping data from 2 websites using BeautifulSoup and Selennium.

2. Using EDA and its Libraries (numpy , pandas, matplotlib , ….) in order to clean data in Jupiter notebook.

3. Using linear regression and its accordance libraries SKlearn in order to operate linear regressions different steps to calculate R^2.

Model Evaluation and Selection

In the conclusion, with the help of the data scrapped for the both websites and I have reached the code through linear regression in order to achieve the real mathematical price of the players that Arsenal FC is potentially could be interested in.

I have provided one of their main targets features and its market value last summer and compared it with predicted value that could be reached with linear regression. Finally I have provided the clubs direction and final decision in their purchase.

Tools

- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting
- Sklearn for performing linear regression
- Mac Keynote and Numbers for the presentation

Communication

There is going to be the code for the data operations and also PDF slides of powerpoint presentation available on my GitHub account.

https://github.com/rezxkoi/English_Premier_League_Valuation