Predicting Player's Value: FIFA19

AKASH KUMAR

25th April, 2020

1. Introduction

1.1 Background:

FIFA 19 is a football simulation video game developed by EA Vancouver as part of Electronic Arts' FIFA series. Announced on 6 June 2018 for its E3 2018 press conference, it was released on 28 September 2018 for PlayStation 3, PlayStation 4, Xbox 360, Xbox One, Nintendo Switch, and Microsoft Windows. It is the 26th installment in the FIFA series. As with FIFA 18, Cristiano Ronaldo initially as the cover athlete of the regular edition: however, following his unanticipated transfer from Spanish club Real Madrid to Italian side Juventus, new cover art was released, featuring Neymar, Kevin De Bruyne and Paulo Dybala. The game features the UEFA club competitions for the first time, including the UEFA Champions League and UEFA Europa League. Martin Tyler and Alan Smith return as regular commentators, while the new commentary team of Derek Rae and Lee Dixon feature in the UEFA competitions mode. Composer Hans Zimmer and rapper Vince Staples recorded a new remix of the UEFA Champions League anthem specifically for the game. The character Alex Hunter, who first appeared in FIFA 17 returns for the third and final installment of "The Journey", entitled, "The Journey: Champions". In June 2019, a free update added the FIFA Women's World Cup as a separate game mode.

1.2 Problem:

A player's value is everything as it gives us a complete background picture of the player's performance as well as his popularity among his fans. Any data that might contain the player's value could be used to predict the player's reputation in the market. This project aims to bring out that reputation through value prediction.

1.3 Interest:

Any club might want to know his player's as well as other club's players market value for competitive advantage and could also bring new emerging talent into the market. Others might also be interested are fans and other business people who are in sports industry.

2. Data Acquisition and Cleaning

2.1 Data Sources:

All the player's data containing his name, photo, player's position etc. can be found in the csv file of Kaggle's FIFA19 dataset here.

2.2 Data Cleaning:

Data downloaded has 18207 rows and 89 columns and has many missing data needed to be filled, redundant attributes that needed to be truncated and some similar attributes to be aggregated. There was no duplicate data in the dataset, all names were unique. Firstly, unnecessary columns like unnamed: 0, ID were dropped and then features like photo, club logo, etc which contains links were dropped. Features which contains player's personal information like: - *'Real Face', 'Joined', 'Loaned From', 'Contract Valid Until', 'Jersey Number', 'Release Clause', 'Special'* were dropped. This led to the formation of Data Frame containing 18207 rows and 77 columns containing 40 numerical attributes and 37 nominal attributes in which some were also ordinal like '*Work Rate',* etc.

Secondly, missing data needed to be handled. I wrote a script to fill the missing values of numeric and nominal data separately. Also, in numerical data there was two separation some attributes were type(int) so they needed to be filled integer values only. I used the imputer method of python to fill these values. Numeric data containing decimal was filled by the mean values and integer data by median of the respective attributes. Nominal data was filled by the most frequent value of the attributes. Some nominal attribute like *'Club'* were given no club value to the missing data while in some attributes like *'Body Type'* there was some abrupt value that needed to be filled right.

Thirdly, there were some nominal attribute like: - *'Height', 'Weight', 'Value', 'Wage'*, which contains numeric type data which needed to be converted into decimal values. I wrote functions to convert these string values into decimal values: Height in inches, Weight in lbs and Value, Wage in euros.

Fourthly, there were many similar attributes that could be aggregated to form a single column which determine the overall functionality of those attributes. I studied about the Football terminologies which provided me method to aggregate them. Features like *'Acceleration', 'Sprint Speed'* comes under the **Pace** attribute and similarly other attributes were also aggregated by taking means into **Shooting, Passing, Technique, Defence, Physical, Goalkeeper, Rating**. In this rating attribute was formed by aggregating player's potential to perform and his overall rating, shooting attribute was formed by aggregating player's shot power, long shots, penalties accuracy, etc. Passing by aggregating vision, accuracy, curves, etc. Technique by aggregating dribbling, balance, reaction, ball control, etc. Defence from tackle capabilities, physical from their strength, jumping, etc. And goalkeeper from aggregating goalkeeper's necessary skills.

Lastly, dropping all the columns that has been aggregated and also dropping In-Game-Stats ratings like: - *'RF', 'CAM'*, etc. because they just determines a player's ratings according to its positioning in the game and it's also obvious that no club would make his attacker to play in defence position and we already has column *'Position'* containing best position in which players mostly play their game. After all the cleaning and aggregating the data we have reduced the Data Frame into 18207 rows and 23 columns containing no missing data.