Analysis of FIFA 2019

Team Name : DASK_2019

1. Subhranil Das PES1201700309
2. Ketan Panwar PES1201700195

PES University , Bengaluru

***Abstract*— This Literature Survey includes analysis of the dataset containing information of football players provided by FIFA 2019. This analysis includes an explanation of the approach used and observations found during the analysis.**

***Keywords*— Fifa 2019, Analysis, EDA Soccer**

I. Introduction

FIFA 19 is a football simulation video game developed by EA Vancouver as part of Electronic Arts' FIFA series. We are using the dataset to help club managers choose players based on potential and skill, grouping/clustering players having similar features. At the same time, we also compare existing teams on different levels.

II. Critical Analysis of Work Done By Others

Citation 1:

https://www.kaggle.com/applecider327/football-analytics-analysing-the-fifa-19-dataset

1. The author drops the column 'Wage' , which might be a deterministic factor for selection of players in clubs.

2. Our goal is to help the club managers select the best possible teams. Teams have to be diverse. Dropping the column 'Preferred Foot' might not be the best option as the club manager may end up choosing a team with most right footed players. Formulating a strategy to defeat such a team would be easy.

3. The author is trying to measure the accuracy of the overall rating given for goalkeepers. For this he uses a self formulated field of 'clean sheet rate' , which counts the number of games where a goalkeeper concedes no goals. A goalkeeper conceding no goal may not necessarily be the goalkeeper's talent. It might also be the case that the other players of the same team rarely allowed the ball to touch the goal line of their side goal. Alternatively, we can use any external data set having information of the goalkeeper's saves when he gets an opportunity to save.

4. The author is trying to measure the accuracy of overall scores of defenders. For this he uses a self formulated field of 'defence score' , which has the same problem as the previous point. The defender is alone responsible for the score.

5. The point worth mentioning is that the author performs the above calculations based on the performance of only 4 groups of 10 members. This statistic may not be enough to judge the accuracy of any field.

Citation 2:

https://www.kaggle.com/roshansharma/fifa-data-visualization/notebook

1. The author is trying to fill missing values of several fields inappropriately.
2. In ‘Contract Valid Until’ the values used is 2019 so, after 2019 the model will show those players available but actually they won’t be.
3. The above problem pertains with ‘joined’ and ‘Club’ field also.

III. Problem Statement

We aim to assist the club managers to select the best possible team under a general budget based on similarity of players in various fields. We also analyse player’s performances based on various fields.

IV. Difference in Approach

We plan to compare the players using various fields. Our main concern is to help the club managers choose the best possible teams while not exceeding a general budget. We wish to cluster players using various fields such as preferred foot, position where he plays, age group, country, etc. Then we form such teams and see if any existing team as close formations. If such teams exist, we see their performance and measure the accuracy of our model.

References

[1] https://en.wikipedia.org/wiki/FIFA_19.
[2] https://www.kaggle.com/applecider327/football-analytics-analysing-the-fifa-19-dataset
[3] https://www.kaggle.com/dude431/interactive-visualizations-among-fifa-players
[4] https://www.kaggle.com/roshansharma/fifa-data-visualization/notebook
[5] https://github.com/sharmaroshan/FIFA-2019-Analysis
[6] https://ieeexplore.ieee.org/document/8697111
[7] *https://scholar.google.com/*