README

November 1, 2021

0.0.1 Dataset:

FIFA 19 complete player dataset 18k+ FIFA 19 players, ~90 attributes extracted from the latest FIFA database

The FIFA 2019 dataset includes information about almost 18K players from more than 160 distinct nationalities and distributed over more than 600 football clubs from around the world, for each player in the dataset there are +80 features describing the characterstics of the players like nationality, age, physical characterstics, wages, position on the pitch, skills and performance measures.

0.0.2 Main Findings:

1. The dataset contains 86 features describing the characterstics of each player like nationality, age, physical characterstics, wages, position on the pitch also skills and performance like:
    - Attacking
    - Ball handling skills
    - Movement
    - Power
    - Mentality
    - Defending
    - Goalkeeping
2. From intial exploration to the datset we find out that are very strong multicolinearity between variables in the dataset, thus to fix the issue and make the dataset more legable and comperhandable will be trimming and engineering the feature and reduce dimensionality from 89 dimension to 29 dimension
3. In regard to relationship between players skill rate and wages we find out that there a significant exponential relationship between the variables with 81% correlation between players skill rate and log10 of players wages
4. In regard to the idea of having a dominant team, after conduting a hypothesis testing on the top 2 team with highest average players skill rate, we concluded that there is no one team which has a siginificance dominance when it comes to players skills
5. To better understand the types of players and football players in the dataset we applied ML clustering algorithm KMeans and after conducting exhaustive test on the right number of clusters in the dataset we found out that there are 4 segements in the data, we can call them as follow (Super players - Elite players - Above average players - Average players)
6. To estimate the players wage based on available variable, we have applied linear regression model and were able to predict the wages of the a test dataset at 95% confidence interval with accuracy 99.7%, with only 17 incorrect predication out of +5.3K predications

0.0.3 Extras

Feedback Received:

1. Received +300 views, 7 upvotes, 2 positive comments and a bronze medal from kaggle on the notebook
2. Received 800 views and 12 likes on LinkedIn

Project Files:

1. fifa-2019.ipynb ---> Complete project Notebook
2. FIFA19-exploration.slides.html ---> Slide deck for communicating exploratory analysis result
3. ExplanatoryAnalysis.slides.html ---> Slide deck for communicating explanatory analysis result
4. README.pdf ---> README file in PDF format with summary of project details
5. README.ipynb ---> README file in ipynb format with summary of project details

Resources:

- Kaggle Notebook : *https://www.kaggle.com/homegahed/fifa-2019/notebook*
- Dataset : *https://www.kaggle.com/karangadiya/fifa19*
- LinkedIn post : *https://bit.ly/2ZLiCNb*
- Sofifa Website: *https://sofifa.com*

```
In [1]: ! jupyter nbconvert README.ipynb --to pdf
```

```
[NbConvertApp] Converting notebook README.ipynb to pdf
[NbConvertApp] Writing 20101 bytes to ./notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', './notebook.tex']
[NbConvertApp] Running bibtex 1 time: ['bibtex', './notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 34423 bytes to README.pdf
```