

Machine Learning Approach to Soccer Player's Overall Rating Prediction

Antai Ouyang #1, Xianglong Li #2, Haochen Li #3

Abstract

A abstract sums up your work in very few sentences: (i) state the problem you are addressing; (ii) say why it's an interesting problem, and which issues are hard to tackle; (iii) give your approach towards solving the problem; (iv) say Why and how well your approach solves the problem.

Introduction

Your introduction briefly explains the problem you address, and what you've achieved towards solving the problem. It's an edited and updated version of your introduction and objective from your topic proposal

Dataset

0.1 Datasets overall

Our project dataset is FIFA 2019 players attributes dataset, which is collected from Kaggle. This dataset includes 89 attributes of a soccer player, and our project only uses some of these attributes concerning only player capability. Besides, considering that goalkeepers have a totally different set of capability metrics compared with other players, we will only consider data records of nongolekeepers in our model.

0.2 Data processing

Now we beginning the data processing. There are 18206 players for which 89 features each are provided. Firstly, we should drop the duplicate entries. Then we check the variables:

ID, Name, Age, Photo, Nationality, Flag, Overall, Potential, Club, ClubLogo, Value, Wage, Special, PreferredFoot, InternationalReputation, WeakFoot, SkillMoves, WorkRate, BodyType, RealFace, Position, JerseyNumber, Joined, LoanedFrom, ContractValidUntil, Height, Weight, LS, ST, RS, LW, LF, CF, RF, RW, LAM, CAM, RAM, LM, LCM, CM, RCM, RM, LWB, LDM, CDM, RDM, RWB, LB, LCB, CB, RCB, RB, Crossing, Finishing, HeadingAccuracy, ShortPassing, Volleys, Dribbling, Curve, FK Accuracy, LongPassing, BallControl, Acceleration, SprintSpeed, Agility, Reactions, Balance, ShotPower, Jumping, Stamina, Strength, LongShots, Aggression, Interceptions, Positioning, Vision, Penalties, Composure, Marking, StandingTackle, SlidingTackle, GK Diving, GK Handling, GK Kicking, GK Positioning, GK Reflexes, ReleaseClause

Now we have to delete the missing value which will disturbing our model.

<i>Index</i>	<i>column name</i>	<i>Total missing</i>	<i>Percent missing</i>
0	<i>Loaned From</i>	16943	0.930576
1	<i>LWB</i>	2085	0.114516
2	<i>LM</i>	2085	0.114516
...
23	<i>RS</i>	2085	0.114516
24	<i>ST</i>	2085	0.114516
25	<i>LS</i>	2085	0.114516
26	<i>RCB</i>	2085	0.114516
27	<i>Release Clause</i>	1564	0.085901
28	<i>Joined</i>	1553	0.085297
29	<i>Contract Valid Until</i>	289	0.015873
..
58	<i>Finishing</i>	48	0.002636
59	<i>Crossing</i>	48	0.002636
60	<i>Positioning</i>	48	0.002636
61	<i>LongPassing</i>	48	0.002636
62	<i>BallControl</i>	48	0.002636
63	<i>Acceleration</i>	48	0.002636
64	<i>SprintSpeed</i>	48	0.002636
65	<i>FKAccuracy</i>	48	0.002636
66	<i>Reactions</i>	48	0.002636
67	<i>ShotPower</i>	48	0.002636
68	<i>GKReflexes</i>	48	0.002636
69	<i>Jumping</i>	48	0.002636
70	<i>Stamina</i>	48	0.002636
71	<i>Strength</i>	48	0.002636
72	<i>LongShots</i>	48	0.002636
73	<i>Aggression</i>	48	0.002636
74	<i>Balance</i>	48	0.002636
75	<i>Interceptions</i>	48	0.002636
76	<i>Flag</i>	0	0.000000
77	<i>Name</i>	0	0.000000
78	<i>Age</i>	0	0.000000
79	<i>Photo</i>	0	0.000000
80	<i>Nationality</i>	0	0.000000
81	<i>Value</i>	0	0.000000
82	<i>Overall</i>	0	0.000000
83	<i>Potential</i>	0	0.000000
84	<i>Club Logo</i>	0	0.000000
85	<i>Wage</i>	0	0.000000
86	<i>Special</i>	0	0.000000
87	<i>ID</i>	0	0.000000

Table 1: Missing value percent

Considering these missing value we have below comments:

- Since a big chunk of players are not loaned to other clubs, hence there are 16943 missing values out of 18206 i.e. they play for their own club.
- There is a consistency (11.45%) in missing values for the position (they play on), this needs to be explored.
- Release clause, joined, contract valid until, club, position, jersey number have missing values and need to be explored.
- There is a consistency (0.26%) in missing values for the players rating attributes.

Now we will drop unnecessary columns, which can address the problem simply, reference the FIFA official guideline (<https://www.fifa.com/fifa-tournaments/technical-factsheets/fs444/2018/06>). we only keep significant values that have highly contributed to the model.

Age, Overall, Potential, WeakFoot, Position, Crossing, Finishing, HeadingAccuracy, ShortPassing, Volleys, Dribbling, Curve, FKAccuracy, LongPassing, BallControl, Acceleration, SprintSpeed, Agility, Reactions, Balance, ShotPower, Jumping, Stamina, Strength, LongShots, Aggression, Interceptions, Positioning, Vision, Penalties, Composure, Marking, StandingTackle, SlidingTackle, GK Diving, GK Handling, GK Kicking, GK Positioning, GK Reflexes

We have to delete the missing value which will disturbing our model and then we will filter data with null values for position:

<i>Index</i>	<i>column name</i>	<i>Total missing</i>	<i>Percent missing</i>
0	<i>Position</i>	60	0.003295
1	<i>GKReflexes</i>	48	0.002636
2	<i>Curve</i>	48	0.002636
...
30	<i>Positioning</i>	48	0.002636
31	<i>Interceptions</i>	48	0.002636
32	<i>Aggression</i>	48	0.002636
33	<i>LongShots</i>	48	0.002636
34	<i>Strength</i>	48	0.002636
35	<i>Stamina</i>	48	0.002636
36	<i>Potential</i>	0	0.000000
37	<i>Overall</i>	0	0.000000
38	<i>Age</i>	0	0.000000

Table 2: Missing value percent in useful attribute

After handling these missing values we have 18147 rows thought, subtracting the 60 missing values. Besides, we have to transform some *string* type values to *int* type, which guarantees our model works.

At last, we partition original player data according to positions, we mainly divide players into 4 positions: **Forward**, **Midfielder**, **Back**, and **Goalkeeper**.

Solution

The solution section covers all of your model design, algorithms, formulas, findings etc. It explains in detail each contribution, if possible with figures/schematics.

Results and Discussion

The results section details your metrics and experiments for the assessment of your solution. It allows you to compare your idea with other approaches you've tested.