Title: Estimate Rating of a Player on the Basis of Other Attributes.

Objective: Selecting players and team players is always a tough task. When organizations need to decide a team for any tournament, it needs to look several parameters of a players and that's the thing which make it challenging. Presented study shows a machine learning model to get overall rating of a Soccer Players.

Tools and Dataset: The data was sourced from:

- http://football-data.mx-api.enetscores.com/ : scores, lineup, team formation and events
- http://www.football-data.co.uk/ : betting odds.
- http://sofifa.com/ : players and teams attributes from EA Sports FIFA games. *FIFA series and all FIFA assets property of EA Sports.*

Table	Total Rows	Total Columns	Columns
Country	11	2	id, name
League	11	3	id, country_id, name
Match	25979	100	id, country_id, league_id, season, stage, date, match_api_id, home_team_api_id, away_team_api_id, home_team_goal, away_team_goal, home_player_X1, home_player_X2, home_player_X3, home_player_X4, home_player_X5, home_player_X6, home_player_X7, home_player_X8, home_player_X9, home_player_X10, home_player_X11, away_player_X1, away_player_X2, away_player_X3, away_player_X4, away_player_X5, away_player_X6, away_player_X7, away_player_X8, away_player_X9, away_player_X10, away_player_X11, home_player_Y1, home_player_Y2, home_player_Y3,

			home_player_Y4, home_player_Y5, home_player_Y6, home_player_Y7, home_player_Y8, home_player_Y9, home_player_Y10, home_player_Y11, away_player_Y1, away_player_Y2, away_player_Y3, away_player_Y4, away_player_Y5, away_player_Y6, away_player_Y7, away_player_Y8, away_player_Y9, away_player_Y10, away_player_Y11, home_player_1, home_player_2, home_player_3, home_player_4, home_player_5, home_player_6, home_player_7, home_player_8, home_player_9, home_player_10, home_player_11, away_player_1, away_player_2, away_player_3, away_player_4, away_player_5, away_player_6, away_player_7, away_player_8, away_player_9, away_player_10, away_player_11, goal, shoton, shotoff, foulcommit, card, cross, corner, possession, B365H, B365D, B365A, BWH, BWD, BWA, IWH, IWD, IWA, LBH, LBD, LBA, PSH, PSD, PSA
Player	11060	7	id, player_api_id, player_name, player_fifa_api_id, birthday, height, weight
Player_Attributes	183978	42	id, player_fifa_api_id, player_api_id, date, overall_rating, potential, preferred_foot, attacking_work_rate, defensive_work_rate, crossing, finishing, heading_accuracy, short_passing, volleys, dribbling, curve,

			free_kick_accuracy, long_passing, ball_control, acceleration, sprint_speed, agility, reactions, balance, shot_power, jumping, stamina, strength, long_shots, aggression, interceptions, positioning, vision, penalties, marking, standing_tackle, sliding_tackle, gk_diving, gk_handling, gk_kicking, gk_positioning, gk_reflexes
Team	299	5	id, team_api_id, team_fifa_api_id, team_long_name, team_short_name
Team_Attributes	1458	25	id, team_fifa_api_id, team_api_id, date, buildUpPlaySpeed, buildUpPlaySpeedClass, buildUpPlayDribbling, buildUpPlayDribblingClass, buildUpPlayPassing, buildUpPlayPassingClass, buildUpPlayPositioningClass, chanceCreationPassing, chanceCreationPassingClass, chanceCreationCrossing, chanceCreationCrossingClass, chanceCreationShooting, chanceCreationShootingClass, chanceCreationPositioningClass, defencePressure, defencePressureClass, defenceAggression, defenceAggressionClass, defenceTeamWidth, defenceTeamWidthClass, defenceDefenderLineClass

Furthermore, Jupyter Notebook (5.4.0) used for the Python code execution and sklearn libraries used to implement regression.

Description of the procedure:

In this method, Multiple Regression is implemented on the dataset. Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables). The player-datasets is divided into training and testing datasets in 3:1 ratio. The independent columns are-

['crossing', 'finishing', 'heading_accuracy', 'short_passing','dribbling', 'curve', 'free_kick_accuracy','long_passing', 'ball_control', 'acceleration', 'sprint_speed','agility', 'reactions', 'balance', 'shot_power', 'jumping', 'stamina','strength', 'long_shots', 'aggression', 'interceptions', 'positioning','vision', 'penalties', 'marking', 'standing_tackle', 'sliding_tackle','gk_diving', 'gk_handling', 'gk_kicking', 'gk_positioning', 'gk_reflexes']

And the dependent variable is-

[‘overall_rating’]

The Root Mean Square Error method is used in order to check the accuracy of the model.

(**Note:** Coding and Figures are represented in Notebook file.)

Result and discussion:

We have successfully implemented the proposed model with following accuracy.

```
Mean Squared Error =  10.9622742619
Root Mean Squared Error =  3.3109325366
```