Predicting Player Ratings by Player Position for European Football

Springboard Intermediate Data Science: Python
Capstone Project Final Report

Prepared by: Ruhama Ahale (ruhama.ahale@gmail.com)
Mentor: Raghunandan Patthar

October 2017

Table of Contents

1. INTRODUCTION

Player transfers in association football are multimillion dollar deals, every year clubs pay millions of dollars to sign new players. The client, a European football club wants a list of top players for each playing position so that they can shortlist desirable candidates from the players available on the transfer list.

The client wants a model that can predict the aggregated overall rating of a player at each age of the player. They want to identify the current top players for each playing position and the significant attributes that affect the player rating for each position. They will use this list of best players to consider during transfer season. In case they need to make a transfer, they will bring in a replacement, a major part of player selection depends upon which position the player will play at. Depending on the need of the team, the managers will consider the up and coming players who will suit their need.

Based on an interview with Mueller-Moehring who is responsible for rating players in FIFA games every year [here], the player ratings are calculated by FIFA based on player attribute stats as well as other factors such as the league the player is associated with, the specific circumstances which resulted in an attribute score and various other factors. The code used for this study is available [here]

Our goal is to fit a model that predicts the overall rating using only the player attributes available to us in the data source.

2. DATA ACQUISITION AND CLEANING

2.1 DATA SOURCE:

The data we will use for this problem is sourced from several websites such as
Football-Data-MX-API: scores, lineup, team formation and events
Sofifa: players and teams attributes from EA Sports FIFA games, FIFA series and all FIFA assets property of EA Sports; It is curated by Hugomathien and made available on Kaggle [here]

We want to identify top players playing at each of the four positions - Forward, Defender, Midfielder and Goalkeeper from the dataset of players playing for different clubs in European Football since 2008 up to 2016.

Below are details of each table:

1. Match: Match data with each row having - match_date, home_team, opponent_team, player_ids of players playing in the match, player_id at each playing coordinate
2. Player: Player details- player_id, player_name, birthdate, height, weight
3. Player_Attributes: Attributes sourced from several sources for each match player has played - player_id, preferred_foot, attacking_work_rate, overall_rating, crossing, finishing, heading_accuracy, short_passing etc.
4. Country: country_id, country_name
5. League: id, country_id, league_name

Data is collected from this database and processed to merge some of these files to get player positions data and player league and country data in a more accessible format. It is not necessary that a player plays at the same position for each match, so we choose the players favored position using the mode function to calculate the player position. These new tables are then pushed back into the database and this new SQLite database is used as source for this analysis. [Here] is the link to the ipynb file which contains the data manipulations code in detail.

```
Data columns (total 39 columns):
player_api_id          52725 non-null int64
player_name            52725 non-null object
player_position        52725 non-null object
overall_rating         52725 non-null float64
age                    52725 non-null float64
height                 52725 non-null float64
weight                 52725 non-null int64
finishing              52725 non-null float64
heading_accuracy       52725 non-null float64
short_passing          52725 non-null float64
volleys                52725 non-null float64
dribbling              52725 non-null float64
curve                  52725 non-null float64
free_kick_accuracy     52725 non-null float64
long_passing           52725 non-null float64
ball_control           52725 non-null float64
acceleration           52725 non-null float64
sprint_speed           52725 non-null float64
agility                52725 non-null float64
reactions              52725 non-null float64
balance                52725 non-null float64
shot_power             52725 non-null float64
jumping                52725 non-null float64
stamina                52725 non-null float64
strength               52725 non-null float64
long_shots             52725 non-null float64
aggression             52725 non-null float64
interceptions          52725 non-null float64
positioning            52725 non-null float64
vision                 52725 non-null float64
penalties              52725 non-null float64
marking                52725 non-null float64
standing_tackle        52725 non-null float64
sliding_tackle         52725 non-null float64
gk_diving              52725 non-null float64
gk_handling            52725 non-null float64
gk_kicking             52725 non-null float64
gk_positioning         52725 non-null float64
gk_reflexes            52725 non-null float64
dtypes: float64(35), int64(2), object(2)
```

Fig 1.1.1

Figure 1.1 shows the information about the data in the consolidated dataset. We use this data for data exploration and analysis.

2.2 HANDLING MISSING VALUES

Since data from 2008-09 season is inconsistent, we exclude that from our analysis. From the remaining data, the proportion of missing values is very small 2.6%, we ignore the missing values and delete them from the dataset as they are less than 5%. The data analysis and modeling is done on this data.

However, in the results, we need to merge the prediction data with the table having players' country and league detail. This table has more missing data, mainly any player who hasn't played since Jan 2013 is not available in this table. It makes sense to show only active players in the list of top players so we exclude these missing values from the final output.

3. METHODOLOGY

This is a prediction problem where using the independent data i.e. player attributes we must predict the dependent variable i.e. overall rating. Fig 3.1 depicts a flow chart of the methodology used for this study.

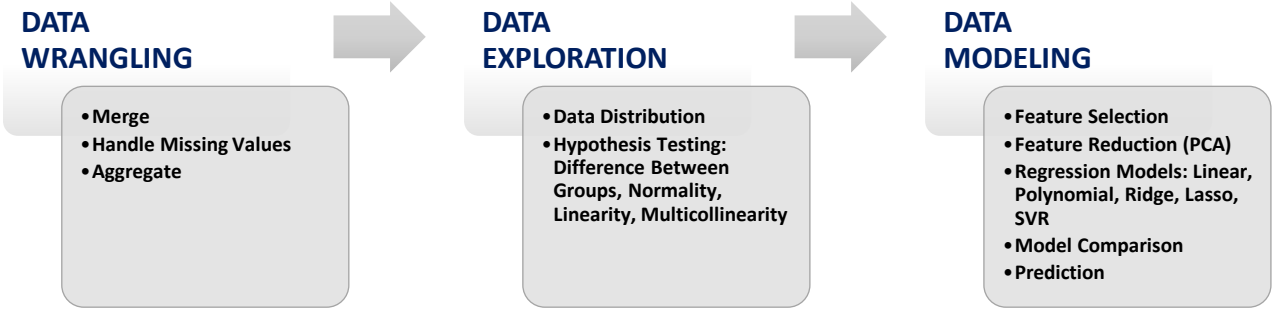

Fig 3.1

3.1 DATA PREPARATION

We have already merged the files and removed missing values using the raw data. To get predicted ratings for players at the next age of player, we want data for ratings and attributes of a player at a given age. The data that we have has ratings and attributes at multiple matches for multiple ages of a player. So, for one age of a player e.g. 25, there could be more than one match details. To get one data point for each age of a player we aggregate the features at that age using the average function.

Data Description:

	player_api_id	overall_rating	age	height	weight	finishing	heading_accuracy	short_passing	volleys
count	52725.000000	52725.000000	52725.000000	52725.000000	52725.000000	52725.000000	52725.000000	52725.000000	52725.000000
mean	128807.768687	68.168710	26.070422	182.010980	169.056918	48.495372	56.690740	61.567312	48.034650
std	130042.194779	6.719322	4.563284	6.401438	15.070259	19.201114	16.882404	14.386956	18.321289
min	2625.000000	37.000000	16.000000	157.480000	117.000000	2.000000	3.000000	3.000000	1.000000
25%	34225.000000	64.000000	23.000000	177.800000	159.000000	32.500000	49.000000	56.750000	34.000000
50%	73841.000000	68.000000	26.000000	182.880000	168.000000	51.000000	60.000000	64.333333	51.000000
75%	183532.000000	73.000000	29.000000	185.420000	179.000000	64.000000	68.000000	71.000000	62.600000
max	750584.000000	94.000000	44.000000	208.280000	243.000000	97.000000	95.000000	97.000000	93.000000

Fig 3.1.1

3.2 DATA EXPLORATION

Player Age Distribution:

Since the relationship of the player age and player rating is of interest to us, as there is a general assumption in football that player rating decreases with age of player, let us explore if this assumption is true for each of the playing positions

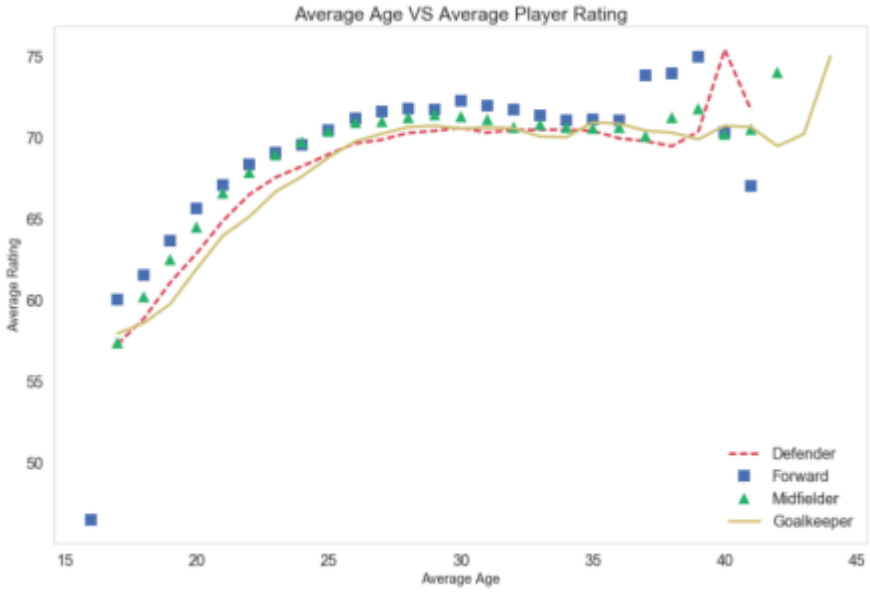

Fig 3.2.1

From the above plot, we can see that the player rating increases with age and for Defender, Midfielder and Forwards around the age range of 30-35 the rating starts to drop. While for goalkeepers, the trend is different with ratings remaining high from ages 25 onwards. So, the assumption that player rating falls as age increases is true only after age range of 30-33 years.

Plot the distribution of ages for players by positions

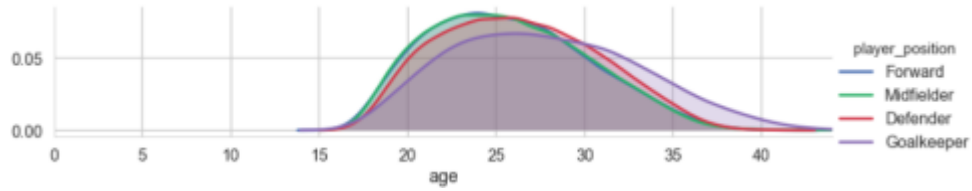

Fig 3.2.2

From the above plot, we can see that the distribution of player age is not normally distributed, with the most players for Defender, Midfielder and Forward positions being around the age range of 22-25. While for goalkeepers, most players are in the age range of 23-32 and there are more older goalkeepers than other players

Difference in ratings between positions

To predict the player rating we can separate the data into train and test datasets and fit a regression model to the data to find a model with a good fit. However, we know that the data has 36 attributes for each player and according to our assumptions about the game, we think that there is a difference between ratings based on which position the player plays at. If this is true, it would make sense for us to separate the data by positions and fit separate models on each position data. For this, we test the following hypothesis:

H0: There is no significant difference between the mean ratings of the different position groups

VS

H1: There is significant difference between the mean ratings of the different position groups

We test the difference between the means using one way anova and at 5% l.o.s we reject the null hypotheses for the alternate. Check Fig 3.2.3, p-value = 6.67….e-54

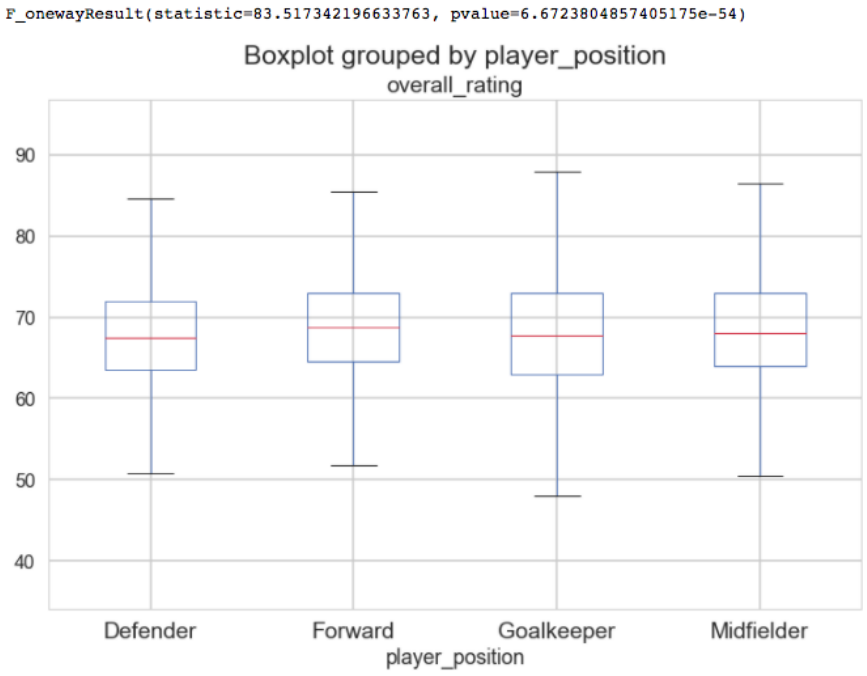

Fig 3.2.3

We carry out Tukey's test using the pairwise_tukeyhsd() function in the statsmodels.stats.multicomp library:

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
Defender	Forward	1.2678	1.0532	1.4825	True
Defender	Goalkeeper	0.2732	-0.0058	0.5523	False
Defender	Midfielder	0.6837	0.5058	0.8616	True
Forward	Goalkeeper	-0.9946	-1.2948	-0.6944	True
Forward	Midfielder	-0.5842	-0.7936	-0.3747	True
Goalkeeper	Midfielder	0.4104	0.1354	0.6855	True

Fig. 3.2.4

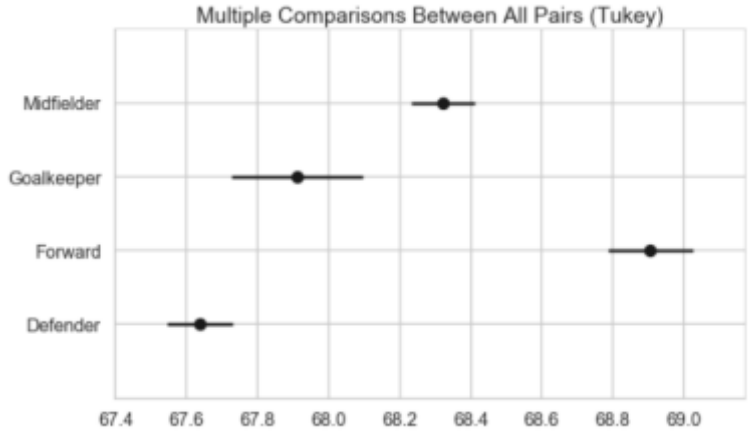

Fig 3.2.5

The output of the Tukey test shows the average difference, a confidence interval as well as whether you should reject the null hypothesis for each pair of groups at the given significance level. In this case, the test suggests we reject the null hypothesis for 5 pairs, with an exception of the groups Defender-Goalkeeper. The 95% confidence interval plot reinforces the results visually: only Defender and Goalkeeper groups' confidence intervals overlap.

We have enough evidence to conclude that there is difference in ratings between the positional groups and we go forward with our suggestion to fit separate models for each position.

Testing Normality, Linear Relationship and Multicollinearity: We tested the data for checking distribution of available features and using the normaltest() function from the scipy.stats.mstats library we concluded that most of the variables fail the normality test. Overall Rating, Height, Weight, Marking and Sliding Tackle are the only variables which are normally distributed. Hence, to fit regression models to this data, we will need to normalize these features.

To check for linear relationship between each of the attributes and the dependent variable 'overall_rating' we made scatter plots. According to these plots the relationship between most of the features and the rating is not linear but curvilinear, according to this a better fit would be Ridge Regression or Support Vector Regression (SVR)

To test for multicollinearity between the variables, we made a heat map of correlation between the variables, the result was as shown in Fig 3.2.6. As we can see most these variables are correlated, so there is multicollinearity present in the data, to deal with this we will use PCA to find independent

components which can be used to predict overall rating. Note that the goalkeeping attributes have high negative correlation with the other attributes. We need to select the features with significance for each of the positions to build an efficient model

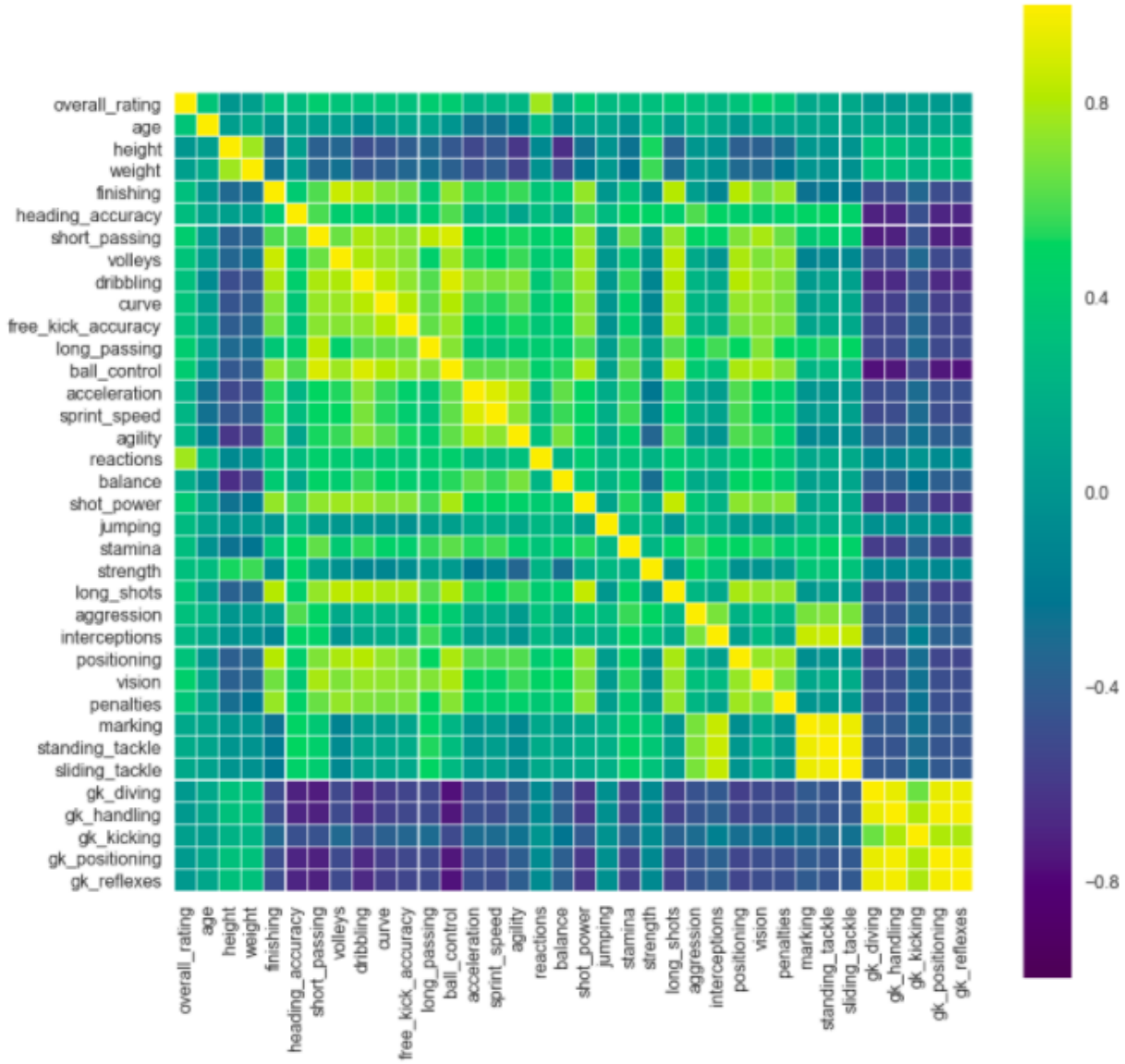

Fig. 3.2.6

3.3 MODELING TECHNIQUES

Feature Selection and Dimensionality Reduction: The dataset is very large with many features (36). Feature selection was done basis player position where if there is no significant relationship between player rating and feature we discard the feature. For e.g. For non-goalkeeper positions some goalkeeper related features aren't significant so we discard them. The selected features are scaled to increase modeling accuracy

Principal Component Analysis is performed on the scaled feature data to reduce dimensionality and the new components are selected basis scree plot. These new components are used in linear and polynomial regression models. While for Ridge, Lasso and SVR we used the original feature data after scaling as these methods inherently deal with multidimensionality issues

Model Selection: We split the data into training (~80%) and test (~20%) and then find the best fitting model for this transformed (train) data to predict the overall ratings for the next ages of the players (test data). We fit the following models and find one with the best fit to predict the rating at the next age for each player:

Linear Regression: Multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables

Polynomial Regression: Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x

Ridge Regression: Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It is hoped that the net effect will be to give estimates that are more reliable

Lasso Regression: The LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method that involves penalizing the absolute size of the regression coefficients

Support Vector Regression: Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. Firstly, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem

We compare the models and select the best fit to predict player ratings. Using these predicted ratings, we find the top x (client can change x as required) players for each position with the highest rating. This will be the data reported to the client to select a suitable player for the upcoming transfer season. As per the requirement, we can filter the players by age, country, league or other performance factors to find the best potential players the club will be interested in.

4. POSITIONAL DATA INSIGHTS

4.1. FEATURE SELECTION

For feature selection for the Position datasets, we find the r, r-squared, p-value for measuring the relationship of each independent feature with the dependent variable 'overall_rating'. The Null Hypothesis is

H0: There will be no significant prediction of overall rating by feature VS H1: There will be significant prediction of overall rating by feature

We test this hypothesis at l.o.s. 1% and use only the significant features for model building

Null Hypotheses is **true** for the following features for the corresponding position data and they are excluded as insignificant features:

Defender:

Insignificant features are:

	features	r_score	r2_score	pvalue
34	gk_reflexes	0.009236	0.001349	0.2215
33	gk_positioning	0.003310	0.000012	0.6613
31	gk_handling	-0.010202	0.000536	0.1769

Forward:

Insignificant features are:

	features	r_score	r2_score	pvalue
31	gk_handling	0.023150	0.000536	0.0195
30	gk_diving	0.021562	0.000465	0.0296
33	gk_positioning	0.003423	0.000012	0.7299

Midfielder:

Insignificant features are:

	features	r_score	r2_score	pvalue
33	gk_positioning	0.003289	0.000012	0.6406
34	gk_reflexes	0.002207	0.001349	0.754
31	gk_handling	0.001654	0.000536	0.8144
1	height	-0.016278	0.000819	0.0208

Goalkeeper:

Insignificant features are:

	features	r_score	r2_score	pvalue
3	finishing	-0.016550	0.583881	0.2484
6	volleys	-0.019294	0.583784	0.1784
21	long_shots	-0.020163	0.532845	0.1596
29	sliding_tackle	-0.028369	0.007979	0.0478
7	dribbling	-0.028393	0.526532	0.0477

Fig 4.1

4.2 DATA PREPARATION

We separate the data into train and test datasets such that 80% data is training and 20% is testing. However, as our aim is to predict rating for current age of player we take maximum age of each player and separate it as test data. For each of the positions the proportion of this test data is found to be approximately 20%

The train-test datasets are then split into x and y datasets such that x is a subset of all independent features to be used for modeling and y is the target variable- player rating. The x datasets are then scaled using StandardScaler() function from the preprocessing library.

4.3 FEATURE REDUCTION

We use Principal Component Analysis (PCA) method for feature reduction. PCA is a learning method where the input data is to principal components to lower the dimensions of the data. A good rule of thumb with PCA is that we should be able to explain 95% variance with the reduced dimensions. If we plot number of components vs variance retained we should see that the explained variance increases steadily and then saturates, see Fig 4.3.1 for the scree plot of Defender dataset:

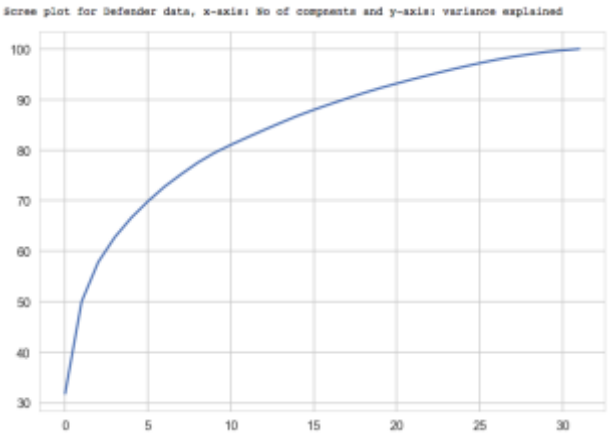

Fig 4.3.1

As we can see, % of variance in data is explained by 24 components, we select these 24 components for model building wherever feature reduction is required. Now we have 24 variables instead of 32. Similarly, we use PCA to find reduced components for Forward, Midfielder and Goalkeeper datasets.

4.4 DATA MODELING

We use the PCA components to fit Linear and Polynomial Regression Models to the data and we use the scaled features to fit Ridge, Lasso and SVR regression methods. We perform K-fold cross validation with k=5 on the train data for each model. We find the optimal parameters using grid search function for Ridge and Lasso regression. The modified models are then trained on the training data and fit on the test data for prediction.

4.5 MODEL COMPARISON

Shown below are the model test accuracy scores for each position:

Defender Model Comparison:

	Model	Test_Accuracy
0	Linear_Regression	75.32
1	Ridge_Regression	89.08
2	Lasso_Regression	89.06
3	SVR	97.79

Forward Model Comparison:

	Model	Test_Accuracy
0	Linear_Regression	56.22
1	Ridge_Regression	91.56
2	Lasso_Regression	91.64
3	SVR	98.21

Midfielder Model Comparison:

	Model	Test_Accuracy
0	Linear_Regression	65.71
1	Ridge_Regression	86.60
2	Lasso_Regression	86.61
3	SVR	97.06

Goalkeeper Model Comparison:

	Model	Test_Accuracy
0	Linear_Regression	-266.57
1	Ridge_Regression	87.14
2	Lasso_Regression	86.61
3	SVR	97.06

Fig 4.5.1

4.6 PREDICTION

We find that SVR is the best fitting technique for all position datasets giving accuracy of 97% and higher. We use SVR to predict player rating.

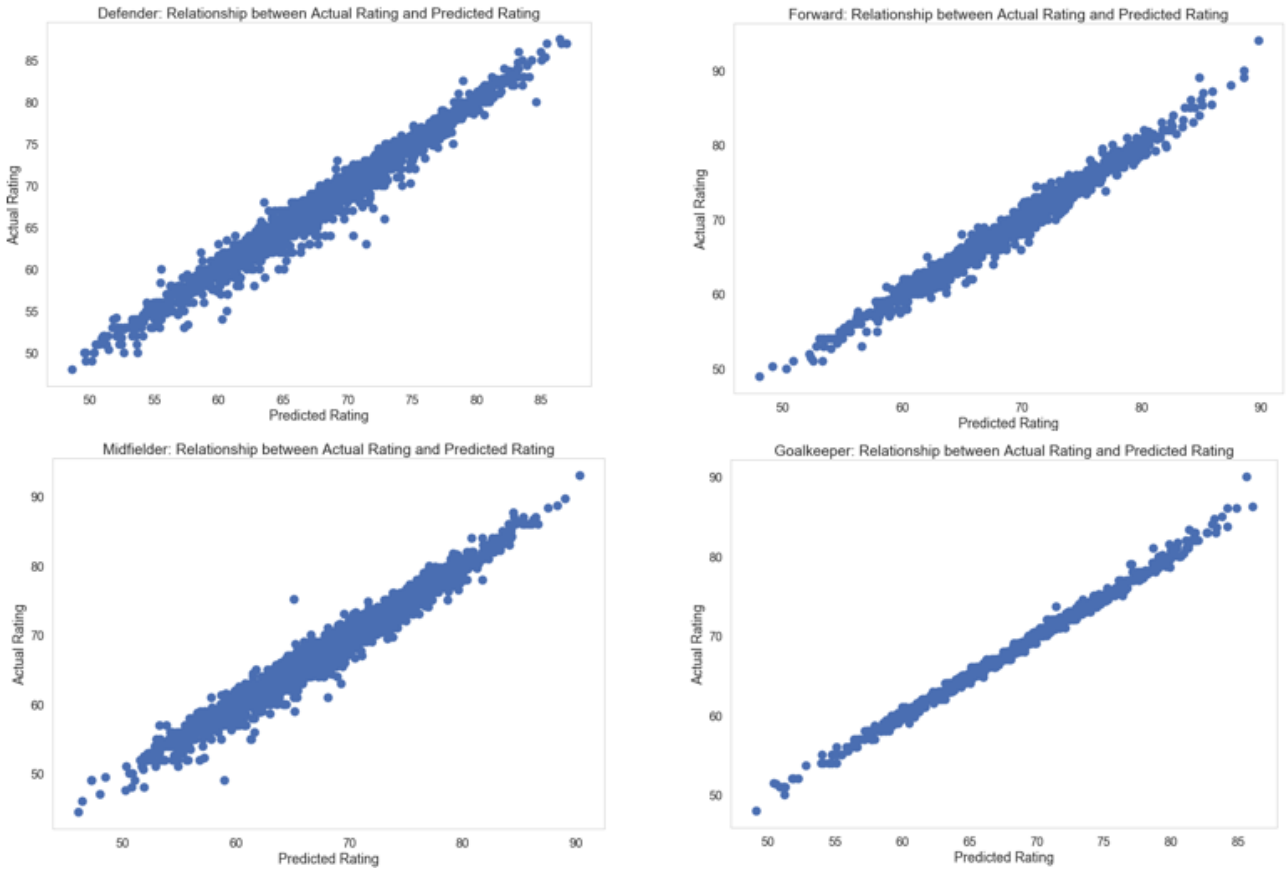

Fig 4.5.1

As we can see in the Fig 4.5.1, the scatter plots of the actual vs predicted ratings for all four positions show almost negligible dispersion. So, if the Actual is 50, the predicted will be reasonably close to 50 too. We can draw a regressed diagonal line through the data and the model will have a high R Square, since all the points would be close to this diagonal line.

5. RESULTS

The table 'players with leagues' has data for active players who have played in matches since 2013 January, if a player does not have any match data since Jan 2013 then he is excluded from this table. We have merged our predicted datasets with this table to get a list of players who are active. This will help weed out non-active players. However, the merge results in the loss of 26% (2771) players from the data. Below are the lists for each position:

DEFENDERS:

TOP 10

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
80562	Defender	Thiago Silva	86	87	2	France Ligue 1	France	31
30962	Defender	Sergio Ramos	86	87	2	Spain LIGA BBVA	Spain	30
30894	Defender	Philipp Lahm	86	87	2	Germany 1. Bundesliga	Germany	32
36183	Defender	Jerome Boateng	85	87	4	Germany 1. Bundesliga	Germany	27
56678	Defender	Diego Godin	85	85	4	Spain LIGA BBVA	Spain	30
36388	Defender	Mats Hummels	84	86	8	Germany 1. Bundesliga	Germany	27
121633	Defender	David Alaba	84	85	8	Germany 1. Bundesliga	Germany	24
31306	Defender	Branislav Ivanovic	84	80	8	England Premier League	England	32
19327	Defender	Miranda	84	85	8	Italy Serie A	Italy	32
39027	Defender	Vincent Kompany	84	85	8	England Premier League	England	30

TOP SIGNIFICANT FEATURES

Standing tackle, interceptions, marking, sliding tackle and reactions

FORWARDS

TOP 10

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
30981	Forward	Lionel Messi	89	94	1	Spain LIGA BBVA	Spain	29
35724	Forward	Zlatan Ibrahimovic	88	89	2	France Ligue 1	France	34
40636	Forward	Luis Suarez	88	90	2	Spain LIGA BBVA	Spain	29
37412	Forward	Sergio Aguero	87	88	4	England Premier League	England	28
38817	Forward	Carlos Tevez	85	85	6	Italy Serie A	Italy	32
164684	Forward	James Rodriguez	85	87	6	Spain LIGA BBVA	Spain	25
93447	Forward	Robert Lewandowski	85	87	6	Germany 1. Bundesliga	Germany	27
30829	Forward	Wayne Rooney	85	85	6	England Premier League	England	30
19533	Forward	Neymar	84	89	9	Spain LIGA BBVA	Spain	24
26166	Forward	Karim Benzema	84	86	9	Spain LIGA BBVA	Spain	28

TOP SIGNIFICANT FEATURES

Ball control, reactions, positioning, finishing, volleys, shot power, long shots, short passing and dribbling.

MIDFIELDERS

TOP 10

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
30893	Midfielder	Cristiano Ronaldo	90	93	1	Spain LIGA BBVA	Spain	31
30834	Midfielder	Arjen Robben	89	89	2	Germany 1. Bundesliga	Germany	32
107417	Midfielder	Eden Hazard	88	88	3	England Premier League	England	25
30955	Midfielder	Andres Iniesta	87	88	4	Spain LIGA BBVA	Spain	31
37459	Midfielder	David Silva	86	86	6	England Premier League	England	30
30924	Midfielder	Franck Ribery	86	86	6	Germany 1. Bundesliga	Germany	33
31921	Midfielder	Gareth Bale	86	87	6	Spain LIGA BBVA	Spain	27
129944	Midfielder	Marco Reus	86	86	6	Germany 1. Bundesliga	Germany	27
39854	Midfielder	Xavi Hernandez	85	86	9	Spain LIGA BBVA	Spain	35
154257	Midfielder	Sergio Busquets	85	86	9	Spain LIGA BBVA	Spain	28

TOP SIGNIFICANT FEATURES
Ball control, reactions, vision, and short passing

GOALKEEPERS
TOP 10

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
182917	Goalkeeper	David De Gea	86	86	1	England Premier League	England	25
27299	Goalkeeper	Manuel Neuer	85	90	2	Germany 1. Bundesliga	Germany	30
170323	Goalkeeper	Thibaut Courtois	84	86	4	England Premier League	England	23
42422	Goalkeeper	Samir Handanovic	84	83	4	Italy Serie A	Italy	32
30859	Goalkeeper	Petr Cech	84	86	4	England Premier League	England	34
215168	Goalkeeper	Bernd Leno	83	83	7	Germany 1. Bundesliga	Germany	24
26295	Goalkeeper	Hugo Lloris	83	85	7	England Premier League	England	29
31432	Goalkeeper	Joe Hart	83	84	7	England Premier League	England	29
30717	Goalkeeper	Gianluigi Buffon	83	84	7	Italy Serie A	Italy	38
37421	Goalkeeper	Claudio Bravo	82	83	10	Spain LIGA BBVA	Spain	32

TOP SIGNIFICANT FEATURES
Goalkeeping positioning, goalkeeping reflexes, goalkeeping diving and goalkeeping handling

A list of top 10 players under 25 years of age is provided in the appendix

6. LIMITATIONS

There are certain limitations in the data that reduce the robustness of the models we have developed:

1. Not all historical data is available for each player. Knowing that the FIFA Player Ranking is based on historical data of player, we do not have this data for each player. A lot of information will not be learned by the model due to the absence of this data.
2. Data related to ranks of leagues, teams and countries is also not available in this data. This information would have made the model more accurate as these factors play an important role in determining player ranking.
3. There is not enough historical data available for each player, if it were, a time series model could be built to predict the future ratings of players
4. There was country and league data missing for few players and more than 20% players from our players' data had not been active since 2013. These players were excluded from featuring in the top players lists.

7. RECOMMENDATIONS

The player rating increases with age and for Defender, Midfielder and Forwards around the age range of 30-35 the rating starts to drop. While for goalkeepers, the trend is different with ratings remaining high from ages 25 onwards. So, the general assumption that player rating falls as age increases is true only after age range of 30-33 years. This information can be used on the best players given by the models to make better selections

We have identified the current best players and this list can be used to filter out potential new additions to the club. We were successful in building a model to predict overall ratings of Goalkeepers with 99% accuracy, Midfielders with 97% accuracy and lastly Defenders and Forwards with 98% accuracy. These models were based only on the numeric attributes available to us. By predicting the ratings for

players by age the client can conveniently gauge the expected player performance for his upcoming matches

8. FUTURE WORK

The FIFA 17 game data is now available, we can add that data to enhance and update the player ranking based on current data. With availability of more future data we can identify patterns in attributes of top players to then identify similar patterns in new players. This will help in scoping out up and coming players, this will help the client to make decisions related to player transfers as they could look at signing younger players with higher potentials at a lower cost.

We could also do additional exploratory analysis on effect of player nationality or age on players' attributes. We can define more specific player positions as Left Back, Right Winger, Attacking Midfielder etc., to give more specific suggestions for player selection. Also, we can identify leagues or countries which produce top players. There is demand in identifying key performance attributes, effect of player's associations with leagues and other such factors. Further studying of attribute differences between top players and bottom ranking players can help us understand attributes better.

If desired, we can further deep dive into analyzing the historical performance trend of the top players so that the client can make a final hiring decision considering the historical performance of the player.

References:

1. NCSS Ridge Regression: Link
2. Statistics solutions- Assumptions of Multiple Linear Regression Link
3. Hamelg Blogspot – Python for data analysis ANOVA Link
4. VG 24/7 Article Link
5. Kernel SVM tripod Link
6. Wikipedia-Linear and Polynomial Regression

APPENDIX

TOP 10 UNDER 25

1. DEFENDER:

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
121633	Defender	David Alaba	84	85	1	Germany 1. Bundesliga	Germany	24
115591	Defender	Ricardo Rodriguez	82	82	2	Germany 1. Bundesliga	Germany	24
282674	Defender	Daniel Carvajal	81	81	3	Spain LIGA BBVA	Spain	24
230982	Defender	Raphael Varane	81	82	3	Spain LIGA BBVA	Spain	23
267365	Defender	Marquinhos	80	80	6	France Ligue 1	France	21
184999	Defender	Shkodran Mustafi	80	82	6	Spain LIGA BBVA	Spain	24
188555	Defender	Stefan de Vrij	80	81	6	Italy Serie A	Italy	24
411617	Defender	Aymeric Laporte	80	81	6	Spain LIGA BBVA	Spain	21
213485	Defender	Matija Nastasic	79	79	9	Germany 1. Bundesliga	Germany	23
195299	Defender	Serge Aurier	79	80	9	France Ligue 1	France	23

2. FORWARD:

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
19533	Forward	Neymar	84	89	1	Spain LIGA BBVA	Spain	24
325916	Forward	Paulo Dybala	82	79	2	Italy Serie A	Italy	22
181276	Forward	Romelu Lukaku	80	81	5	England Premier League	England	23
213501	Forward	Alvaro Morata	80	80	5	Italy Serie A	Italy	24
303824	Forward	Memphis Depay	80	80	5	England Premier League	England	22
241825	Forward	Francisco Alcacer	80	80	5	Spain LIGA BBVA	Spain	23
364520	Forward	Domenico Berardi	80	81	5	Italy Serie A	Italy	22
194165	Forward	Harry Kane	80	81	5	England Premier League	England	23
354467	Forward	Yannick Ferreira-Carrasco	79	79	9	Spain LIGA BBVA	Spain	23
202443	Forward	Vincent Aboubakar	79	78	9	Portugal Liga ZON Sagres	Portugal	24

3. MIDFIELDER:

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
248453	Midfielder	Paul Pogba	84	86	1	Italy Serie A	Italy	23
184536	Midfielder	Philippe Coutinho	83	84	2	England Premier League	England	24
177714	Midfielder	Mario Goetze	83	84	2	Germany 1. Bundesliga	Germany	24
184533	Midfielder	Koke	82	83	6	Spain LIGA BBVA	Spain	24
191315	Midfielder	Isco	82	84	6	Spain LIGA BBVA	Spain	24
243164	Midfielder	Julian Draxler	82	82	6	Germany 1. Bundesliga	Germany	23
190972	Midfielder	Marco Verratti	82	84	6	France Ligue 1	France	23
157723	Midfielder	Christian Eriksen	82	83	6	England Premier League	England	24
174850	Midfielder	Erik Lamela	81	79	9	England Premier League	England	24
242709	Midfielder	Roberto Firmino	81	81	9	England Premier League	England	24

4. GOALKEEPER:

player_api_id	player_position	player_name	predicted_overall_rating	overall_rating	player_rank	league_name	country_name	age
170323	Goalkeeper	Thibaut Courtois	84	86	1	England Premier League	England	23
215168	Goalkeeper	Bernd Leno	83	83	2	Germany 1. Bundesliga	Germany	24
184554	Goalkeeper	Marc-Andre ter Stegen	81	82	3	Spain LIGA BBVA	Spain	24
181910	Goalkeeper	Mattia Perin	80	81	5	Italy Serie A	Italy	23
177126	Goalkeeper	Jan Oblak	80	81	5	Spain LIGA BBVA	Spain	23
212815	Goalkeeper	Timo Horn	80	80	5	Germany 1. Bundesliga	Germany	23
245555	Goalkeeper	Geronimo Rulli	79	79	7	Spain LIGA BBVA	Spain	24
288880	Goalkeeper	Jack Butland	79	78	7	England Premier League	England	23
287894	Goalkeeper	Loris Karius	78	79	9	Germany 1. Bundesliga	Germany	23
210164	Goalkeeper	Alphonse Areola	77	77	10	Spain LIGA BBVA	Spain	23