

OBJECTIVES

The following work tries to predict the overall ratings of FIFA players based on various attributes present in the dataset using the techniques of **Regression**. These are attributes based on actual data of the latest EA's FIFA 18 game. Attributes include on all player style statistics like Dribbling, Aggression, GK Skills, etc., and also player personal data like Nationality, Photo, Club, Age, Wage, Salary, etc. It can be usually seen that the overall rating of a participant could be influenced by the age of the player, the value of the player, and various other factors.

MATERIALS & METHODS

The Regression task was performed on the FIFA-19 dataset collected from KAGGLE.

- Data pre-processing and Exploratory Data Analysis(EDA) presents us with correlation among various attributes in the data
- The regression techniques applied help us in predicting the overall rating of the players based on attributes. The regression techniques considered for our study

– Linear Regression

In this, we try to model a linear relation between input variables and a target variable, in which target variable is calculated as a linear combination of independent variables

– Lasso Regression

It is basically a regularized version on linear regression. It adds an L1 penalty to the loss function, which shrinks the coefficients of the attributes that do not contribute significantly to the prediction task.

INTRODUCTION

The **Overall** rating indicates the current potential of a player and the **Potential** indicates the potential rating of a player in the future. Overall rating of a player is dependent on various attributes like Age of the player, Value of the player, it also varies based on the position of player on the field, Weekly wage of the player, etc. Calculating the Overall rating of a player helps in understanding the worth of a player and assigning an optimal position to the player based on their skillset. In this study we wish to predict overall rating of a player based on above attributes.

RESULTS 2

The following table presents us with the error values when the two techniques considered (Linear Reg. and Lasso Reg.) were used on our data set. The error metrics considered were Mean Squared Error(MSE), Root Mean Squared Error(RMSE), R-Squared Score

$$MSE = \sum_{i=1}^D (x_i - y_i)^2 \quad (1)$$

$$RMSE = \sqrt{MSE} \quad (2)$$

$$R^2 = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2) \cdot (n \sum y^2 - (\sum y)^2)}} \quad (3)$$

Method	MSE	RMSE	R-squared
Linear Reg.	3.3862	1.8402	0.9310
Lasso Reg.	3.9890	1.9972	0.9188

Table 1: Error Values for various Regression techniques

RESULTS 1

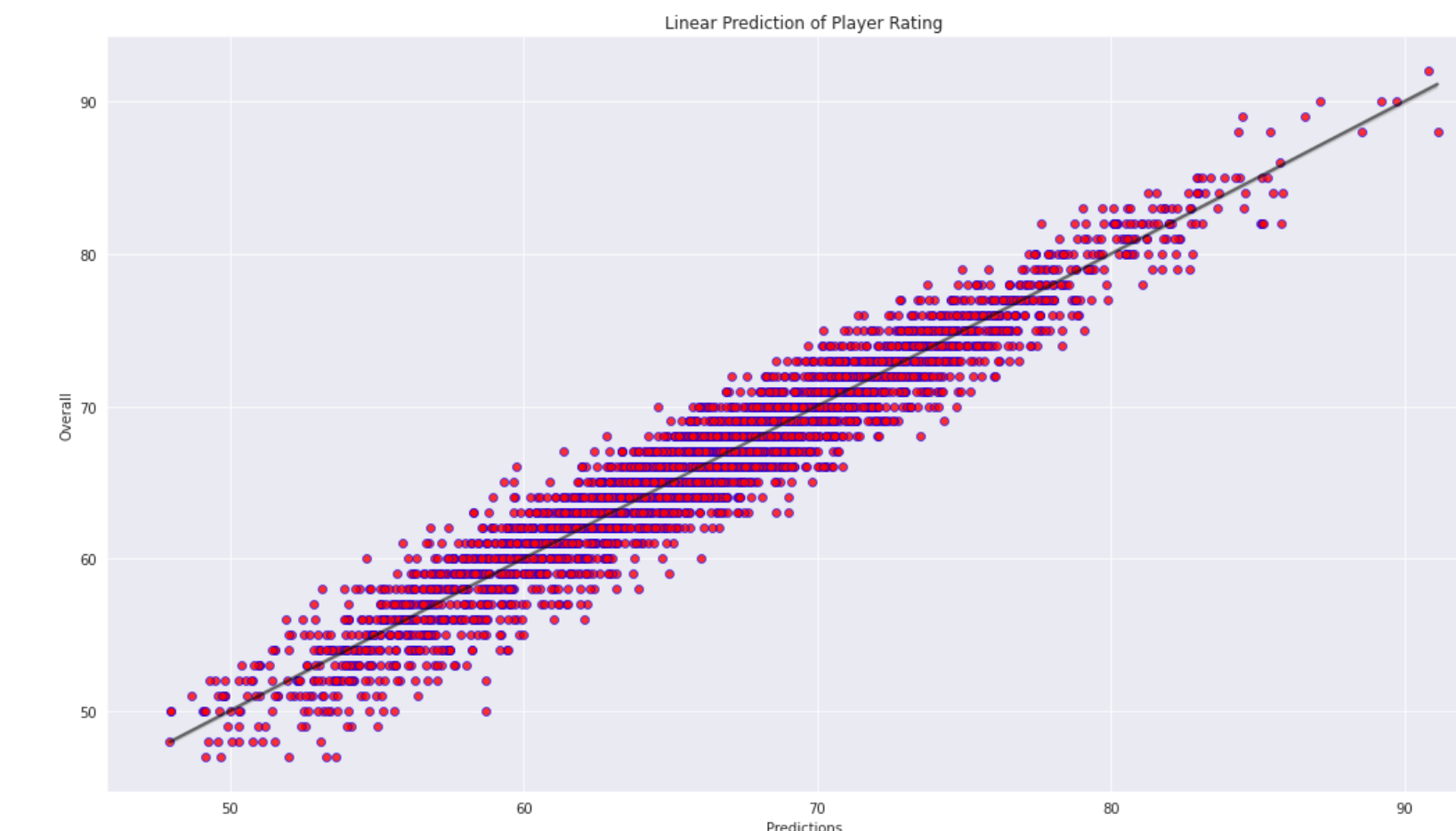


Figure 1: Linear Regression Plot

We have performed EDA and dropped the variables that do not affect the target variable significantly.

Lasso Regression works similar to Linear regression except that it includes L1 penalty to the loss function. This shrinks the coefficients of input variables that do not contribute much to the prediction task. MSE, RMSE and R-Squared values are used to compare the performance of linear regression and lasso regression. It is found that linear regression performs better than the lasso regression in case of predicting overall potential of a player

cantly. Few attributes are linearly correlated with target variable, making linear regression more effective for prediction. The error metrics MAE, MSE and RMSE present us with a figure of deviation of predicted values from the actual values. We require them to be as small as possible. The errors recorded were presented in Results 2 section. The R-Squared, measures how close the data fits to the regression model, the more the number approaches 1 the more it shows that a significant percentage of the values are explained by the linear regression.

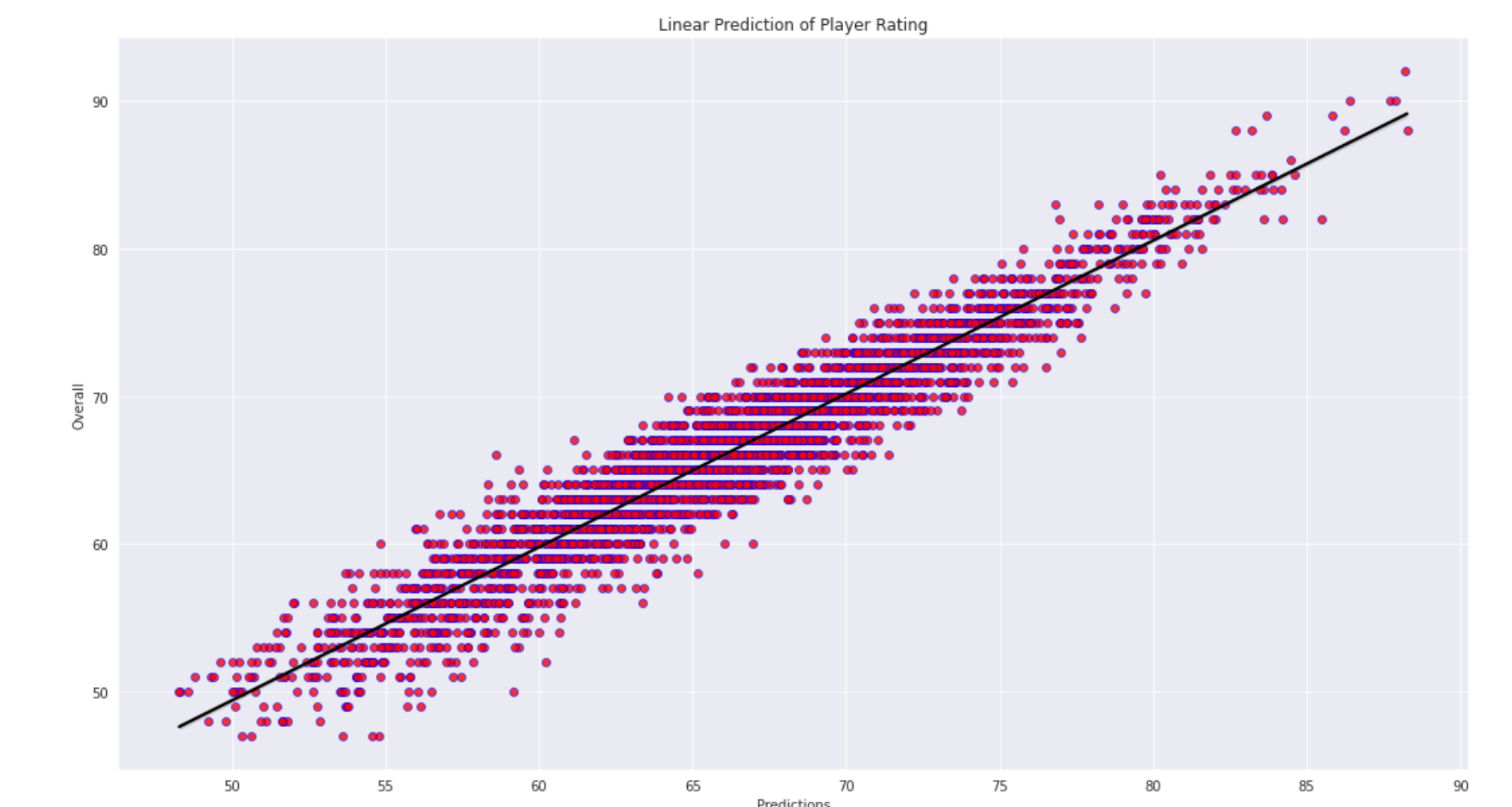


Figure 2: Lasso Regression Plot

CONCLUSION

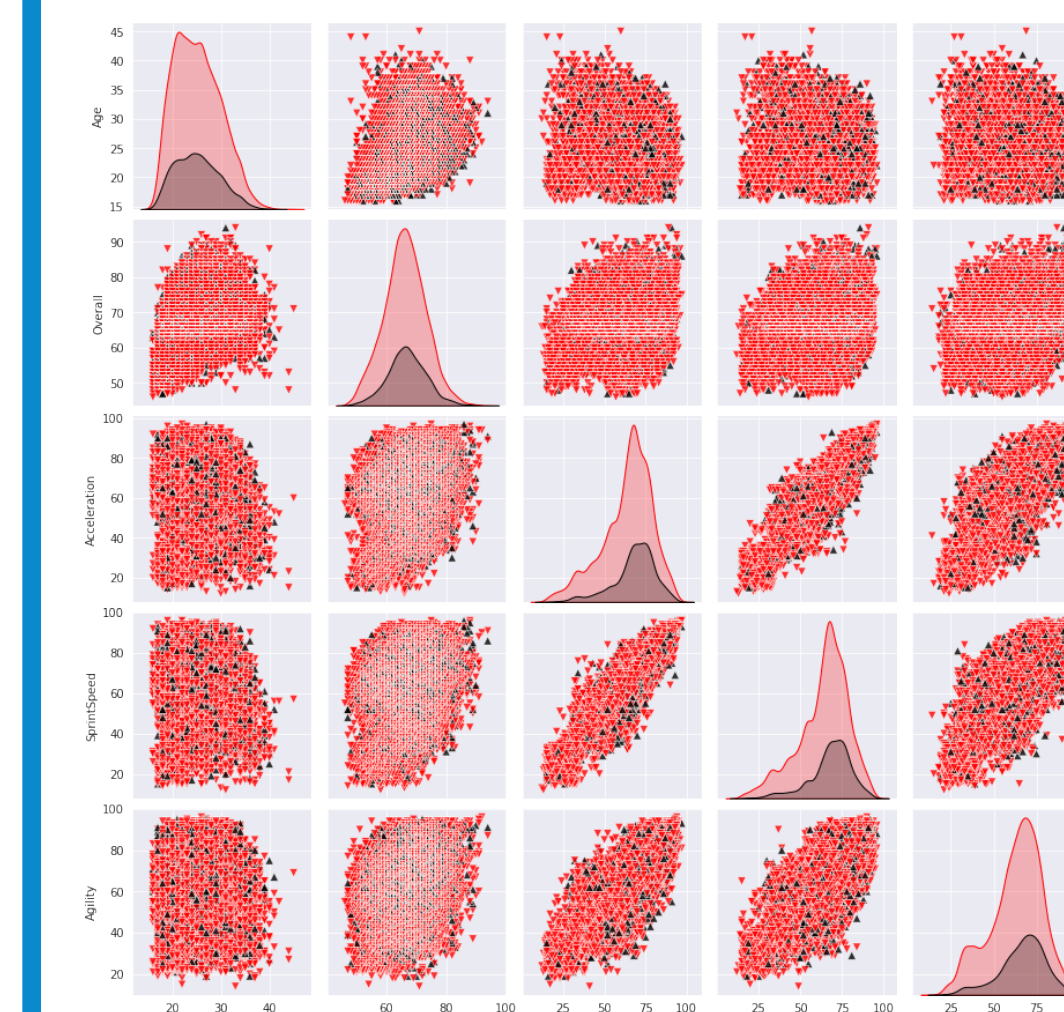


Figure 3: Pair Plot of attributes

- Based on EDA we can

identify that the Potential rating and the sprint speed tend to decrease as the age of player increases

- There is a pairwise linear relationship between Acceleration, Agility and Sprint Speed
- Most of the players are usually right footed
- The attributes which significantly contribute to the

Overall rating based on the linear regression include ['Age', 'Nationality', 'Preferred Foot', 'Weak Foot', etc]

- Most of the players are from England
- height of most players are ranging from 5'6 to 6'3 foot per inch
- Work rates of most players is Medium/ Medium

REFERENCES

- [1] Kaggle.
- [2] Junlong Zhao, Hansheng Wang, Wei Lan, Yingying Ma and Chih-Ling Tsai. Article Title. *Journal title*, 101(476):1418–1429, Jan 2006.

FUTURE RESEARCH

Analysis of Match in FIFA has predominately utilised notational analysis, a statistical summary of events based on video footage, to study the sport and prepare teams for competition.

Machine learning is a relatively new concept in football, and little is known about its usefulness in identifying performance metrics that determine match outcome.

CONTACT INFO. OF MEMBERS

Email fmml20210329@ihub-data.iiit.ac.in

Name Samarth Meghani

Email fmml20210264@ihub-data.iiit.ac.in

Name Sai Mahathi P V N