PREDICTING MARKET VALUE OF A FOOTBALL PLAYER

GROUP MEMBERS

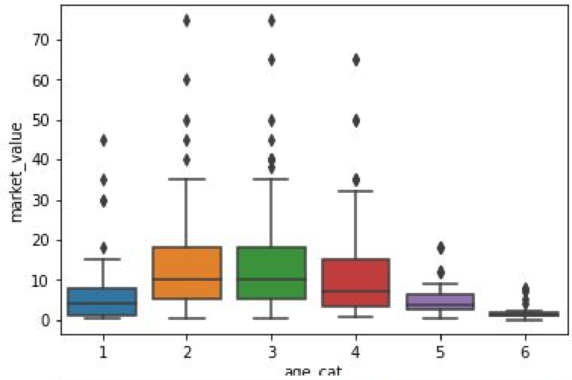
- Nupur Kapur
- Malika Jain
- Ishpreet Kaur Gulati
- Anchit Srivastava
- Navneet Kaur

PROBLEM DESCRIPTION:

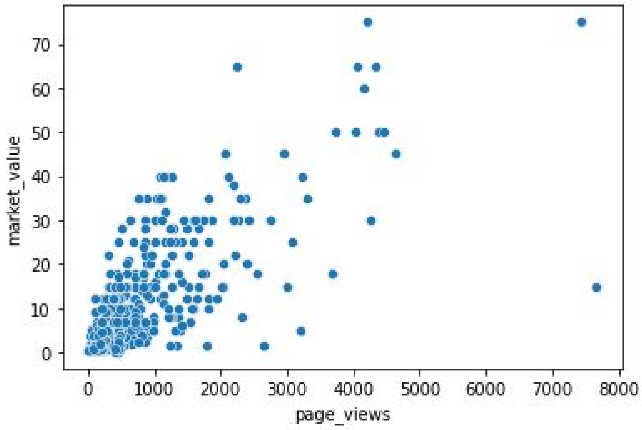
Predicting fair market value of football players with the help of features present in the dataset. Deploy the model as a RESTful Web Service.

EXPLORATORY DATA ANALYSIS (EDA):

We used Seaborn library to explore the data with the help of graphs and analysed which features influenced the market value and were important in prediction of market value of the player and remove those features which were not influencing market value.

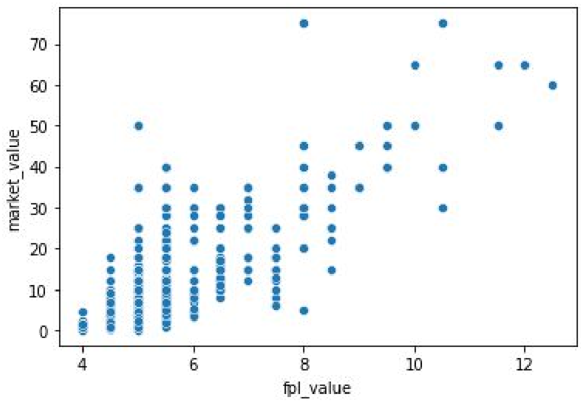


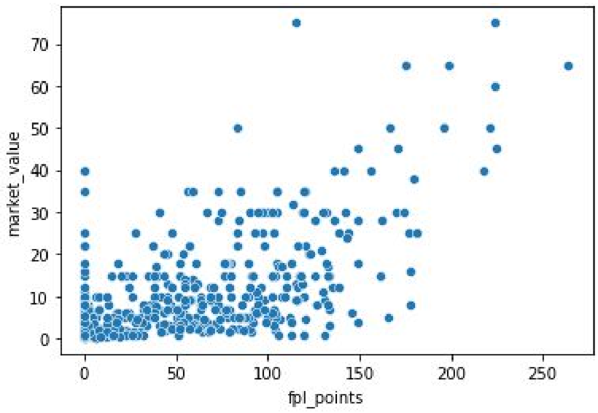
In the above graph we can see, players falling in the age category 2-4 are more popular and get a good price for themselves as compared to category 1, 5 and 6. The new players and the older players get less price as compared to other players. We do have a few outliers that get a higher price than others.



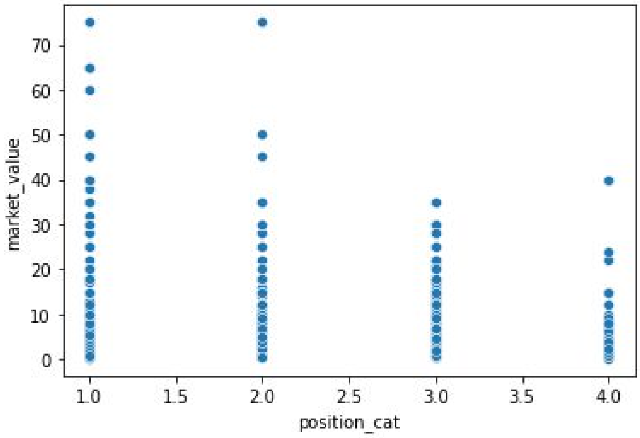
In this graph we can see that the players that have a higher number of the page views get a much better price as compared to other players with less number of views.

There are some outliers like some players with a huge number of page views have got a comparatively less price for themselves.





In the above graphs, fpl_value and fpl_points have a high correlation with the market value. The higher the fpl value and fpl points the higher the price for the player.



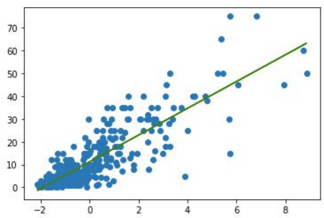
This graph shows that the players that come under position category 1 and 2 get a much better price as compared to the other categories.

We have taken age_cat instead of age, position_cat instead of position and region instead of nationality because these variables were highly correlated and also to reduce the dimensionality.

Models Tried:

Features Used: age_cat, page_views, position_cat, fpl_value, fpl_sel, fpl_points, region, club_id, big_club, new_signing.

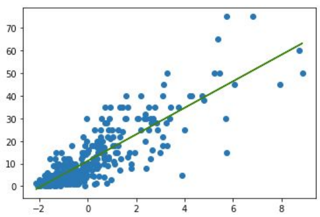
1. Linear Regression:



Mean_squared_error:

r2_score:

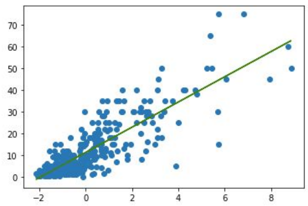
2. Lasso Regression



Mean_squared_error: 28.69

r2_score: .8101

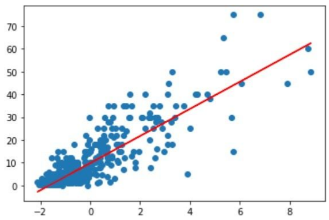
3. Ridge Regression:



Mean_squared_error: 28.36

r2_score: .8102

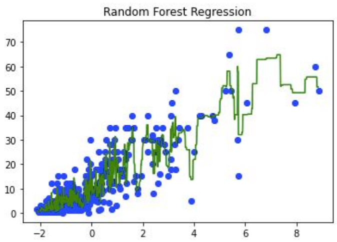
4. Support Vector Regression:



Mean_squared_error: 25.27

r2_score: 0.8332

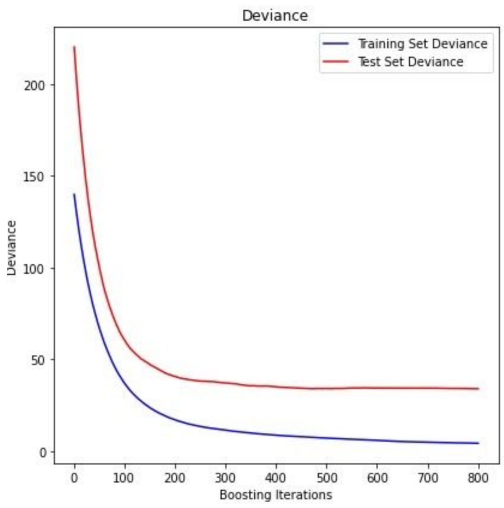
5. Random Forest:



Mean_Squared_error: 20.9

r2_Score: 0.859

6. Gradient Boosted Regression:



Mean_squared_error: 34

r2_score: 0.83

7. Nearest Neighbour Regression:

   r2_score: 0.71

8. Tree Regression:

   Mean_squared_error: 79

   r2_score: 0.62

Random Forest is chosen because it has the highest r2_score and least Mean_squared_error.

Restful web Service