Football player market value predictor and fantasy team recommender

1)Ajeya B S PES1201701604 *CSE dept. PES University.*
2)Sirajahamed N D PES1201701496 *CSE dept. PES University.*
3)Viraj C Bukitagar PES1201701503 *CSE dept. PES University.*

***Abstract*—Recent progress on predicting the value of the player in the market has been done using huge data obtained from various sources all over the world. Taking into account of various attributes of the player which determines performance and which in turn influences his value in the market. Many models have been developed using these features values for a particular year like linear models, Neural Network models etc and have been fairly successful. We aim at developing model which takes into the account of the players various features development over the years. Also giving the goalkeepers their fair share of importance.**

I. INTRODUCTION

Football is the most popular sport in the world, and thus there is no surprise that footballers earn a lot of money. The market value of the ones at the pinnacle hold a huge market value.

With all teams not having a big budget, the ones with a smaller budget have to manage the finances and also should ensure the best team possible with those finances. So, in this project, we aim to tackle this problem by predicting the market value of footballers given various attributes of them like age, team, overall rating and their market value from the years before, and also recommend the best possible team, given the money allotted to buy players .

Analysis of the data has been done in various fields in today's world be it business, health etc. But in recent years there has been its impact lately in the Sports which statisticians have taken a liking to. Football, most popular and watched sport in the world, has been affected too. But it was difficult to analyze the football data until recently because of difficulty to collect data and organize as it is the sport which is played all around the world.

Sure, football is mostly about the goals and cups but it also dependent on hugely on the quality of the players. When its quality of the players, their value in the market also plays a vital role. Huge transfers occur every year where top players are always involved. Recently, Neymar was transferred from FC Barcelona (Spanish Team) to PSG (French Team) for a record-breaking transfer fee of the 222 million euros.

This project mainly focuses on predicting market value of the players using modelling techniques. Various steps are involved to get this job done. Data has to be collected and organized into the features and target value, model this data on various models, test and evaluate the models, and finally make predictions based on the best model and check its accuracy.

Various papers have been published on the similar lines. Few of them have been reviewed to compare what is the current state of the problem and what we approach to do.

II. RELATED WORK

A. Linear Models

Yuan He [1], he has predicted the market value of the soccer(football) players using linear modeling techniques. He has collected the data from online resources like transfermarkt.de and Wikipedia. He consolidated the data to contain the personal information of the player and performance data of the player. Here he has tried to fit a linear model to the data which gives the best accuracy. Initially he did some summary statistics between the predictors and dependent variables to know the relationship between them. Then ANOVA was used to get the features which had acceptable p values.

Now he fitted the data with the selected features for various linear models such as OLS(Ordinary Least Squares), KNN (K-Nearest Neighbors),Ridge Regression for various lambdas and PCR(Principal Component Regression). He obtained the results which gave him the least RMS error for the PCR model with 16 predictor variables.

B. Neural Network Models

Another approach for this task used by Sourya Dey [2] is Multilayer deeply connected Neural Network. He has used the FIFA 2017 dataset which contains roughly about 15000 players data. He has created a 3 layered dense Neural Network. He has cross-validated the number of neurons for each layer and has chosen the combination which gives the highest accuracy. He tried

using different combinations of activation functions like Sigmoid + Softmax etc. for different number of neurons in the hidden layer. Also the accuracy for various learning rate was also calculated. All the optimum results were used for various parameters for modelling. He used the momentum for the gradient descent to avoid the local minimum.

C. *Recommendation Systems*

A recommendation system was created by Jaka, Karsten and Kostas [3] for recommending the starting lineup 11 players. They have used Collaborative Filtering Recommendations which uses the preferences or tastes of many users to recommend the new user. It first determines the worst player available based on age, overall and position. And it gets players which are at least better than worst player and selects the top 11 players.

D. *Players' Selection*

A model for selecting the players in the football team was developed by Onwuachu Uzohukwu C and P Enyindah [4] in their paper. They have used multiple Neural Network Models for Player's Resistance, Speed, Physical and Technique. They have used the PES stats dataset. The mean of the output of each Network is taken which is used as a criterion for the selection of the player.

E. ***Our Approach***

In various paper which we have surveyed above majority of the models compare players with different positions which seldom is case in real world. So, we will try to make different use different models for corresponding positions.

We are taking into account of players with high value which may seem like outliers but are integral part. We split the data on the overall performance value to separate them out.

Other models try to predict the transfer value of the player based on only the single year data which doesn't give much insight of the caliber of the player. But analysis of his performance over the years tells a lot about the player which has high effect in the value of the player.

From dataset of 2014-15 to 2018-19 the market value of the goalkeepers was undervalued. But in the present year, market value of the goalkeepers has risen. None of these above models consider this, but we aim to rectify it by introducing the bias.

Models are predicting the value on the new set of the attribute values which are given by the user which may or may not be real values. But we are trying to predict the value of the existing player taking into the considerations of his performance over the years.

III. DATA

The data used is the fifa dataset from the years 2015 to 2020 , courtesy Kaggle.

There are 104 columns in each dataset, with the number of rows, i.e , the number of players being about 17,000 in each year.

The following data preprocessing has been done:

- The numerous positions which a player might play in is changed to just his best position, for easier computation.
- The columns which have gk attributes have been set to NULL for players who are not goalkeepers, and the attributes which do not matter for a goalkeeper have been set to NULL for players who are not goalkeepers.
- The data which initially has 104 columns is brought down to 65 columns, as the attributes of the players' weak positions would not be of any use to predict his market value . (This might change in the future, as while making the best team, some players 'might' fall out of their favourite position.).
- The players are grouped into continents based on the nationality , as the region in general, rather than nationality has a higher influence on a footballer.
- The players are grouped into 4 categories: forwards, midfielders, backs, and goalkeepers , based on their best position.
  Care is taken for the above changes to the dataset such that when a new column is created ,the actual data is not lost.

IV. EXPERIMENTS

OLS is used which gives a general idea as to what features are significant.

Data transformation and cleaning is done such that residuals do not follow any pattern.

The data has been divided based on the position of the players and can be used for better modelling of the

transfer value by comparing against it's position counter part.
An adjusted R^2 of 0.97 is obtained, which leads to the initial belief that there is overfitting in this model. Stripcharts have been plotted between the target variable and the predictors which are assumed to have a huge impact on a player's performance (for example , age, position , continent , reputation), and the assumptions turn out to be true.

V. Future works

- Try other models which might provide a better result in predicting the market value, like, PCR,Ridge and Lasso Regression, ANN, etc.
- Recommend the best team with the funds available using techniques like knowledge based filtering.

References

[1] Predicting Market Value of Soccer Players Using Linear Modeling Techniques by Yuan He
[2] Pricing Football Players using Neural Networks by Sourya Dey in Neural Learning and Computational Intelligence April 2017, University of Southern California
[3] FIFARecs: A Recommender System for FIFA18 by Jaka, Karsten and Kostas
[4] A Machine Learning Application for Football Players' Selection by Onwuachu Uzohukwu C and P Enyindah