AN INTELLIGENT FRAMEWORK FOR PREDICTING THE VALUE OF FOOTBALL PLAYERS

G.Bodhini 1, K.Vengamamba 2, V.Sreevani 3, M.Satyam Reddy 4

1,2,3 Student, Department of CSE, Narasaraopeta Engineering College,Narasaraopeta,Guntur(D.T) A.P, India

4 professor, Department of CSE, Narasaraopeta Engineering College, Narasaraopeta, Guntur(D.T) A.P,India

golibodhini2002@gmail.com[1], kandulavengamba@gmail.com, vallurisreevani @gmail.com[3],

---

Abstract

As we all know that football is a very popular and a trending game across the globe and the football players like Christiano Ronaldo,Lionel Messi are became very popular in recent games. We all know their names and origin of the famous players, but many of us don't know their net value.Market values also play a vital role. Generally the market values are predicted by football experts. Actually the expert decisions are incorrect and not transparent. Now, we are going to propose a method to determine the football player's net value. This is completely based on machine learning algorithms. Here we are going to use a fifa 20 dataset, which is collected from kaggle.com. In this approach, we are going to use 4 models like Linear Regression, Multiple Regression, Decision Tree, Random forest. Here, we will take the most important factors that will help in predicting the player's market value. The results will be highly accurate, good performance and less errors. These results will help in between the foot ball clubs and player's agents. Hence, from this we can predict the football player's market value.

Keywords - player's value prediction, Linear Regression, Multiple Regression, Decision Tree, Random forest, machine learning.

I. INTRODUCTION

The football is one of the tremendous game in the world.The popularity for football players are increasing drastically day by day. The experts are paying keen observation on the market value of the players.So to determine the value we are taking different categories such as player characteristics, player performance and player popularity. Nowadays machine Learning is used in every domain, such as finance, disease prediction, value prediction etc. Here we are using FIFA 20 data set collected from kaggle.com.In this dataset, we have approximately more than 17,000 players. By using this dataset, we can predict the value of the player accurately and efficiently.Now, we only consider the attributes that will help in estimating the net value of football players such as height, weight,age, passing etc. We are using four models such as Linear Regression, Multiple Regression, Decision Tree and Random forest.After processing the data with various models, we conclude that Random forest is the best model. It requires less inputs and gives best result. The results are accurate and efficient.

II. EXISTING SYSTEM

Previously, judgments are made by agents and experts based on their experience and knowledge. This will result in leading many errors, takes a lot of time to calculate and more expensive which affects the value of football players. Football experts will calculate the value based on the player's characteristics, player's performance and player's popularity. These expert decisions are sometimes incorrect and not efficient. They also consume lot of time.Hence, lacking of accuracy may results in lack of value of football players.

III. PROPOSED SYSTEM

Now, we are going introducing machine learning in our approach. Here, we are using four machine learning algorithms like Linear Regression, Multiple Regression,Decision Tree and Random Forest. Here we will consider three important factors such as player characteristics, player performance and player popularity.

1. Player Characteristics:

This is one of the most important characteristic which is to consider. These include Age, player height, weight and player position. Age is an attribute which reflects in experience and ability.

Height, which helps in increase the score and preventing the goals. Weight of the player will help in estimating the value of the football player. Player position like defender, midfielder, goal keeper etc are useful in predicting the value of football player.

2. Player Performance:

Player performance which includes passing, shooting, dribbling and yellow and red cards.Passing, it represents the passing of a ball intentionally from one player to another player in the same team. Shooting represents the hitting of a ball in an attempt to score the goal. Dribbling I t represents the passing of a ball in a given direction and avoiding the defender's attempts to intercept the ball. Yellow card and Red card represents the number of warnings and mistakes they have committed.

3.Player Popularity :

The popularity will also helps in determining the value of football player.It means the crowd pulling power and the image of the player they show on the pitch. Hence, we can say international reputation will play a major role in determining the market value of a player.By using these algorithms we can reduce the complexity in predicting the value the football players, reduce errors, enhance accurate results. It is easier to predict the value and it will also help the club agents to make quick decisions. The main advantages are Generate accurate and efficient results, Computation time is greatly reduced, Reduces manual work, Efficient and transparent results.

IV. ALGORITHMS

Here, we are using 4 algorithms to predict the value of a football players. The four algorithms are Linear Regression, Multiple Regression, Decision Tree and Random Forest.

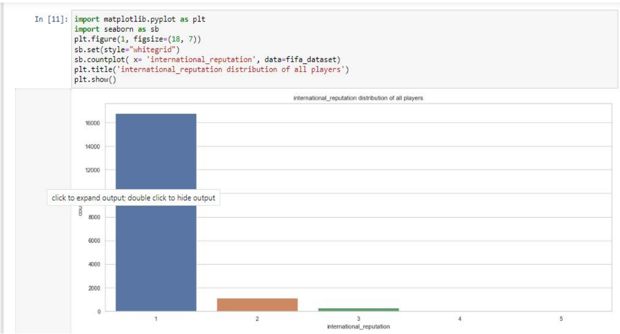
- Linear Regression is the simplest form of algorithm, it is used to predict the relationship between two variables. There are 2 types of variables such as Independent and dependent variable. Suppose Y=a+X, here X is Independent variable. Since for a given value of X, Y is changing accordingly. Hence Y is a dependent variable. These two variables are used to predict the target variable.

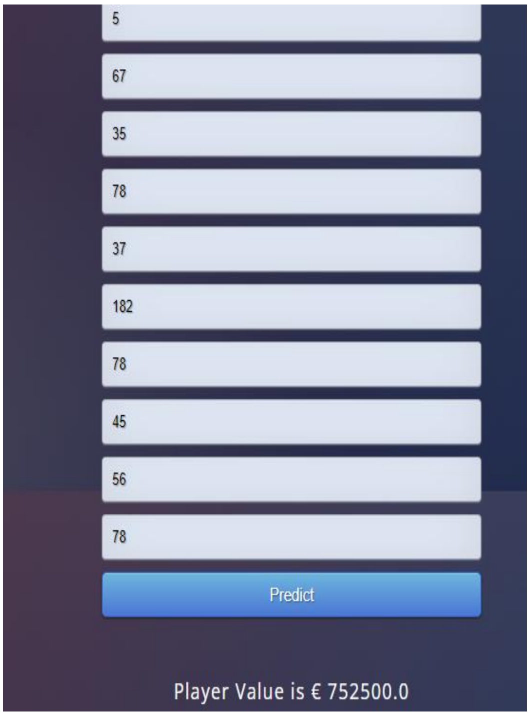
- Multiple Regression is the extension of Linear Regression. It is used to find the relationship between two or more number of independent variables and one dependent variable. For example, we are making survey for finding the reasons for lung cancer. The factors include smoking, drinking etc. Here, we are having two or more variables as the reason for lung cancer. In this scenario, we will use multiple regression.

- Decision Tree is a supervised learning technique. It is having a flow-chart structure. It is having internal nodes, branches and leaf nodes.Internal nodes represents the attributes of the dataset, branches represent the decision rules and leaf nodes tells about the outcome. The decisions rules are taken from the dataset. Here the attributes are compared among the decisions, if it matches it will show you the outcome else it will skip the respective condition and jump to the next node.

- Random Forest is the group of decision trees.It will split the data into subparts and solve the complex problem.It will predict the accuracy of the dataset. The more number of decision trees in the forest leads to higher accuracy,good transparency and it avoids the problem of overfitting.The Random Forest is having the highest coefficient of determination where as linear regression is having the lowest coefficient of determination. Hence, we conclude Random Forest is the best model.

V. CODE IMPLEMENTATION





VI. CONCLUSION

N	Classifier	MAE	RMSE	R^2
1	Linear Regression (Baseline)	5,468,144	5,468,144	0.47
2	Multiple Linear Regression	2,618,108	4,662,630	0.61
3	Regression Tree	835,935	2,713,452	0.87
4	Random Forest Regression	576,874	1,649,921	0.95

Here, we have used 4 algorithms. We make use of 3 metrics such as, Mean Absolute Error(MAE) , Root Means Square Error (RMSE), Coefficient of Determination(R2). The above table shows the errors between the actual and predicted values . From the above table the random forest shows the least root mean square between the actual and predicted values, where as Linear Regression provided the highest root mean square values.Here we also calculated the coefficient of determination. The value which is close to 1 indicates with zero error. If the value is close to 0, It means it shows the error.From the above table we can say that the random forest algorithm provides the highest coefficient of determination and linear regression provides the least coefficient of determination. Hence we conclude the Random forest is the best suitable mode for modelling.Ultimately, we can say that expert judgments are not accurate and it also consumes lots of time. The results are not transparent and inefficient. So, By using these machine learning algorithms we will predict the value of football players accurately and efficiently.

REFERENCES

[1] R. Asif, ''Football (soccer) analytics: A case study on the availability and limitations of data for football analytics research,'' Int. J. Comput. Sci. Inf. Secur., vol. 14, no. 11, p. 516, 2016.

[2] R. Poli, L. Ravenel, and R. Besson, ''CIES football observatory,'' Annu. Rev., pp. 1–83, 2014. [Online]. Available: https://www.footballobservatory.com/IMG/pdf/ar2013_exc-2.pdf.

[3] S. Kiefer, ''The impact of the Euro 2012 on popularity and market value of football players,'' Inst. Org. Econ., Berlin, Germany, 2012.[Online].Available:https://www.wiwi.unimuenster.de/io/forschen/downloads/DPIO_07_2020.

[4] S. Leone. (2020). FIFA 20 Complete PlayerDataset. [Online]. Available: https://www.kaggle.com/stefanoleone992/fififa-20-complete-playerdataset.

[5] J. Dufour, ''Coeffificients of determination,'' Dept. Econ., McGill Univ., Montreal, QC, Canada, 2011, pp. 1–14. [Online]. Available:https://monde.cirano.qc.ca/~dufourj/Web_Site/ResE/Dufour_1983_R2_W.pdf

[6] J. L. Felipe, A. Fernandez-Luna, P. Burillo, L. E. de la Riva, J. Sanchez-Sanchez, and J. Garcia-Unanue, ''Money talks: Team variables and player positions that most inflfluence the market value of professional male footballers in Europe,'' Sustainability, vol. 12, no. 9, p. 3709, May 2020.

[7] L. Cotta, ''Using fififa soccer video game data for soccer analytics,'' in Proc. Workshop Large Scale Sports Anal., 2016, pp. 1–4.

[8] P. Siuda, ''Sports gamers practices as a form of subversiveness–the example of the FIFA ultimate team,'' Crit. Stud. Media Commun., vol. 38, no. 1, pp. 75–89, 2021.

[9] P. Awasthi, A. Beutel, M. Kleindessner, J. Morgenstern, and X. Wang, ''Evaluating fairness of machine learning models under uncertain and incomplete information,'' in Proc. ACM Conf. Fairness, Accountability, Transparency, Mar. 2021, pp. 206–214.

[10] J. Gareth, An Introduction to Statistical Learning: With Applications in R. Springer, 2013.

.