

PREDICTIVE ANALYSIS OF FOOTBALL PLAYER MARKET VALUE USING MACHINE LEARNING

Gokul Krishna*¹, Dr. Anjana S Chandran*²

*¹Department of Computer Applications, SCMS School of Technology and Management, Muttom, Aluva ,683106

*²Associate Professor, Department of Computer Applications, SCMS School of Technology and Management, Muttom, Aluva, 683106

ABSTRACT

Football (also referred as 'soccer') is one of the most played sports around the world. Many countries have football players contributing to both regional and national championships. Football clubs all over the world are ready to spend a huge amount of money on buying highly skilful and match winning players in order to reach their target. So, football players market value is a key parameter which is used for determining the price and quality of a football player. Football clubs can make proper budgeting on transfers with the help of these market values. So, football club can make best use of their funds on buying apt players. Market values are determined by analysing factors such as age, height, number of minutes played, number of passes provided, number of goals scored etc. So, in this paper, a dataset which contain all information regarding the players are used as input and it gets processed to generate market value of the football player as output with the help of machine learning techniques.

Keywords: Machine Learning, Soccer, Market Value, Transfer, Predictive Analysis.

I. INTRODUCTION

The Predictive analytics is becoming ever more prevalent within the business world as companies explore for ways to realize a bonus over their competitors. In football, clubs can make advantage of predictive analysis as they will properly budget their funds for transfers. Predicting football players market price can determine who are overvalued or undervalued. Agents may manipulate market price for getting extra money from clubs. So, with the assistance of Predicted market values, clubs can prevent fraud from player agents.

At present, Football clubs are able to spend huge amount of cash on buying players. Transfer amount has been increased subsequently over the years. So, transfers play an important role in club development. Predictive analytics allows football clubs to create decisions supported performance of the players. So, clubs can scout or shortlist the specified players which can fit under their budget. Generally, the value is decided by ratings from football pundits, ex-players, fans, journalists etc. The accuracy is very questionable during this case. So, with the assistance of machine learning algorithms, we will determine the accurate rating of the athlete. this may help the clubs to seek out the apt player for his or her clubs at right cost.

Clubs can budget wisely with the assistance of this market price. Club can analyze the abilities and rating of the player and might choose the most effective fit one. If the club is forced to pay huge transfer fee to a player who doesn't deserve that value, it can affect the condition of the club very badly. So, with proper rating and accurate value, clubs are going to be ready to find the apt player which is able to fit under their budget. The Jupyter notebook is an environment that enables computations within the browser of the user's system. Regression models are accustomed predict a continual value. This is often one in all the foremost common variety of Machine Learning Algorithms. Python packages like NumPy, Pandas, Beautiful Soup, matplotlib and sklearn are used for prediction of market values.

II. METHODOLOGY

Here the project is to predict the market price using machine learning and that we use used regression for the prediction. Regression methods predicts a continual value. This can be one among the foremost common kind of Machine Learning Algorithms. Here we predict a target variable Y (Value) supported the input variable X. A linear relationship should exist between target variable and predictor so comes the name Linear Regression.

Our project goes through 7 stages: -

1) Define the problem

Aim of this paper is to predict the market price of the footballer and to seek out who is undervalued and overvalued.

2) Web Scraping

Web scraping could be a technique that collecting data from the web and parsing it into meaningful form. If there's no direct thanks to download data, you would like to extract the information into a meaningful form. during this paper BeautifulSoup.

3) Data Cleansing

Data cleansing may be a critically important step in any machine learning project. Cleaning data can give more accurate results. It removes unwanted characters, splits columns, deleting blank rows etc.

4) Exploratory Data Analysis

Exploratory Data Analysis helps you to grasp data, exploring structure of information and recognizing the link between the variables. In short, exploratory data analysis tells almost everything about data.

5) Building a Model

In this paper regression model is employed. Regression methods are used to predict a nonstop value. This can be one in all the foremost common kind of Machine Learning Algorithms.

6) Evaluating Model's Result

Python's Matplotlib package can used for plotting the lead to graphical format. Attributes like overall, Age, International Reputation, Growth, Release Clause, Height and Weight are used for analyzing data.

7) Conclusion

Market value of footballer is predicted and comparison on actual market values and predicted market values is performed.

III. MODELING AND ANALYSIS

Prediction is sometimes done by using regression techniques. If you'd prefer to predict something like price, sales etc. regression are going to be a decent solution for you. The characteristics of the target or variable quantity is dichotomous, which suggests there would be only two possible classes. In simple words, the variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts $P(Y=1)$ as a function of X. It's one amongst the most effective Machine Learning algorithms which will be used for various classification problems like spam detection, Diabetes prediction, cancer detection etc. In Binary or Binomial sort of classification, a quantity will have only two possible types either 1 or 0. for example, these variables may represent success or failure, yes or no, win or loss etc. In Multinomial classification, dependent variables can have 3 or more possible unordered types or the kinds having no quantitative significance.

The aim was to predict the market price of footballers using machine learning techniques. The information is scrapped from sofifa website. Sofifa contains data of footballers playing around different leagues with detailed specifications like goals scored, games played, assist provided, number of passes given etc. Web scrapping is completed with the assistance of BeautifulSoup Package. Then Data cleansing helps to get rid of issues like duplicate entries, unwanted characters, split lines etc. Exploratory data analysis is conducted to grasp and explore the structure of information. With the assistance of regression model, relationship between the target and predicted variable is discovered. In line with the experimental results, change was clearly visible between actual and predicted market values. Results are visualized in graphical form.

```

===== OLS Regression Results =====
Dep. Variable:          Value      R-squared (uncentered):      0.948
Model:                  OLS      Adj. R-squared (uncentered):  0.948
Method:                 Least Squares      F-statistic:      6.528e+04
Date:                  Mon, 07 Jun 2021      Prob (F-statistic):  0.00
Time:                  10:23:47      Log-Likelihood:    -39545.
No. Observations:      17779      AIC:               6.110e+04
Df Residuals:          17774      BIC:               6.114e+04
Df Model:              5
Covariance Type:      nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Overall	0.0541	0.002	35.303	0.000	0.051	0.057
Age	-0.1762	0.004	-43.908	0.000	-0.184	-0.168
Int_Reputation	1.6952	0.039	43.487	0.000	1.619	1.772
Growth	-0.1016	0.002	-47.658	0.000	-0.106	-0.097
Release_Clause	0.4540	0.001	338.421	0.000	0.451	0.457

Figure 1: Regression Result

SETTING UP THE SYSTEM

Jupyter Notebook is our environment used for computation. NumPy is a Python library that provides a multidimensional array object, various derived objects and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, input/output, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more. NumPy can be installed by using ‘pip install numpy’ command. Python should be installed in the system to get pip command accessible. Pandas is a python library. Main use of pandas is to perform data analysis and data manipulation. Pandas can be installed by using ‘pip install pandas’ command. Matplotlib is used in our project for data visualization. Matplotlib can be installed by using ‘pip install matplotlib’

IV. RESULTS AND DISCUSSION

Market value for footballers is predicted with the help of regression models. After analyzing the result, differences in actual and predicted values can be clearly seen. Biggest negative difference is € 8.48 million for Sergio Ramos. His actual market value is € 32.5 million. But predicted value by the model is € 39.98 million whereas biggest positive difference is € 39.04 million for M. Icardi. Actual market value of Icardi is € 53 million but model predicted value is € 13.96 million.

Table 1. Comparison of actual vs predicted market values

Name	Actual Market Value (In million Euros €)	Predicted Market Value (In million Euros €)	Difference (In million Euros €)
Neymar JR	105.5	88.4	17.09
M.Politano	26.5	9.22	17.28
R. Nainggolan	27.5	8.74	18.76
Sergio Ramos	31.5	39.98	-8.48
Luka Modric	39.0	46.35	-7.35
Vinicius JR	22.5	29.09	-6.59
Coutinho	40.5	13.65	26.85
M.Icardi	53.0	13.96	39.04

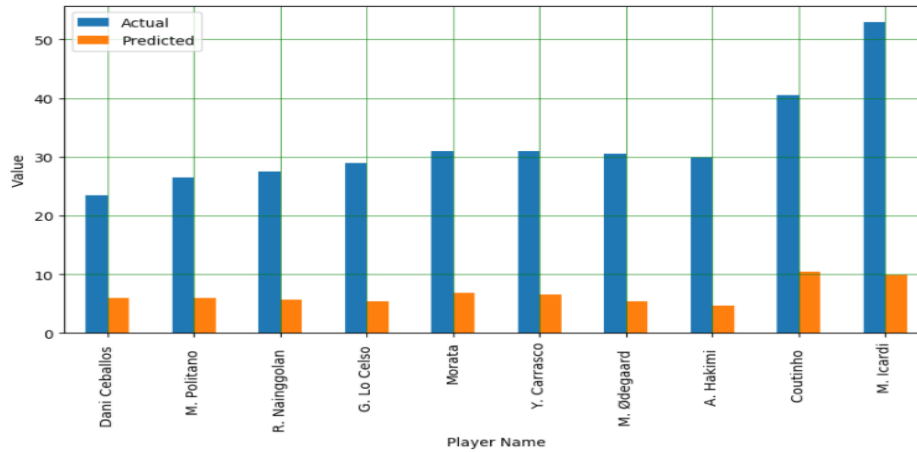


Figure 2: showing values for underpredicted players

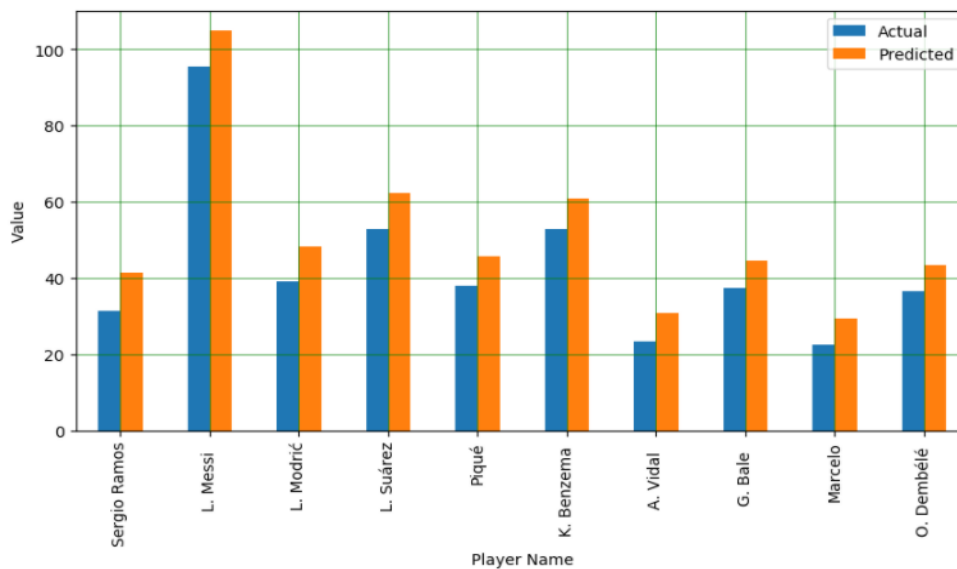


Figure 3: showing values for overpredicted players

V. CONCLUSION

From these we can conclude that prediction of market value with the help of machine learning can prevent manipulations over market values as huge differences were seen over actual and predicted results. The clubs can utilize this technology to crosscheck the players value and to pay what they really deserve. This technology can be implemented widely for preventing such malpractices over market values in the upcoming future.

VI. REFERENCES

- [1] Sidarrth Mahadevan. PREDICTING MARKET VALUE OF FOOTBALL PLAYERS USING MACHINE LEARNING ALGORITHMS. *IEEE*.
- [2] Salil Chincholikar, Niharika Sadul, Parth Thakkar, Ashutosh Amrutkar and Nikhil Dhavase(Dec 2020). PREDICTIVE ANALYTICS OF FOOTBALL.IJRET
- [3] Jongho Shin and Robert Gasparyan(2015).A NOVEL WAY TO SOCCER MATCH PREDICTION.IEEE
- [4] Dwijen Rudrapal, Sasank Boro, Jatin Srivastava and Shyamu Singh(Jan 2021). A DEEP LEARNING APPROACH TO PREDICT FOOTBALL MATCH RESULT.
- [5] Sourya Dey(April 2017).PRICING FOOTBALL PLAYERS USING NEURAL NETWORKS.University Of Southern California.
- [6] Miao He, Ricardo Cachucho, and Arno Knobbe. FOOTBALL PLAYER'S PERFORMANCE AND MARKET VALUE.

- [7] P. Singh, P. S. J. J. o. D. M. S. Lamba, and Cryptography. INFLUENCE OF CROWDSOURCING, POPULARITY AND PREVIOUS YEAR STATISTICS IN MARKET VALUE ESTIMATION OF FOOTBALL PLAYERS. vol. 22, no. 2, pp. 113-126, 2019.
- [8] I. Behravan And S. M. J. S. C. Razavi. A NOVEL MACHINE LEARNING METHOD FOR ESTIMATING FOOTBALL PLAYERS' VALUE IN THE TRANSFER MARKET. PP. 1-13, 2020.
- [9] K. Sujatha, T. Godhavari and Nallamilli P G Bhavani, FOOTBALL MATCH STATISTICS PREDICTION USING ARTIFICIAL NEURAL NETWORKS, Volume 3, 2018 International Journal of Mathematical and Computational Methods
- [10] Drs Raffaele Poli, Loïc Ravenel and Roger Besson, SCIENTIFIC ASSESSMENT OF FOOTBALL PLAYERS' TRANSFER VALUE, CIES Football Observatory October 2018
- [11] Milad Keshtkar Langaroudi, Mohammad Reza Yamaghani, SPORTS RESULT PREDICTION BASED ON MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE APPROACHES: A SURVEY, J. ADV COMP ENG TECHNOL, 5(1) Winter 2019
- [12] Anthony C. Constantinou, Norman E. Fenton And Martin Neil, PI-FOOTBALL: A BAYESIAN NETWORK MODEL FOR FORECASTING ASSOCIATION FOOTBALL MATCH OUTCOMES, February 2012
- [13] Joaquin Lago-Ballesteros, Carlos Lago-Peñas, PERFORMANCE IN TEAM SPORTS: IDENTIFYING THE KEYS TO SUCCESS IN SOCCER, Journal Of Human Kinetics volume 25 2010, 85-9
- [14] Qing Yi, Miguel-Ángel Gómez-Ruano, Hongyou Liu⁵, Shaoliang Zhang, Binghong Gao², Fabian Wunderlich and Daniel Memmert, EVALUATION OF THE TECHNICAL PERFORMANCE OF FOOTBALL PLAYERS IN THE UEFA CHAMPIONS LEAGUE, International Journal of Environmental Research and Public Health January 2020.
- [15] Bartosz C'wiklinski, Agata Giełczyk and Michał Choras, WHO WILL SCORE? A MACHINE LEARNING APPROACH TO SUPPORTING FOOTBALL TEAM BUILDING AND TRANSFERS, entropy January 2021.
- [16] Rory P. Bunker and Fadi Thabtah, A MACHINE LEARNING FRAMEWORK FOR SPORT RESULT PREDICTION, Applied Computing and Informatics 15 (2019) 27–33
- [17] Johannes Stübinger, Benedikt Mangold and Julian Knoll, MACHINE LEARNING IN FOOTBALL BETTING: PREDICTION OF MATCH RESULTS BASED ON PLAYER CHARACTERISTICS, MDPI Applied Sciences 2019
- [18] Oliver Müller, Alexander Simons, Markus Weinmann, BEYOND CROWD JUDGMENTS: DATA-DRIVEN ESTIMATION OF MARKET VALUE IN ASSOCIATION FOOTBALL, European Journal of Operational Research 263 (2017) 611–624
- [19] Pedro Moreira Silva, Júlio Garganta, José António Maia and Pedro Manuel Santos, TRACKING PERFORMANCE IN FOOTBALL – AN EXAMPLE USING GOAL SCORING DATA, The Open Sports Science Journal, 2012, 5, 181-187
- [20] Iman Behravan, Seyed Mohammad Razavi, A NOVEL MACHINE LEARNING METHOD FOR ESTIMATING FOOTBALL PLAYERS' VALUE IN THE TRANSFER MARKET, March 2021
- [21] Victor Chazan Pantzalis and Christos Tjortjis, SPORTS ANALYTICS FOR FOOTBALL LEAGUE TABLE AND PLAYER PERFORMANCE PREDICTION, Jan 2020.
- [22] Niek Tax and Yme Joustra, PREDICTING THE DUTCH FOOTBALL COMPETITION USING PUBLIC DATA: A MACHINE LEARNING APPROACH, September 2015.
- [23] J. Goddard, WHO WINS THE FOOTBALL?, Significance, vol. 3, no. 1, pp. 16–19, 2006.
- [24] Kerys Harrop and Alan Nevill, PERFORMANCE INDICATORS THAT PREDICT SUCCESS IN AN ENGLISH PROFESSIONAL LEAGUE ONE SOCCER TEAM", International Journal of Performance Analysis in Sport April 2017
- [25] Kringstad, M. and Olsen, T.-E, CAN SPORTING SUCCESS IN NORWEGIAN FOOTBALL BE PREDICTED FROM BUDGETED REVENUES?, European Sport Management Quarterly Volume 16, 2016 -Issue-1.