Predicting Market Value of Football Players using Machine Learning Algorithms

Sidharrth Mahadevan   
 *Faculty of Science and Technology*  
*Bournemouth University*Bournemouth, United Kingdom   
sidharrthnix@gmail.com

*Abstract*—Football clubs spend a huge amount of money every year to buy professional football players, during the transfer window. Predicting how much players value in the transfer market is one of the difficult tasks for managers of the club.Trasnfermarkt.com determines the players’ value using football experts. As crowdsourcing is needed is updated infrequently because of the participation of many users and humans are prone to errors and data scientists attracted to this topic since then, around the world to create datasets and estimating methods using data-driven. The data-driven method seems to be an alternative approach for crowd driven websites for estimating players’ market value. I have constructed a dataset using 4250 players from 150 teams playing in European competitions and a period of five seasons, I have estimated players’ market value using regression analysis.

Keywords—Football, Transfermarkt, data-driven, regression

# Introduction

Football referred to as soccer is one of the most played sports around the world. Many countries have football players contributing to both regional and national championships. The English premier league is the highly watched sports league in the world, broadcasting to 643 million homes in 212 territories. A BBC article has estimated the football betting market to be around $70 0billion to $1 trillion [1]. More recently, researchers have started to consider the players' market value. Predictive system modeling for football matches is hugely significant in terms of economic values but not merely an interest in academia. The players' market value has taken much attention from the researchers, More recently. Players contracts can be sold from one team to another, in which the market value of a player can be estimated [2]. although the transfer fees are represented by the actual price paid, the transfer negotiations is an important role in the market value provided on transfer fees. Despite the fact that the values in the transfer market has been extended assessed by football specialists, for example, managers of team and sportswriters, sites like Transfermarkts have given convenience in assessed the market an incentive in the previous few years Nevertheless, the estimation value of market by information-driven methodologies have not yet gotten up to speed in expert football. Because of the impediment in the site information researcher have made datasets and information-driven assessing strategies, for individual sports like tennis or snooker, which is more straight forward to assess the players value, independent of the team, however, to assess a team of players is not as straightforward and need more complex data-driven approach.

Due to many uncertain factors which include short career span, risk of injury, club budgets, and variable form throughout the career, both the club and the player seek to continuously maximize their performance. Besides the attractiveness of the game, there are more attractive issues for the fans. One of them is transferring a player during transfer windows. The fan will be interested in knowing which player the club will sign and how much money the club will spend in the transfer window period. [3]

# Market values in professional football

## Estimation of market value By Crowd-Based method

Transfermarket is the most leading website in the transfer market of football. This site offers information in football such as the latest updates, live information about matches and results on football, and transfer rumors. The website was launched in the year 2001 and the English version of the website was released in 2009. Ever since the website has been made live in all European countries.

Transfermarkt was built on the idea that the user can build the estimated market value together or better than football experts. “wisdom of crowds.” Is a term invented by Surowiecki. Transfermarkt has provided several studies that build the foundation for the transfer market in football [4].

Many social challenges like social impact, control endeavors, absence of information, and experience are faced by crowdsourcing which is estimated accurately by Transfermarkt[5].

The market values are not calculated by Transfermarkt by using the mean or median of individual gauges however using the community members referred to as "Judges". The user estimates are reviewed by the judges accordingly in which they weigh them and make their choices. In this way, they can increment or decline the impacts of the client whom they consider to be recognized.

Despite the demonstrated accuracy and the arguable benefits of the transfer market the crowd-driven method to estimate the transfer fees comes with its limitations. The estimation of arbitrary indicators was based on the first community members which may happen unconsciously, so they lack objectivity. Secondly, the final market values can be independently determined by the judges based on personal evaluations of user estimates, so these are not reproducible. Thirdly many user participations are required for crowd estimation, the transfer fees are not updated on a day-to-day basis in which the accuracy after games will no longer be accurate. The accuracy of the players who tend to be more popular in a sufficiently large audience tends to be more accurate than the player, playing in minor leagues, which tends to be a huge back draw in the crowd estimation.

## Data-driven estimation of market value

The primary club to utilize data analytics to enlist players was Major League Baseball (MLB). the senior supervisor of Oakland Athletics Billy Beane began utilizing measurable information before the finish of the 1990s for settling on choices about group program, a story well known through the bestseller, “Moneyball” [6]. For the vast majority of the little or medium-sized clubs, purchasing costly players is not a reasonable technique, the information examination was utilized deliberately by a couple of clubs. For instance, the factual models to assess groups and players started by the Danish Superliga club FC Midtjylland [7]. The proprietor of German Bundesliga goliath TSG Hoffenheim and fellow benefactor of SAP, Dietmar Hopp has distributed the utilization of factual investigation at Hoffenheim. Roberto Firmino who cost Hoffenheim just € 4million was sold for the high exchange charge of around 41 million, gone through in the year 2015 in the record-breaking most elevated exchange of FC Liverpool. Hopp recognized two achievement factors for progress: distinguishing youthful gifts and create them, so they can contribute both on the monetary record and pitch and being an early adopter of inventive innovations.

# Indicators of market value

## Player Characteristics

Player characteristics are conceptualized as players’ demographic and physical attributes. Age is a significant indicator of market value, as it affects experience and viability [8]. Using terms like age to allow for nonlinear relationships, as players’ value usually increase in their youth and declines gradually [9]. The taller the player, the better heading ability to score or preventing a goal so players’ height is found to significantly increase salary. Player footedness has also been studied to affect market value.

## Goals Scored

Baio et al [10] proposed the Gamma distribution mixture model on how the amount of goal scored can influence players’ market value. The most “In-Form” players are said to have higher market value Miljkovic et al [11]

## Fatigue

Constantinou et al [12] has used fatigue as one of the features in his prediction model. Fatigue influences players’ fitness levels. The number of days since the previous match, the toughness of the previous match, and the amount of rest the individual gets cause fatigue.

## Football skills

Min et al [13] proposed that highly skilled players are having high market prices. He grades the skill level for defense, offense, and possession-based on expert opinion. Games such as FIFA give an accurate rating of skill level with 5 being the highest and 1 is the lowest.

## Strategy

Palomino et al [14] showed the influence of different playing strategies using a game-theoretic approach (e.g., defending style, attacking style) on match outcome.

## Position

Several Researchers have found players’ position – forward, midfielder, defender, or goalkeeper is important in estimating the market value and salaries as they reflect players’ crowd-pulling capacity and degree of specialization. Goalkeeper earns the least because of his less involvement in the field [15]. Attackers receive higher attention than other positions, as they are visible to the wider people and higher audience bringing power[16].

## Player Performance

Player performance has been constantly evaluated using playing time as it shows how well the player performs on the pitch. For example, appearances in European competitions, domestic tournaments, and on the intercountry league has an impact positively on market values and fees used for transfer [8]. Another performance measure used to predict players’ values are goals scored, penalties, headers, dribbling indicates players’ scoring ability, so there is an unambiguous performance measure [17]. Only a handful of data-scientists have trained machine learning models other than assist and goals to predict value because of the everlasting unattainability of detailed performance data in professional football. Dueling like clearance, blocks, and interceptions [4], passing [2], committed fouls [18], and Bookings [19] are infrequently used. Researchers have added commerce effects in their model of players’ market value Because of significant varying performance indicators by position [20] For example. Defenders start to improve their defending skills; midfielders are expected to attack and defend well, and forwards are expected to score goals.

## Player Popularity

Players' fairly estimated worth independent of what they show on the pitch and relies upon their fans bringing power, as this force can sell their club's seats and pullover. The occasions a players' name is referenced in the German games magazine Kicker gauges the compensation. Thus, media presence hugely affects the player's worth. [21].

# Data Collection

I have assembled information on players' prominence, execution, and qualities from various web sources, including Reddit, google, transfermarkt,wikipedia, Onefootball, and youtube. I have collected data for five years, from the 2010/11 season to 2015/16 season for the player from top leagues in Europe. The dataset consisted of 9,450 observations from 4250 players on 150 teams. The accuracy of Transfermarkt’s prediction has been verified by several market enthusiasts and because of the unattainability of other reliable sources, I have used Transfermarkt’s prediction on training and validation dataset. The average value of player across all competition is around €8 million and players’ value classed from €30,000 to €200 million with €10 million deviations.

I have segregated the player characteristics using players’ Height(cm), Age(years), Nationality, Footedness (two-footed or single footed), and Position on the pitch. The average height was 183cm and 28 years old. Nine percent of all

**Variables used to build the OLS Regression Model**: These are the finalized variables that tend to have a positive correlation with players’ market value

players were double footed, 78 percent of them were from Europe and 53 percent of them were midfielders. I have

| SL.NO | Indicators of Market Value | | | |
| --- | --- | --- | --- | --- |
| Variable | Mean | Median | St. Dev |
| 1. | TransferMarkt’s value | 6000000 | 3000000 | 8000000 |
| 2. | Age | 28.3 | 28 | 4 |
| 3. | Height | 183.51 | 183 | 6.30 |
| 4. | No. of minutes played | 1700 | 1782 | 890 |
| 5. | Passes | 30 | 28.85 | 12 |
| 6. | Assists | 1.85 | 2 | 1.50 |
| 7.. | Goals | 2.4 | 1.00 | 3.85 |
| 8. | Skills | 4 | 4 | 1 |
| 9. | Fatigue | 30 | 25 | 5 |
| 10. | Tackles | 3.1 | 2.9 | 0.3 |
| 11. | Wiki views | 80,000 | 73,000 | 320,000 |
| 12. | Reddit score | 16 | 12 | 40 |
| 13. | Youtube videos | 40,000 | 1000 | 150,000 |
| 14. | Google T.S.I | 14 | 12 | 12 |

measured players' performance by Goals, minutes played, and Bookings per season; the number and accurate proportion of Dribbles, Passes, Tackles, and Aerial duels per game. I have estimated player popularity utilizing four web measurements: the occasions a player's name has been looked on Google, how regularly the player's Wikipedia page was seen, how successive player's name was looked on Reddit, and the number of perspectives on youtube. The normal player has in excess of 40,000 recordings transferred on youtube and has around 80,000 Wikipedia sees. Google patterns search record was 12 and his name has showed up in 15 posts on Reddit.

# Modeling:

| SL. NO. | Model Evaluation | | |
| --- | --- | --- | --- |
|  | RMSE | MAE |
| 1. | Trasnfermarkt’s prediction | 5,800 | 3100,820 |
| 2. | Model’s prediction | 5,992 | 329,743 |
| 3. | Relative diffrenece | +4.5% | +4.1% |

To build a Machine Learning Model to Predict the Market value of Football players, I have used regression models, the dependent variable (Y) is the market value to predict and independent variables (X) are players’ performance measure, characteristics, and popularity. Common evaluation metrics for predicting models like k-fold validation due to spanning our samples for several playing seasons [26]. They would have leaked the future data to train the model in the past. Therefore time-series based evaluation is used to predict the

Note: The positive value depicts the crowd has an advantage over model prediction. This is due to skewness in costlier players.

market value as training dataset is aware of data only up to that point in time. I have used Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) for performance measures. They are calculated by the difference between the model prediction and the current market value. The model tends to be less accurate for high-prices players, while comparatively more accurate for low-medium priced-players. The result from the model did not swerve considerably from the current market value. The mean deviation was around €5 million which is marginal compared to the huge market value of players. The model performed more accurately for 85% of the transfer. Although the model estimate was disproportionally inaccurate for a smaller share of players, the final result tilted the mean, so the crowd was more precise than the machine predicted model. This is due to the inability to predict the value of superstars because of the shortage of important intangible indicators. In other words, the crowd has total freedom to set prices for popular players while the model is limited to the predictors used.

# Related Work

* Behravan and Razavi have designed a machine learning model using FIFA 20 dataset. FIFA is a series of association football simulation video games developed and released annually by Electronic Arts under the EA Sports label. The main novelties of his research are:
* 1.Dividing the data points into automatic clustering, by segregating players with the same roles in the same cluster.
* 2.Feature selection with the help of PSO (Particle Swarm Optimization) and using grid-search for hyperparameter tuning of SVR -model. Feature selection has helped to identify players’ market value and hyperparameter tuning has helped to increase the accuracy of SVR model [22].
* In another research Herm et al. Estimating the transfer fee of football players based on players’ talent and judgment variables. Talents used are age, success, assertion, flexibility, and precision. His model shows age is inversely proportional to players’ market value. In this research, he used fans to estimate the value. So, the main drawback is the biased result and lacking sufficient knowledge.
* Frank and Nuesh [4] predicted players’ market value by investigating the effect of talent and popularity. They have used nearly 20 parameters in correlation with their teams’ success
* Singh and Lamba [23] used FIFA 18 dataset to predict market values. They have taken a different approach by predicting the player's value using in-game player prices. They were able to produce an accurate result as gamers can predict the values of in-game players easily.
* Yigit et al. predicted the market value using a dataset from a football manager simulation game. The data contains 49 attributes, each attribute is a number between 0 and 20 determined by professional football scouts [24].
* Felipe et al. has proposed a model that predicts the market value by analyzing the effect of team variable and player position [25].
* Rade & Laszlo has used InStat and transfermarkt datasets. They have applied PCA (principle component analysis) on their dataset and have calculated the difference between transfermarkt.com market value estimate (TMVE) and performance-driven market value estimate (PDMVE) [3].

# Conclusion and Future Work:

The use of the data-driven method is reproducible and more transparent than crowd judgment, as the coefficient has a direct influence on players’ market value. The model I have proposed can be utilized by fantasy-league websites as players’ value is determined by performance data. Although my model has incorporated 85% of the market prediction, commercial giants have the resources to capture more than two hundred metrics per player per game which I do not have access to. Comparing my model with crowd estimates shows that the data-driven approach has overcome several limitations and able to produce more accurate results

# Acknowledgments

I want to thank Philip Davies and Festus Adedoyin for guiding me towards my field of studies and in helping me find my potential. I also want to thank Philip Davies and Festus Adedoyin for the useful discussions, we had every week. I want to thank my professors for their endless support.

# Reference

[1] R. Baboota and H. J. I. J. o. F. Kaur, "Predictive analysis and modelling football results using machine learning approach for English Premier League," vol. 35, no. 2, pp. 741-755, 2019.

[2] S. Herm, H.-M. Callsen-Bracker, and H. J. S. M. R. Kreis, "When the crowd evaluates soccer players’ market values: Accuracy and evaluation attributes of an online community," vol. 17, no. 4, pp. 484-492, 2014.

[3] R. Stanojevic and L. Gyarmati, "Towards data-driven football player assessment," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 2016, pp. 167-172: IEEE.

[4] E. Franck and S. J. E. I. Nüesch, "Talent and/or popularity: what does it take to be a superstar?," vol. 50, no. 1, pp. 202-216, 2012.

[5] J. Lorenz, H. Rauhut, F. Schweitzer, and D. J. P. o. t. n. a. o. s. Helbing, "How social influence can undermine the wisdom of crowd effect," vol. 108, no. 22, pp. 9020-9025, 2011.

[6] M. Lewis, *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.

[7] J. J. M. R.-t. M. Murtagh, "Moneyball FC: how Midtjylland harnessed the power of stats to set up Euro showdown with Southampton," vol. 14, p. 2017, 2015.

[8] F. Carmichael and D. J. A. E. Thomas, "Bargaining in the transfer market: theory and evidence," vol. 25, no. 12, pp. 1467-1476, 1993.

[9] A. Bryson, B. Frick, and R. J. J. o. S. E. Simmons, "The returns to scarce talent: Footedness and player remuneration in European soccer," vol. 14, no. 6, pp. 606-628, 2013.

[10] G. Baio and M. J. J. o. A. S. Blangiardo, "Bayesian hierarchical model for the prediction of football results," vol. 37, no. 2, pp. 253-264, 2010.

[11] D. Miljković, L. Gajić, A. Kovačević, and Z. Konjović, "The use of data mining for basketball matches outcomes prediction," in *IEEE 8th International Symposium on Intelligent Systems and Informatics*, 2010, pp. 309-312: IEEE.

[12] A. C. Constantinou, N. E. Fenton, and M. J. K.-B. S. Neil, "pi-football: A Bayesian network model for forecasting Association Football match outcomes," vol. 36, pp. 322-339, 2012.

[13] B. Min, J. Kim, C. Choe, H. Eom, and R. B. J. K.-B. S. McKay, "A compound framework for sports results prediction: A football case study," vol. 21, no. 7, pp. 551-562, 2008.

[14] F. Palomino, L. Rigotti, and A. Rustichini, *Skill, strategy and passion: An empirical analysis of soccer*. Tilburg University, 1998.

[15] B. J. S. J. o. P. E. Frick, "THE FOOTBALL PLAYERS'LABOR MARKET: EMPIRICAL EVIDENCE FROM THE MAJOR EUROPEAN LEAGUES," vol. 54, no. 3, pp. 422-446, 2007.

[16] P. Garcia‐del‐Barrio, F. J. M. Pujol, and D. Economics, "Hidden monopsony rents in winner‐take‐all markets—sport and economic contribution of Spanish soccer players," vol. 28, no. 1, pp. 57-70, 2007.

[17] F. Carmichael, D. Forrest, and R. J. B. o. E. R. Simmons, "The labour market in association football: who gets transferred and for how much?," vol. 51, no. 2, pp. 125-150, 1999.

[18] S. Kiefer, "The impact of the Euro 2012 on popularity and market value of football players," Diskussionspapier des Instituts für Organisationsökonomik2012.

[19] S. J. A. E. L. Medcalfe, "English league transfer prices: is there a racial dimension? A re-examination with new data," vol. 15, no. 11, pp. 865-867, 2008.

[20] B. Gerrard and S. J. J. o. E. S. Dobson, "Testing for monopoly rents in the market for playing talent–evidence from English professional football," 2000.

[21] E. E. Lehmann and G. G. J. A. E. Q. Schulze, "What does it take to be a star?-The role of performance and the media for German soccer players," vol. 54, no. 1, p. 59, 2008.

[22] 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.

[23] 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.

[24] A. T. Yiğit, B. Samak, and T. Kaya, "Football Player Value Assessment Using Machine Learning Techniques," in *International Conference on Intelligent and Fuzzy Systems*, 2019, pp. 289-297: Springer.

[25] J. L. Felipe, A. Fernandez-Luna, P. Burillo, L. E. de la Riva, J. Sanchez-Sanchez, and J. J. S. Garcia-Unanue, "Money Talks: Team Variables and Player Positions that Most Influence the Market Value of Professional Male Footballers in Europe," vol. 12, no. 9, p. 3709, 2020.