Research Article

Lalitaditya Charapanjeri*

COMPREHENSIVE ANALYSIS AND ESTIMATION OF CURRENT FOOTBALL PLAYER'S VALUE

https://doi.org/10.1515/sample-YYYY-XXXX
Received Month DD, YYYY; revised Month DD, YYYY; accepted Month DD, YYYY

Abstract: Football is one of the most popular sport in the world. The use of data analysis has been vital especially in the last few years.The best teams regularly invest millions into this field in order to achieve competitive edge over its rival teams.We aim to develop a method for accurately valuing players based on their on-field contributions. Our approach involves identifying key performance metrics, determining their importance, and combining them into a formula to calculate a player's overall value. This methodology aims to provide football clubs, agents, and analysts with valuable insights for making informed decisions in player recruitment, contract negotiations, and squad management

Keywords: Performance Metrics,Data Analysis

1 Introduction

Football, often regarded as the world's most popular sport, is not only a source of entertainment but also a multi-billion-dollar industry. In the modern era, the sport has witnessed a significant shift towards data-driven decision-making, with clubs, agents, and analysts increasingly relying on statistical analysis to gain a competitive edge [1][2]. One crucial aspect of this data-driven approach is the estimation of football players' values, a process that plays a pivotal role in player recruitment, contract negotiations, and overall squad management.[3] Traditionally, player valuation has been a subjective and somewhat opaque process, heavily influenced by factors such as reputation, market demand, and personal preferences. However, with the advent of advanced statistical techniques and the availability of vast amounts of player performance data, there has been a growing interest in developing more objective and data-driven methods for player valuation.[4] This project seeks to address this need by proposing a systematic methodology for estimating the value of football players based on their performance statistics.[5][6] By leveraging machine learning algorithms and statistical modelling techniques, the project aims to identify the key performance metrics that correlate with a player's value and determine the optimal weights for these metrics. These weights will then be used to construct a weighted average formula, providing a quantitative measure of a player's worth.This quantitative measure provides as a tool for coaches and scouts to accurately evaluate their players.

2 Literature Review

This section of the paper briefs about all the other research work done by people regarding data analysis of a football player's performance.

***Corresponding author: Lalitaditya Charapanjeri,** Vellore Institute of Technology,Chennai, School of Computer Science(SCOPE), Chennai, India, e-mail: lalitadityareddyc@gmail.com

Data analysis in football is not a new approach at all, there have been many teams that have used this powerful and effective method to identify key strengths and weaknesses and also how to rectify them.Now more than ever do teams need to adopt this strategy in order to remain competitive and not upstaged by their rivals.

There are several different research papers on how a player's performance could correlate with their market value.[7][8][9] This approach is the most popular and very relevant to today's footballing economy as football has never been 'more business" than before.Every year, billions of dollars are being invested into football and would only continue to grow as the popularity of the sport is at an all time high as FIFA reported the number of fans to be in billions.

There also other factors that could influence a player's performance such as physiological aspects which was one of the factors deemed relevant among many by Chelsea Football Club[10].The right state of mind would act as a catalyst to improve the player's performance.Chelsea have identified not just footballing statistic would shape the player but also their physical capabilities.Features such as sprint speed,acceleration,height,stamina etc. were identified as major influences to a player's style of play and ultimately his footballing position.Many pattern recognition algorithms along with clustering analysis was used to predict how effective a player could be.

But it is very difficult to predict an outcome accurately in football especially in recent years as there have many surprise team performances as well as players that have made it difficult for the so-called big teams.

The goal of this paper is not to predict the effectiveness of a player or an outcome of a match but it is to estimate a player's value through his performances.It does not involve use of market value as any indicator to judge a player's performance but solely through his contributions for his team throughout the season.

3 Proposed Work

This section of the paper identifies and explains the working of the model that was used to calculate the value of a player relatively.

The model used, has no reference to the current market value of the player and is solely evaluated based on the player's performances.

The section is divided into three subsections were each part is clearly defined and shown how it

3.1 Data Source

These were the parameters that were included for the comprehensive analysis of the players.

Parameters
Player Name
Team Name
Year
Age
Position
Total played
Started
Minutes per game
Team of the week
Goals
Scoring frequency
Assists
Goals per game
Assists per game
Shots per game
Shots on target per game
Touches
Big chances created
Key passes
Accurate passes per game
Own half passing(%)
Accurate Longballs
Clean sheets
Tackles per game
Possession won per game
Possession lost per game
Errors led to shot
Errors led to goal
Rating

These were the parameters of the data source that were taken for the model, the rest were deemed irrelevant based on personal football knowledge.

These parameters were then loaded into a correlation matrix for all positions independently where the target variable was "Rating" column. The parameters with the highest correlation were deemed as the most relevant for the performance analysis for a certain footballing position.These final parameters which were chosen are shown in the results section.

The dataset then is loaded into a model where the priority was to calculate the weights for the selected parameters with respect to the "Rating" column. The gradient descent algorithm is used to optimize the calculation of weights by using a loss function which was mean square error in this case. L2 regularization was used to avoid overfitting so that the calculated weight does not get a high magnitude to avoid poor generalization of the model which can further affect training in a negative manner.

3.2 Equations

Below is the primary equation that is used to calculate the value of a player.

$$X = \frac{x_g * w_g + x_a * w_a + x_k * w_k + x_t * w_t + x_b * w_b + x_s + w_s}{n} \tag{1}$$

where,

$$g = goals, a = assists, k = keypasses, t = touches, b = bigchances, s = shotspergame \tag{2}$$

$$x_i = statistics, w_i = weights, n = paramters \tag{3}$$

The above equation is an example on how the value of a player will be calculated.Please note that the above equation is only for the forward players and not fit for all the models that is midfielders and defenders where the features and weights would change based on the features extracted from the correlation matrix and the gradient descent execution respectively.

3.3 Algorithms, Program codes and Listings

Function gradientdescent(features, target, initialweights, learningrate, numiterations, lambdareg)
weights := COPY(initialweights)
losses := [] // List to store loss values at each iteration
for i := 1 to numiterations do: grad := gradient(features, weights, target, lambdareg) weights := weights - (learningrate * grad)

if i MOD 100 == 0 then: // Print weights every 100 iterations print("Weights at iteration", i, ":", weights) loss := lossfunction(features, weights, target) append(loss, losses)

print("Optimized weights:", weights)
showplot(losses)
return weights, losses

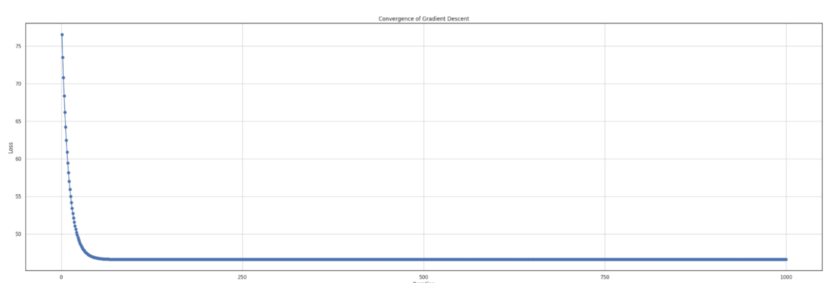

Fig. 1: Convergence graph for forwards

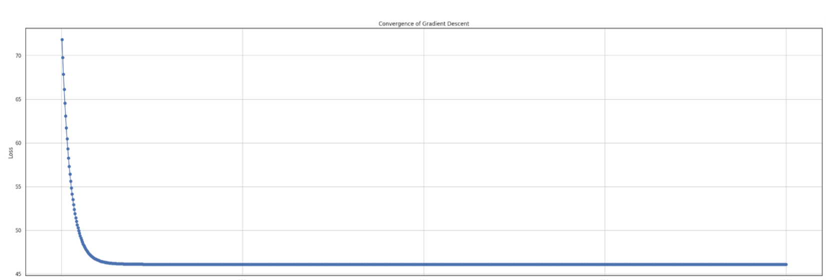

Fig. 2: Convergence graph for midfielders

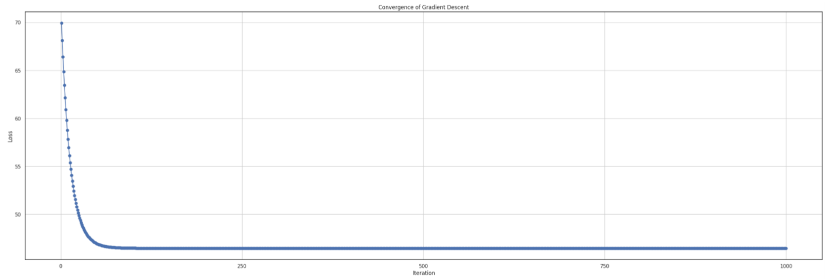

Fig. 3: Convergence graph for defenders

4 Results

The main objective of this project was to give football analysts, scouts and coaches to easily analyze player?s performance irrespective of position. Previously given critics ratings were taken as target variable to improve the quality of weights which would factor into a player's quality of factors.The most important features and weights for the respective positions are found to be, (refer Tables 1 and 2 and 3).

From Table 3, it is implied that the features were the most prevalent for each position respectively, The critics ratings were biased to the attacking returns of the defenders thus the prevalent features were mostly attacking returns.Though the attacking parameters were preferred for performance, the greatest impactful parameter turned out to be clean sheets which is when the defending side does not concede a goal for the entire match as the major role of the defenders is to not concede goals.An interesting feature weight is the games started one as it has a negative weight.My model concludes this due to drop in consistency of player's performance as he faces more games.

From breaking down Table 2,there are a lot of midfielder positions such as Centre Attacking midfielders(CAM), Centre Midfielder(CM) and Centre Defensive midfielder (CDM), their roles are fairly different, but the most common important parameter is the passing accuracy based on results obtained.In a future development, the positions can be clustered out to pick much more accurate results.Also Possession lost is most negative weights among all as ball distribution and handling is key aspect of a midfielder' role.

And finally Table 1 shows that the forwards have even weights among all the features.Key passes is found to be the most impactful feature among all, this could be due to a number of different positions available in the forward department such as Right winger (RW), Left Winger (LW) , Striker (ST) and Center Forward (CF) and key passes is important for all the positions relatively except the striker position. Shots on target is given a high preference as the forward is threatening the opposition's defense at regular times.Goals and assists also make in the list for most important features as this generally the sole aim for the forwards which is to get attacking returns.

Tab. 1: Most important features for forwards

Features	Weights (After 1000 iterations)
Big chances created	0.03411182
Key passes	0.0698413
Shots per game',	0.02092933
Goals	0.04315087
Touches	0.01850832
Shots on target per game	0.04572533
Assists	0.03272361

Tab. 2: Most important features for midfielders

Features	Weights (After 1000 iterations)
Accurate passes per game%	0.2856798
Own half passing %	0.04051005
Touches	-0.06455698
Key passes	0.07154584
Minutes per game	0.05071722
Accurate passes per game	0.0262352
Possession lost	-0.14198515
Assists	0.03334138

Tab. 3: Most important features for defenders

Features	Weights (After 1000 iterations)
'Clean sheets'	0.07436156
'Touches'	0.01793559
'Accurate passes per game'	0.05014182
'Started'	-0.01419658
'Total played'	0.02691812
'Shots per game'	0.05069822
'Acc. long balls'	0.01789451

Now with the use of these weights, each player's value was calculated by running their statistics into the model. There is also a column included that is titled 'Normalised value' which is scales the calculated value from 0-1000

The following parameters are included in the description

1. Name of the player
2. Team Name
3. Calculated value
4. Average rating of player in the season
5. Normalized score(Scaled from 0-1000)

Tab. 4: Performance Analysis of forwards in the 19/20 Premier League season

Name	Team Name	Calculated value	Rating	Normalised value
Mohamed Salah	Liverpool	2.6918	7.41	1000.0
Riyad Mahrez	Manchester City	2.5127	7.53	933.4527
Sadio Mane	Liverpool	2.4717	7.29	918.2216
Raheem Sterling	Manchester City	2.4079	7.22	894.5387
Raul Jimenez	Wolverhampton	2.3853	7.19	886.1266
Willian	Arsenal	2.31424	7.49	859.72566
Heung-Min Son	Tottenham	2.272648	7.08	844.2717
Jamie Vardy	Leicester City	2.15376	7.23	800.1087
Pierre-Emerick Aubameyang	Arsenal	2.11815	7.16	786.8771
Anthony Martial	Manchester United	2.0729	7.11	770.0733
Richarlison	Everton	2.05477	7.09	763.3340
Roberto Firmino	Liverpool	2.0072	7.04	745.6789
Danny Ings	Southampton	1.8414	7.04	684.0931
Harry Kane	Tottenham	1.7906	7.24	665.2114
Adama Traore	Wolverhampton	1.7614	7.27	654.3768
Christian Pulisic	Chelsea	1.7572	6.94	652.7731
Michail Antonio	West Ham United	1.7037	7.08	632.8988
Gabriel Jesus	Manchester City	1.6979	7.12	630.7487
Troy Deeney	Watford	1.63412	7.04	607.0661
Nicolas Pepe	Arsenal	1.5955	7.02	592.7338
Wilfried Zaha	Crystal Palace	1.58011	6.67	586.9998
Jordan Ayew	Crystal Palace	1.488	6.9	552.817
Ayoze Perez	Leicester City	1.486	6.85	552.054
Joshua King	Watford	1.4602	6.76	542.4892
Tammy Abraham	Chelsea	1.4394	6.84	534.7480

Tab. 5: Performance Analysis of Defenders in the 19/20 Premier League season

Name	Team Name	Calculated value	Rating	Normalised value
Virgil van Dijk	Liverpool	7.2797	7.28	1000.0
John Stones	Manchester City	6.6355	7.22	911.5072
Caglar Soyuncu	Leicester City	5.931	7.08	814.7435
Kurt Zouma	Chelsea	5.8637	6.97	805.49163
Harry Maguire	Manchester United	5.8519	7.13	803.8653
Andrew Robertson	Liverpool	5.8127	7.17	798.488
Trent Alexander-Arnold	Liverpool	5.7664	7.38	792.1210
Kyle Walker	Manchester City	5.70765	6.97	784.0423
Jonny Evans	Leicester City	5.6357	6.99	774.17123
Joe Gomez	Liverpool	5.6197	6.92	771.9732
Aymeric Laporte	Manchester City	5.5535	7.03	762.8857
Antonio Rudiger	Chelsea	5.5340	7.0	760.1964
Cesar Azpilicueta	Chelsea	5.5000	7.15	755.5319
Lewis Dunk	Brighton & Hove Albion	5.3334	7.05	732.643
Victor Lindelof	Manchester United	5.2098	6.87	715.6623
Oleksandr Zinchenko	Manchester City	5.17026	6.87	710.2303
Ben Chilwell	Chelsea	5.0909	6.84	699.3386
Ricardo Pereira	Leicester City	5.0212	7.08	689.7574
Toby Alderweireld	Tottenham	4.7855	6.97	657.3840
Adam Webster	Brighton & Hove Albion	4.7049	6.91	646.3084
Andreas Christensen	Chelsea	4.6892	6.9	644.1508
Conor Coady	Wolverhampton	4.6520	6.84	639.0411
Benjamin Mendy	Manchester City	4.6493	6.74	638.6725
Aaron Wan-Bissaka	Manchester United	4.6212	7.08	634.8169
Romain Saiss	Wolverhampton	4.3477	6.96	597.2473

Tab. 6: Performance Analysis of midfielders in 19/20 Premier League season

Name	Team Name	Calculated value	Rating	Normalised value
Kevin De Bruyne	Manchester City	14.329344076251408	8.0	1000.0
Jack Grealish	Aston Villa	14.29637734907303	7.39	997.6993554622633
Dwight McNeil	Burnley	14.20082316609979	6.94	991.0309285988449
John McGinn	Aston Villa	14.160628665839818	6.95	988.2258804371089
Ismaila Sarr	Watford	13.97545338210342	6.66	975.3030779172577
Mason Mount	Chelsea	13.654986932794372	6.92	952.9387290954459
Marcus Rashford	Manchester United	13.596079176539948	7.21	948.8277414646824
Ryan Fraser	Newcastle United	13.579221461168522	6.89	947.6512943585399
James Maddison	Leicester City	13.554240799863388	7.34	945.9079723214526
Miguel Almiron	Newcastle United	13.512372638752025	6.75	942.9861246158935
Gylfi Sigurdsson	Everton	13.496561861666459	6.86	941.8827400505266
John Fleck	Sheffield United	13.323381231356771	7.01	929.7970067895949
Jeff Hendrick	Newcastle United	13.252010711278656	6.55	924.8162819428518
Ashley Westwood	Burnley	13.216350863996205	7.01	922.3276929960939
Nathan Redmond	Southampton	13.190265129056622	6.94	920.5072513345096
Tariq Lamptey	Brighton & Hove Albion	13.176843750546269	6.8	919.5706154048444
James Ward-Prowse	Southampton	13.173038756682393	7.2	919.3050768118965
Joao Moutinho	Wolverhampton	13.117593517625853	7.21	915.4357274012397
Luke Thomas	Leicester City	13.105735802813236	7.1	914.6082146588897
Jonjo Shelvey	Newcastle United	13.099322929951777	7.14	914.1606803665079
Conor Hourihane	Aston Villa	13.07059540023059	7.0	912.1558761292506
Daniel James	Manchester United	13.002018760399116	6.74	907.3701274259914
Pascal Gross	Brighton and Hove Albion	12.986083192460502	7.05	906.2580340982149
Jarrod Bowen	West Ham United	12.946790401719271	6.92	903.5159134168955
Domingos Quina	Watford	12.815213174385281	6.47	894.3335512212624

5 Conclusion and Discussion

This section defines and explains the results of the model executed and how it compares to the real scenario.

The model clearly shows the best 25 players for each footballing position except the goalkeeper, in the 2019/2020 Premier League season. To compare the model with reality, we can evaluate the players with the help of any awards.

Kevin De Bruyne won the PFA Player of the Year for this contributions to Manchester City after a stellar season, he also leads the charts in the value calculated of midfielders in the 19/20 season.

Bruno Fernandes won the player of the month twice but does not make the top 25 as he joined late in the season which would not help his consistency stats for example - Key passes,Assists.Since the model is for all types of midfielders which include both attacking and defending midfielders, his goal contribution is not deemed important enough to raise his player value which shows the limitations of the model which can be improved upon in future works.

Trent Alexander-Arnold,Michail Antonio and Jamie Vardy are other Player of the month winners that rate fairly high in their calculated value showing the consistency of the model.

Another award that could support the model's evaluation is the Balon'dor award rankings which shows Virgil Van Dijk at 2nd position in the rankings and he leads the defenders table in calculated value and Mohamed Salah and Sadio Mane in 5th aand 4th positions respectively for the Balon'dor ranking as they were ranked 1st and 3rd respectively by the model.There are other influences on the Balon'Dor rankings as their UEFA Champions League(UCL) performances are taken into account and Liverpool were crowned the UCL winners which helps their prospects.

These points help in the support of the model as most of the award winners were ranked fairly high by the model as well.Although there are a few limitation of the model, the model gets fairly close to what experts think as most of the award winners coincide with players that were ranked high by the model

References

[1] Guimarães, J. H. M. M. (2018). *Data Analytics Applied to Football and Football Players* (Doctoral dissertation).
[2] Herberger, T. A., & Litke, C. (2021). The Impact of Big Data and Sports Analytics on Professional Football: A Systematic Literature Review. In *Digitalization, Digital Transformation and Sustainability in the Global Economy: Risks and Opportunities* (pp. 147-171).
[3] Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., ... & Hassabis, D. (2021). Game Plan: What AI Can Do for Football, and What Football Can Do for AI. *Journal of Artificial Intelligence Research*, 71, 41-88.
[4] Majewski, S. (2016). Identification of Factors Determining Market Value of the Most Valuable Football Players. *Central European Management Journal*, 24(3), 91-104.
[5] Li, C., Kampakis, S., & Treleaven, P. (2022). Machine Learning Modeling to Evaluate the Value of Football Players. arXiv preprint arXiv:2207.11361.
[6] Yiğit, A. T., Samak, B., & Kaya, T. (2020). Football Player Value Assessment Using Machine Learning Techniques. In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making: Proceedings of the INFUS 2019 Conference, Istanbul, Turkey, July 23-25, 2019* (pp. 289-297). Springer International Publishing.
[7] Inan, T., & Cavas, L. (2021). Estimation of Market Values of Football Players through Artificial Neural Network: A Model Study from the Turkish Super League. *Applied Artificial Intelligence*, 35(13), 1022-1042.
[8] Behravan, I., & Razavi, S. M. (2021). A Novel Machine Learning Method for Estimating Football Players' Value in the Transfer Market. *Soft Computing*, 25(3), 2499-2511.
[9] He, M., Cachucho, R., & Knobbe, A. J. (2015, June). Football Player's Performance and Market Value. In *Mlsa@ pkdd/ecml* (pp. 87-95).
[10] De Silva, V., Caine, M., Skinner, J., Dogan, S., Kondoz, A., Peter, T., ... & Smith, B. (2018). Player Tracking Data Analytics as a Tool for Physical Performance Management in Football: A Case Study from Chelsea Football Club Academy. *Sports*, 6(4), 130.