
DECONSTRUCTING PLAYER PERFORMANCE: A PCA AND RANDOM FOREST BASED APPROACH

A PREPRINT

Keelan Evers

School of Mathematical and Statistical Sciences
University of Galway
Galway, H91 TK33, Ireland
k.eyers1@universityofgalway.ie

July 18, 2024

Abstract

Player performance evaluation in professional football is essential for informed decision-making by coaches, scouts, and analysts. Traditional methods often rely on subjective assessments, leading to inconsistent and biased evaluations. This study aims to provide a more objective evaluation by utilizing Principal Component Analysis (PCA) and Random Forest regression. The PCA approach reduces dimensionality and creates composite performance scores from a comprehensive set of metrics, offering a holistic view of player performance. Random Forest regression identifies key performance metrics influencing player market value, providing insights into which attributes are most valued in the market. The dataset, sourced from FB REF and Transfermarkt, includes performance metrics for 569 players from the 2022/2023 Premier League season. Results show that different metrics drive performance and market value for defenders, midfielders, and forward. The analysis highlights the complexity of market valuation and identifies over- and undervalued players. This study contributes to football analytics by presenting a robust methodology for objective player evaluation and market value prediction, offering valuable insights for talent acquisition and player development. Future research should incorporate broader datasets and additional factors to enhance model performance and predictive power. Availability: Code available at: <https://github.com/eyersk98/HDS> Contact: k.eyers1@universityofgalway.ie

1 Introduction

Player performance evaluation in professional football is a critical aspect that significantly influences decisions made by coaches, scouts, analysts, and even fans. Traditional methods of evaluating player performance are often prone to bias and subjectivity, relying heavily on personal opinions and qualitative assessments. (McIntosh et al., 2019) (Salabun et al., 2020)(Oliveira and Newell 2024) This inherent subjectivity can lead to inconsistent and potentially unfair evaluations, which ultimately affect player market values, team selections, and career advancements. As the sport evolves and becomes increasingly data-driven, the need for more objective, reliable methods to evaluate player performance becomes imperative. (McHale et al., 2012) One of the primary issues with subjective performance evaluation is the variability and bias it introduces. Different scouts, coaches, and analysts may have varying opinions on what constitutes good performance. These differences can be due to personal preferences, biases, or a lack of comprehensive understanding of all aspects of the game. For instance, one scout might prioritize a midfield player's goal-scoring ability, while another might value defensive contributions more highly. This inconsistency can result in players being undervalued or overvalued, affecting their career opportunities and market values. (Dutch et al., 2010)(McCormack et al., 2021)

Football is a complex sport where various skills and attributes contribute to overall performance. These include not only goal-scoring and defensive actions but also passing accuracy, dribbling, positioning, and

many other metrics. Traditional methods that focus on isolated performance indicators do not provide a holistic view of a player's contributions on the field. Consequently, there is a pressing need for an evaluation system that can integrate multiple performance metrics into a single, objective score that reflects a player's overall performance. (Merzah et al., 2024) (Mann et al., 2022)

This research aims to address these shortcomings by employing advanced statistical techniques, specifically Principal Component Analysis (PCA) and Random Forest regression, to objectively measure player performance and determine the factors influencing a player's market value. By reducing dimensionality through PCA, we can distill a large set of performance metrics into principal components that retain the most critical information about player performance. An approach inspired by (Oliveira and Newell 2024) allows us to create composite scores that reflect a player's overall performance more accurately. Furthermore, using Random Forest regression, we can identify the specific performance metrics that most significantly impact a player's market value, providing insights into how different aspects of performance are valued in the market. (Yi et al., 2020) (Ramnath and Priya 2024)

The need for an objective player performance evaluation system is paramount in modern football for several reasons. Firstly, subjective assessments often lead to inconsistent evaluations, as different scouts, coaches, and analysts may have varying opinions on what constitutes good performance. This inconsistency can result in players being undervalued or overvalued, affecting their career opportunities and market values. Secondly, a data-driven approach can help in identifying undervalued talent, providing teams with a competitive edge in player acquisitions and squad development. (McIntosh et al., 2019) (Salabun et al., 2020) (Oliveira and Newell 2024)

Moreover, an objective evaluation system benefits not only the teams but also the players themselves. By providing a transparent and unbiased measure of performance, players can better understand their strengths and areas for improvement. This feedback can be crucial for their development and progression in the sport. Additionally, such a system can enhance the overall quality of the game by ensuring that players are evaluated and rewarded based on their actual performance rather than subjective opinions. (Merzah et al., 2024) (Mann et al., 2022)

1.1 Aims

This study fills a significant gap in the existing literature and practice of player performance evaluation by introducing a holistic, data-driven methodology. The primary contributions of this research are twofold:

1. Objective Measurement of Player Performance: By applying PCA to a comprehensive set of performance metrics, this study creates a composite performance score for each player. This score provides a holistic measure of performance, integrating various aspects of the game such as shooting, passing, defense, and possession into a single metric. This approach not only reduces the dimensionality of the data but also ensures that the most important performance characteristics are retained.
2. Identification of Key Performance Metrics Influencing Market Value: Using Random Forest regression, this study identifies which performance metrics are most predictive of a player's market value. By doing so, it highlights the attributes that scouts and teams consider most valuable in players. This insight can inform player development programs, scouting strategies, and market negotiations.

1.2 PCA

PCA is a statistical technique used to simplify a complex dataset by transforming it into a set of uncorrelated variables called principal components. These components are linear combinations of the original variables and are ordered such that the first few retain most of the variation present in the original dataset. In this study, PCA is used to reduce the dimensionality of the performance metrics while preserving the most significant information. The principal components derived from PCA serve as the basis for creating composite performance scores for players. (Maćkiewicz & Ratajczak, 1993)

PCA helps in identifying the key factors that contribute to the overall variance in the dataset, thereby reducing the complexity of the data while retaining the most important information. This technique is particularly useful in handling large datasets with many correlated variables, as it transforms the data into a smaller set of uncorrelated components. By focusing on these principal components, we can gain insights into the underlying structure of the data and identify the main drivers of player performance. (Pino-Ortega et al., 2021) (Rodionova et al., 2021) (Oliveira and Newell 2024)

1.3 Composites

Composite scores are created by combining multiple performance metrics into a single score that represents a player's overall performance. In this study, the composite scores are calculated using the principal components obtained from PCA. The use of composite scores allows for a more comprehensive evaluation of player performance, as it integrates various aspects of the game into a single metric. (UmaMaheswari & Rajaram, 2009)

The composite score provides a holistic measure of performance, taking into account various dimensions of the game. This approach ensures that players are evaluated based on their overall contributions rather than isolated metrics. By aggregating multiple performance indicators into a single score, we can provide a more accurate and balanced assessment of a player's abilities. This composite score can then be used to compare players objectively, identify top performers, and inform decision-making processes. (Song et al., 2013) (Oliveira and Newell 2024).

1.4 Random Forest

Random Forest regression is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction value for regression. It is particularly useful for handling large datasets with complex interactions among variables. In this study, Random Forest regression is employed to predict player market values based on their performance metrics. The model identifies the most important variables influencing market value, providing insights into which aspects of performance are most highly valued in the market. (Breiman, 2001) (Groll et al., 2018)

Random Forest regression is advantageous because it can handle both linear and non-linear relationships between variables. It also provides measures of variable importance, which indicate how each predictor contributes to the model's predictions. This feature is particularly valuable in our study, as it allows us to identify the key performance metrics that drive player market values. By understanding these relationships, we can provide actionable insights to teams and scouts about the factors that influence player valuation.(Pantzalis & Tjortjis, 2020) (Morciano et al., 2022).

Player market value is influenced by a variety of factors, including age, performance metrics, position, and market dynamics. Performance metrics such as goals scored, assists, defensive actions, and passing accuracy are key indicators of a player's value. However, other factors such as a player's age, injury history, and market demand also play a significant role. This study focuses on the performance metrics that can be quantified and analyzed to predict market value, providing a data-driven approach to understanding how different aspects of performance are valued. (Patnaik et al., 2019)

Market value is a complex construct that reflects a player's perceived contribution to a team's success. It is influenced by both objective performance metrics and subjective factors such as reputation, popularity, and marketability. By focusing on quantifiable performance metrics, our study aims to provide a more objective and transparent basis for player valuation. This approach can help mitigate the influence of biases and provide a clearer understanding of how performance translates into market value. (Jishnu et al., 2022) (Frick, 2007)

2 Methods

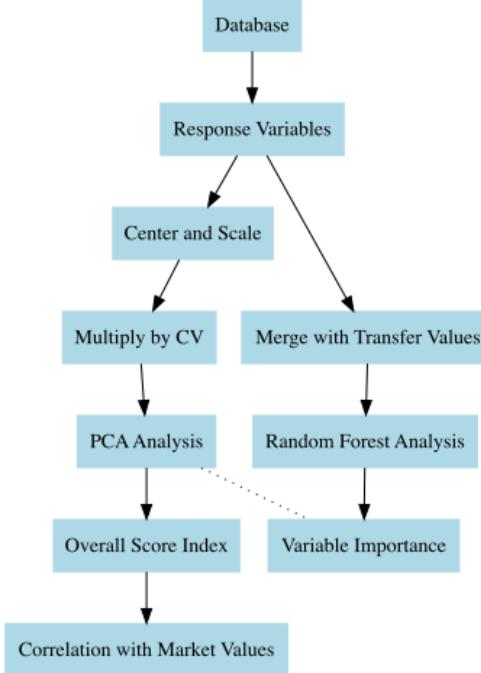


Figure 1: Flow Chart of the project workflow, the process of the two analysis and the comparisons made

2.1 Data acquisition

The dataset used in this analysis was obtained from the FB REF database (<https://fbref.com/en/>), a publicly accessible source of Football data. The data obtained were in separate groups characterised as shooting, passing, pass types, goal and shot creation, defensive actions, possession and miscellaneous player performance metrics. The database that resulted from this compilation is available at (<https://github.com/eyersk98/HDS>)

This database contains 569 players from the 2022/2023 premier league season and 126 player performance metrics, 24 shooting, 30 passing, 22 pass types, 23 goal and shot creation, 23 defensive actions, 29 possession and 23 miscellaneous metrics. Duplicate variables such as crosses, interceptions and passes attempted were removed after merging. Minimal filtering was applied to remove players with low matches played ($90s < 3$) and the remainder were used as input for analysis. Players were assigned to their primary positions by selecting the first listed position in the cases where multiple positions were listed. Goalkeepers, 39 in total, were removed from the dataset and then players were separated into 3 distinct position groups: 165 Defenders, 133 Midfielders and 121 Forwards.

Within each position group variables with NA values were excluded, this is due to missing value imputation or similar methods not being appropriate as variables with NA values typically included ones such as “shots on target %” or “one vs one %”, wherein the percentage value could not be calculated as the parent value was zero.. Also variables with zero variance such as penalty attempts for defenders were excluded as they did not provide relevant information.

2.2 EDA

Exploratory data analysis was conducted to understand the structure and characteristics of the datasets. The sample mean, standard deviation and coefficient of variation for the response variables for each position group were calculated. R package pheatmap was used to construct sample heatmaps showing the Pearson correlation coefficient between each pair of variables together with hierarchical clustering. (Sedgwick, 2012) Using exploratory data analysis to understand my data gave a sense of how variables relate to each other, if they are negatively or positively correlated, helped guide me to understand player component scores and if

reversing of components is needed these variables appear in. (Casal et al., 2021) (Sedgwick, 2012). All EDA visualisations used during this project are available in the supplementary section.

2.3 Scaling

Non-continuous variables such as player team, name, and squad leaving us with only continuous response variables, some of which may be correlated. The sample means and standard deviations of each variable were calculated to standardise each variable using the relevant sample standard deviation and mean. Next we multiply each variable value by its respective coefficient of variation (CV), a measure of the dispersion of a probability distribution, as performed by (Oliveira and Newell 2024). Is defined as a ratio of the standard deviation to the mean.(Reed et al., 2002) This scaling by CV ensures that the variability of each response variable is suitably accounted for. Defining:

$$z_i^* = z_i \frac{s_i}{\bar{r}_i} = \frac{r_i - \bar{r}_i}{s_i} \cdot \frac{s_i}{\bar{r}_i} = \frac{r_i - \bar{r}_i}{\bar{r}_i}, \quad |\bar{r}_i| > 0.$$

2.4 PCA

PCA was performed on the pre-processed data to reduce dimensional while retaining the most important variance in the dataset. The PCA process involved eigenvalue decomposition, extraction of component loading values and player principal component scores. Principal components are linear combinations of a set of uncorrelated variables, represented by y , derived from z variables. This can be expressed as:

$$y_c = \alpha_{c1} z_1^* + \alpha_{c2} z_2^* + \dots + \alpha_{cn} z_n^*, \quad c = 1, 2, \dots, n.$$

The c -th principal component is the linear combination $y_c = \alpha_c^\top z^*$ that has the largest sample variance. A denotes the vectors that define the n principal components. $A = [\alpha_1, \alpha_2, \dots, \alpha_n]$ and $\lambda^\top = [\lambda_1, \lambda_2, \dots, \lambda_n]$ a vector of length n containing their corresponding eigenvalues. On the basis of visual examination of scree plots, a five component interpretation was deemed appropriate for defenders which accounted for 75% of the total variation in the data, two component interpretation for midfielders (89%) and one component interpretation for forwards (95%). Visual plots of the components, the percentage of variation account for by the component and only the predominant loadings (i.e. top five highest absolute loading values) were extracted.(Oliveira and Newell 2024)(Mackiewicz & Ratajczak, 1993)(Newell et al., 2006)

2.4.1 Composite score

Following on from the methods used by (Oliveria and Newell 2024), as a result of the weighted scaling method, it allows us to use all the components derived from PCA. Using A and λ a weighted matrix W can be calculated and defined as :

$$W = (1_n^\top \otimes \lambda) \odot A^\top S^{-1}, \text{ with } S = \sum_{c=1}^n \lambda_c$$

As a result principal components with high variances contribute more to the weighted mean than those with low variances.(Oliveira and Newell 2024)The overall score vector as a result of encompassing all scores for players can be expressed as:

$$\gamma = (W^\top 1_n) Z^{*\top}$$

2.4.2 Min-max scaling

Min-max scaling transformation is a technique used to normalise data and make it more interpretable. (Patro & sahu, 2015) Data is rescaled usually between fixed values of 0 and 1 preserving original relationships between data points. Min-max scaling was applied to the players overall composite scores to make them more interpretable and more comparable. (Oliveira and Newell 2024) In this study we set the fixed values between 0 and 100 so rescaled scores will range between 0 and 100. With higher values indicating better performance. The min-max scale formula can be defined as:

$$\gamma = a + \frac{(\gamma - \min(\gamma))(b - a)}{\max(\gamma) - \min(\gamma)}$$

2.5 Transfer Values

Player market values were scraped from the Transfermarkt website (<https://www.transfermarkt.com/>) using Python's BeautifulSoup and HTML parsing libraries.(Lotfi et al., 2021) Scrapped data was merged with the raw player performance metrics database using R package fuzzyjoin to account for slight discrepancies in player names between the two datasets.

2.6 Random forest

The datasets were split into training and testing sets to evaluate the performance of the Random Forest model. Specifically, 70% of the data was allocated to training set and the remaining 30% was reserved for testing. Partitioning was done using the R package caret to create an unbiased split of the data.

A random forest model was employed to predict player market values based on their performance metrics. Random forest algorithm is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction value for regression of the individual trees. Training of the model was performed using 10-fold cross-validation to prevent overfitting. This cross-validation process involves dividing the training set into 10 subsets, training the model on nine of the folds and validating it on the remaining fold. This processes is repeated 10 times, with each fold used once for validation.

After model training, variable importance was assessed to identify which performance metrics had the most significant impact on predicting player market values. Variable importance is typically measured by the mean decrease in impurity (residual sum of squares for regression), this measures the change in the model's performance when the values of a give variable are permuted. Variable importance was extracted using the caret package in R. This allows us to see a ranked list of variables based on their importance score.

Performance of the random forest model was evaluated using several key metrics and visualizations. The RMSE, R-squared value and the relative error rate of the models were calculated. Predicted vs observed and residuals plot were created to help identify any systematic errors or bias. Predictions on the full dataset were done to obtain residuals for each data point, this helped identify over or under valued players according to the model. (Breiman 2001)(Biau 2012)

2.7 Overall Score and Market Value

To explore the relationship between the composite performance score derived from PCA and the player market values we integrated the two datasets. Composite scores, representing the holistic performance value for each player, was merged with their respective market values. We utilised Pearson's correlation coefficient, a measure of linear correlation between two variables to quantify this relationship, where $r = 1$ indicates a perfect positive linear relationship, $r = -1$ indicates a perfect negative linear relationship and $r = 0$ indicates no linear relationship. Visual inspection was also carried out where scatter plots of composite score vs market value fitted with a linear regression line to visualise the trend and direction of the relationship.

3 Results

3.1 PCA

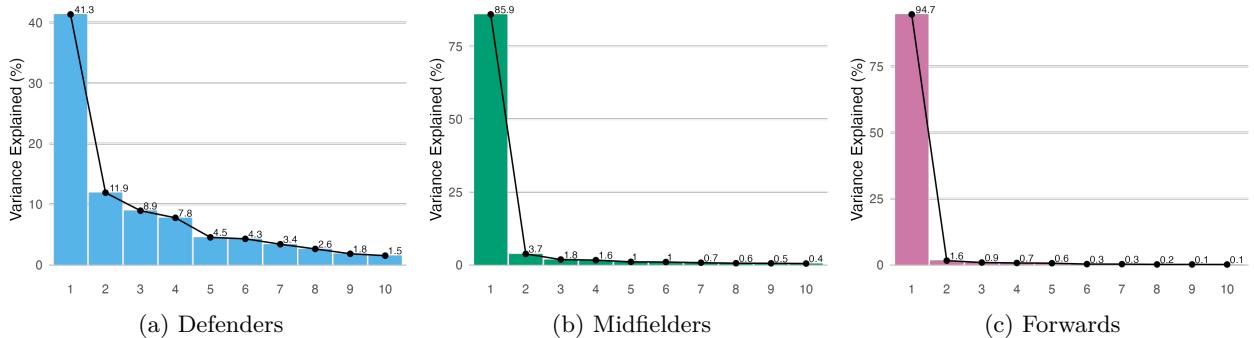


Figure 2: Scree plots, showing the variance explained (eigenvalues) of the first 10 principal components for each position group

Eigenvalues of the principal components from PCA values were extracted to generate scree plots illustrating the variance explained by each principal component. While all components are incorporated into a player's composite score, the variance explained helps us decide which components we should interpret for understanding the main driving force of a player's overall score. As seen in Figure 2a, the first principal component for defenders accounts for approximately 41% of the total variance, and the first five component together explain about 75% of the variance, so were chosen for interpretation. For midfielders the first principal component alone explains around 86% of the variance, the second component adds another 4%, these two components were chosen for interpretation accounting for around 90% of the variation seen in Figure 2b. In Figure 2c, the first principal component for forwards explains about 95% of the variance, meaning we can just interpret this component to understand most of the data.

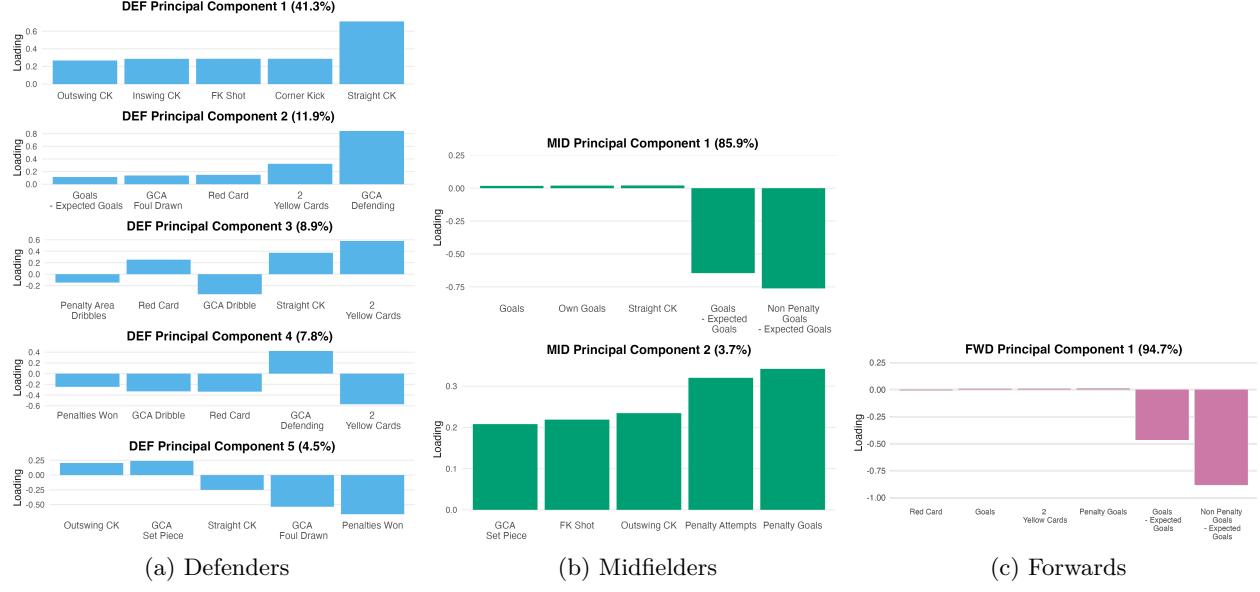


Figure 3: Principal component loadings, depicting the loading values variables of the 5 highest absolute values for each position group

Interpreting the components is crucial to understanding what the main drivers are of a players overall score. The component loadings provide insights into which performance metrics contribute most to the interpreted components. In Figure 3a we see defenders principal component 1 is dominated by variables related to set-piece situations with straight, inswinging and outswinging corner kicks present along with free-kick shots and so we will call this component set pieces. Component 2 highlights both defensive actions and drawing fouls leading to goals along with disciplinary action seen with red card and 2 yellow cards, so we will call this component actions leading to goals and aggressive defending. Principal component 3 similarly focuses on disciplinary action accompanied with penalty area dribbles and dribbling actions that lead to goals , we will label this component as aggressive defending and dribbling threat. Component 4 focuses on aggressive defending and defensive and dribbling actions leading to goals so as such we will call this aggressive defending and actions leading to goals. Finally principal component 5 for defenders we can see fouls drawn that lead to penalties or goals and set pieces so this will be characterised as drawing fouls for goals and set piece taking. In Figure 3b we can see midfielders principal components loadings. The first component is strongly influenced by offensive metrics goals minus expected goals and non penalty goals minus expected goal, which essentially represents if players are scoring more or less than they are expected so we will call this component taking their chances. The second component includes penalty goals, penalty attempts and other set piece actions leading to goals so this can be called penalties and set pieces. Figure 3c shows the loadings for forwards component 1 with goals minus expected goals and non penalty goals minus goals seen here similarly to midfielders, this component can be classed as forwards taking their chances. These loadings help us understand the factors driving the composite player performance scores.

3.2 Player Performance

The top-ranked players for defenders, midfielders and forwards based on their overall composite scores can be seen in Figures 4, 5 and 6, showing their scores for the interpreted components, their overall composite score and their min-max scaled overall composite score.

Rank	Player	Set Pieces	Aggressive, GCA Def	Aggressive, Drib GCA	Aggressive, GCA	Set Pieces, Draw Fouls	Overall Score	Scaled Overall Score
1	Kieran Trippier	160.28	-7.19	-20.10	-10.10	7.46	133.23	100.00
2	Ivan Perišić	22.00	3.92	22.31	14.91	-11.30	79.36	66.91
3	Pervis Estupiñán	3.31	5.51	12.36	9.41	14.34	56.72	53.00
4	Andrew Robertson	23.18	3.77	17.86	10.78	-16.19	53.32	50.91
5	Fabian Schär	3.92	62.95	7.43	-27.63	3.73	50.96	49.46
6	Tyrone Mings	-1.81	2.02	3.20	3.72	15.33	45.30	45.98
7	Ben Chilwell	10.37	1.16	9.32	9.65	10.71	43.90	45.12
8	Antonee Robinson	1.33	3.72	14.22	10.27	4.49	40.91	43.28
9	Trent Alexander-Arnold	40.73	6.98	25.76	17.18	-17.08	40.68	43.15
10	James Justin	-1.88	1.34	9.40	10.32	15.69	40.16	42.83

Figure 4: Top 10 Defenders based on overall composite score and key component scores for each player contributing to overall score

In Table 4 we can see Kieran Trippier ranks highest, driven by his high score in set-pieces, while players like Ivan perisic and Pervis Estupinan show more balance through their scores from defensive and offensive contributions.

Rank	Player	Taking Their Chances	Set Pieces	Overall Score	Scaled Overall Score
1	James Ward-Prowse	102.46	28.78	174.96	100.00
2	Martin Ødegaard	126.94	5.93	129.42	83.08
3	Pascal Groß	92.93	18.80	118.90	79.17
4	Jefferson Lerma	61.77	-0.11	101.99	72.89
5	Mathias Jensen	72.93	15.38	78.87	64.30
6	Declan Rice	53.01	2.32	72.44	61.91
7	Jacob Ramsey	26.71	3.16	65.66	59.40
8	James Maddison	50.36	18.76	65.36	59.28
9	Granit Xhaka	59.03	0.61	64.15	58.83
10	Philip Billing	53.92	-0.90	56.07	55.83

Figure 5: Top 10 Midfielders based on overall composite score and key component scores for each player contributing to overall score

In Table 5, James Ward_prowse leads due to his strong contributions in taking his chances and set piece scores, while players like Jefferson Lerma rely purely on their high score in taking their chances to overcome their poor set piece score to have them high up the rankings.

Rank	Player	Taking Their Chances	Overall Score	Scaled Overall Score
1	Harry Kane	345.86	327.38	100.00
2	Erling Haaland	263.18	228.39	81.44
3	Roberto Firmino	228.17	224.91	80.78
4	Gabriel Martinelli	236.66	215.42	79.00
5	Phil Foden	211.83	205.29	77.10
6	Rodrigo	196.98	191.80	74.58
7	Luis Sinisterra	55.57	188.46	73.95
8	Harvey Barnes	170.59	172.00	70.86
9	Dwight McNeil	146.00	138.53	64.58
10	Taiwo Awoniyi	141.49	133.01	63.55

Figure 6: Top 10 Forwards based on overall composite score and key component scores for each player contributing to overall score

In Table 6, we see players like Harry Kane and Erling Haaland top the ranks as they are taking their chances and scoring more than they are expected.

3.3 Random Forest

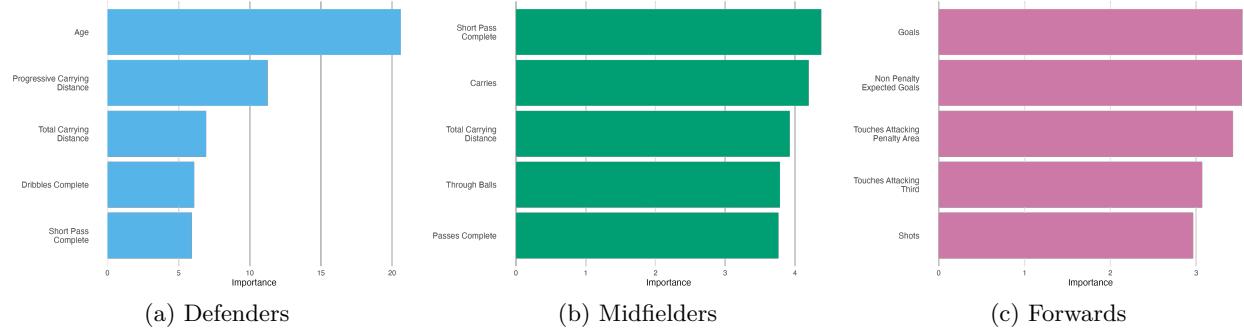


Figure 7: Variable Importance extracted from random forest models depicting variables that result in the most loss of model performance when removed, determined by mean gini index

Variable importance was extracted from the random forest model for each position. Figures 7a, 7b and 7c highlight the most critical predictors of player market value, based on the random forest model's mean decrease in impurity (Gini index).

For defenders in Figure 7a we can see age, progressive and total carrying distance, dribbles and short passes complete are deemed most important. This highlights that a defender's age and their ball playing ability contribute to their market value. Figure 7b highlights the most important variables for midfielders as we see short pass and passes complete, through balls and total carrying distance and carries. This tells us a midfielder's passing and dribbling ability influence their market value. Figure 7c shows the most important variables for forwards as we see goals, non penalty expected goals, shots and touches in attacking penalty area and third. This suggests that a forward's market value is determined by their general attacking threat. It is interesting to note that when comparing these important variables to those with high loadings found in

the first few important components there is a lack of overlap regarding defenders and midfielders but some similarities seen with forwards as attacking threat and taking their chances can be comparable.

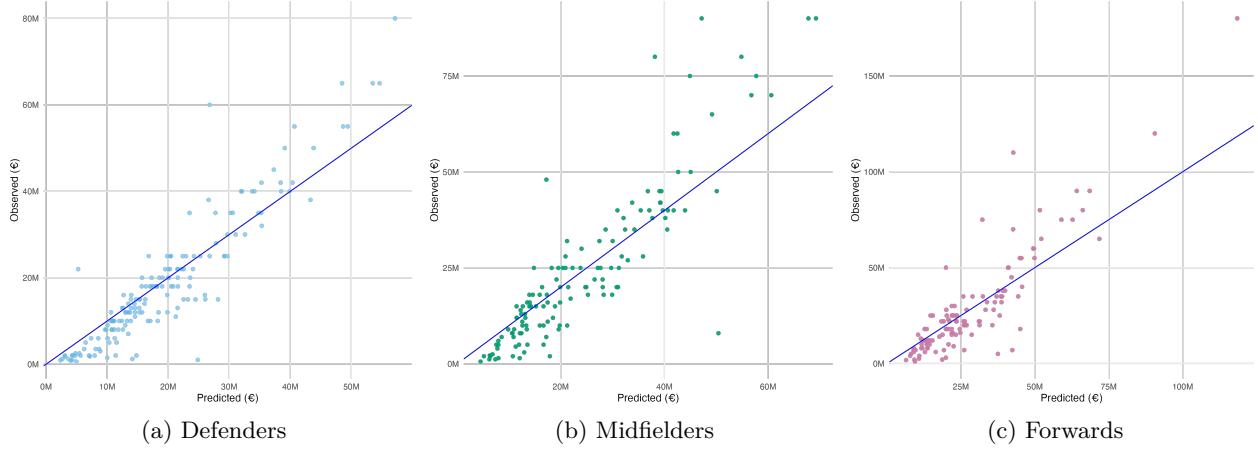


Figure 8: Observed vs predicted values for each random forest model predictions within each position group

Model	Rsquared	RMSE	RelativeError
Defender	0.33	8766833	46.59
Midfielder	0.57	15952030	64.56
Forward	0.45	20671930	72.54

Figure 9: Random Forest model performance

To help visualise model performance we can look at the observed vs predicted market values seen in Figure 8 fitted with a line of equality to assess model accuracy and identify over or undervalued players according to the model. There is a pattern seen across all three groups in that decreased accuracy for high-valued players is seen compared to lower valued ones.

Looking at Figure 9 we can see the model performance values in terms of RMSE, R-squared and relative error rate. For defenders we can see an R squared value of 0.41, an RMSE of 8,766,833 and relative error rate of 5%. For midfielders we see 0.57 R squared value, 15,952,030 RMSE and a relative error of 65%. For forwards we see an R squared of 0.48, RMSE of 20,671,930 and a relative error rate of 73%. Across all three groups we can see that the random forest model captures some of the variance in player market values, but there is substantial room for improvement. The R squared values indicate moderate predictive power but the high RMSE and relative error rates indicate significant prediction errors.

In Figure 10 we can see player market value residuals which tells us what their predicted minus observed values are to see. We can use this to determine over and under valued players according to the model. Despite the models low to moderate performance we can use these residuals to identify under or over valued players that will help determine factors for why the model may be under performing.

Player	Residual	Type	Player	Residual	Type	Player	Residual	Type	Player	Residual	Type	Player	Residual	Type	Player	Residual	Type
Cristian Romero	33221460	Defender	Declan Rice	42833165	Midfielder	Phil Foden	67293533	Forward	Tim Ream	-23843503	Defender	Pascal Groß	-42401483	Midfielder	Danny Welbeck	-35374276	Forward
Rúben Dias	22854233	Defender	Bernardo Silva	41850152	Midfielder	Erling Haaland	61562249	Forward	Nelson Semedo	-13151170	Defender	Tom Cairney	-15887409	Midfielder	Jordan Ayew	-32523067	Forward
Mathias Jørgensen	16751703	Defender	Richarlison	30784677	Midfielder	Luis Diaz	42753375	Forward	Aaron Cresswell	-12930127	Defender	Harrison Reed	-11898241	Midfielder	Roberto Firmino	-27072778	Forward
Reece James	16524200	Defender	Moisés Caicedo	30100975	Midfielder	Mykhailo Mudryk	30047875	Forward	Thiago Silva	-12796183	Defender	Idriissa Gana Gueye	-11668393	Midfielder	Michail Antonio	-19121551	Forward
Wesley Fofana	14338700	Defender	Enzo Fernández	25172174	Midfielder	Bukayo Saka	29423433	Forward	Kenny Tete	-11061600	Defender	Jordan Henderson	-11225827	Midfielder	Willian	-16917383	Forward

(a) Over-valued

(b) Under-valued

Figure 10: Player Residuals

3.4 Score vs market value

score index vs market value

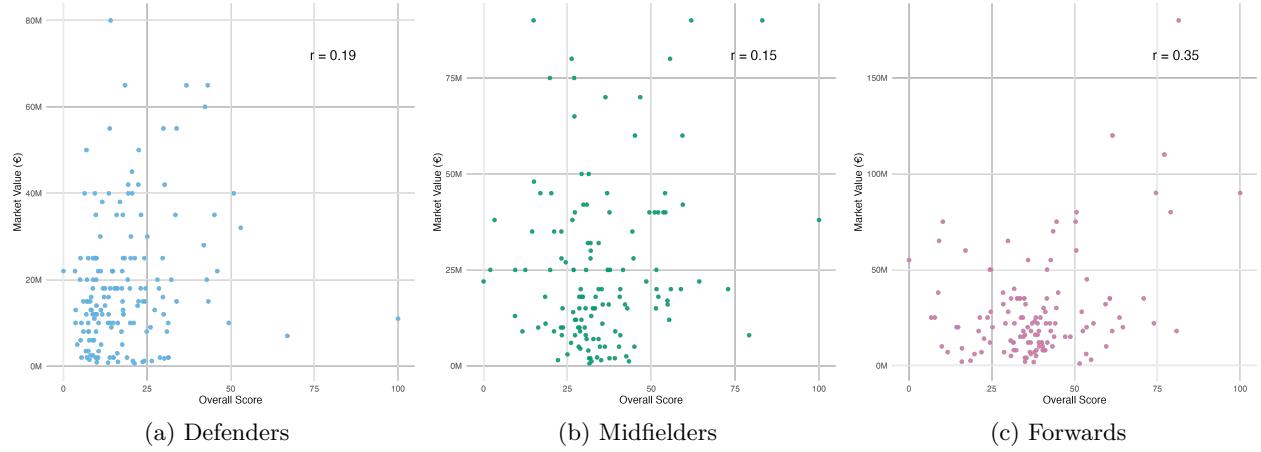


Figure 11: Overall score vs Market Value, allowing us to see the relationship between the two values for a player and understand them better.

The Pearson correlation coefficients (r) between composite scores and market values for defenders, midfielders and forwards are 0.19, 0.15 and 0.36 respectively and the relationship can be seen in Figure 11. The weak correlation for defenders indicate that factors beyond performance metrics significantly influence market value evaluations. Similarly for midfielders the complexity of market valuation is not decided by performance metrics alone. A moderate correlation seen in forwards suggests that performance metrics are more directly related to market values but other factors still play a key role.

4 Discussion

This study aimed to address the inherent subjectivity in traditional player performance evaluations by utilizing Principal Component Analysis (PCA) and Random Forest regression. The objective was to generate a composite performance score for each player and determine which performance metrics most significantly impact player market value. Our findings reveal important insights and further understanding of the complexity of player performance evaluation and determining a player's market valuation.

The PCA allowed us to reduce the dimensionality of the dataset while retaining the most critical information about player performance. By interpreting the principal components, we identified the key performance metrics that drive overall player performance for each position group. The results indicate that different metrics are important for defenders, midfielders, and forwards:

Defenders: Metrics related to set-pieces, defensive actions, and disciplinary actions were significant. This highlights the importance of both defensive solidity and contribution to set-pieces in evaluating defender performance.

Midfielders: Offensive metrics, particularly those related to goals and set-pieces, were crucial. This underscores

the multifaceted role of midfielders in both attacking and set-piece situations.

Forwards: Metrics related to goal-scoring efficiency (goals minus expected goals) were predominant, emphasizing the primary role of forwards in converting chances into goals.

The composite scores provide a holistic measure of player performance by integrating multiple performance metrics into a single score. The top-ranked players in each position group were identified based on these composite scores, revealing the key contributors in each category. This approach ensures a more comprehensive and balanced evaluation of player abilities, moving beyond isolated performance indicators. (Oliveira and Newell 2024) (O'Donoghue, 2008)

The Random Forest model identified the most important performance metrics influencing player market value. These metrics varied across positions, reflecting the different skill sets valued for defenders, midfielders, and forwards:

Defenders: Age, carrying distance, dribbles, and short passes were crucial. This indicates that a defender's ability to progress the ball and age significantly impact their market value.

Midfielders: Passing and dribbling metrics were key, highlighting the importance of ball distribution and carrying ability in determining a midfielder's value.

Forwards: Goal-scoring metrics and touches in the attacking areas were most important, emphasizing the critical role of attacking efficiency in forward valuation

The Random Forest model's performance varied across position groups, with moderate predictive power (R-squared values between 0.41 and 0.57). However, the high RMSE and relative error rates indicate substantial prediction errors, particularly for high-value players. The Pearson correlation coefficients between composite scores and market values were weak to moderate (0.15 to 0.36), suggesting that factors beyond performance metrics significantly influence market value evaluations. (Biau, 2012)

4.1 Complexity of Market Valuation

The weak to moderate correlations between composite scores and market values highlight the complexity of market valuation in football. While performance metrics play a role, other factors such as age, contract length remaining, market dynamics, reputation, and subjective assessments by scouts and analysts also significantly impact market value. This complexity suggests that a purely data-driven approach, while valuable, cannot fully capture the nuances of player valuation. (Singh & Lamba, 2019)(Majewski, 2016)

4.2 Identifying Over- and Undervalued Players

Despite the model's limitations, the analysis of residuals (predicted minus observed values) allows us to identify over- and undervalued players. These insights can be crucial for teams looking to acquire talent that is undervalued by the market, providing a competitive edge in player acquisitions. Similarly, understanding why certain players are overvalued can help teams avoid costly mistakes. (He et al., 2015) (Biau, 2012)

4.3 Implications for Football Analytics

This study contributes to football analytics by providing a robust methodology for objectively evaluating player performance and understanding the factors influencing market value. The use of PCA to create composite performance scores and Random Forest regression to identify key performance metrics offers a comprehensive framework for player evaluation. This approach can be adopted by teams, scouts, and analysts to enhance decision-making processes and reduce the subjectivity in player assessments. (Rojas-Valverde et al., 2020) The composite score accompanied with player market value assessment can help teams target key players that objectively performing well according to data driven analysis but have low market values.

Similiarities or differences between high loading variablea and variable importancea age/contract etc differences between processes, variation and what makes a player stand out vs predictive power Reasons for poor model performance More variation in defenders compared to midfielders and forwards

4.4 Limitations

The analysis done in this study relies on the quality and completeness of the data from FB REF and Transfermarkt. Missing values and discrepancies between datasets can affect results. The analysis was performed on one league in one season, a more comprehensive approach could include various leagues at low

and high levels. This study did not include available factors such as contract length and player injury history, with the aim to be a purely data driven and performance metric evaluation these were left out.

4.5 Potential Improvements

Future research should focus on improving model performance by incorporating more comprehensive datasets that include subjective factors such as player reputation, marketability and potential along with excluded known factors such as injury history and contract length. Additionally, exploring other machine learning techniques and ensemble methods could provide better predictive power and insights into the complex relationships between performance metrics and market value such as boosting or more comprehensive feature engineering. Longitudinal studies that track player performance and market value over multiple seasons could also offer deeper insights into the dynamics of player valuation.

Acknowledgements

I would like to acknowledge and thank my supervisor, John Newell for his continued guidance and feedback throughout the project.

References

- Aditya Ramnath and R. Priya (2024) ‘Football player analysis for identifying best team using machine learning’, International Journal of Advanced Research in Science, Communication and Technology, pp. 248–254. doi:/<https://doi.org/10.48175/ijetir-1245>{10.48175/ijetir-1245}.
- Biau, G. (2012) Analysis of a random forests model, arXiv.org. Available at: <https://doi.org/10.48550/arXiv.1005.0208> (Accessed: 18 July 2024).
- Breiman, L. (2001) Machine Learning, 45(1), pp. 5–32. doi:/<https://doi.org/10.1023/a:1010933404324>{10.1023/a:1010933404324}.
- Casal, C.A. et al. (2021) ‘Multivariate exploratory comparative analysis of LaLiga teams: Principal Component Analysis’, International Journal of Environmental Research and Public Health, 18(6), p. 3176. doi:/<https://doi.org/10.3390/ijerph18063176>{10.3390/ijerph18063176}.
- de Paula Oliveira, T. and Newell, J. (2024) ‘A hierarchical approach for evaluating athlete performance with an application in elite basketball’, Scientific Reports, 14(1). doi:/<https://doi.org/10.1038/s41598-024-51232-2>{10.1038/s41598-024-51232-2}.
- Duch, J., Waitzman, J.S. and Amaral, L.A. (2010) ‘Quantifying the performance of individual players in a team activity’, PLoS ONE, 5(6). doi:/<https://doi.org/10.1371/journal.pone.0010937>{10.1371/journal.pone.0010937}.
- Football statistics and history (no date) FBref.com. Available at: <https://fbref.com/en/> (Accessed: 18 July 2024).
- Frick, B. (2007) ‘The football players’ labor market: Empirical evidence from the major European Leagues’, Scottish Journal of Political Economy, 54(3), pp. 422–446. doi:/<https://doi.org/10.1111/j.1467-9485.2007.00423.x>{10.1111/j.1467-9485.2007.00423.x}.
- Groll, A. et al. (2018) Prediction of the FIFA World Cup 2018 - a random forest approach with an emphasis on estimated team ability parameters, arXiv.org. Available at: <https://doi.org/10.48550/arXiv.1806.03208> (Accessed: 18 July 2024).
- He, M., Cachorro, R., & Knobbe, A.J. (2015). Football Player’s Performance and Market Value.
- Jishnu, V.B. et al. (2022) ‘Football player transfer value prediction using advanced statistics and FIFA 22 data’, 2022 IEEE 19th India Council International Conference (INDICON) [Preprint]. doi:/<https://doi.org/10.1109/indicon56171.2022.10040117>{10.1109/indicon56171.2022.10040117}.
- Komorowski, M. et al. (2016) ‘Exploratory Data Analysis’, Secondary Analysis of Electronic Health Records, pp. 185–203. doi:/https://doi.org/10.1007/978-3-319-43742-2_15{10.1007/978-3-319-43742-2_15}.
- Lotfi, C., Srinivasan, S., Ertz, M., Latrous, I. (2021). Web scraping techniques and applications: A literature review. SCRS CONFERENCE PROCEEDINGS ON INTELLIGENT SYSTEMS, 381–394. doi:10.52458/978-93-91842-08-6-38}
- M. Merzah, B., S. Croock, M. and N. Rashid, A. (2024) ‘Intelligent classifiers for Football Player Performance Based on Machine Learning Models’, International journal of electrical and computer engineering systems, 15(2), pp. 173–183. doi:/<https://doi.org/10.32985/ijeces.15.2.6>{10.32985/ijeces.15.2.6}.
- Majewski, S. (2016) ‘Identification of factors determining market value of the most valuable football

- players', *Journal of Management and Business Administration. Central Europe*, 24(3), pp. 91–104. doi:/<https://doi.org/10.7206/jmba.ce.2450-7814.177>} {10.7206/jmba.ce.2450-7814.177}.
- Mann, J.B. et al. (2022) 'Momentum, rather than velocity, is a more effective measure of improvements in Division IA Football Player Performance', *Journal of Strength and Conditioning Research*, 36(2), pp. 551–557. doi:/<https://doi.org/10.1519/jsc.0000000000004206>} {10.1519/jsc.0000000000004206}.
- Maćkiewicz, A. and Ratajczak, W. (1993) 'Principal Components Analysis (PCA)', *Computers & Geosciences*, 19(3), pp. 303–342. doi:/[https://doi.org/10.1016/0098-3004\(93\)90090-r](https://doi.org/10.1016/0098-3004(93)90090-r)} {10.1016/0098-3004(93)90090-r}.
- McCormack, S. et al. (2021) 'Coaches' assessment of players physical performance: Subjective and objective measures are needed when profiling players', *European Journal of Sport Science*, 22(8), pp. 1177–1187. doi:/<https://doi.org/10.1080/17461391.2021.1956600>} {10.1080/17461391.2021.1956600}.
- McHale, I.G., Scarf, P.A. and Folker, D.E. (2012) 'On the development of a soccer player Performance Rating System for the English Premier League', *Interfaces*, 42(4), pp. 339–351. doi:/<https://doi.org/10.1287/inte.1110.0589>} {10.1287/inte.1110.0589}.
- McIntosh, S., Kovalchik, S. and Robertson, S. (2019) 'Comparing subjective and objective evaluations of player performance in Australian Rules Football', *PLOS ONE*, 14(8). doi:/<https://doi.org/10.1371/journal.pone.0220901>} {10.1371/journal.pone.0220901}.
- Morciano, G. et al. (2022) 'Machine learning prediction of the expected performance of football player during training', 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE) [Preprint]. doi:/<https://doi.org/10.1109/metroxraine54828.2022.9967621>} {10.1109/metroxraine54828.2022.9967621}.
- Newell, J. et al. (2006) 'Using variable reduction techniques and tolerance intervals to summarise a fitness testing battery in soccer', *Journal of Quantitative Analysis in Sports*, 2(4). doi:/<https://doi.org/10.2202/1559-0410.1037>} {10.2202/1559-0410.1037}.
- O'Donoghue, P. (2008) 'Principal components analysis in the selection of key performance indicators in sport', *International Journal of Performance Analysis in Sport*, 8(3), pp. 145–155. doi:/<https://doi.org/10.1080/24748668.2008.11868456>} {10.1080/24748668.2008.11868456}.
- Pantzalis, V.C. and Tjortjis, C. (2020) 'Sports analytics for football league table and player performance prediction', 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA) [Preprint]. doi:/<https://doi.org/10.1109/iisa50023.2020.9284352>} {10.1109/iisa50023.2020.9284352}.
- Patnaik, D. et al. (2019) 'A study of prediction models for football player valuations by Quantifying Statistical and economic attributes for the global transfer market', 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) [Preprint]. doi:/<https://doi.org/10.1109/icscan.2019.8878843>} {10.1109/icscan.2019.8878843}.
- Patro, S.G.K. and Sahu, K.K. (2015) 'Normalization: A preprocessing stage', *IARJSET*, pp. 20–22. doi:/<https://doi.org/10.17148/iarjset.2015.2305>} {10.17148/iarjset.2015.2305}.
- Pino-Ortega, J. et al. (2021) 'Training design, performance analysis, and talent identification—a systematic review about the most relevant variables through the principal component analysis in soccer, basketball, and rugby', *International Journal of Environmental Research and Public Health*, 18(5), p. 2642. doi:/<https://doi.org/10.3390/ijerph18052642>} {10.3390/ijerph18052642}.
- Reed, G.F., Lynn, F. and Meade, B.D. (2002) 'Use of coefficient of variation in assessing variability of quantitative assays', *Clinical and Vaccine Immunology*, 9(6), pp. 1235–1239. doi:/<https://doi.org/10.1128/cdli.9.6.1235-1239.2002>} {10.1128/cdli.9.6.1235-1239.2002}.
- Rodionova, O., Kucheryavskiy, S. and Pomerantsev, A. (2021) 'Efficient tools for principal component analysis of complex data—a tutorial', *Chemometrics and Intelligent Laboratory Systems*, 213, p. 104304. doi:/<https://doi.org/10.1016/j.chemolab.2021.104304>} {10.1016/j.chemolab.2021.104304}.
- Rojas-Valverde, D. et al. (2020) 'A systematic review of methods and criteria standard proposal for the use of principal component analysis in team's sports science', *International Journal of Environmental Research and Public Health*, 17(23), p. 8712. doi:/<https://doi.org/10.3390/ijerph17238712>} {10.3390/ijerph17238712}.
- Safabun, W. et al. (2020) 'A fuzzy inference system for players evaluation in multi-player sports: The Football Study Case', *Symmetry*, 12(12), p. 2029. doi:/<https://doi.org/10.3390/sym12122029>} {10.3390/sym12122029}.
- Sedgwick, P. (2012) 'Pearson's correlation coefficient', *BMJ*, 345(jul04 1). doi:/<https://doi.org/10.1136/bmj.e4483>} {10.1136/bmj.e4483}.
- Singh, P. and Lamba, P.S. (2019) 'Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players', *Journal of Discrete Mathematical Sciences and Cryptography*, 22(2), pp. 113–126. doi:/<https://doi.org/10.1080/09720529.2019.161080>}.

- 1576333} {10.1080/09720529.2019.1576333}.
- Song, M.-K. et al. (2013) ‘Composite variables’, *Nursing Research*, 62(1), pp. 45–49. doi:/<https://doi.org/10.1097/nnr.0b013e3182741948>} {10.1097/nnr.0b013e3182741948}.
- UmaMaheswari, P. and Rajaram, M. (2009) ‘A novel approach for mining association rules on sports data using Principal Component Analysis: For Cricket Match Perspective’, 2009 IEEE International Advance Computing Conference [Preprint]. doi:/<https://doi.org/10.1109/iadcc.2009.4809163>} {10.1109/iadcc.2009.4809163}.
- Yi, Q. et al. (2020) ‘Evaluation of the technical performance of football players in the UEFA Champions League’, *International Journal of Environmental Research and Public Health*, 17(2), p. 604. doi:/<https://doi.org/10.3390/ijerph17020604>} {10.3390/ijerph17020604}.
- Ćwiklinski, B., Giełczyk, A. and Choraś, M. (2021) ‘Who will score? A machine learning approach to supporting football team building and transfers’, *Entropy*, 23(1), p. 90. doi:/<https://doi.org/10.3390/e23010090>} {10.3390/e23010090}.

5 Supplementary

Response	Mean	SD	CV	Response	Mean	SD	CV	Response	Mean	SD	CV
Aer.	54.90	13.51	0.25	LongPassCmp	74.19	62.69	0.85	Sh.FK	0.27	1.17	4.39
AerL	23.32	14.62	0.63	MedPass.	84.50	8.33	0.10	Sh.p90	0.58	0.38	0.66
AerW	32.91	25.28	0.77	MedPassA	462.18	303.87	0.66	ShBlocked	10.99	9.94	0.90
Age	26.02	3.93	0.15	MedPassCmp	401.64	284.39	0.71	ShortPass.	88.87	4.11	0.05
Ast	0.85	1.53	1.79	Mis	13.45	10.23	0.76	ShortPassA	413.84	252.55	0.61
Blocks	24.00	14.93	0.62	OG	0.18	0.43	2.37	ShortPassCmp	372.05	233.91	0.63
CK	5.13	21.31	4.15	Off	1.11	1.74	1.57	SoT	2.59	2.46	0.95
Carries	680.28	429.93	0.63	OutCK	2.33	10.28	4.41	SoT.p90	0.15	0.16	1.05
CarryProgDist	1832.27	1269.48	0.69	PA.Carries	2.69	4.36	1.62	StCK	0.12	1.18	10.26
CarryTotDist	3338.44	2149.41	0.64	PA.Crosses	3.87	5.61	1.45	Switches	6.79	8.68	1.28
Clr	62.12	41.22	0.66	PA.Passes	10.41	13.17	1.27	Tackled	6.44	7.48	1.16
Crs.x	27.64	48.21	1.74	PKcon	0.37	0.65	1.75	TacklesA	32.73	19.51	0.60
Dead.Passes	105.32	103.08	0.98	PKwon	0.07	0.25	3.75	TacklesW.x	19.32	11.77	0.61
Dis	6.85	6.84	1.00	Pass.	80.16	7.35	0.09	Throws	74.70	84.70	1.13
DribA	16.74	16.70	1.00	PassA.x	1056.53	636.73	0.60	Thruballs	1.33	2.72	2.05
DribCmp	8.00	8.04	1.00	PassBlocked.x	15.87	13.41	0.85	Tkl.Int	54.38	29.89	0.55
Err	0.90	1.04	1.16	PassCmp.x	862.15	548.51	0.64	TklAtt3rd	2.52	2.69	1.07
F3.Carries	16.52	15.42	0.93	PassOff	3.01	2.53	0.84	TklDef3rd	19.87	12.20	0.61
F3.Passess	63.08	48.18	0.76	PassProgDist	5566.08	3889.21	0.70	TklMid3rd	10.35	6.96	0.67
Fld	12.61	11.01	0.87	PassTotDist	15293.64	10341.65	0.68	Touches	1248.38	709.57	0.57
Fls	16.17	10.51	0.65	PassesFK	23.15	23.00	0.99	Touches.Att3	183.98	181.73	0.99
G.xG	-0.28	0.79	-2.83	Past	12.39	9.25	0.75	Touches.AttPA	22.04	16.86	0.76
GCA	2.21	2.71	1.23	Prog.Carries	22.10	20.95	0.95	Touches.Def3	521.13	335.74	0.64
GCA90	0.12	0.12	1.07	Prog.Passes	67.82	48.38	0.71	Touches.DefPA	124.78	99.07	0.79
GCADef	0.03	0.20	6.74	Prog.Rec	36.89	44.42	1.20	Touches.Live	1248.38	709.57	0.57
GCADrib	0.05	0.25	4.64	R.Crd	0.07	0.26	3.58	Touches.Mid3	551.27	347.54	0.63
GCAFld	0.08	0.28	3.29	RecA	771.02	491.91	0.64	X1v1.	59.07	12.83	0.22
GCAPassDead	0.16	0.58	3.53	Recov	102.82	57.67	0.56	X1v1A	29.46	18.95	0.64
GCAPassLive	1.71	2.21	1.29	SCA	24.32	22.16	0.91	X1v1W	17.07	10.87	0.64
GCASh	0.16	0.40	2.46	SCA90	1.30	0.82	0.63	X2Y.Crd	0.03	0.17	5.67
Gls	0.66	0.83	1.26	SCADef	0.39	0.66	1.70	X90s	18.77	9.56	0.51
InCK	2.09	9.24	4.42	SCADrib	0.55	1.13	2.06	Y.Crd	3.33	2.36	0.71
Int.x	21.65	12.75	0.59	SCAFld	0.70	0.98	1.40	npxG	0.94	0.84	0.89
Key.Passes	10.03	12.89	1.29	SCAPassDead	2.58	8.55	3.31	xA	1.02	1.41	1.38
Live.Passes	948.19	588.09	0.62	SCAPassLive	18.76	14.83	0.79	xAG	1.07	1.60	1.50
LongPass.	52.31	12.21	0.23	SCASh	1.34	1.61	1.20	xG	0.94	0.84	0.89
LongPassA	138.41	109.95	0.79	Sh	10.44	8.44	0.81	NA	NA	NA	NA

Response	Mean	SD	CV	Response	Mean	SD	CV	Response	Mean	SD	CV
Aer.	44.22	15.31	0.35	MedPass.	83.66	7.20	0.09	Sh.p90	1.24	0.70	0.56
AerL	19.31	16.13	0.84	MedPassA	313.37	235.61	0.75	ShBlocked	5.02	4.43	0.88
AerW	17.56	17.50	1.00	MedPassCmp	266.92	210.25	0.79	ShortPass.	87.26	4.34	0.05
Age	25.47	3.91	0.15	Mis	26.96	18.57	0.69	ShortPassA	406.55	265.03	0.65
Ast	2.08	2.56	1.23	OG	0.06	0.24	3.97	ShortPassCmp	357.93	240.73	0.67
Blocks	24.09	14.48	0.60	Off	1.63	2.72	1.67	SoT	7.30	7.39	1.01
CK	13.05	27.33	2.09	OutCK	4.28	10.77	2.52	SoT.p90	0.36	0.27	0.75
Carries	602.80	399.93	0.66	PA.Carries	6.33	8.39	1.33	StCK	0.29	0.98	3.33
CarryProgDist	1423.13	1038.80	0.73	PA.Crosses	3.12	3.96	1.27	Switches	7.00	9.30	1.33
CarryTotDist	3012.00	2047.81	0.68	PA.Passes	17.59	17.65	1.00	Tackled	13.71	12.28	0.90
Clr	20.60	17.84	0.87	PK	0.17	0.69	4.16	Tackled.	36.39	14.46	0.40
Crs.x	32.89	46.82	1.42	PKA	0.21	0.78	3.70	TacklesA	38.71	25.36	0.66
Dead.Passes	48.07	64.89	1.35	PKcon	0.14	0.39	2.74	TacklesW.x	22.43	14.46	0.64
Dis	20.17	14.99	0.74	PKwon	0.15	0.45	3.01	Throwins	8.60	19.52	2.27
Drib.	50.00	14.76	0.30	Pass.	79.74	6.42	0.08	Thruballs	4.70	5.80	1.23
DribA	35.52	28.70	0.81	PassA.x	877.96	600.77	0.68	Tkl.Int	56.72	38.15	0.67
DribCmp	17.40	14.27	0.82	PassBlocked.x	18.50	12.19	0.66	TklAtt3rd	5.50	4.31	0.78
Err	0.45	0.77	1.71	PassCmp.x	709.31	508.49	0.72	TklDef3rd	16.14	11.80	0.73
F3.Carries	24.98	19.70	0.79	PassOff	3.31	3.14	0.95	TklMid3rd	17.08	11.91	0.70
F3.Passes	70.70	56.52	0.80	PassProgDist	3257.73	2570.49	0.79	Touches	1076.83	687.24	0.64
Fld	21.52	16.37	0.76	PassTotDist	11651.14	8970.52	0.77	Touches.Att3	300.50	229.79	0.76
Fls	23.51	14.68	0.62	PassesFK	19.85	25.81	1.30	Touches.AttPA	35.08	31.78	0.91
G.xG	-0.06	1.50	-24.89	Past	20.66	12.99	0.63	Touches.Def3	224.94	159.97	0.71
GCA	5.05	5.02	0.99	Prog.Carries	29.68	25.72	0.87	Touches.DefPA	40.53	33.22	0.82
GCA90	0.25	0.19	0.77	Prog.Passes	90.13	68.21	0.76	Touches.Live	1076.62	686.95	0.64
GCADef	0.12	0.37	3.08	Prog.Rec	57.49	56.31	0.98	Touches.Mid3	563.50	384.26	0.68
GCADrib	0.26	0.64	2.49	RCrd	0.08	0.30	3.66	X	212.58	120.16	0.57
GCAId	0.29	0.54	1.91	RecA	716.93	491.16	0.69	X1v1.	43.32	12.54	0.29
GCAPassDead	0.34	0.86	2.54	Recov	113.71	69.82	0.61	X1v1A	36.51	22.77	0.62
GCAPassLive	3.68	4.04	1.10	SCA	50.41	40.29	0.80	X1v1W	15.85	11.21	0.71
GCASh	0.37	0.68	1.84	SCA90	2.59	1.14	0.44	X2YCrD	0.05	0.21	4.62
Gls	2.21	2.83	1.28	SCADef	0.95	1.15	1.21	X90s	18.57	10.43	0.56
InCK	5.88	12.28	2.09	SCADrib	2.25	2.89	1.28	YCrd	3.86	3.04	0.79
Int.x	18.02	14.32	0.79	SCAFld	2.10	2.42	1.16	npG.xG	-0.05	1.57	-29.33
Key.Passes	22.02	20.94	0.95	SCAPassDead	4.89	10.17	2.08	npxG	2.10	2.12	1.01
Live.Passes	826.59	564.19	0.68	SCAPassLive	37.61	29.00	0.77	xA	1.96	2.10	1.07
LongPass.	59.71	12.78	0.21	SCAsh	2.62	2.58	0.98	xAG	2.32	2.59	1.12
LongPassA	101.50	92.41	0.91	Sh	23.64	20.35	0.86	xG	2.27	2.40	1.06
LongPassCmp	61.06	59.48	0.97	Sh.FK	18 ⁹	3.01	2.34	NA	NA	NA	NA

Response	Mean	SD	CV	Response	Mean	SD	CV	Response	Mean	SD	CV
Aer.	34.45	13.51	0.39	MedPass.	72.62	9.30	0.13	Sh.p90	2.27	0.78	0.35
AerL	31.23	26.11	0.84	MedPassA	129.55	103.78	0.80	ShBlocked	1.55	1.91	1.23
AerW	19.17	20.47	1.07	MedPassCmp	96.83	82.31	0.85	ShortPass.	80.27	7.90	0.10
Age	25.25	4.08	0.16	Mis	40.37	26.47	0.66	ShortPassA	237.64	163.48	0.69
Ast	2.43	2.67	1.10	OG	0.02	0.13	7.75	ShortPassCmp	196.25	143.67	0.73
Blocks	14.88	10.21	0.69	Off	7.10	6.12	0.86	SoT	14.16	11.40	0.81
CK	10.02	22.86	2.28	OutCK	1.95	6.22	3.19	SoT.	36.42	11.51	0.32
Carries	396.07	276.72	0.70	PA.Carries	19.27	18.56	0.96	SoT.p90	0.84	0.40	0.48
CarryProgDist	1155.56	949.11	0.82	PA.Crosses	3.02	4.18	1.39	StCK	0.21	0.76	3.51
CarryTotDist	2463.02	1902.43	0.77	PA.Passes	16.40	15.58	0.95	Switches	2.95	4.27	1.45
Clr	10.70	9.58	0.89	PK	0.43	1.16	2.70	Tackled	23.30	17.96	0.77
Crs.x	32.70	46.28	1.42	PKA	0.59	1.46	2.49	Tackled.	45.42	12.76	0.28
Dead.Passes	28.68	39.28	1.37	PKcon	0.07	0.25	3.77	TacklesA	17.80	13.43	0.75
Dis	26.02	18.15	0.70	PKwon	0.36	0.66	1.81	TacklesW.x	10.36	8.22	0.79
Drib.	41.31	11.03	0.27	Pass.	71.00	7.61	0.11	Throwins	6.49	9.97	1.54
DribA	51.65	39.16	0.76	PassA.x	457.31	330.09	0.72	Thruballs	2.98	3.13	1.05
DribCmp	21.28	16.65	0.78	PassBlocked.x	17.36	13.36	0.77	Tkl.Int	23.74	18.69	0.79
Err	0.09	0.32	3.48	PassCmp.x	331.64	249.71	0.75	TklAtt3rd	4.86	4.66	0.96
F3.Carries	25.36	20.31	0.80	PassOff	2.07	2.13	1.03	TklDef3rd	5.93	5.94	1.00
F3.Passes	20.75	17.12	0.83	PassProgDist	1165.72	1056.11	0.91	TklMid3rd	7.02	5.05	0.72
Fld	22.27	17.89	0.80	PassTotDist	4817.05	3926.94	0.82	Touches	649.02	430.45	0.66
Fls	18.64	12.06	0.65	PassesFK	4.43	10.94	2.47	Touches.Att3	339.91	248.58	0.73
G.xG	-0.05	2.21	-41.82	Past	11.62	8.87	0.76	Touches.AttPA	75.45	55.26	0.73
GCA	5.98	5.61	0.94	Prog.Carries	42.29	36.32	0.86	Touches.Def3	72.10	52.96	0.73
GCA90	0.34	0.22	0.66	Prog.Passes	42.59	35.31	0.83	Touches.DefPA	14.83	12.10	0.82
GCADef	0.11	0.31	2.89	Prog.Rec	117.17	90.08	0.77	Touches.Live	648.43	430.00	0.66
GCADrib	0.67	0.96	1.44	RCrd	0.03	0.18	5.43	Touches.Mid3	246.83	153.17	0.62
GCAFld	0.47	0.88	1.86	RecA	467.79	313.23	0.67	X	239.37	134.09	0.56
GCAPassDead	0.26	0.77	2.92	Recov	58.91	39.77	0.68	X1v1A	18.15	13.42	0.74
GCAPassLive	3.82	3.77	0.99	SCA	47.32	35.95	0.76	X1v1W	6.53	5.35	0.82
GCASh	0.65	1.05	1.62	SCA90	2.80	1.02	0.36	X2YCrD	0.01	0.09	11.00
Gls	5.17	5.74	1.11	SCADef	0.63	0.85	1.35	X90s	15.91	9.31	0.59
Gls.pSh	0.11	0.07	0.66	SCADrib	4.23	4.24	1.00	YCrd	2.17	2.03	0.93
InCK	5.66	14.79	2.61	SCAFld	3.31	3.39	1.03	npG.xG	-0.03	2.14	-78.42
Int.x	5.93	6.18	1.04	SCAPassDead	2.97	7.06	2.38	npxG	4.76	4.09	0.86
Key.Passes	20.37	16.67	0.82	SCAPassLive	31.91	23.91	0.75	npxG.pSh	0.13	0.05	0.36
Live.Passes	426.55	300.44	0.70	SCASh	4.28	3.54	0.83	xA	1.90	1.76	0.92
LongPass.	53.44	17.12	0.32	Sh	37.07	26.55	0.72	xAG	2.46	2.23	0.91
LongPassA	41.97	50.67	1.21	Sh.Dist	1951	2.73	0.18	xG	5.22	4.88	0.93
LongPassCmp	21.40	24.03	1.12	Sh.FK	0.69	1.68	2.45	NA	NA	NA	NA

