Dean&Francis
ISSN 2959-6157

The Value of Players in the Top Five Football Leagues for the 24-25 Season Based on Multiple Factors

Xiang Fang

College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350000, China
Corresponding author: 3226101148@stu.fafu.edu.cn

Abstract:

The top five football leagues, as one of the most high-profile sporting events, entail enormous commercial value, making player valuation in these leagues a focal point. Analyzing the influencing factors of transfer values helps clubs formulate strategies and promotes the development of the football market. This study leverages descriptive statistics and the XGBoost regression model, analyzing a Kaggle dataset of 2024-25 season player performance. Key findings show 20-28-year-olds command higher transfer values, with a general decline after 30. The Premier League leads in player valuation, with other leagues 40-60% lower. Attacking positions far outpace defensive roles in market price. The eXtreme Gradient Boosting (XGBoost) weight analysis for four positions—forwards, midfielders, defenders, and goalkeepers—reveals that offensive-type data tailored to each position (such as scoring opportunities created for forwards, average carrying distance for defenders, etc.) are critical determinants of player transfer values. Specifically, defensive players now require offensive integration to achieve high valuation, with key factors including forward ball carries and long-pass organization. Traditional defensive metrics have lost weight in valuation models, while offensive attributes have emerged as primary value drivers.

Keywords: Top Five Football Leagues; Multiple Factors; XGBoost

1. Introduction

With the vigorous development of the global sports industry, professional football—particularly the five major European leagues—has demonstrated enormous commercial and social influence. The valuation of football players has emerged as a focal point of widespread attention, carrying profound significance. The total market value of a team's players can effectively impact match outcomes, fully highlighting the critical role of player value in a team's developmental process [1]. Analyzing these influencing factors not

ISSN 2959-6157

only assists clubs in formulating strategies and evaluating investments but also plays a pivotal role in promoting the healthy development of the football market, especially in the context of club transfers and market investments.

Existing research has shown that factors such as age, club affiliation, and league characteristics exert varying degrees of influence on player valuation [2]. For instance, Ma's findings indicate that age, scoring ability, organizational skills, and offensive prowess are core determinants of a football player's value, with valuation typically peaking between the ages of 25 and 28 [3]. Building on this, Liao et al. have further highlighted that key influencing factors differ by position: goalkeepers' value hinges on shot-stopping skills, while midfielders rely on creative organization [4]. A separate study by Rehemjiang explores the correlation between the relative age effect and player valuation, revealing a significant positive correlation between the number of players participating in international A-level competitions and valuation, whereas the relative age effect exhibits a negative correlation [5].

Some research has focused on the impact of commercial value and transfer systems on player valuation. In terms of commercial value, an athlete's brand appeal is shaped by the popularity of their sport and the maturity of its commercial development. Athletes in high-profile sports often secure more lucrative endorsement contracts, directly influencing their market value [6]. Regarding transfer systems, Lucifora and Simmons' study on the Italian league has shown that a player's early-career salary level significantly affects long-term valuation [7]. Further research on professional football leagues indicates that imperfections in the current transfer system have posed numerous challenges to player valuation [8].

Although previous studies have explored football player valuation multidimensionally, they lack a systematic analysis of influencing factors for the 2024-2025 season and a position-specific refined valuation system. This study integrates multi-source data, uses descriptive analysis and Extreme Gradient Boosting (XGBoost) model to construct evaluation systems for different positions from in-game performance and off-field factors, identifies core valuation elements, and establishes a scientific dynamic valuation framework.

2. Methods

2.1 Data Source and Introduction

The player data used in this article comes from the five major European leagues in the 2024-2025 season. The data comes from the Kaggle website, which contains 2,793 observational data points and 268 variables in the format. CVS [9]. This data set provides key information for exploring the factors that affect the value of professional football players. Due to the presence of missing values for some players in the dataset, these missing values were removed during the data analysis process, and ultimately, 2,335 observations were used. When selecting variables, considering the four positions under study, combining multiple factors of data quality and referencing relevant literature, a total of 54 variables were selected, including the number of goals scored per game by players (Gls), the number of assists per game (Ast), the number of penalty kicks per game (PK), and the number of shots per game (Sh). Each variable has a different range of values and reflects different performance data of players in games.

2.2 Method Introduction

XGBoost suits studying player value-influencing factors. It handles multi-factor, heterogeneous data and nonlinear couplings, solves high-dimensional feature overfitting via model complexity regularization, ensures generalization, and its feature importance ranking quantifies contributions, aiding clubs in recruitment and salary talks.

As a gradient - boosting - based ensemble algorithm, XGBoost starts from raw input data, sequentially builds decision tree weak learners (each trained on the prior tree's errors, outputting function results). Summing these gives prediction y, realizing boosting (combining weak learners into a strong one). The model, composed of decision trees and prediction, trains by optimizing the objective function with loss and regularization terms via the gradient-boosting framework.

In this study, 2,335 players were divided into strikers, midfielders, defenders, and goalkeepers according to their position on the court. All player data included the player's basic information variable indicators (outfield factors) and the player's season data variable indicators (on-field factors). Descriptive statistical analysis is used for the basic information data to be related to the player's value. The player's season data is analysed using the XGBoost regression model.

3. Results and Discussion

3.1 Descriptive Analysis

An analysis of player value distribution across four positions in the top five leagues (Fig. 1), where the x-axis represents player market value and the y-axis represents probability density) reveals significant right-skewness and positional disparities, Defenders (DF) mostly cluster in the €0-10M range due to limited offensive contributions and complex evaluation systems, Forwards (FW), driven by goal-scoring ability and high market demand—have the highest values, mostly concentrated at €0-25M with 20% exceeding €30M, Goalkeepers (GK) see most values in the €0-5M range and are overall undervalued due to the implicit nature of defensive impact, while Midfielders (MF), shaped by their dual offensive-defensive roles, show a dispersed distribution with most values at €0-25M and 15% exceeding €30M. These characteristics are collectively influenced by positional responsibilities, quantifiable performance, and market dynamics.

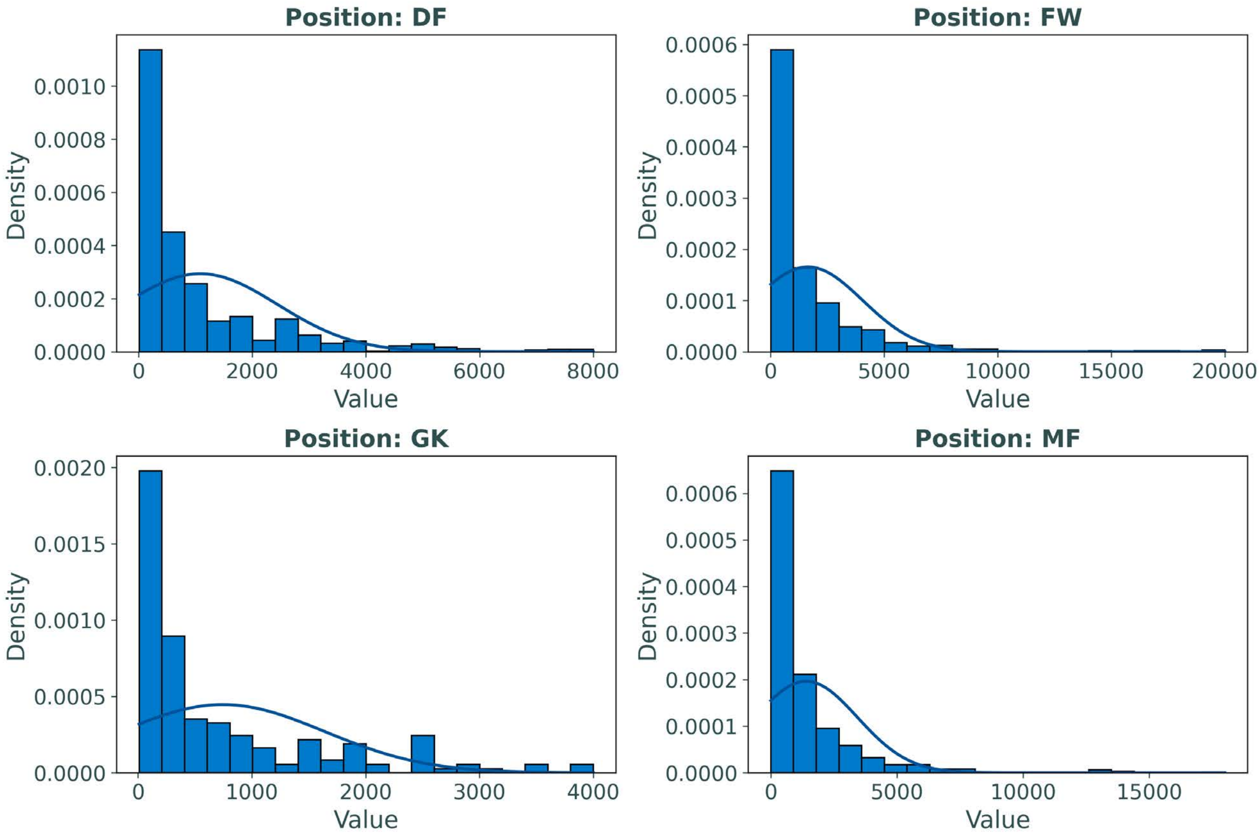

Fig. 1 The histogram of the normal distribution curve of the value column (picture credit: original)

Fig. 2 (the x-axis represents age and the y-axis represents player value) illustrates the age-value relationship of players in four positions across the Top Five Leagues. In the golden age (20-28), defender values concentrate due to mature physique and stable contributions, while midfielder and forward values scatter owing to talent, competition, and media attention. High-value goalkeepers remain rare and sporadic. After 30, some forwards and midfielders maintain high value through strength or experience, but defender values drop sharply due to physical decline. Goalkeepers show few high-value cases with gentle fluctuations. Forward values vary with goal-scoring ability and influence; midfielders rely on tactical adaptability and experience; defenders struggle to preserve value due to functional decline; goalkeepers demonstrate stable but limited growth as age barely impacts performance.

ISSN 2959-6157

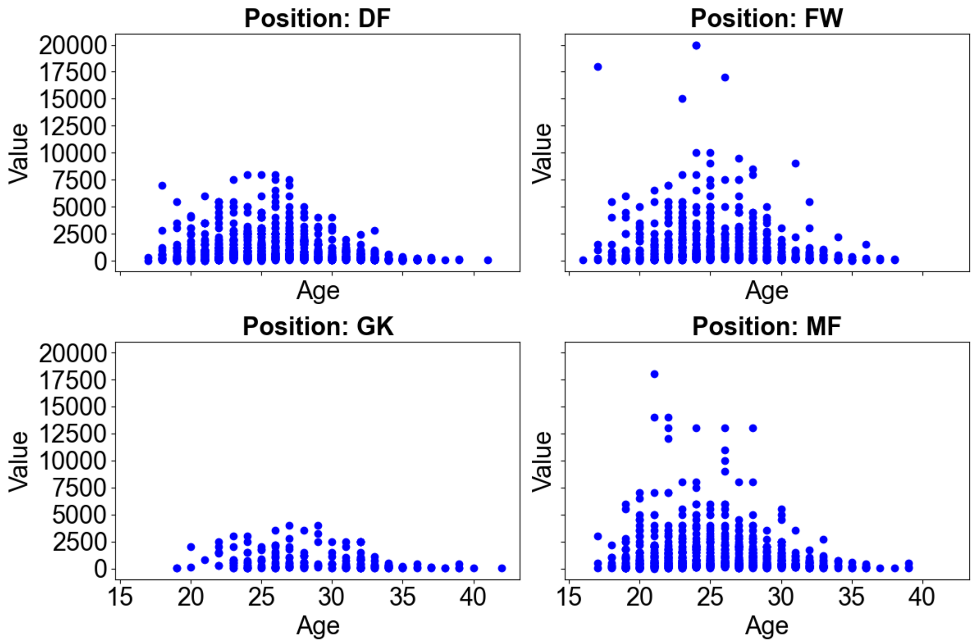

Fig. 2 The scatter map of the relationship between age and value of football players (picture credit: original)

Fig. 3 shows the average player values by position across the top five leagues. The x-axis represents the league, and the y-axis represents the player's average value. Premier League leads overall, with the highest position-specific values: FW (2865), MF (2687), DF (2080). La Liga follows, with FW (1478), MF (1233), and DF (814), driven by elite clubs' investments in key positions but lagging behind the Premier League in overall averages. Bundesliga, Ligue 1, and Serie A have lower average values due to more cautious spending and gaps in commercial revenue/ broadcast shares.

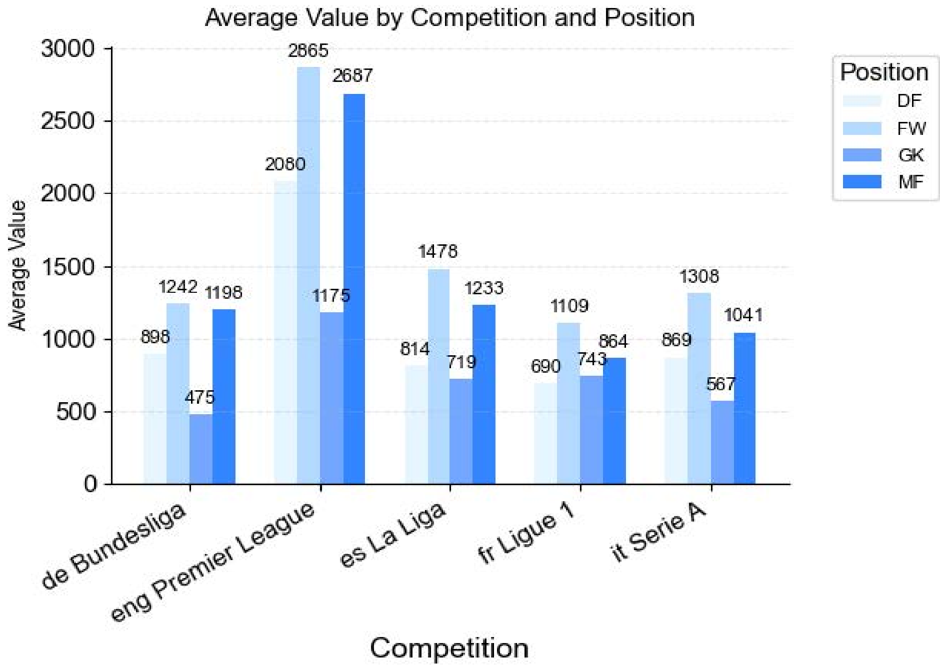

Fig. 3 Bar chart of the average value of football players in different competitions (picture credit: original)

3.2 Xgboost Analysis

The model fitting indicators in Table 1 show that the data for each position has achieved a theoretically good fit—the independent variables can well explain the variation of the dependent variable. This indicates that XGBoost has strong capabilities in mining the relationship between features and target variables, and its feature combinations can fully explain the target variables.

In this paper, using the complete dataset as the training set for variable importance calculation is feasible because the core of this process lies in weight allocation rather than prediction accuracy: the complete dataset allows the model to learn all features, deeply excavate the relationships between features, and conduct more accurate evaluations. At the same time, it avoids the distortion of importance judgments caused by feature distribution bias resulting from sampling.

Table 1. Fitting metrics for models of four player positions

Term	RMSE	MAE	MAPE
FW	8.704	5.497	1.746
DF	19.301	12.763	4.308
MF	15.792	10.446	3.248
GK	2.279	0.689	0.16

From Fig.4, the same axis representations apply to subsequent Figs. 5-7. The x-axis represents importance, and the y-axis represents metrics. Carries per Game (21.9%) are the most critical. Excellent dribbling enables defenders to initiate attacks, break through pressure, and create opportunities. TotDist (Total Distance, 11.9%) ranks second. A longer running distance reflects defensive coverage and support for transition. Secondary factors include PKwon (Penalty Kick Win, 7.4%), SCA (Shot Creating Action, 6.2%), and Att Pen touches (Attacking Penalty Area, 6.0%). Traditional defensive stats like recoveries (4.1%), clearances (4.1%), and blocks (2.5%) play relatively minor roles. Currently, in evaluating defenders' market value, besides considering defensive responsibilities, offensive involvement (such as build-up play in attacks) is also prioritized.

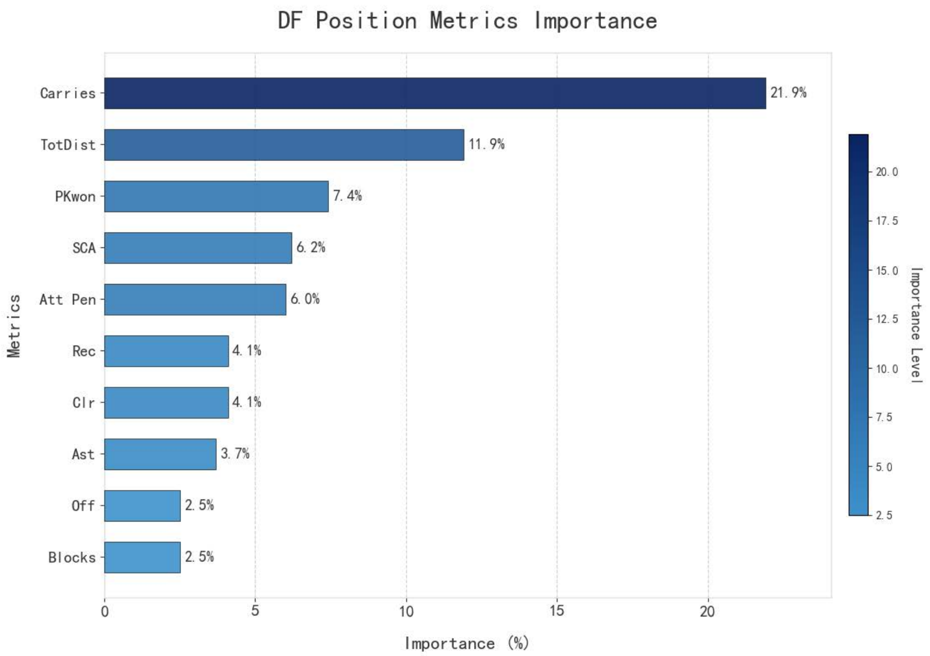

Fig.4 Bar chart of metrics importance for DF position (TOP 10) (picture credit: original)

Regarding the market value influencers of strikers (as shown in the top 10 XGBOOST factors in Fig. 5), attacking penalty area touches (Att Pen, 41.2%) emerge as the dominant determinant. As dangerous-area involvement defines scoring threat. Shots per Game (16.4%) rank second, directly linking shooting volume to goal impact. Secondary factors: assists (5.5%), goals (4.2%), ball control (4.1%), progressive movement (3.8%); defensive penalty

ISSN 2959-6157

area contributions (Def Pen, 3.4%) and PKcon(Concede a Penalty 1.8%) play minor roles, Core conclusion is the valuation of top - level players depends on their presence in the penalty area, shooting efficiency, and the consistency of organizing attacks.

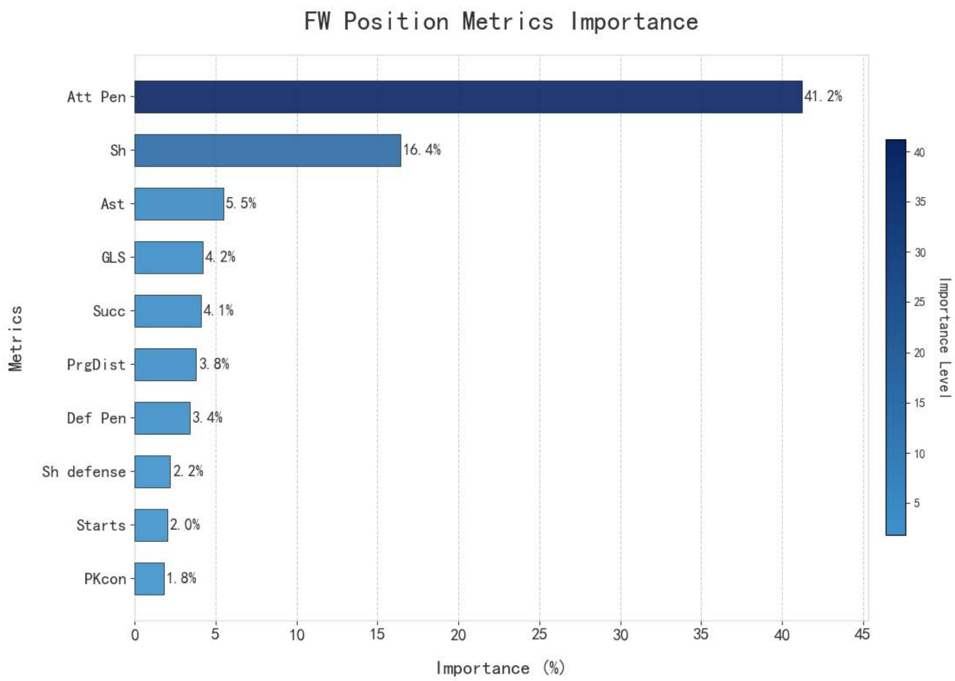

Fig. 5 Bar chart of metrics importance for FW position (TOP 10) (picture credit: original)

Midfielder Market Value Influencers (From Fig. 6 XGBOOST Top 10), SCA (Shot Creating Action) per Game (19.2%) dominates as chance creation defines offensive organizing value. TotDist (19.0%) reflects dual offensive-defensive roles; higher distance signals broader coverage. Secondary factors: Att Pen (9.4%), SoT (Shots on Target 4.7%), recoveries (4.3%); finishing (GLS/Sh, 2.6%/2.3%) and defensive stats (tackles, 3.4%) play smaller roles. The offensive creativity and stamina are core, with defensive/finishing contributions as secondary drivers.

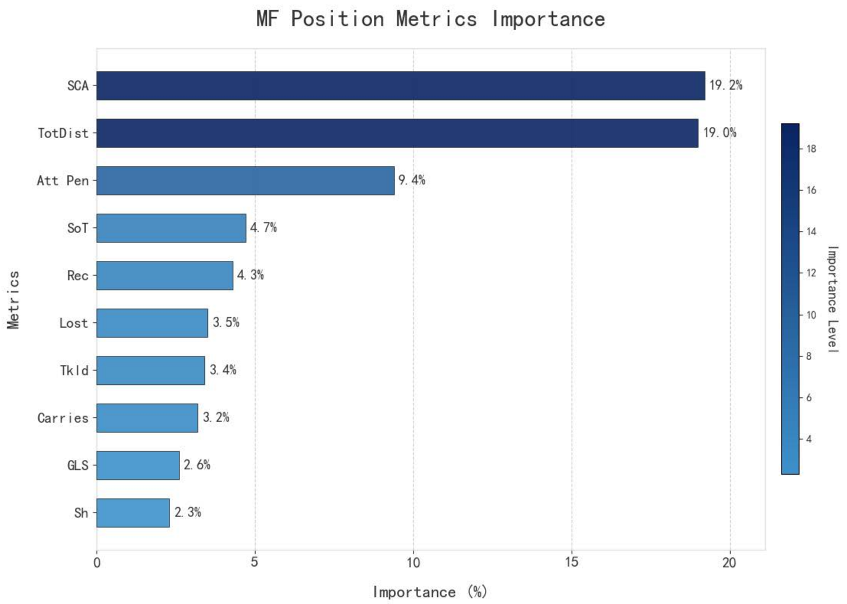

Fig.6 Bar chart of metrics importance for DF position (TOP 10) (picture credit: original)

Goalkeeper Market Value Influencers (From Fig. 7 XG-BOOST Top 10), Avg Kick Distance (17.2%), is pivotal for initiating counterattacks and breaking defenses. Def Pen (12.1%) reflects play organization and threat neutralization. Errors (8.9%) directly impact value due to goal-conceding risks. Secondary factors: carries (7.4%), SCA (7.3%), offsides (5.4%); traditional saves (4.5%) now baseline. Modern valuation of goalkeepers prioritizes build-up passing, defensive organization, and error control over traditional shot-stopping.

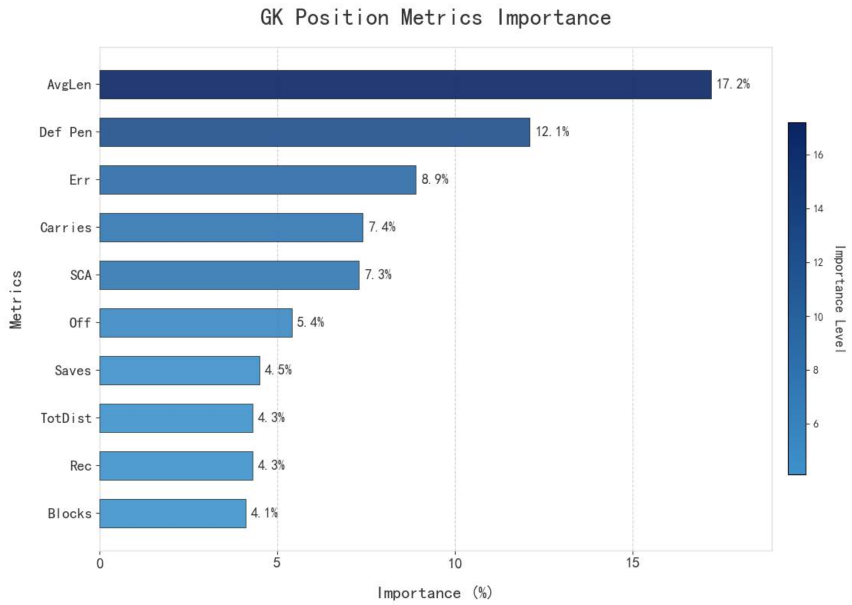

Fig. 7 Bar chart of metrics importance for DF position (TOP 10) (picture credit: original)

4. Discussion

This study still has certain limitations. At the data level, although multi-source league data was integrated, due to limitations in data collection channels, some non-public data (such as players' psychological states and quantitative indicators of team collaboration) could not be included in the analysis, which may affect the completeness of the valuation model. In terms of model application, although the XGBoost machine learning model can effectively screen features, it has weak interpretability for some non-linear complex relationships, and does not fully consider the impact of dynamic factors such as changes in football match rules and the innovation of tactical systems on player value. In addition, the study focuses on the top five European leagues, and the applicability of the valuation of players in other leagues and youth training players remains to be verifiedFuture research can be deepened and expanded in multiple directions. At the technical level, exploring the integration of deep learning and causal inference models is feasible to improve the predictive accuracy for dynamic changes in player values. Meanwhile, natural language processing (NLP) technologies can be used to mine media reports and social media data, enriching the evaluation dimensions of players' off-field influence. Additionally, reference can be made to the ensemble model based on Dempster-Shafer Theory and Fourier Amplitude Sensitivity Testing (FAST) proposed by Shen to further optimize feature selection and model generalization capabilities [10]. In terms of research scope, the research can build a global player valuation model, incorporate data of players from different leagues and different age groups, and enhance the universality of the model. The research should also strengthen interdisciplinary research, integrate economic and psychological theories, and provide more comprehensive theoretical support and decision-making references for the development of the football industry.

5. Conclusion

This study integrates multi-source data, applies descriptive statistics and an XGBoost model to build evaluation systems for different positions from on-field and off-field perspectives, reaching the following conclusions. The value of attack-dominated positions is significantly superior. Offensive positions have higher value due to their direct correlation with match outcomes and commercial appeal. Forwards and midfielders gain an advantage in player value assessment through offensive performances such as creating scoring opportunities and efficient shooting,

ISSN 2959-6157

with their market prices far exceeding those of defensive positions. Defensive positions need to transform to impact high value, breaking through functional limitations and integrating offensive elements to enhance market recognition. Defenders and goalkeepers can improve their value through offensive participation behaviors such as carrying the ball forward and long-pass organization. The weight of traditional defensive data has decreased, and offensive attributes have become new value-added points. Off-field factors also drive value differences.

In terms of the price comparison of the four positions in different leagues, the Premier League leads with its viewing appeal and commercialization; La Liga relies on star players but is slightly weaker; the player values in the Bundesliga, Ligue 1, and Serie A are 40-60% lower.

Regarding age factors, players in all four positions have higher market prices between the ages of 20 and 28. After the age of 30, the value of each position generally declines gradually, but the decline rate varies slightly depending on the position.

References

[1] Yue Quan, Cheng Cheng, Chen Zhihong. The Influence of the Team's Market Value on Sports Performance Based on Least Absolute Deviations Regression and Its Management Implications—Data from the 2020 UEFA European Football Championship. Journal of Sichuan Vocational and Technical College, 2022, 32(6): 112-117.

[2] Yunus Koloğlu, Hasan Birinci, Sevde Ilgaz Kanalmaz, Burhan Özyılmaz. A Multiple Linear Regression Approach for Estimating the Market Value of Football Players in Forward Position. 2018.

[3] Ma Rui. Research on the Market Value of Professional Football Players and Its Influencing Factors. Tianjin University of Finance and Economics, 2020.

[4] Liao Bin, Wang Zhining, Li Min, Sun Ruina. A Method for Predicting the Market Value of Football Players and Analyzing Their Characteristics by Integrating the XGBoost and SHAP Models. Computer Science, 2022, 49(12): 691-702.

[5] Rehemujiang Repukati. A Study on the Correlation between the Relative Age Effect of Football Players and Their Market Value—Taking the Men's Youth Football Teams of 39 Countries as an Example. Beijing Sport University, 2021.

[6] Xu Zhong. Analysis of the Factors Influencing the Commercial Value of Outstanding Athletes in China. Journal of Sports Adult Education, 2014, 30(5): 50-52.

[7] Claudio Lucifora, Rob Simmons. Superstar effect in sport Evidence from Italian soccer. Journal of Sports Economics, 2003, 1(4):35 - 55.

[8] Zhang Lei. A Study on the Current Situation and Development Countermeasures of Domestic Transfers of Players in the Chinese Professional Football League. Beijing Sport University, 2009; Zheng Lu. A Study on the Transfer System of Chinese Professional Football Players. Sports World (Academic Edition), 2010, (5): 104-106.

[9] Shen Qijie. Predicting the value of football players: machine learning techniques and sensitivity analysis based on FIFA and real-world statistical datasets. Applied Intelligence. 2025, 55.

[10] Hubert. Sidorowicz. Players_data-2024_2025(Dataset). 2025-4-25. 2025-4-18. https://www.kaggle.com/datasets/hubertsidorowicz/football-players-stats-2024-2025/data