Construction and Application Research of Big Data Model for Player Value Evaluation in Professional Sports Clubs

Hanwen Liu

Westfield Secondary School, Toronto, Canada

lhw08618@Outlook.com

Keywords: Professional Sports Clubs; Player Value Evaluation; Big Data Model; Indicator System

Abstract: In response to the subjectivity and single-dimensionality of traditional player value evaluation in professional sports clubs, this study constructs a practical evaluation model by combining big data technology. Firstly, clarify the three-dimensional connotations of player value, including competitive, commercial and potential aspects. Follow the principles of relevance, accessibility and quantifiability to establish an index system, and use the simplified analytic hierarchy process to determine the weights. Among them, competitive value accounts for 60%, commercial value for 25%, and potential value for 15%. Then, it designs a three-step data processing flow of collection, cleaning, and standardisation, combined with a hybrid algorithm of multiple linear regression and weighted summation. Verified with data from mainstream leagues, the model's error rate is below 15%, demonstrating validity. The study further applies the model to scenarios such as player recruitment, contract renewal negotiations, and commercial development, and establishes an iterative mechanism of quarterly review, departmental feedback, and environmental adaptation to provide objective support for club decision-making. At the same time, it points out the model's limitations in non-quantitative indicators and adaptation to minor leagues, indicating directions for subsequent optimisation.

1. Introduction

1.1. Research Background

At present, the global professional sports industry is developing towards a high degree of commercialisation, and the commercial value of mainstream leagues such as football and basketball continues to rise. Players, as the most core assets of clubs, have their value assessments directly influencing key decisions such as transfer operations, salary negotiations, and commercial development. However, traditional player value evaluation models excessively rely on coaches' subjective experience, media reputation, or single data (such as goals scored or assists), lacking systematic quantitative standards. They are easily affected by personal preferences and short-term performance, making it difficult to fully reflect a player's comprehensive value. For example, some defensive players may not have outstanding offensive data, yet they are crucial to the team's tactical system, and traditional evaluation often underestimates their value.

The application of big data technology in the sports field is gradually becoming widespread. It can integrate multidimensional information such as official match statistics, club training monitoring data, and social media fan data [1]. This provides technical support to break through the limitations of traditional evaluation and achieve objective quantification of player value, promoting the transformation of professional sports club player evaluation models from experience-driven to data-driven.

1.2. Research Significance

From a theoretical perspective, this study can make up for the deficiencies of traditional player value evaluation methods. Traditional research mostly focuses on players' competitive value, neglecting the correlation between commercial value and potential value, and lacks systematic exploration of the application of big data technology. By constructing an evaluation model covering

  DOI: 10.25236/gemmsd.2025.093

the three dimensions of competitive, commercial, and potential value, and deeply combining big data technology with player value evaluation, this study can improve the theoretical framework of player value evaluation in the field of professional sports, enrich research content in the interdisciplinary field of sports economics and data science, and provide a basic reference for subsequent related studies.

From a practical perspective, the research results can directly serve the operational decision-making of professional sports clubs: in player recruitment, helping clubs accurately judge the cost-effectiveness of target players and avoid expensive signings of short-term performance players; in contract renewal negotiations, determining reasonable salaries based on objective data to reduce labour disputes; in commercial development, quickly identifying players with high commercial value and formulating targeted development strategies. Ultimately, it helps clubs reduce decision-making risks, improve operational efficiency, and enhance competitiveness in the league.

2. Basic Cognition of Player Value and Big Data Evaluation in Professional Sports Clubs

2.1. Connotation and Dimensions of Player Value in Professional Sports Clubs

The value of players in professional sports clubs is a comprehensive reflection of their current contribution and long-term value to the club. The core connotation is the organic combination of the three dimensions of competitive, commercial, and potential value [2].

Competitive value is the foundation, referring to the player's direct or indirect contributions in matches, such as goals, defensive interceptions, and running coverage, which determine the team's competitiveness on the field. Commercial value is the extension, reflected in the player's impact on shirt sales, endorsement cooperation, and fan traffic, which influences the club's commercial revenue. Potential value is the support, related to the player's age, injury history, and room for technical improvement, which determines long-term cost-effectiveness. The three together constitute the complete system of player value, and none can be omitted.

2.2. Core Role of Big Data Technology in Player Value Evaluation

Big data technology provides full-dimensional, objective, and dynamic support for player value evaluation. It can break down data barriers and integrate multi-source information such as official match statistics (e.g., pass completion rate), club training data (e.g., physical fitness reserves), and social media data (e.g., fan activity), avoiding the problem of fragmented data in traditional evaluation [3]. It replaces subjective experience with quantitative indicators, such as using the number of key passes per game instead of vague judgments of strong passing ability, reducing human error. It supports dynamic evaluation by updating data in real time, tracking fluctuations in players' performance (such as competitive performance after injury recovery), making value evaluation shift from static results to dynamic tracking, and better fitting club decision-making needs.

2.3. Basic Requirements of Big Data Evaluation Models

Big data evaluation models for professional sports clubs must meet three basic requirements: practicality, comprehensiveness, and stability [4]. Practicality is the premise; the model should avoid complex algorithms and adopt logic easily understood by club staff (such as weighted summation), and even be converted into an Excel template to lower the application threshold. Comprehensiveness is the core; it must cover the three value dimensions of competitive, commercial, and potential, without omitting key indicators (such as ignoring injury history, which would overestimate potential value). Stability is the guarantee; in different scenarios (such as evaluating forwards and defenders, or league and cup matches), the model's output must maintain reasonable consistency, avoiding excessive evaluation deviation caused by scenario changes, and ensuring that clubs can make stable decisions based on the model.

3. Construction of the Big Data Model for Player Value Evaluation in Professional Sports Clubs

3.1. Construction of the Evaluation Indicator System

The construction of the evaluation indicator system must follow the three principles of relevance, accessibility, and quantifiability to ensure that the indicators fit the actual needs of clubs and are easy to operate. First, the specific indicators are decomposed from the three dimensions of player value: competitive, commercial, and potential. The competitive value dimension prioritises the selection of core data available on official match websites, such as average running distance per game, number of key passes, shot success rate, and number of defensive interceptions, while including the number of matches missed due to injuries as a reverse indicator (the more matches missed, the lower the score). The commercial value dimension selects publicly available statistical indicators, including annual shirt sales, number of social media followers (taking the total from mainstream platforms), and number of commercial endorsements. The potential value dimension focuses on quantifiable growth-related data, such as player age (23 to 28 years set as the golden range for bonus points) and competitive performance growth rate over the last three seasons (calculated based on changes in the average of core competitive indicators).

The indicator weights are determined by a simplified analytic hierarchy process, combined with opinions from the club's transfer, coaching, and marketing departments, determining that competitive value accounts for 60%, commercial value for 25%, and potential value for 15%, which highlights core value while also ensuring comprehensive evaluation.

3.2. Design of the Big Data Processing Flow

The big data processing flow is divided into three steps: collection, cleaning, and standardisation, to ensure data quality and usability. The first step, data collection, clarifies multi-channel sources: match data is obtained in real time from the official league database (such as statistics during live matches), training data is extracted from the club's internal training monitoring system (such as fitness training app data), and commercial data connects with social media platform interfaces (such as Weibo, Douyin follower data) and partner merchant sales systems (such as shirt sales data). The collection frequency is distinguished by data type—match data collected in real time, commercial data updated daily, and potential value data summarised monthly.

The second step, data cleaning, focuses on handling outliers and missing values: removing statistical errors in match data (such as duplicated pass counts), filtering out fake followers in social media data (by screening fan activity), and filling missing data using the average of players in the same position (for example, if a player's single-match data is missing, it is replaced by the average data of players in the same position for that match). The third step, data standardisation, adopts the Z-score standardization method, converting indicators of different units (such as counts, numbers of people, percentages) into unified interval data with a mean of 0 and a standard deviation of 1, avoiding the impact of unit differences on calculation results.

3.3. Selection of the Core Algorithm for the Big Data Evaluation Model

The selection of the core algorithm follows the principles of being simple, interpretable, and easy to implement, avoiding complex deep learning algorithms, and instead adopting a combination of multiple linear regression models and weighted summation. First, the basis for algorithm selection is clear: the multiple linear regression model can verify the correlation between each indicator and player value through statistical analysis (such as verifying whether the average number of goals per game has a significant effect on competitive value), ensuring indicator validity; the weighted summation method can directly combine indicator weights to calculate the comprehensive score, with intuitive logic that is easy for club staff to understand.

The specific operation is divided into two steps: the first step uses training set data (70% of player data) to build the multiple linear regression model, determining the importance of each indicator through regression coefficients and eliminating indicators with low correlation (for example, if player height is not significantly correlated with value, it is deleted). The second step, based on the determined indicators and weights, performs weighted summation of the standardised scores of each dimension indicator to obtain the comprehensive player value score. The formula is simplified as:

Player comprehensive value score = (sum of standardized scores of competitive value dimension indicators × corresponding weights) × 60% + (sum of standardized scores of commercial value dimension indicators × corresponding weights) × 25% + (sum of standardized scores of potential value dimension indicators × corresponding weights) × 15%.

To reduce the application threshold, the algorithm can be embedded into an Excel template, where staff input standardized data and scores are automatically generated.

3.4. Validation of Model Effectiveness

The validation of model effectiveness is achieved through three steps: data segmentation, result comparison, and parameter adjustment, ensuring that the evaluation results fit reality. The first step is data selection for validation, involves selecting player data from the past two seasons of a mainstream professional league (such as the CSL or CBA) as samples, excluding abnormal players who retired or transferred more than three times in a season, and dividing the sample into a 7:3 ratio as the training set (for model construction) and the validation set (for effectiveness testing).

The second step is validation method, inputs the standardised data of validation set players into the model to obtain predicted comprehensive value scores, and then compares them with actual market performance data. Competitive value corresponds to actual transfer fees (error rate = |predicted score corresponding transfer fee − actual transfer fee| /actual transfer fee × 100%), and commercial value corresponds to actual endorsement income (error rate calculated similarly). If both error rates are below 15%, the model is considered effective.

The third step is model adjustment, is necessary if the error rate exceeds 15%, requiring targeted optimization: if the competitive value error is high, key performance indicators in important matches (such as playoff scoring proportion) can be added; if the commercial value error is high, the weight of the social media follower activity indicator can be adjusted. After adjustment, the validation set data is re-input for testing until the error rate meets the standard, ensuring that the model can provide reliable support for club decision-making.

4. Application of the Big Data Model for Player Value Evaluation in Professional Sports Clubs

4.1. Application Scenarios Based on the Evaluation Indicator System

Based on the three-dimensional indicator system of competitive, commercial, and potential values, the model can directly serve the core decision-making scenarios of clubs [5]. In the player introduction stage, clubs can use competitive value indicators (such as average interceptions per game, pass success rate) to judge the compatibility of target players with the tactical system. For example, teams focusing on counter-attacks prioritize defenders whose defensive interception indicator scores rank in the top 20%; at the same time, combined with potential value indicators (such as the golden age of 23 to 25, growth rate of performance in recent seasons), the long-term cost-effectiveness can be evaluated to avoid high-priced introduction of declining veteran players.

In player contract renewal negotiations, the comprehensive value score can be used as the salary basis: if a player’s competitive value score in recent seasons increases by 15% and commercial value score grows by 10%, the salary increase can be controlled within the range of 10% to 15%, which both reflects the growth of player value and avoids salary bubbles. In commercial development scenarios, focusing on commercial value indicators (such as top 10 in shirt sales, more than 5 million social media followers), customised development plans can be made for high-scoring players, such as launching personal merchandize in collaboration with brands or connecting with high-end endorsement resources, to maximize commercial revenue.

4.2. Application Guarantee Based on the Big Data Processing Flow

The processes of big data collection, cleaning, and standardization provide full-chain guarantees for model application. In the data collection stage, real-time connection with official match databases and training monitoring systems generates dynamic dashboards of player value. For example, when a player's number of key passes in a match reaches the target, the competitive value score is updated in real time, and the coaching team can immediately judge the player's contribution in that match and adjust tactical rotations. Daily collection of commercial data can timely capture fan consumption trends [6]. For example, when a player's shirt sales surge, the marketing department can quickly replenish stock to reduce revenue loss.

Data cleaning removes anomalies (e.g., fake followers, match errors) to prevent biased evaluations, while standardization converts metrics like goals and defensive success into unified scores, allowing clubs to compare players across positions and make clearer reinforcement decisions [7]. Data permission management is set in the process (for example, coaching staff can only view competitive data, and the marketing department can only view commercial data) to ensure the security of core data and avoid information leakage.

4.3. Application Optimisation Based on the Core Algorithm

Focusing on the core algorithms of multiple linear regression and weighted summation, the application adaptability is optimized according to the actual needs of clubs. For players in different positions, the weights of competitive value indicators are adjusted: when evaluating forwards, the weight of average goals per game is increased from 10% to 20%, and the weight of defensive interceptions is reduced to 5%; when evaluating defenders, the adjustment is reversed to ensure that indicator weights match positional functions, avoiding bias of evaluating defenders by forward standards.

For different competition scenarios, the input indicators of the algorithm are optimized: in league scenarios, conventional competitive indicators are retained, while in key cup matches, an additional key match scoring proportion indicator (weight 10%) is added to more accurately evaluate players' value in important competitions. For example, a player may perform average in league matches but have a high scoring rate in decisive cup matches; the optimized model can accurately capture this "big-match attribute." To reduce the application threshold, the algorithm is embedded in visualization tools (such as Excel templates, simple data platforms): staff only need to input player standardized data, and the template can automatically calculate the comprehensive score through weighted summation without the need to understand algorithm principles, achieving zero technical threshold application; at the same time, explanatory notes of algorithm results are provided (such as scores above 80 indicating high-value players, scores between 60 and 80 indicating potential players), helping non-professionals quickly understand evaluation results [8].

4.4. Application Improvement Based on Model Validation

Based on model validation results, the model is continuously iterated and improved during application. A quarterly review mechanism is established: each quarter, the latest player data (such as transfer fees, endorsement income of the season) is included in the validation set, and the error rate is recalculated. If the competitive value error rate rises to 18% (exceeding the 15% standard), after analysing the reason, a new tactical adaptability indicator (such as the compatibility score of players with the team's attacking tactics) is added. After adjustment, revalidation is conducted until the error rate falls back within the standard. Feedback from various club departments is collected to optimize the model: if the coaching team suggests that potential value indicators fail to reflect players' psychological quality, the technical completion rate under high-pressure training can be collected through the training monitoring system and converted into a quantifiable indicator to be included in the potential value dimension; if the marketing department believes that commercial value indicators ignore fan interaction rate, a new social media comment interaction rate indicator is added to improve the accuracy of commercial value evaluation [9].

The model is also improved to adapt to changes in the external environment: if the league introduces salary cap policies, a salary cap adaptation coefficient can be added to the algorithm to adjust the correspondence between comprehensive value scores and salaries; if the league introduces new statistical indicators (such as expected goals), they are included in the competitive value dimension to ensure that the model always fits industry rules and development trends, providing long-term effective support for club decision-making [10].

5. Conclusion

This study focused on the practical needs of evaluating the value of professional sports club players, and constructed and explored the construction logic and application path of the big data model. The main research conclusions are as follows: Firstly, it clarified the core connotation of player value as a three-dimensional integration of competition, business, and potential, and based on this, it constructed an evaluation index system following the principles of relevance, availability, and quantifiability, and determined the weights using the simplified analytic hierarchy process. This not only highlights the core position of competitive value (60% weight) but also takes into account the supplementary role of commercial value and potential value, solving the problem of a single evaluation dimension. Secondly, it designed a three-step data processing flow of collection, cleaning, and standardization, combined with the combination algorithm of multiple linear regression and weighted summation, to ensure the objectivity and accuracy of the model. Through methods such as embedding in Excel templates, it lowered the application threshold and achieved the dual goals of technical feasibility and simplicity of operation. After validation with the test set, the error rate of the model was lower than 15%, demonstrating practical application value. Thirdly, through scenario-based application design, the model was deeply integrated with club core decisions such as player recruitment, contract renewal negotiations, and commercial development. An iterative mechanism of quarterly review, department feedback, and environment adaptation was established to ensure that the model can dynamically respond to actual needs and external changes, providing effective support for clubs to reduce decision-making risks and improve operational efficiency. This study still has limitations: Firstly, the index system does not fully cover dimensions such as teamwork ability and psychological quality that are difficult to quantify, which may affect the comprehensiveness of the evaluation. Secondly, the data sources rely on public data from mainstream leagues, and the adaptability to lower-level leagues or niche projects is insufficient. Thirdly, the algorithm mainly uses simplified models, and there is room for improvement in prediction accuracy compared to deep learning algorithms. Future research can expand the quantification indicators by combining AI video analysis technology, promote the establishment of a unified data sharing platform for professional sports leagues, and moderately introduce simple machine learning algorithms to further improve the evaluation accuracy while maintaining the practicality of the model, providing a more comprehensive solution for the value evaluation of professional sports clubs.

References

[1] Lv X , Li M .Application and Research of the Intelligent Management System Based on Internet of Things Technology in the Era of Big Data[J].Mobile Information Systems, 2021, 2021(16):1-6. DOI:10.1155/2021/6515792.

[2] BRODIE Patricia. The invisibility of midwifery: will developing professional capital make a difference?[D]. Sydney: University of Technology Sydney, 2003.

[3] Nieh J , Lam M S .The Design, Implementation and Evaluation of SMART: A Scheduler for Multimedia Applications[J].ACM SIGOPS Operating Systems Review, 1997, 31(5):184-197. DOI:10.1145/269005.266677.

[4] Barbierato E , Gribaudo M , Iacono M .Performance evaluation of NoSQL big-data applications using multi-formalism models[J].Future Generation Computer Systems, 2014. DOI:10.1016/j.future.2013.12.036.

[5] Baumer B S , Matthews G J , Nguyen Q .Big ideas in sports analytics and statistical tools for their investigation[J].WIREs: Computational Statistics, 2023, 15(6).DOI:10.1002/wics.1612.

[6] Yin S , Kaynak O .Big Data for Modern Industry: Challenges and Trends [Point of View][J]. Proceedings of the IEEE, 2015, 103(2):143-146.DOI:10.1109/JPROC.2015.2388958.

[7] Ganti V , Sarma A D .Data Cleaning:A Practical Perspective[J].Morgan & Claypool, 2013. DOI:10.2200/S00523ED1V01Y201307DTM036.

[8] Verginelli, Iason, Yao, et al. An Excel((R))-Based Visualization Tool of Two-Dimensional Soil Gas Concentration Profiles in Petroleum Vapor Intrusion[J].Ground Water Monitoring & Remediation, 2016.

[9] Low W L , Lee M L , Ling T W .A knowledge-based approach for duplicate elimination in data cleaning[J]. Information Systems, 2001, 26(8):585-606.DOI:10.1016/S0306-4379(01)00041-2.

[10] Woodhead V .How does coaching help to support team working? A case study in the NHS[J]. Oxford Brookes University, 2011.