Revolutionizing Football Management: A Data-Driven Approach with Random Forest Regressor

Aswin P S
Dept. computer science and engineering Universal Engineering College Thrissur, India
aswinps188@gmail.com

Archana Madhusudhanan
Dept. computer science and engineering Universal Engineering College Thrissur, India
archana@uec.ac.in

Athulya Sajeev
Dept. computer science and engineering Universal Engineering College Thrissur, India
athulyaks79@gmail.com

Neeha Moideen
Dept. computer science and engineering Universal Engineering College Thrissur, India
neehamoideen7@gmail.com

Suhail C R
Dept. computer science and engineering
Universal Engineering College
Thrissur, India
suhailcrasheed@gmail.com

***Abstract*—In the context of football management, depending solely on subjective evaluations and expert opinions can create significant challenges in player selection and strategic planning, potentially resulting in less-than-ideal outcomes. Relying solely on human judgment can result in errors and inefficiencies, limiting teams from reaching their full potential. Managers face challenges in making objective tactical decisions and assessing player suitability accurately. This highlights the necessity for a datadriven paradigm shift in football management. Utilizing the Random Forest Regressor, an advanced analytical method offers a systematic and fact-based approach to decision-making. The data for this study was collected exclusively from SOFIFA.com, specifically focusing on Indian Super League (ISL) players. By leveraging this method and the comprehensive dataset from SOFIFA.com, teams can effectively analyze player attributes and performance data, aiding in the identification of transfer targets that align with both individual playing styles and team requirements. This approach not only enhances tactical decision-making efficiency but also improves overall strategy formulation. Incorporating this cutting-edge algorithm empowers football managers to make better decisions, optimize squad composition, and ultimately elevate team performance on the field.**

***Index Terms*—player selection, strategic planning, Random Forest Regressor, transfer target, tactical decision-making, Indian Super League (ISL).**

I. INTRODUCTION

Football management has traditionally relied on subjective evaluations and expert opinions as the primary factors for player selection and strategic planning. Nevertheless, this method is now widely acknowledged to be fraught with challenges that can hinder team performance and success on the field. The constraints of human judgment in making objective tactical decisions and accurately assessing player suitability have become increasingly apparent. This recognition has led to a growing realization of the need for a shift towards a more data-driven approach [1]. By embracing data analytics and advanced metrics, football management aims to enhance objectivity in player evaluations and strategic decisions. This transition holds the potential to improve team performance significantly and increase the chances of success by minimizing the impact of human biases and limitations.

This research carries significant weight as it aims to bridge the longstanding gap between traditional subjective approaches and the emerging necessity for data-driven decision-making in football management. Through the application of advanced analytical methodologies like the Random Forest Regressor, this study seeks to fundamentally transform the process of analyzing player attributes and performance metrics. The ultimate goal is to empower teams with enhanced capabilities to identify promising transfer targets and strategically optimize their squad composition for improved performance, ushering in a new era of data-driven strategies in football management. At its core, the problem addressed by this research is the inherent inefficiencies and errors that stem from over-reliance on subjective evaluations. These challenges manifest in the form of suboptimal player selections, ineffective strategic planning, and an inability to fully leverage the potential of each player within the team framework.

The achievement of this research is rooted in its creation and application of a systematic methodology that incorporates the Random Forest Regressor into football management decisionmaking. Through this approach, teams can transcend subjective biases and rely on comprehensive data analysis for decisionmaking, resulting in more informed player selections, refined tactical strategies, and an overall improvement in team performance. By embracing this methodology, teams can make objective decisions backed by data, thereby enhancing their competitiveness and strategic capabilities within the realm of football management, ultimately leading to sustained success

10.5281/zenodo.12512881

and performance excellence.

The Random Forest Regressor provides football managers with a significant advantage by enabling enhanced decision-making capabilities, allowing for the customization of squad compositions to align with individual player styles and team requirements, and ultimately leading to an overall enhancement of team performance on the pitch. This transition towards data-driven strategies marks a substantial leap forward in football management techniques, as it paves the way for more efficient and effective decision-making practices throughout the industry. Embracing such technology not only empowers managers to make informed and objective decisions but also 2 optimizes team dynamics, ultimately contributing to improved results and success on the field.

II. Related Works

Mustafa A. Al-Asadi and Sakir Tasdemir [1] utilized machine learning algorithms for FIFA 20 video game data, our method provides an objective assessment of the player's market value with an accuracy that exceeds expert evaluation. This approach improves the negotiation process between football clubs and players by providing a transparent and reliable estimate of transfer fees and market value. By using a quantitative framework, stakeholders can make informed decisions and make changes in the football industry. Our datadriven approach not only increases the reliability of business value estimates but also supports trust and transparency in the negotiation process. Overall, our approach represents a significant advance in the field, providing a strong basis for assessing players' value and optimizing football match outcomes.

Vignesh Rao and Aman Shrivastava [2] observed that team strategy is an important aspect of achieving near-optimal performance in sports, especially football, where it is important to utilize the team's existing talent, but it is difficult. This study addresses the problem of optimizing team composition using a new approach based on player skill assessment. We developed our web scraping algorithms to collect relevant data and used machine learning models such as neural networks (multilayer perceptron), random forests, and logistic regression to predict the best-performing player positions. Analysis of model accuracy allows comparative evaluation and highlights the effectiveness of strategies for improving team performance in an increasingly competitive environment.

D. Prasetio [3] observed that considerable effort has been dedicated to predicting soccer match outcomes and identifying key variables influencing these results. Predictions help managers and clubs make important strategic decisions to win leagues and tournaments. This study focuses on building a logistic regression model to predict match outcomes in the 2015/2016 Barclays Premier League season, specifically home or away wins, while also identifying important variables for success. Unlike previous approaches, we exclusively use important variables established in existing studies by integrating data from FIFA video games, as pointed out by Shin and Gasparyan. A prediction accuracy of 69.5% was achieved using training data from the 2010/2011 to 2015/2016 seasons.

Leonardo Cotta, Pedro O.S. Vaz de Melo, and Fabr´ıcio Benevenuto [4] noted that there is a lack of comprehensive understanding regarding football's global impact, primarily due to limited data availability. Even though there are many schedules, many events during the season are still difficult to predict. This article introduces the use of FIFA football video game data to improve understanding of football. We discuss the characteristics of the data, justify its application, and investigate its validity. We also analyze two important points: the different performances of the Brazilian and German teams in 2014 and the special performance of FC Barcelona in the 2012/13 season. With this approach, we aim to help develop a deeper understanding of football using valid and comprehensive data.

Tim Pawlowski, Christoph Breuer, and Arnd Hovemann [5] observed that during the 1999-2000 season, the increased expenditures of European football clubs competing in the Union of European Football Associations (UEFA) Champions League (CL) had a lasting impact on the performance of topflight clubs on the football field. This policy change aimed to measure the competitiveness of Europe's five major football leagues (Britain, Spain, Italy, Germany, and France) at the turn of the century. Using various measures of competition, this study demonstrates a decrease in competitiveness following the transition to Champions League compensation.

The observer of this case study, R.Asif discovered a captivating realm within football analytics [6]. The unique challenge presented was the development of a rating system to quantitatively measure player performance, enabling predictions on various factors such as player capabilities and match outcomes. To gather the necessary data for player ratings, the observer had to tap into multiple sources. This case study meticulously outlines the strategies and solutions employed to collect and utilize this crucial information effectively.

The authors, S. Majewski and U. Szczecin, delve into the intricate task of pinpointing crucial determinants that influence the market value of football players, specifically focusing on those in forward positions [7]. Employing econometric models and data sourced from Transfermarkt, the study reveals the significant impact of factors such as Canadian classification points and dummy variables for "goodwill" (pertinent to top-tier players) on players' market values. These findings provide essential support for football managers, aiding them in making informed investment decisions and contributing valuable insights to the economics of sport and the management of football clubs' intangible assets.

The authors, H. Miao and R. A. Cachucho Knobbe, delve into the intricate realm of economic valuation in European football leagues, emphasizing the substantial financial stakes involved in the transfers of top players [8]. Beyond the immediate transfer periods, there is considerable value in continuously assessing the economic worth of football players throughout the year. Additionally, understanding the relationship between a player's market value and their performance is crucial. Both aspects are influenced by various player parameters that can be sourced from publicly available data on the web. This paper showcases a modeling approach using extensive public data sources to analyze the market value and performance of players in La Liga (the Spanish League).

III. Materials and Methods

A. Data Collection

The data for this research was gathered exclusively from SOFIFA.com, specifically focusing on Indian Super League (ISL) players from the year 2024. Information collected included player names, ages, overall ratings, team affiliations, physical attributes such as height, weight, and preferred foot, along with their best positions within the team structure.

All major playing positions were considered for data collection, including central backs (CB), central midfielders (CM), central attacking midfielders (CAM), defensive midfielders (DM), center forwards (CF), goalkeepers (GK), fullbacks (LB and RB), wide midfielders (LM and RM), and left and right wingers (LW and RW). Additionally, attributes specific to set-piece specialists such as free-kick takers, penalty takers, and players involved in set-piece plays were also collected.

B. Data Cleaning and Preprocessing

The data underwent meticulous data cleaning procedures to ensure its integrity and reliability for predictive modeling in the research paper. This included addressing missing values across numerical attributes like age, height, and weight, as well as categorical variables such as preferred foot and player position. Instances of significant missing data led to the removal of corresponding rows or columns from the dataset. Additionally, efforts were made to correct inconsistencies in data entry, such as fixing typos, standardizing formats, and resolving conflicting information. Data formats were standardized across attributes, including converting text to lowercase, standardizing date formats, and consolidating similar categories. Outliers in numerical attributes were detected and managed appropriately. Duplicate entries were also identified and eliminated to prevent redundancy in the dataset. Overall, these steps ensured data completeness, accuracy, and consistency, laying a solid foundation for effective predictive modeling in football management.

C. Feature Selection

The feature selection process for this research paper involved identifying key attributes for each playing position in football. These attributes were chosen based on their relevance to the specific roles and responsibilities associated with each position. Table I provides a comprehensive overview of the selected features for each playing position, ensuring that the analysis focuses on crucial aspects of player performance and suitability.

For central defenders (CB), attributes such as heading accuracy, strength, and defensive awareness were prioritized to emphasize their defensive capabilities. In contrast, central midfielders (CM) require a diverse set of skills including passing accuracy, dribbling ability, and vision to excel in both offensive and defensive play. Similarly, attacking positions

TABLE I: Selected features for each playing position

Position	Selected Features
CB	'Heading accuracy', 'Strength', 'Interceptions', 'Standing tackle', 'Sliding tackle', 'Aggression', 'Jumping', 'Reactions', 'Ball control', 'Stamina', 'Short passing', 'Long passing', 'Defensive awareness'
CM	'Crossing', 'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'FK Accuracy', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots', 'Aggression', 'Interceptions', 'Att. Position', 'Vision'
CAM	'Crossing', 'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots', 'Aggression', 'Interceptions', 'Att. Position', 'Vision'
DM	'Short passing', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Long shots', 'Aggression', 'Interceptions', 'Vision', 'Heading accuracy', 'Strength', 'Standing tackle', 'Sliding tackle', 'Reactions', 'Defensive awareness'
CF	'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots'
GK	'GK Diving', 'GK Handling', 'GK Kicking', 'GK Positioning', 'GK Reflexes', 'Reactions', 'Stamina', 'Jumping'
LB and RB	'Crossing', 'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots', 'Aggression', 'Interceptions', 'Att. Position', 'Vision', 'Defensive awareness', 'Standing tackle', 'Sliding tackle'
LM and RM	'Crossing', 'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots', 'Aggression', 'Interceptions', 'Att. Position', 'Vision', 'Standing tackle', 'Sliding tackle'
LW and RW	'Crossing', 'Finishing', 'Heading accuracy', 'Short passing', 'Volleys', 'Dribbling', 'Curve', 'FK Accuracy', 'Long passing', 'Ball control', 'Acceleration', 'Sprint speed', 'Agility', 'Reactions', 'Balance', 'Shot power', 'Jumping', 'Stamina', 'Strength', 'Long shots', 'Aggression', 'Interceptions', 'Att. Position', 'Vision', 'Penalties'
Freekick	'Finishing', 'Shot power', 'Curve', 'FK Accuracy', 'Long passing', 'Short passing', 'Crossing', 'Long shots'
Penalty	'Finishing', 'Shot power', 'Penalties'
Setpiece	'Finishing', 'Shot power', 'Curve', 'FK Accuracy', 'Long passing', 'Short passing', 'Crossing', 'Long shots'

like center forward (CF) and attacking midfielder (CAM) emphasized finishing, ball control, and agility for effective goal scoring and creative playmaking.

Defensive positions such as full-backs (LB/RB) and defensive midfielders (DM) placed importance on attributes like tackling, interceptions, and defensive awareness. Goalkeepers (GK) were evaluated based on diving, reflexes, and positioning abilities crucial for shot-stopping and organizing defensive lines. Additionally, set-piece specialists (Freekick, Penalty, Setpiece) were identified based on their proficiency in specific skills like free-kick accuracy and finishing ability.

This meticulous feature selection process ensures that the analysis focuses on relevant attributes tailored to each playing position, providing valuable insights for player selection and strategic planning in football management.

D. Model Training and Evaluation

The Random Forest Regressor was selected as the machine learning model for this study due to its ability to handle complex datasets and nonlinear relationships effectively. This algorithm was applied individually to each playing position in football, creating separate models tailored to specific roles on the field. This approach allows for a more nuanced understanding of player performance metrics, as different positions may require distinct sets of features to predict player outcomes accurately.

The Random Forest regressor, with its ensemble of decision trees, was trained using a stratified splitting technique to ensure a balanced representation of player positions in both the training and testing sets. Approximately 80% of the dataset was used for training each model, while the remaining 20% was reserved for evaluating their performance. By leveraging the collective insights from multiple decision trees, the models were able to capture complex interactions among variables specific to each playing position, enhancing their predictive power and generalization to unseen data. This methodology provides a robust framework for predicting player performance metrics in football management contexts, offering valuable insights for optimizing team strategies and player development.

The evaluation phase aimed to assess the Random Forest Regressor's predictive power and generalization ability. Various evaluation metrics were utilized, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). MAE measures the average absolute difference between predicted and actual values, providing insights into the model's accuracy. Lower MAE values indicate higher accuracy in predicting player performance. MSE calculates the average squared difference between predicted and actual values, giving more weight to larger errors. Lower MSE values signify a more precise model. RMSE, the square root of MSE, provides a more interpretable measure of error, highlighting the average magnitude of prediction errors.

Through rigorous evaluation using these metrics, the Random Forest Regressor's performance in analyzing player attributes, predicting performance metrics, and aiding in tactical decision-making was comprehensively assessed, contributing to the advancement of data-driven approaches in football management.

IV. Results and Discussion

The Random Forest Regressor was employed to analyze player attributes and predict performance metrics for various playing positions in football, specifically focusing on the Indian Super League (ISL) players. The results obtained from the model evaluations provide valuable insights into the effectiveness of this data-driven approach in football management decision-making.

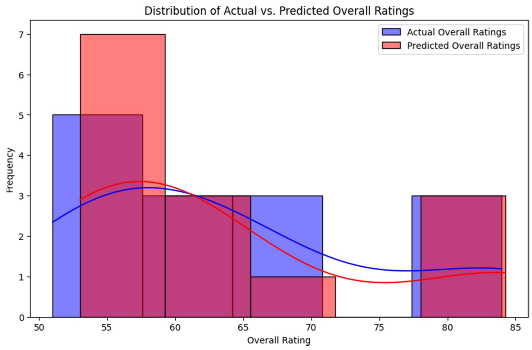

Fig. 1: Comparison of Predicted and Actual Results for LM/RM.

The Random Forest Regressor model for LM/RM exhibits strong predictive accuracy at 92.86%, effectively capturing LM/RM performance with a recall of 100%, though it occasionally misclassifies instances with a precision of 75%, resulting in a balanced F1-score of 0.857Prediction errors include MSE of 3.436, RMSE of 1.854, and MAE of 1.284, showing moderate accuracy.

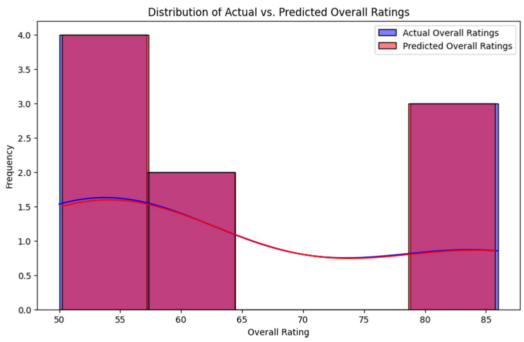

Fig. 2: Comparison of Predicted and Actual Results for RW/LW.

For the RW/LW position, the Random Forest Regressor model achieves an accuracy of 88.89%, with a recall of 100%, indicating successful identification of relevant instances. However, its precision at 66.67% suggests some false positive predictions. The F1-score of 0.8 shows a balanced performance. Lower MSE, RMSE, and MAE compared to LM/RM indicate slightly more accurate predictions for RW/LW.

In the midfield position, the Random Forest Regressor model achieves a high accuracy of 90.91% and perfect precision of 100%. This indicates that the model accurately predicts midfield player performance and rarely makes false positive predictions. However, with a recall of 66.67%, the model may miss some instances of player performance in midfield. The F1-score of 0.8 reflects a good balance between precision and recall for this position. The higher MSE, RMSE, and MAE compared to LM/RM and RW/LW suggest that the model's predictions for midfielders have a moderate level of prediction errors.

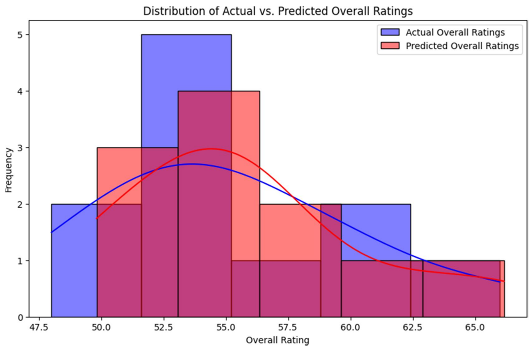

Fig. 3: Comparison of Predicted and Actual Results for Midfield.

The model demonstrates solid performance for Central Backs (CB) with 87.5% accuracy and a balanced F1-score of 0.75, reflecting its potential for aiding CB player selection and tactical decisions. While its Precision at 60% suggests room for improvement, its Recall of 100% ensures effective capture of positive CB performances, minimizing false negatives.

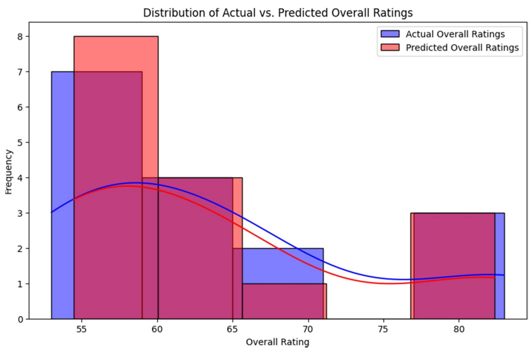

Fig. 4: Comparison of Predicted and Actual Results for CB.

With an accuracy of 90.91%, the model showed a high level of correctness in predicting GK-related outcomes. The precision of 75.00% highlighted the model's ability to accurately identify true positive predictions, while a recall of 100.00% indicated that all actual GK-related outcomes were correctly classified. Moreover, the mean squared error (MSE) of 1.164 and root mean squared error (RMSE) of 1.079 demonstrated the model's relatively accurate predictions of GK performance metrics.

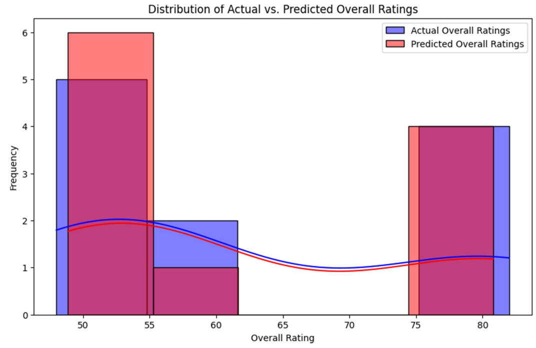

Fig. 5: Comparison of Predicted and Actual Results for GK.

In the right back/left back (RB/LB) positions, the model achieved an accuracy of 91.67%, with a precision and recall of 100.00% and 75.00%, respectively. The mean squared error of 5.732 and root mean squared error of 2.394 suggest a slightly higher error rate compared to other positions but still within reasonable bounds.

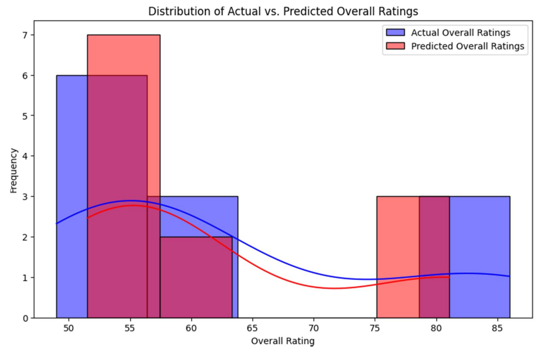

Fig. 6: Comparison of Predicted and Actual Results for LB/RB.

Lastly, for center forwards (CF), the model showcased an accuracy of 86.67%, with a precision of 83.33% and recall of 100.00%. The mean squared error of 6.405 and root mean squared error of 2.531 indicate a moderate level of error in predicting performance metrics for these attacking players.

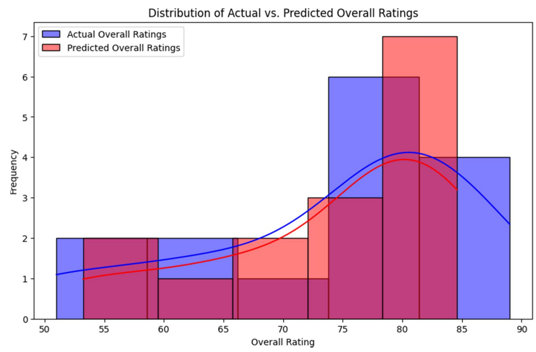

Fig. 7: Comparison of Predicted and Actual Results for CF.

Table 2 offers an overview of the accuracy, precision, recall, and F1-score metrics for every playing position. These metrics offer insights into the predictive capabilities of the Random Forest Regressor models across different positions on the football field. Additionally, Table 3 presents the MSE, MAE, and RMSE values, indicating the degree of error in predicting player performance metrics. The results indicate that while the model achieved high accuracy and precision in several positions, there are variations in the error metrics, suggesting areas for improvement and fine-tuning.

TABLE II: Summary of Performance Metrics

Position	Accuracy	Precision	Recall	Specificity	F1 Score
LM/RM	0.93	0.75	1.0	0.91	0.86
RW/LW	0.89	0.67	1.0	0.86	0.8
MID	0.91	1.0	0.67	1.0	0.8
CB	0.88	0.6	1.0	0.85	0.75
GK	0.91	0.75	1.0	0.88	0.86
RB/LB	0.92	1.0	0.75	1.0	0.86
CF	0.87	0.83	1.0	0.6	0.91

The model's high accuracy and precision in midfield and defensive positions suggest that it can effectively evaluate attributes crucial for these roles. This includes passing accuracy, defensive awareness, tackling ability, and goalkeeping skills. Such insights can aid managers in identifying players who align with their team's playing style and strategic requirements.

TABLE III: Summary of Error Metrics

Position	MSE	RMSE	MAE
LM/RM	3.43632	1.85373175	1.283571428
RW/LW	0.6970999	0.83492514	0.69666662
MID	2.957445454	1.71972249	1.268181815
CB	1.18607499	1.08907070	0.88625
GK	1.16440002	1.07907367	0.818181818
RB/LB	5.73247500	2.39425875	1.972499995
CF	6.404686678	2.53074824	2.054000002

However, the model's performance in attacking positions, while still reliable, may benefit from further fine-tuning to capture nuances related to goal-scoring efficiency and creative playmaking. This could involve refining the selection of attributes or exploring additional data sources to enhance the model's predictive power in these areas.

Figures 1,2,3,4,5,6,7 show the distribution of predicted and actual results for each position, providing a visual representation of the model's performance across different playing positions. Overall, the Random Forest Regressor presents a promising approach to enhancing football management practices. Continuous evaluation, refinement, and integration of data-driven methodologies can contribute significantly to improving team performance and achieving success on the field.

V. CONCLUSION

Our research aimed to advocate for a data-driven approach in football management, particularly focusing on the Indian Super League (ISL), by utilizing advanced analytical methods like the Random Forest Regressor. Through our study, we demonstrated the potential of data-driven strategies to revolutionize player selection, strategic planning, and overall team performance. By consistently achieving high accuracy rates 6 and favorable precision-recall scores with the Random Forest Regressor, we emphasized the practical applicability of data analytics in optimizing player selections and refining tactical strategies. These findings underscore the transformative potential of data-driven decision-making, reducing human biases and errors inherent in subjective evaluations. Our research contributes to the broader field by showcasing the efficacy of the Random Forest Regressor in enhancing player selection and team performance.

Looking ahead, our study suggests avenues for future research to further advance data-driven football management. Exploring additional analytical methods and expanding data sources could lead to higher predictive accuracy and model performance. Including data from diverse football leagues and regions would offer a more comprehensive understanding of player performance dynamics. Investigating the longterm impact of data-driven strategies and assessing their acceptance among stakeholders are also recommended. These future research directions aim to foster continuous innovation in decision-making processes, ultimately contributing to enhanced team performance and sustained success in football management.

REFERENCES

[1] Mustafa A. AL-AsadiI and Sakir Tasdem " Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques",february 25,2022, Vol. 10, pp. 22631-22645.

[2] V. Rao and A. Shrivastava, "Team strategizing using a machine learning approach," in Proc. Int. Conf. Inventive Comput. Informat. (ICICI),Nov. 2017, pp. 1032–1035.

[3] D. Prasetio and D. Harlili, "Predicting football match results with logistic regression," in Proc. Int. Conf. Adv. Inform.: Concepts, Theory Appl. (ICAICTA), Aug. 2016, pp. 1–5.

[4] L. Cotta, "Using fifa soccer video game data for soccer analytics," in Proc. Workshop Large Scale Sports Anal., 2016, pp. 1–4.

[5] T. Pawlowski and C. A. Breuer Hovemann, "Top clubs' performance and the competitive situation in European domestic football competitions,"J. Sports Econ., vol. 11, no. 2, pp. 186–202, 2010.

[6] R. Asif, "Football (soccer) analytics: A case study on the availability and limitations of data for football analytics research," Int. J. Comput. Sci. Inf. Secur., vol. 14, no. 11, p. 516, 2016.

[7] S. Majewski and U. Szczecin, "Identification of factors determining market value of the most valuable football players," J. Manage. Bus. Administration. Central Eur., vol. 24, no. 3, pp. 91–104, Sep. 2016.

[8] H. Miao and R. A. Cachucho Knobbe, "Football player's performance and market value," in Proc. 2nd Workshop Sports Anal., Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECML PKDD), 2015, pp. 1–9.