Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/BA08.20250366
Published: December 20, 2025

ML Based Player Selection Assistant for Franchise Cricket Leagues: A Regional Performance Analysis

Jubayer Ahamed, Arnob Aich Anurag, Shaikat Das Joy, MD Sajid Bin-Faisal, Dip Nandi and Md. Asraf Ali
Department of Computer Science
American International University-Bangladesh Dhaka, Bangladesh
jubayer@aiub.edu, aicharnob88@gmail.com, skdas@aiub.edu, sajid@aiub.edu, dip.nandi@aiub.edu, asrafali@aiub.edu

Abstract

Franchise cricket leagues rely heavily on effective player selection at auctions, but human bias too frequently leads to poor performance and missed talent. Using a subjective judgment and a historical pricing model to determine winning teams at auction has led to much greater financial inefficiencies, poor franchise team composition around the globe, and undervalued high-potential emerging players with little to no visibility. This research presents a machine learning based Player Selection Assistant that uses international cricket performance statistics to predict suitability and auction value on a regional basis. The research evaluated four models: XGBoost, Random Forest, Decision Tree and Artificial Neural Networks using regional performance metrics which include batting strike rates, Bowling economy and fielding. The results show that ensemble methods provided superior performance over traditional approaches, with XGBoost (RMSE: 3.458, R^2: 0.846) and Random Forest (RMSE: 3.475, R^2: 0.870) outperforming other methods. The framework has regional ability suitability scoring/indices as well as a pressure index to provide consistent and transparent data-driven recommendations to reduce human bias and improve player selection performance model for auction decisions in franchise cricket leagues.

Keywords
Player selection, Machine Learning, Auction value prediction, Regional performance analysis.

1. Introduction
Franchise cricket leagues have transformed modern cricket by integrating the elements of sport, entertainment and marketing. The emergence of Twenty20 (T20) cricket has led to the widespread popularity and success of faster game forms, particularly the Indian Premier League (IPL), resulting in unprecedented access to information regarding player performance and club management (A. Shenoy et al. 2022) .This bias results in teams frequently underperforming, while untested talents who may outperform their predecessors are overlooked (S. Hundekari et al. 2024).The objective of machine learning (ML) is to authenticate data and enhance educational contributions to decision-making in sports (K. Passi and N. Pandey 2018) . In cricket, machine learning has been employed to choose match winners, evaluate player's performance potential and establish base pricing for auction participants. Models such as Random Forest, Support Vector Machines (SVM) and Gradient Boosting demonstrate reliable performance in classification and regression tasks connected to performance metrics. Ensemble methods such as XGBoost and Random Forest provide superior performance compared to single-model approaches in predicting batting metrics, including strike rate and batting average.

This research contributes a new machine learning-based approach to identify players for auction in franchise cricket leagues, grounded in player's international performance statistics. It incorporates players' batting strike rates, boundary percentages, death overs, bowling economy, wickets per game, and fielding data, as well as contextual parameters like pitch type and quality of opposition. Following the works (K. Passi and N. Pandey 2018) ,(B. Bhattacharjee et al. 2024), in the proposed approach, the data will be adjusted to correct the distortions from the international and franchise formats. The goal is to invent a multi-tiered approach that estimates suitability, auction value, and classification of a role. Initial findings indicate that ensemble models like XGBoost and Random Forest, which recorded RMSEs of ≈ 3.4 and R^2 of over 0.84, performed the best, in contrast to overfitting Decision Trees and Artificial Neural Networks which had high residuals and were unsatisfactory. This agrees with sports data literature citing ensemble models as the best approach due to their sophisticated design (K. Passi and N. Pandey 2018).

We propose a Player Selection Assistant that employs supervised learning, clustering and recommendation techniques. The regression module predicts the auction value of players, and the classification module determines the suitability of the player for a temperamental role. Additional features that we derive from contextual statistics are: The League Suitability Score, a Pressure Index, and an Auction Value Predictor (S. Hundekari et al. 2024). Reinforcement learning techniques will be applied to simulate the auction sequences and to determine the optimal selections. The assistant will provide explainable recommendations concerning the consistency, adaptability and role-fit of the players it selects, thus providing transparency and reducing human bias (B. Bhattacharjee et al. 2024).

In section 2, we show the Literature Review section of our study. The Methodology will be described in Section 3. In section 4, the Results and Discussion section will be shown. Finally, we set Conclusion and Future work as part of our study.

1.1 Objectives

The research objectives are given below:

- Design and compare models based on machine learning for player selection in franchise cricket leagues, using data on the player's performance at the international level.
- Integrate many statistical characteristics including averages and strike rates in batting and bowling and a 'match impact' score, to predict the player's suitability for the franchise league.
- Enhance team selection decisions by minimizing any one of the many biases from traditional scouting approaches.
- Use machine learning techniques, including regression and classification models such as XGBoost and Random Forest, to predict auction value for the player and gauge player suitability for a team role.
- Implement a recommendation system for player selection so that the decisions are transparent and consistent.

2. Literature Review

Alaka (Alaka et al. 2021) examined the difficulty of creating meaningful feature representations for cricket analytics in the data-scarce IPL. Traditional machine-learning methods generally failed to capture inter-player and inter-team links needed for reliable prediction, according to the report. A deep representation-learning approach employing Siamese networks and adaptive embeddings allowed the model to learn latent vectors for batting and bowling characteristics. Their multi-modal approach used pre-match features and pitch-report embeddings to show how contextual information improved prediction accuracy. Contrastive-loss representation learning outperformed cross-entropy-based categorization with up to 95% accuracy. This highlighted the importance of learnt embeddings, multi-modal fusion and contrastive learning for cricket prediction tasks, laying the groundwork for more generalizable and interpretable sports data analysis ML models.

Malhotra (G. Malhotra et al. 2022) proposed a comprehensive approach to estimating player value in IPL auctions by incorporating not only previous auction prices but also key features such as performance, age, popularity, and team-specific role suitability. The study emphasizes the importance of hedonic factors, highlighting their impact on ticket sales and franchise profitability. Various machine learning algorithms, including Bayesian Ridge Regression, Random Forest, and Linear Regression, were explored, with Linear Regression yielding the best performance based on R^2 scores. However, the study lacks analysis on player performance differences in home vs. away matches and contributions in international or high-pressure games.

In another work, Md. Ashiqur Rahman Khan et al. illustrated how Machine Learning (ML) algorithms can be used to select cricket players for national and franchise teams (Md. Robel et. al. 2024). Their work mainly focused on national team players, using data collected from ESPNcricinfo and Cricbuzz. They considered both physical fitness and player performance, and classified players into roles such as batsman, bowler, all-rounder, and wicketkeeper. The study applied three ML algorithms: Linear Regression, Support Vector Machine (SVM), and Random Forest. Among these, SVM outperformed others. Based on the model outputs, two national squads were generated. However, the study had some limitations. The data sources lacked coverage of domestic league performances, which are important for spotting emerging talent. Additionally, leadership qualities and performance in critical match situations were not considered, even though these factors are crucial for selecting players who can handle pressure and lead the team in important moments.

Chandru and Jayanthi conducted a study to predict player performance and match scores in the IPL using the Random Forest algorithm (M. Chandru et. al. 2022). Their work focused on team performance driven by individual player roles, such as hard-hitter, finisher, consistent performer, wicket-taker, and economical bowler. They also considered environmental and ground conditions. However, the work lacks integration of real-time player form and psychological factors, which are critical in high-pressure scenarios and can significantly affect performance.

Kasande and Jahirabadkar (M. Kasande et. al. 2023) presented a comprehensive survey of various machine learning approaches for cricket team selection. The authors reviewed algorithms used in other sports like baseball and football before focusing on methods developed specifically for cricket team selection. The survey covered techniques such as neural networks, classification algorithms, regression models, clustering, and support vector machines. The authors identified that most existing methods focused on shorter formats of cricket (ODIs and T20s) while test cricket team selection remains largely unexplored. Their analysis revealed that team selection is typically treated as a classification or regression problem, with random forest classifiers generally yielding the best accuracy. Research gaps identified include the lack of methods for test cricket team selection and the limited availability of ball-tracking data, which could significantly enhance model accuracy and provide deeper insights into player performance.

Manju and Philip (M. K. Manju et. al. 2023) proposed a novel framework for ranking batsmen in the Indian Premier League (IPL) based on individual performance, role in the team, and team interactions. They first derived key metrics to evaluate batsmen's individual performance, including batting average, strike rate, hard hitting score, big innings score, opener index, and death over specialist index. The authors then evaluated these metrics using supervised machine learning algorithms such as random forest, gradient boosting, and classification trees to determine their importance in ranking. Their results identified David Warner as the best IPL batsman across seasons 2008-2022. However, the study is limited by its focus on batsmen only, and the framework could be enhanced by extending it to bowlers and incorporating more detailed match context factors like pitch conditions.

This paper reviewed cricket analytics research and highlights the expanding use of machine learning to forecast match results and evaluate player performance (S. K. C et al. 2021). Using data mining, previous studies predicted emerging players, ball-tracking estimation, social-network-based player ranking and IPL results. Many studies had used decision trees, Naïve Bayes, MLP and SVM to evaluate player value or match results, with moderate accuracy. Toss choices, venue effects and home-field advantages had been studied. Other research predicted runs, wickets and batting skills using classification approaches, with Random Forest often outperforming alternatives. This literature showed that incorporating matches, player and contextual variables improves cricket-match prediction algorithms (S. K. C et al. 2021).

The evaluated research (R. Bunker et al. 2022) examined how machine learning predicted team sports outcomes over two decades. It found common algorithms, evaluation techniques, and research gaps in 1996–2019 studies. The authors observed that Artificial Neural Networks had been widely employed but did not always outperform Decision Trees, SVMs, or ensemble approaches. Richer feature engineering, detailed experimental design, and significant validation were common in successful investigations. Soccer was harder to model despite massive datasets, according to the review. The report concluded that larger datasets, better feature selection, and advanced ML approaches are needed for future research (R. Bunker et. al. 2022). Other related studies used historical data to forecast outcomes and choose teams by employing methods like regression and association rule mining. The pressure index was also mentioned in the literature to measure how the match was changing, but a complete, context-aware metric for each player's contribution was not fully explored (Muhammad Sohaib Ayub et al 2023).

The literature study (Ijraset.com, 2023) indicated that prior studies have extensively utilized machines and deep learning in cricket analytics. Shah et al. (Basit et. al. 2020) employed models such as Random Forest to forecast high run chases in T20 matches, but Fadi et al. (Bunker et. al. 2027) obtained significant accuracy in score prediction for hundred-ball cricket utilizing an MLP regressor. (Kapadia et al. 2020) and (Singh et al. 2025) utilized distinct algorithms to forecast the outcomes of IPL and ODI matches, respectively.

3. Methodology

This research focuses on four assorted machine learning techniques: random forest, XGboost, decision trees, and artificial neural networks, to create a model predicting player selection framework for franchise cricket auctions. XGboost is a modified gradient boosting algorithm that accommodates non-linear relationships for different region performance data. All techniques will be amalgamated using K-Fold cross validation to evaluate the model without bias and form a selection for the final model. All algorithms are coded for and implemented on Google collab using the scikit, and tensor flow libraries. All models will be subjected to RMSE, R squared, and feature importance to carve out gaps, and model a good random forest as the final model to add for predictions of player values for auctions, and then predicting regions players may be most suitable for the auction.

3.1 Dataset

The research drew on information gathered from ESPNcricinfo.com, an extensive database of cricket statistics that includes records of player performances on an international level. The data sets used contained records of regional performance metrics of cricket players based in several different geographical areas, including Asia; South Africa/Zimbabwe (Africa); England/Wales (Europe); and Australia/New Zealand (Oceania). Performance measures examined in this study include Batting Average (AVR), Strike Rate (SR).

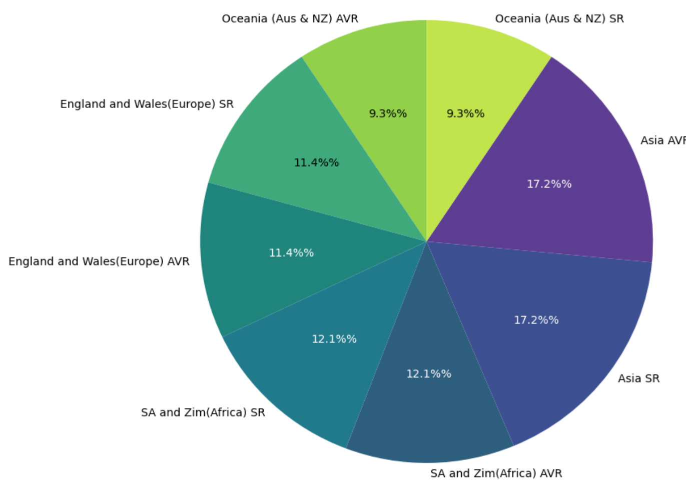

Figure 1. Distribution of Non-Missing Regional Performance Data Across Columns

Figure 1 depicts the distribution of the non-missing values from the dataset across each regional performance metric. The data seems to be evenly represented across the regions shown, with Asia offering the most at 17.2% on both AVR (Average) and SR (Strike Rate) metrics. England and Wales (Europe), SA and Zim (Africa), and Oceania (Aus & NZ), each had similar contributions at roughly equal distribution for each respective AVR and SR measure, i.e., some were as low as 9.3% and as high as 12.1%. This suggests that the dataset is comprehensive across multiple cricket regions with balanced representation.

3.2 Steps for the implementation of the proposed approach

The following steps shown in Figure 3 were performed to generate player recommendations.

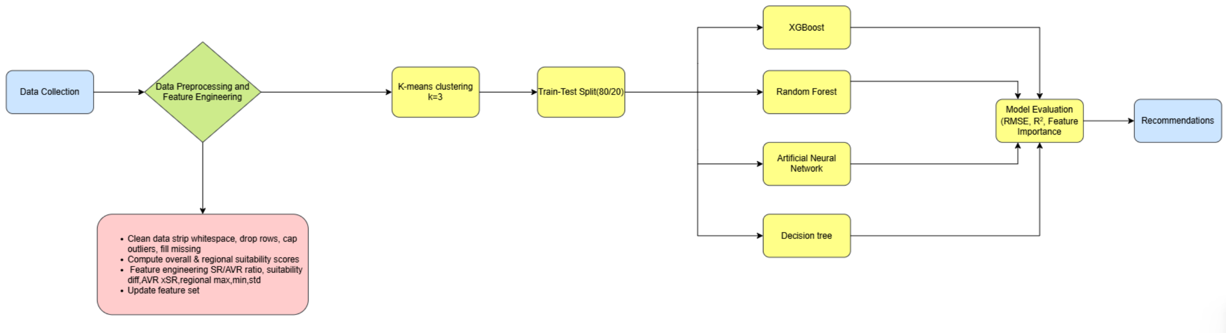

Figure 2. Workflow of Methodology

Figure 2 depicts the overall processes and actions needed for training the player selection machine learning model as part of franchise cricket leagues. The first step is player data collection followed by collection of data preprocessing and feature engineering stages. Next, the data is cleaned, outliers are managed, and the additional features of strike rate and batting averages are calculated. The data is then run through k-means clustering to cluster players into performance clusters. After partitioning the data into training and testing sets, we complete training of models XGBoost, Random Forest, ANN and Decision Trees are fully trained. The models are evaluated using RMSE, R^2 and Feature Importance metrics. Following that players with highest scores in the auction selection are selected. Finally, players are selected and the best models with the best results are taken.

4. Results and Discussion

The four machine-learning models evaluated in depth exhibited vastly different predictive capabilities when evaluating player auction values and regional impact evaluations. Analyzing these models, the comparison focused primarily on how Ensemble Models perform compared to individual traditional model approaches across three different geographical areas. Results also suggest that there are considerable regional variations in player performance, which ultimately affects a player's auction value. There are some players performing exceptionally well in certain regions, while underperforming in others. Results from this study provide practical applications for the removal of human bias in making player selections, while improving prediction accuracy when roster-building via auction.

4.1 Model Performance Comparison

In the research on the performances of the XGBoost, Random Forest, Decision Tree and Artificial Neural Networks on the different regions, there were differences in performances on the regions of the XGBoost which had the overall lowest RMSE with the 3.458 and R^2 of 0.846, and Random Forest which had the RMSE of 3.475 and R^2 of 0.870. The Decision Tree had a poor performance of the RMSE being 5.343 and the R^2 being 0.681 while the Artificial Neural Net were a failure with 14.618 RMSE and negative 1.810 R^2 which means their prediction were worse than simply void of the average values (Table 1).

Table 1. Cross-Validation Results Table

Index	**Model**	**Avg Overall RMSE**	**Avg Overall R^2**	**Avg Asia RMSE**	**Avg England and Wales (Europe) RMSE**	**Avg Oceania (Aus & NZ) RMSE**	**Avg SA and Zim (Africa) RMSE**
0	XGBoost	3.458	0.846	5.046	3.137	2.328	3.541
1	Random Forest	3.475	0.870	5.199	3.249	2.135	3.332
2	Decision Tree	5.343	0.681	6.364	4.323	2.706	5.904
3	Artificial Neural Network	14.618	-1.810	11.999	3.499	9.985	8.760

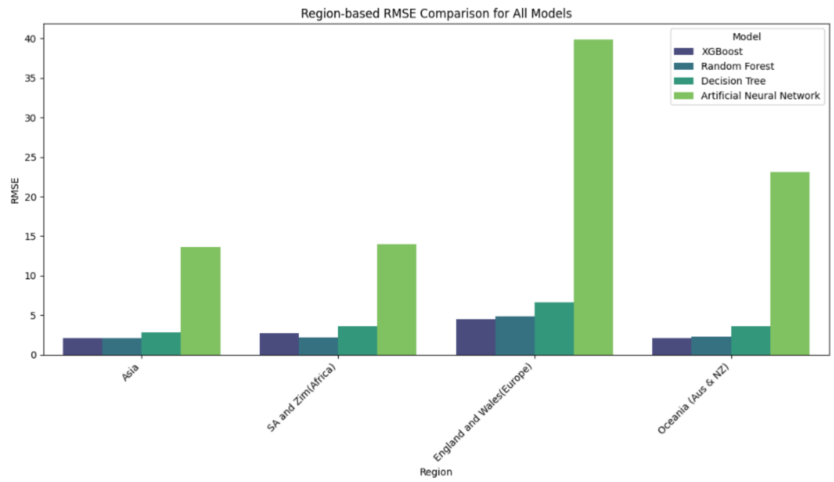

Figure 3. Region-based RMSE Comparison for All Models

Figure 3 portrayed provides a comparison of the Root Mean Square Error (RMSE) between 4 machine learning models (XGBoost, Random Forest, Decision Tree and Artificial Neural Network) in 4 different areas.From the chart, we can see that traditional ensemble methods (XGBoost and Random Forest) present very good performance across every area investigated, not reporting any RMSE values higher than 4 in every area. In comparison, the Decision Tree method shows moderate performance, with RMSE values falling between 4 and 7 across all areas, while the Artificial Neural Network method shows clear underperformance across all areas, which is even more severe in the England and Wales area with RMSE values falling around 40, indicating possible overfitting of the training data or insufficient training for this method of model architecture.

4.2 Player Suitability Analysis by Region

Wide regional variation was observed in player suitability scores. As a result, individual athletes had different performance levels based on geographic locale (i.e., high performing athletes had vastly varying results based on geographic regions). For example, an example of extremely high-fitting player suitability for England and Wales was Brandon King (with a player suitability score of > 100). But on the other end of the spectrum in terms of low-fitting suitability across geographic regions, Tony Ura consistently had among the lowest player suitability scores (and low-fitting overall player performance). Meanwhile, certain players have a strong (or very suitable) performance only in certain geographic regions.

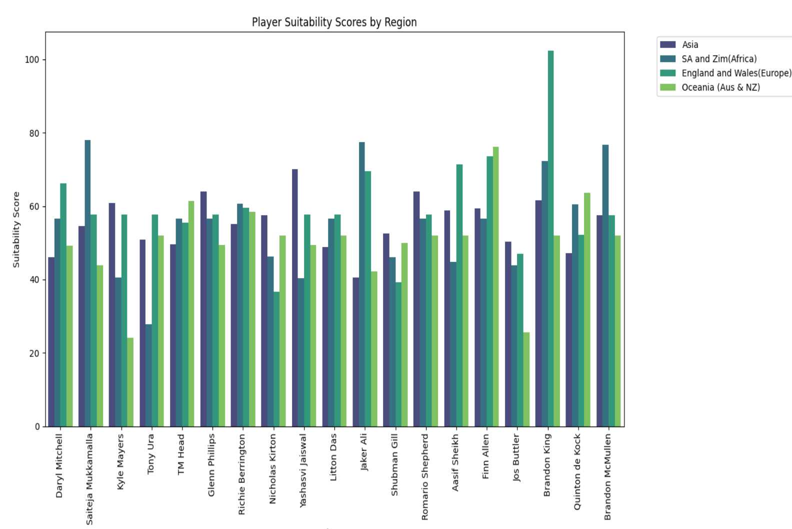

Figure 4. Player Suitability Scores by Region

The suitability scores for numerous cricket players over four geographic areas have been presented visually in Figure 4. Players earned suitability ratings from a minimum of approximately 25, to maximum scores of above 100 (i.e., a match rating greater than 100); one such player with a very high suitability rating in the England & Wales area is Jos Buttler, with a score above 100. Conversely, Tony Ura holds the lowest overall suitability score across all four regions. The player-performance variability is demonstrated by the regional suitability scores shown on the chart. Several players ranked high in one regional suitability, for instance, Saiteja Mukkamalla and Jaker Ali performed very well overall in South Africa & Zimbabwe. Conversely, other players such as Finn Allen have multiple regions where they can perform well. Since each player can perform well in more than one region, there is a demonstrated regional dependency on players to assess their perceived value depending on their geographic area. For example, Daryl Mitchell has his highest suitability scores for South Africa & Zimbabwe and Asia regions; these scores are significantly lower for the non-regional areas. The polarized nature of the player suitability scores as exemplified by Jos Buttler further demonstrates the importance of regional context when evaluating the potential suitability of a player.

4.3 Feature Correlation Analysis

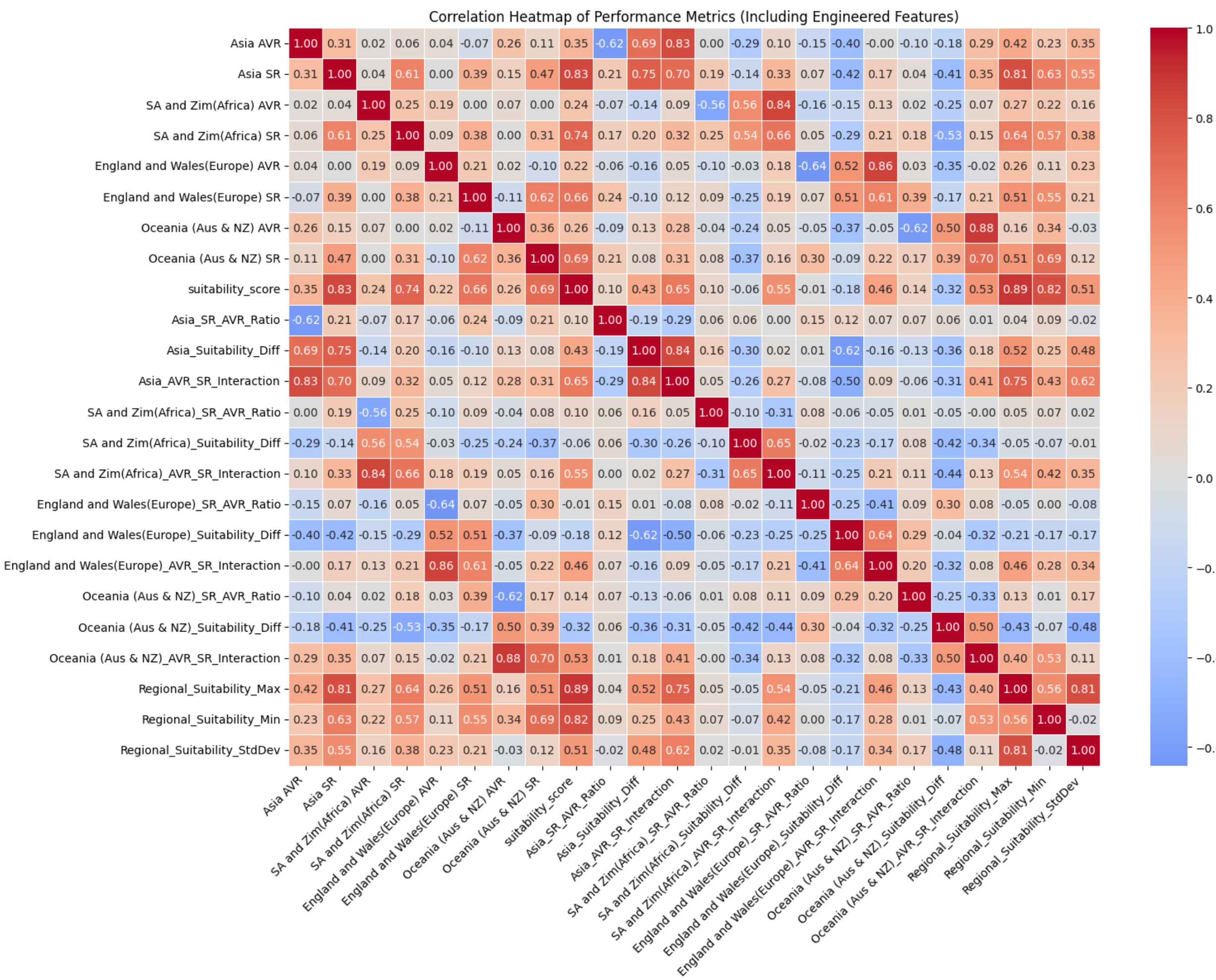

Figure 5. Correlation Heatmap of Performance Metrics

4.4 Individual Model Performance Visualizations

In Figure 5, the correlation matrix as presented highlights the relationships between several performance metrics, which include, regional AVR, SR scores, suitability measures and engineered features. There are a few strong positive correlations (indicated by darker red with values near 1.0) between certain regional metrics and regional suitability scores, especially noted in the diagonal regions. Correlations of interest include Asia_SR_AVR_Interaction and Asia_Suitability_Diff where the correlation between the two is .84 and where correlations between Regional_Suitability_Max and regional metrics range from .81-.89. There are correlations of negative relationships (indicated by the blue shades) which demonstrate inverse relationship between suitability differences and specific regional metrics, and it could indicate that there are greater complexities with respect to regional player performance interdependence across different regions. For example, the combined relationship between the SA and Zimbabwe's (African) metric is negatively correlated approximately -0.54 to -0.62 with the base metrics throughout the region for Asia, which measures average yield for Asia-South Africa and average yield for South Africa and Zimbabwe (both in Africa). The average suitability index (suitability_score) is moderately positively correlated at around 0.62 to 0.74 with both England and Wales (European) as well as Australasia (Australia and New Zealand) metrics for all regions within each of those regions, indicating that, in general, a greater degree of regional success correlates positively with greater overall suitability of that region. These correlations between regional suitability scores and the measures of regional agricultural yield performance are important for refining the overall suitability model by focusing upon the features that show high levels of correlation and/or by identifying potential trade-offs among regions based on observed performance of those regions.

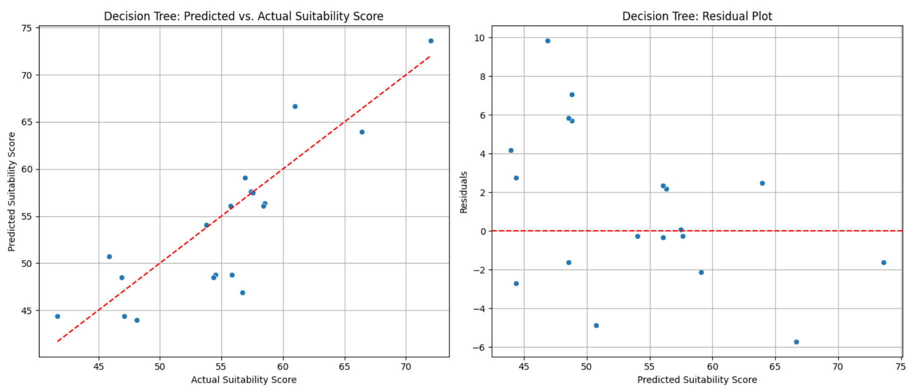

Figure 6. Decision Tree: Predicted vs. Actual Suitability Score

In Figure 6, the left scatter plot indicates the balance between Decision Tree model's predictions presented against the actual suitability scores. Ploiting the differences of observed values and the fitted values presented in a scatter plot, the points show some alignment around the diagonal red dashed line (denoting perfect prediction). An indication of reasonable performance of the fitted model. The predicted scores range between approximately 44 and 74, with majority of the evaluations hovering between scores of 48 to 60. The right residual plot presents the residuals (errors) ranging, with the residuals clustered around zero, and providing some residuals between approximately -6 to +10. It should be noted that some heteroscedasticity is present where the higher predicted model scores rates present slightly larger error terms. Overall, the model presents an acceptable level of performance with errors randomly distributed and unreasonably larger errors at predicted values low.

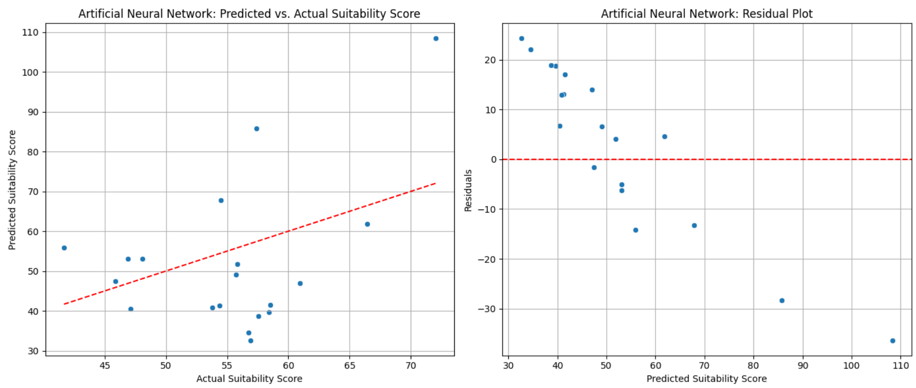

Figure 7. Artificial Neural Network: Predicted vs. Actual Suitability Score

Figure 7 reveals serious performance issues with the ANN model. The left scatter plot indicates very poor prediction performance, with predicted values from 30-110 and actual values from 43-72. There is a very weak relationship between the predicted and actual values, and the predictions are not well aligned with the ideal diagonal line. The residual plot (right) substantiates this conclusion, as we can clearly see large error values between ≈ -30 and +25, while indicating systematic bias as the model over-predicts for some instances and under-predict for others, suggesting inherent problems with the model architecture, training, or hyper-parameter selection or other problems.

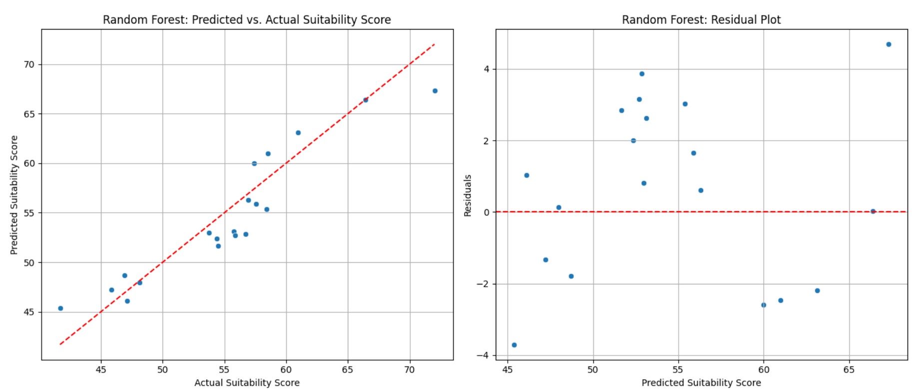

Figure 8. Random Forest: Predicted vs. Actual Suitability Score

In Figure 8, the Random Forest model displays strong predictive performance with points closely hugging the diagonal line in the predicted vs actual plot, showing points ranging from 45 to 72 with actual points matching better than predicted points. The residual plot shows the largest errors mostly contained within the error values of -4 to +5. Most residual errors are compacted around the zero line, again suggesting good accuracy in the model. There is a single residual outlier with a value near +5, but overall, the residuals appear random and evenly spaced without a significant systematic bias in the residuals, suggesting that the Random Forest model has successfully learned the underlying patterns in player suitability as they apply to the geographical regions.

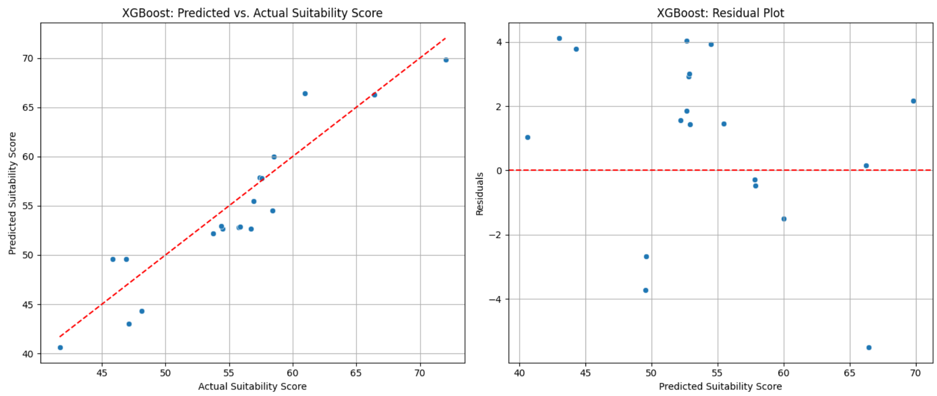

Figure 9. XGBoost: Predicted vs. Actual Suitability Score

From Figure 9, we can see that XGBoost has great predictive performance like Random Forest, with predicted values mapping quite well to actual suitability scores on the diagonal line. The prediction values ranged from about 40 to 70, which covered good ground over the target variable's range. Residuals are reported in a residual plot and cover errors from -4 to +4, indicating most residuals are clustered between -2 and +3. The distribution is reasonably random and does not appear to show any clear-influenced patterns, with a few moderate outliers. Overall, this suggests that XGboost has adequately learned the relationship between features and their effect on player suitability, with minimal systematic bias in predicting the outcome above or below the estimate.

4.5 Overall Model Performance Summary

The results from performance metrics gathered in this study suggest that ensemble machine learning (EML) techniques outperform the other methods used to create models for predicting player auction values within Franchise Cricket Leagues. Both XGBoost and Random Forest had RMSE values of 3.4 to 3.5 and R^2 values greater than 0.84, which indicate a high level of accuracy and a strong explanation for predicting the player values. The Decision Tree produced RMSE values of approximately 5.3 and an R^2 of 0.681; however, these results provide a useful comparison baseline for evaluating the strength of other techniques as well as ensemble methods. The ANNs produced RMSE values of 16.5 and an R^2 of -4.5, which indicates that more tuning of hyperparameters will be required for success with respect to prediction for this task. The regional performance metrics were highly positively correlated (r = 0.81-0.89) to the player suitability scores, which provides strong evidence that the assumptions made about how players are selected have been validated by this framework

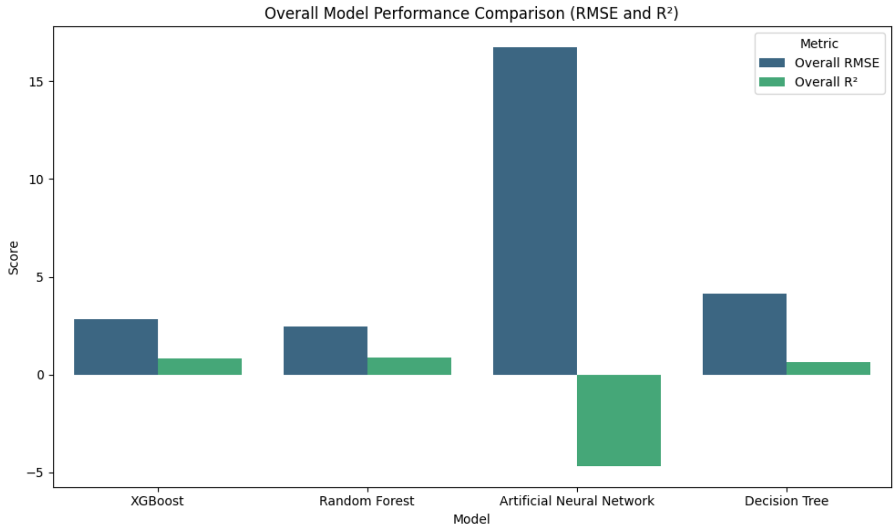

Figure 10 Overall Model Performance Comparison

Figure 10 shows both the RMSE and R^2 scores across the four models. The values of RMSE for XGBoost, for Random Forest, and for Decision Tree were similar or around 2-4, which would suggest they predicted values accurately comparably. The RMSE for ANN, however, is now much higher at approximately 16.5, and so its predictive ability was terrible. For R^2 values (as shown as the green bars), XGBoost and Random Forest had positive R^2 values of around 0.7-0.8, so they had good explanatory power. The Decision Tree had a relatively lower value of around 0.6. However, the ANN's R^2 value was near -4.5, meaning it was predicting worse than just predicting the mean value, so it was catastrophically bad, so the architecture is not likely fit for this prediction task.

5. Conclusion and Future Work

This research successfully establishes a machine learning approach for the identification and assessment of cricket players for franchise league auctions out of international performance statistics. The full study compared a range of machine learning models and showed ensemble methods (XGBoost and Random Forests) produced significantly better output than traditional methods. XGBoost produced the lowest root mean square error of 3.458 with a R^2 value of 0.846; Random Forest produced a root mean square error of 3.475 and R^2 of 0.870. Conversely, Decision Tree and the artificial neural network produced much weaker results, with the ANN producing disastrous results (RMSE: 14.618 and R^2: -1.810). The regional results showed substantial variability of player suitability by geography, with strong positive correlations (0.81-0.89) between regional performance metrics and player suitability scores. The proposed Player Selection Assistant, using ensemble machine learning approaches, accepts a data-driven, transparent and explainable framework for valuing players and classifying them into roles mitigating human bias in cricket auction decision-making.

In future research, there are many areas able to contribute towards model enhancement to improve performance and practical applications. First, hyperparameter tuning, as well as experimentation with more sophisticated deep learning models including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers, would need to be completed to improve the current performance. Second, the models must be improved to account for player form in real-time, psychological components and performance related to pressure situations, necessary to select players able to handle significant moments within a game. Third, reinforcement learning should also be utilized in the modeling of realistic auction sequences, as well as selecting players based on factors that are optimal, and allow teams to simulate bidding in a competitive manner. In addition to this, time-series modeling allowing temporal and seasonal trend investigations; contributing progressively more advanced ball-tracking data; and consulting cricket experts to aid feature engineering would substantially improve the prediction and practical application of the framework.

The validation and expansion of the framework, we suggest, should not stop with examination of the current region, but should also test the model on a few international professional franchise leagues (e.g., Big Bash League, Caribbean Premier League, etc.) to assess model generalizability. Additionally, we suggest that using explainable AI frameworks such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) will help establish a transparent rationale for player "draft" decisions to stakeholders, and coach or franchise management teams will trust their rationale. We also suggest that game-theoretic biases could model competitive bidding behaviors, or if social media sentiment analysis can accurately reflect hedonic value of dormancy, or we could build out a continuous learning framework to develop predictive constructions based on actual player league performance. All these research opportunities support developing a dynamic, all-inclusive framework that grows itself. Each of these efforts together could aim to translate current research into a practical, realized decision-support tool that combines machine learning sophistication, cricket subject area, and real-world auction dynamics.

References

A. Shenoy, “Prediction of the outcome of a Twenty-20 Cricket Match: A Machine Learning Approach,” arXiv preprint arXiv:2209.06346, 2022.

S. Hundekari, “Prediction System for Indian Premier League using Machine Learning Algorithms,” Int. J. Recent Psychol. Res., vol. 5, no. 4, 2024.

K. Passi and N. Pandey, “Increased Prediction Accuracy in the Game of Cricket Using Machine Learning,” arXiv preprint arXiv:1804.04226, 2018.

B. Bhattacharjee, “Optimizing Fantasy Sports Team Selection with Deep Reinforcement Learning,” arXiv preprint arXiv:2412.19215, 2024.

S. Alaka, “Efficient Feature Representations for Cricket Data Analysis Using Deep Learning-Based Multi-Modal Fusion Model,” arXiv preprint arXiv:2108.07139, 2021.

G. Malhotra, "A comprehensive approach to predict auction prices and economic value creation of cricketers in the Indian Premier League (IPL)", Journal of Sports Analytics, vol. 8, no. 3, pp. 149–170, 2022, IOS Press.

Md. Robel, Md. A. R. Khan, I. Ahammad, Md. M. Alam and K. Hasan, "Cricket players selection for national team and franchise league using machine learning algorithms", Cloud Computing and Data Science, vol. 5, pp. 108–139, 2024.

M. Chandru and B. Jayanthi, "A Comprehensive Analysis of Predicting Individual Cricket Players' Performance: A Survey", Indian Journal of Natural Sciences, vol. 12, no. 70, Feb. 2022.

M. Kasande and S. Jahirabadkar, "Automated Player Selection for a Cricket Team using Machine Learning," Grenze International Journal of Engineering and Technology, Jan. 2023.

M. K. Manju and A. O. Philip, "Novel method for ranking batsmen in Indian Premier League," Data Science and Management, vol. 6, pp. 158-173, 2023.

S. K. C, A. Khetan, B. Kumar, D. Tolani, and H. Patel, “Prediction of IPL Match Outcome Using Machine Learning Techniques,” arxiv.org, Sep. 2021, Available: https://arxiv.org/abs/2110.01395.

R. Bunker and T. Susnjak, “The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review,” Journal of Artificial Intelligence Research, vol. 73, pp. 1285–1322, Apr. 2022, doi: https://doi.org/10.1613/jair.1.13509.

“Cricket Win Prediction using Machine Learning,” Ijraset.com, 2023. https://www.ijraset.com/research-paper/cricket-win-prediction-using-machine-learning.

Basit, Abdul, Muhammad Bux Alvi, Fawwad Hassan Jaskani, Majdah Alvi, Kashif H. Memon, and Rehan Ali Shah, "ICC T20 Cricket World Cup 2020 winner prediction using machine learning techniques." In 2020 IEEE 23rd International Multitopic Conference (INMIC), pp. 1-6. IEEE, 2020.

Bunker, Rory & Thabtah, Fadi, “A Machine Learning Framework for Sport Result Prediction”, 2017.

Kapadia, Kumash, Hussein Abdel-Jaber, Fadi Thabtah, and Wael Hadi, "Sport analytics for cricket game results using machine learning: An experimental study." Applied Computing and Informatics ahead-of-print, 2020.

Singh, Tejinder, Vishal Singla, and Parteek Bhatia, "Score and winning prediction in cricket through data mining." 2015 international conference on soft computing techniques and implementations (ICSCTI). IEEE, 2015.

Biographies

Jubayer Ahamed completed his B.Sc. in Computer Science from American International University-Bangladesh, Dhaka, Bangladesh. He obtained his MSc from American International University-Bangladesh, Dhaka, Bangladesh. Currently, Jubayer Ahamed is working as a Lecturer in the Department of Computer Science (CS) at American International University- Bangladesh (AIUB). He is also associated with Artificial Intelligence Research & Innovation Lab (AIRIL) as a research associate. His research interests include Software Engineering and Machine Learning.

Arnob Aich Anurag is currently pursuing his Bachelor of Science in Computer Science and Engineering from American International University-Bangladesh, Dhaka, Bangladesh. He serves as a Research Intern at Advanced Machine Intelligence Research Lab while actively contributing as a Web Developer in the IEEE AIUB Student Branch. Currently he is working as a reviewer for International Journal of Electrical and Computer Engineering (IJECE) and IEEE Access.

Shaikat Das Joy is a Lecturer in the Computer Science Department at American International University-Bangladesh (AIUB). He completed his M.Tech in CSE from IIT Ropar as an ICCR scholar (2021-23) and earned his B.Tech in CSE with specialization in Information Security from VIT Vellore. He received the 2nd Best Paper Award at ICA23. His research interests include Natural Language Processing, Machine Learning and Data Science.

MD Sajid Bin-Faisal is working as an Assistant Professor in the department of Computer Science under the Faculty of Science and Technology at American International University-Bangladesh (AIUB). His research interests include network security, graph theory and Internet of Things. He was awarded as the best presenter on CITIC 2025 held in Malaysia.

Prof. Dr. Dip Nandi pursued his B.Sc. in Computer Science from American International University-Bangladesh, Dhaka, Bangladesh. He obtained his MSc from the University of Melbourne and Ph.D. from RMIT University, Australia. Currently, Dr. Dip Nandi is the Associate Dean of the Department of Computer Science, at American International University-Bangladesh. His research interests include Software Engineering, Technology E-learning, Theory of Learning, and Human-Computer Interaction.

Prof. Dr. Md. Asraf Ali is a man of Jessore and currently working as a Professor of Computer Science at American International University-Bangladesh (AIUB). He is actively involved in research in the fields of Biological Signal Processing, Bioinformatics, and Machine Learning. As for his research outcomes, he has published more than 50 research articles (Journal/Conference) indexed in ISI or SCOPUS. As the community service related to his research, he is working as a Reviewer, Technical Committee member (TPC), TPC Chair, and Session Chair of many International Journals and Conferences. He is the principal investigator of Artificial Intelligence Research & Innovation Lab (AIRIL). Moreover, he is also working as an External Examiner for evaluating PhD Thesis of several universities in India.