International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

Performance Analysis and Optimization of Player Selection Using Machine Learning Algorithms

Prof. Ms. Pooja Pohare[1], Krishna M. Kashikar[2], Gaurav B. Wankhade[3], Aniket V. Rajguru[4], Prashik T. Ingle[5]

[1]*Asst. prof. Dept. of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon-444203, India*

[2]*UG Scholar, Dept. of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon-444203, India*

[3]*UG Scholar, Dept. of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon-444203, India*

[4]*UG Scholar, Dept. of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon-444203, India*

[5]*UG Scholar, Dept. of CSE, Shri Sant Gajanan Maharaj College of Engineering, Shegaon-444203, India*

Abstract

Games like Cricket and Football are very much popular in the world. In India IPL is very much popular, in the similar way games like Pro-Kabaddi, Hockey are also become very popular. Player selection is one the difficult and crucial task in such sports, where the performance of the team is heavily depends on the selection of right players. This research aims to develop an application that employs various algorithms which involves RF, KNN and LR, in order to predict the optimal player line-up based on historical performance data. The study also emphasizes the application of optimization techniques like hyper parameter tuning and cross validation to enhance the accuracy for the prediction of the models. The performance of these models is evaluated using metrices such as Accuracy, Precision, Recall and F1-Score, with results compared to assess the most effective approach for player selection in each sport depending on the selected dataset under one roof.

The results of the study show that all three Machine Learning models – RF, KNN, and LR – can predict player selection with varying degrees of accuracy. After optimization with Grid Search method, Random Forest emerged as the most accurate model for football and cricket, while KNN performed well in Pro-Kabaddi. Expanding beyond traditional Machine learning models to deep learning techniques like Neural Networks could capture more patterns in player performance data.

Keywords: Grid Search Method; KNN; Logistic Regression; Machine Learning; Random Forest

1. INTRODUCTION

An easy Machine learning algorithm, such as Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression, have emerged as powerful tools for predicting player selection in sports analytics. These algorithms leverage historical performance data to forecast the best player line-up for upcoming matches in various sports such as cricket, hockey, Pro-Kabaddi, and football. The ability to analyse vast amounts of player data and make accurate predictions makes these algorithms highly valuable in optimizing team performance. This research focuses on applying these algorithms to predict player selection, with an emphasis on optimizing the models through Grid Search to improve prediction accuracy. By fine-tuning hyper parameters, the research aims to develop an application that can be used in real-time by sports teams to make informed decisions about player selection.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

1.1. Random Forest Algorithm

One well-liked algorithm that is typically utilized for classification and regression problems is Random Forest. One type of supervised learning algorithm in machine learning is the random forest algorithm. the fundamentals of Random Forest, in which every tree makes a prediction based on subsets of the data. With the Random Forest technique, several decision tree classifiers are used. Initially, every decision tree undergoes independent training. The random forest then forecasts the average of these outcomes after taking the predictions from these trees.

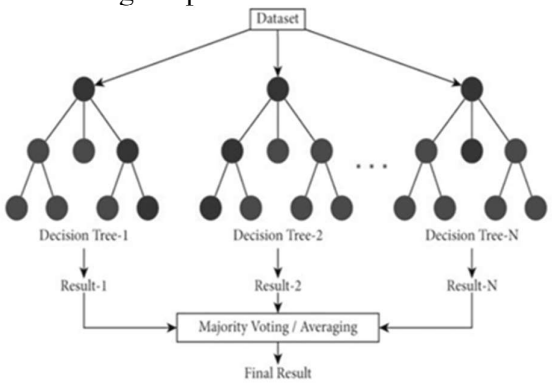

Figure 1. Random Forest Trees

1.2. Logistic Regression Algorithm

A binary classification system called logistic regression forecasts one of two probable outcomes, such as yes/no or 0/1. It uses a linear equation to model the link between the input features and the output. The outcome is converted into a probability between 0 and 1 by passing it via a sigmoid function. The model predicts class 1 if the likelihood is more than or equal to 0.5, and class 0 otherwise.

The model learns the optimal weights and bias during training, typically using gradient descent to minimize a loss function like binary cross-entropy. Essentially, logistic regression uses a linear combination of features to predict probabilities and classify inputs based on a threshold.

1.3. KNN Algorithm

K-Nearest Neighbors (KNN) is a basic classification algorithm that assigns a class to an input based on the 'k' nearest data points, using a distance metric like Euclidean distance. The input is classified by the most frequent class among these nearest neighbors. The number of neighbors, 'k', affects the classification, with results determined by majority voting (or averaging in regression). KNN doesn't require a separate training phase, as it makes predictions directly from the dataset during testing. The non-parametric, supervised learning k-nearest neighbours (KNN) algorithm is mostly helpful for classification and prediction.

2. LITERATURE SURVEY

A literature survey can highlight the methodologies others have used and guide you to design your own research work. The application of machine learning for player selection prediction across sports like cricket, football, Pro Kabaddi, and hockey has gained significant attention in recent years. Several studies (2022-2024) have explored the effectiveness of various algorithms such as RF, KNN, and Logistic Regression.

In order to analyze and predict data various research is already made in the field of Cricket, Kabaddi, Football, Hockey, depending on the various parameter. Literature Survey basically guides to solve the unsolved problems.

S. Biradar [5] Et.al. Says that Kabaddi match result is predicted using Logistic Regression Supervised Learning Algorithm. The implementation of Logistic regression will give accuracy of 78%.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

BosuBabu [3] Et.al says that finding the particular position for a player is difficult hence the problem of finding a player's suitable position in a football team is sort out with the help of Random Forest Algorithm and binary relevance technique.

Manish S [4]. Et.al says that Football match result depends totally on the player performance and using Multiple Regression algorithm players performance will be conducted with high efficiency.

S. Teja [2] Et.al. works on cricket player selection using machine learning algorithms like Random Forest, AdaBoost, Support Vector Machines, etc. where Random Forest and AdaBoost achieved highest accuracy of 98.68%.

Harshit Barot [1] Et.al works on Analysis and prediction for IPL league in order to find which algorithm gives the best accuracy. Various algorithm likes SVM, Logistic Regression, Decision Tree, Naive Bayes were used for prediction out which Logistic Regression give higher accuracy 95% for match predications.

3. PROPOSED SYSTEM BLOCK DIAGRAM

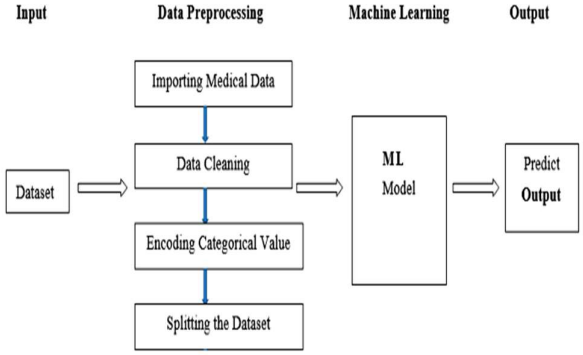

Figure 2. Block Diagram

To start the prediction process, we begin by importing essential libraries such as pandas, NumPy, and machine learning models like Random Forest Classifier, Logistic Regression, or K-Neighbours Classifier from Sklearn. After that, we load the dataset using pandas. read_csv() to read the data. We then examine the dataset with methods like df. head() and df.info() to understand its structure and identify both the features and the target variable. The next step is to separate the features from the target by selecting the appropriate columns

We split the dataset into training and testing sets using train_test_split from sklearn. model selection, typically dividing the data in an 80/20 or 70/30 ratio. Once split, we initialize the chosen model (e.g., RandomForestClassifier, Logistic Regression, or K-Neighbors (Classifier) and train it on the training data using model. Fit (X_train, y_train). After training, we make predictions on the test set with model. predict (X_test). Finally, we evaluate the model's performance by comparing the predicted results with the actual test labels and calculating metrics such as accuracy using accuracy_score from sklearn. metrics.

Grid Search is a technique used to optimize a model by evaluating multiple hyperparameter combinations to find the best performing ones. To begin, we import GridSearchCV from sklearn.model_selection and specify a grid of hyperparameters to test (e.g., the number of trees for Random Forest or the regularization strength for Logistic Regression). We then set up GridSearchCV with the model, the hyperparameter grid, and the number of cross-validation folds. After fitting the model with grid_search.fit (X_train, y_train), Grid Search evaluates all possible hyperparameter combinations and selects the one that performs best based on metrics like accuracy. The optimized model can then be used for predictions and further evaluation.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

4. DATASETS AND TECHNOLOGIES

In this proposed model we have used 4 different dataset viz. cricket_players.csv, football_player.csv, hockey_player.csv and kabaddi_player.csv. Each data set contains 200 rows of player information containing various fields. All these datasets are dummy datasets created just for model working. Different Machine learning algorithms like RF, KNN and LRclassification are used in Python and Flask environment in order to predict and display the result on webpage.

Python Flask is a lightweight and flexible web framework used for building web applications. It provides essential features like routing, request handling, and templating through Jinja2, allowing developers to create dynamic web pages with ease. Flask is designed to be minimalistic, offering simplicity and flexibility, making it ideal for smaller projects or when developers want full control over their app's components. It supports the addition of extra features through extensions, such as authentication, database integration, and form validation. Its modularity and ease of use make it a popular choice for creating scalable and efficient web applications.

5. METHODOLOGY

The methodology of the Flask application follows a structured process to predict player selection for various sports using machine learning models. Here's a step-by-step breakdown of the methodology:

5.1. Data Loading

The application begins by loading datasets for four different sports (Cricket, Kabaddi, Football, and Hockey) using the load_dataset function. These datasets are read into pandas Data Frames from CSV files.

5.2. Data Preprocessing

The train model function preprocesses the data by selecting relevant features (e.g., batting average, fitness, etc.) and standardizing them using StandardScaler to ensure all features are on the same scale.

5.3. Model Training

A Random Forest model is trained for each sport using the preprocessed data. The data is split into training and testing sets using train_test_split from sklearn. Training of model is carried out using the trained data, and predictions are made on the testing data.

5.4. Evaluation

The model's performance is evaluated using common classification metrics such as accuracy, precision, recall, F1 score, and confusion matrix. These metrics help assess how well the model is performing in distinguishing between selected and non-selected players.

5.5. User Interface

The Flask app serves as the user interface. Users can select a sport (Cricket, Kabaddi, Football, or Hockey) and input relevant performance metrics (e.g., batting average, fitness). The app dynamically displays the input fields based on the selected sport.

5.6. Prediction

After the user submits the input data, the app applies the trained model for the selected sport. The model predicts whether the player is "Selected" or "Not Selected" based on the input features

5.7. Results Display

The app displays the prediction, along with the evaluation metrics (accuracy, precision, recall, F1 score), and a confusion matrix visualized as a heatmap. The confusion matrix provides insight into the model's true positives, true negatives, false positives, and false negatives

5.8. Visualizations

The confusion matrix is generated using seaborn to create a heatmap, which is then encoded as a base64 image and sent to the frontend for display. This visualization aids in understanding the model's performance in terms of classification errors.

6. RESULTS AND DISCUSSION

The analysis of different machine learning models—KNN, Logistic Regression (LR), and Random

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

Forest (RF)—on player selection prediction for Cricket, Football, Pro-Kabaddi, and Hockey demonstrated that Random Forest consistently outperformed the other models across all metrics, including precision, recall, accuracy, and F1 score

Logistic Regression showed stable performance with high recall and F1 scores, while KNN performed well but with slightly lower metrics in comparison. Optimization improved the models, especially KNN and Random Forest, enhancing their precision and recall. Overall, Random Forest proved to be the most robust model for predicting player selection, while Logistic Regression and KNN also provided reliable results.

Table 1. Results For Cricket

	Cricket				
	Predic-tion	Preci-sion	Accu-racy	Re-call	F1 Score
KNN	Selected	0.8666	0.9074	0.9423	0.9245
KNN-OPT	Selected	0.8833	0.9245	0.9423	0.9333
LR	Selected	0.8833	0.8813	1	0.9369
LR-OPT	Selected	0.8833	0.8813	1	0.9369
RF	Selected	0.9166	0.9272	0.9807	0.9532
RF- OPT	Selected	0.9166	0.9272	0.9807	0.9532

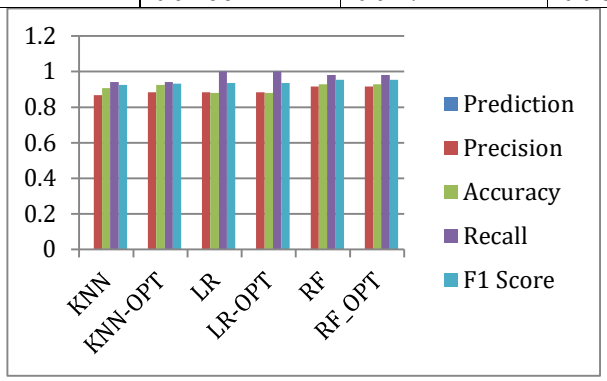

Figure 3. Analysis for Cricket Game

Table 2. Results For Pro-Kabaddi

	Pro-Kabaddi				
	Predic-tion	Preci-sion	Accu-racy	Re-call	F1 Score
KNN	Selected	0.9166	0.96	0.9411	0.9504
KNN-OPT	Selected	0.9166	0.96	0.9411	0.9504
LR	Selected	0.9166	0.96	0.9411	0.9504
LR-OPT	Selected	0.85	0.9565	0.8672	0.9072
RF	Selected	0.9166	0.9791	0.9215	0.9494
RF- OPT	Selected	0.9333	0.9795	0.9411	0.96

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

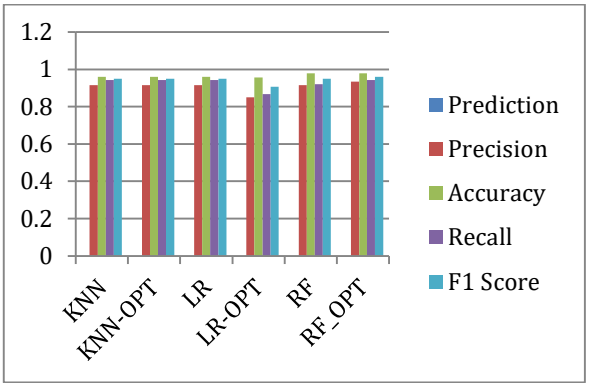

Figure 4. Analysis for Kabaddi Game

In the similar manner we also find the analytical performance for football and hockey game also. Here also we find the predictions using RF, KNN and LR classification method. We also used Grid search method for optimization.
The confusion matrix for the cricket and kabaddi game is given herewith.

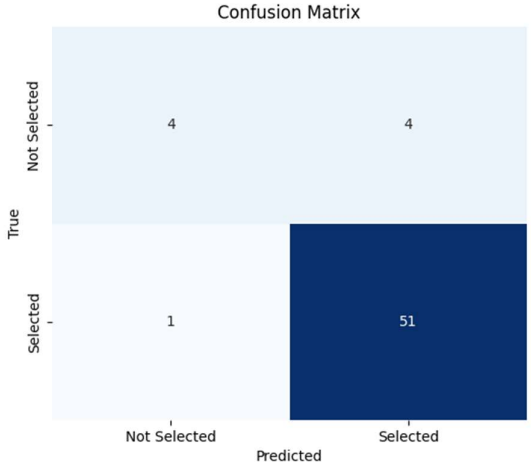

Figure 5. CM for Cricket using RF

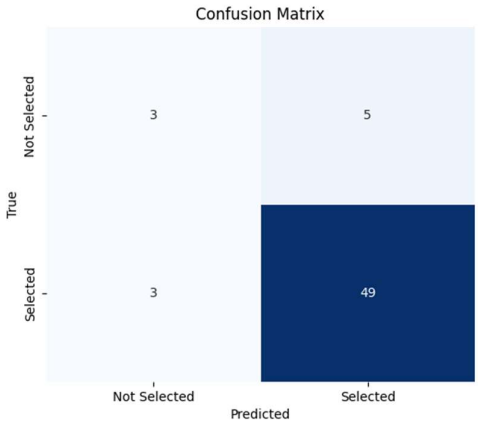

Figure 6. CM for Cricket using KNN

We got the Confusion Matrix values for others as follows

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

Table 3. CM Values

Algorithm	CM Values
KNN-OPT	4 4 3 49
LR	1 7 0 52
LR-OPT	1 7 0 52
RF_OPT	4 4 1 51

7. CONCLUSION

In summary, the player selection prediction model, using algorithms such as Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression, displayed different performance levels. Through optimization with grid search to adjust hyperparameters, there was a significant enhancement in metrics like accuracy, precision, recall, and F1 score. The confusion matrix further illustrated the models' capabilities in making accurate predictions. Overall, these optimized algorithms show great potential for improving the player selection process.

8. FUTURE SCOPE

The future scope of the player selection prediction model offers several exciting opportunities for improvement and expansion. One potential advancement is exploring other algorithms, such as Support Vector Machines or Neural Networks, to compare their performance against the current models. The dataset is expanded in order to include a greater range of variables, like player stats, past performance, and even psychological traits, could lead to more accurate predictions

Additionally, the model could be adapted to work with real-time data, allowing for continuous updates and better decision-making. Exploring more advanced hyper parameter tuning techniques, like Random Search or Bayesian Optimization, could also further optimize the models. As the model evolves, it could be integrated into sports analytics platforms, offering coaches and analysts valuable tools for data-driven player selection.

REFERENCES

[1]. Harshit Barot, Arya Kothari, Pramod Bide, Bhavya Ahir, Romit Kankaria "Analysis and Prediction for the Indian Premier League"*2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020*

[2]. Sri Ram Teja, T. Pavan Kalyan, V. Akhil Kumar Reddy "Cricket Player Selection using Machine Learning" *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

[3]. S. BosuBabu, V. Vivek, D. M. T. Kumar, K. Prathyusha, and G. P. Teja, “Predicting football player’s position,” *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, no. 5, May 2022.

[4]. M. S., V. Bhagat, and P. R. M., “Prediction of football players' performance using machine learning and deep learning algorithms,” in *2021 2nd International Conference for Emerging Technology (INCET)*, 2021, pp. 1–6, doi: 10.1109/INCET51464.2021.9456424

[5]. Shrinivas Biradar, Kavya M, Mahima K. N, Manasa Roopini N, and Meghana M, “Pro Kabaddi winner prediction using logistic regression,” *Int. J. Adv. Res. Sci. Commun. Technol. (IJARSCT)*, vol. 3, no. 10, May 2023.

[6]. M. K. Mahbub, M. A. M. Sikder, M. S. Islam, S. Sorna, S. Hossain, and M. Biswas, “Best eleven

DOI:10.46647/ijetms.2025.v09i02.041 ISSN: 2581-4621

forecasts for Bangladesh cricket team with machine learning techniques," in *2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2021, pp. 1–6, doi: 10.1109/ICEEICT53905.2021.9667862.
[7]. A. Basit, M. B. Alvi, F. H. Jaskani, M. Alvi, K. H. Memon, and R. A. Shah, "ICC T20 Cricket World Cup 2020 winner prediction using machine learning techniques," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–6, doi: 10.1109/INMIC50486.2020.9318077.
[8]. S. Singla and S. S. Shukla, "Integer optimization for Dream 11 cricket team selection," *Int. J. Comput. Sci. Eng.*, vol. 8, no. 11, Nov. 2020, doi: 10.26438/ijcse/v8i11.16.
[9]. N. M. Patil, B. H. Sequeira, N. N. Gonsalves, and A. A. Singh, "Cricket Team Prediction Using Machine Learning Techniques,"
[10]. M. Ishi, J. Patil, and V. Patil, "An efficient team prediction for one-day international matches using a hybrid approach of CS-PSO and machine learning algorithms," *Array*, vol. 14, p. 100144, Apr. 2022, doi: 10.1016/j.array.2022.100144.
[11]. B. Sakthivel, A. A. Abinaya, J. R. M. Juraidha, M. T. Madhuvarshini, and T. B. Tamilsudar, "Cricket team prediction using machine learning," *Int. J. Curr. Eng. Sci. Res. (IJCESR)*, vol. 10, no. 4, pp. 102–106, 2023, doi: 10.21276/ijcesr.2023.10.4.10.
[12]. N. Dhonge, S. Dhole, N. Wavre, M. Pardakhe, and A. Nagarale, "IPL cricket score and winning prediction using machine learning techniques," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 3, no. 5, pp. 1723–1728, May 2021.
[13]. S. K. Gupta and S. K. Sharma, "Analyzing and predicting the performance of players using machine learning," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Coimbatore, India, 2023, pp. 1–5.
[14]. S. K. Gupta and S. K. Sharma, "PlayerRank: Leveraging learning-to-rank AI for player positioning in cricket," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Coimbatore, India, 2023, pp. 1–5.
[15]. S. K. Gupta and S. K. Sharma, "Optimal cricket team selection using machine learning," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Coimbatore, India, 2023, pp. 1–5.