10.48047/jocaaa.2024.33.08.341

CRICKET PLAYER RATING USING MACHINE LEARNING ALGORITHMS

Bh.V. RamaKrishna
Associate Professor,
Department of Computer Science and Engineering (AI & DS),
Vignan Institute of Technology and Science, Deshmukhi, Telangana-508284
Email: drbvrk78@gmail.com

B. Sushma
Assistant Professor,
Department of Information Technology,
MLR Institute of Technology, Medchal, Telangana-508284
Email:bhattiprolusushma@gmail.com

Dr Bobba BasaveswarRao
Head of Computer Center,
Acharya Nagarjuna University,
Guntur, Andhra Pradesh-522510
Email:bbrao@gmail.com

***Abstract*—Cricket is a popular highest fan following game in the world. Cricket play is a team play each player has his own skill to support his team. Assessment of player quality traditionally depended over coach observations, historical statistics and subjective evaluation. There is a bias of inconsistency exist in traditional validations. Sports Analytics is a new dimensional approach provides computer system analysis to comprehensive cricket player assessment. The novel knowledge learning techniques incorporated with several statistical measures and algorithms which perform classification, scoring, predicting and trend analysis. Potential application of machine learning techniques assists coaches, selectors and stakeholders to make better decision making in choosing the players towards win rate optimality.**

***Keywords*—Machine Learning, KNN, Clustering, Classification, Prediction, Algorithms.**

I. INTRODUCTION

The application of Machine Learning algorithms conduct data analysis in various knowledge dimensions [5]. Sports data analytics is becoming an interest to study the player statistics in knowledge features extraction perspective [7]. With increased application of IoT devices in modern cricket playgrounds multivariate data extraction is possible [9]. The digital computers interlaced with smart sensors providing precision enriched data collection with accuracy [3]. Multi dimensional data modeling allow large-scale features analyzed for complex relationships understanding [11]. The modern ML algorithms used to predict match outcomes, player rating and performance [1]. Some of the major factors like batting, bowling, fielding efficiency based on situational performance [12]. Predicting the player performance in upcoming games made possible with machine learning algorithms like Random Forest, Gradient Boost, SVM and KNN [2]. A Deep Player Performance Index (DPPI) proposed as performance metric to identify the player strengths and role in team for effective game play [10]. Deep learning techniques like CNN, GRU and GAN are widely accepted for image processing based knowledge extraction to dynamically identify the batsmen strike rate and pressure conditions influencing boundary shots [4]. Using Gradient algorithms with loss functions over CNN layers to study live cricket player strategies and game play skills. The positioning of players in a match can be predicted towards optimal match win rate identified by probabilistic learning algorithms. Neural networks or multilayer perceptron are the fundamental ranking models [15][16].

In this paper Section II describes machine learning algorithms application. Section III showcases the proposed model layout and description of each module role in player rating. Section IV reveals the methodology of evaluation on data for knowledge patterns.

II. MACHINE LEARNING ALGORITHMS

In this, a wide range of efficient approaches are included for detecting the diabetes mellitus with high accuracy and performance metrics.

A. Logistic Regression (LR)

A statistical binary categorization method for predicting the probability of target variable using a logistic function [9]. It is used to predict the categorical dependent variable and a set of independent variables [9]. The probabilistic values lie between 0 and 1 [9]. It takes numerical values as input and gives binary value [9].

B. K-Nearest Neighbors (KNN)

A Data points are classified using a non-parametric approach that takes majority class of close neighbors[10]. To place new data into related category by comparing it with current data [10]. It uses Euclidean distance to classify [10].

Euclidean distance is given by

$$D = \left[\sum_{j=1}^{n}(u_j - v_j)^2\right]^{05} \quad (1)$$

C. Support Vector Machine (SVM)

An algorithm that builds hyperplanes to separate the classes [11].It improves accuracy of classification by rising the range of classes through support vectors [11]. Kernel functions are used for non-linear separations [11].

D. Decision Tree (DT)

A tree-structured algorithm that splits data based on feature importance by calculating the entropy for effective classification tasks [12].The branches connecting nodes of data are used to determine the estimates [12].
Entropy of Data is given by

10.48047/jocaaa.2024.33.08.341

$$E(Ins) = -\sum_{i=1}^{n} p_{a_j} \log_2 p_{a_j} \quad (2)$$

Information Gain is given by

$$Gain = E(Ins) - \sum_{i1}^{n} \frac{|Ins_i|}{|Ins|} E(Ins_i) \quad (3)$$

E. Random Forest (RF)

An approach that creates many choice trees and classify the data according to Aggregates of predictions [13]. It is robust to overfitting and provides high accuracy[13].

F. Gradient Boosting Machine (GBM)

A sequential boosting technique that optimizes the model by minimizing a loss function [14]. It is effective for handling complex datasets [14]. It is improving DTs by including poor learners gradually to form a strong prediction model [14]. It can deal a variety of data types and give robust predictions with excellent performance [14].

G. XGBoost

It is said to be an advanced form of gradient boosting that utilises regularization and parallel processing for better performance and efficiency [15].

H. Clustering

Clustering is an unsupervised learning technique which groups similar data points into clusters based on some similarity measures. Clusters are categorized as Hard and Soft clusters. Hard clustering strictly binds each data point to exactly one cluster widely accepted for segmentation tasks. Soft Clustering on other hand assigns each data point to multiple clusters depending on probability degree of membership useful for fuzzy classifications and overlapped boundaries.

I. Outlier Analysis

Outlier Analysis explore the significantly distinct data points that deviates from distribution anomaly. In our work collective outliers, contextual outliers and global outliers helps to identify player discrepancies during normal game play. Detects exceptional performances by players. Analyze player suitability to match during risk transitions.

Each algorithm undergoes cross validation and hyperparameter tuning to get optimal performance. Evaluation metrics provide the analysis of model performance. This Systematic approach helps in the choosing of best-Performing algorithm for prediction.

III. Proposed Model

Figure 1 shows the proposed model for cricket player skill assessment and rating using machine learning algorithms.

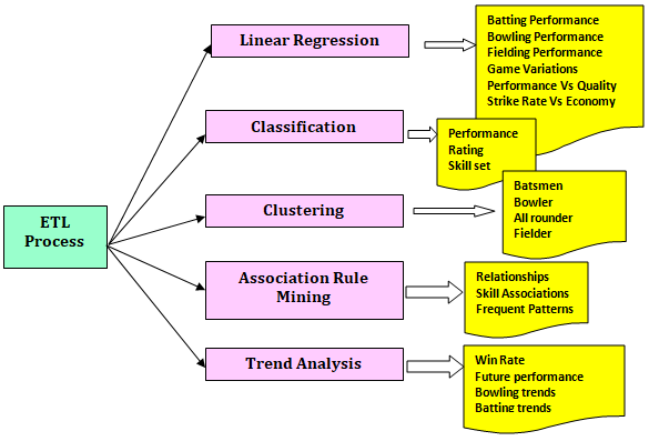

Fig. 1. Flowchart of Player Skill Assessment Model.

The model is a collection of phased process to evaluate player skill assessment using variety of machine learning techniques.

A. ETL Process

In ETL process a staged transformation takes place to raw cricket data into clean, structured, feature-rich data that can be used for training ML models such as Random Forest, XG-Boost, Regression, Clustering, or Rating Algorithms.

B. Knowledge Learners

In this stage various machine learning algorithms are applied over training data which is prepared using cross validation process for model building. Several interesting measures are identified to estimate skill rates and player behavior assessment during matches.

C. Assessment patterns

For both batsman and bowler performance analysis summaries are generated. Batsman strike rate and economy based on skill sets applied over regression models for variation estimates. Classification of players using different contexts can be done.

IV. Methodology

In this work we collected wide variety of data for cricket player skill set prediction. The player's performance, roles and historical data with features like batting, bowling, fielding, wicket keeping and all rounder abilities considered in skill evaluation by learning models.

Cricket player dataset from Kaggle related to T20 cricket statistics. This includes 1260 instances with 12 features among which 5 considered as features and 1 target column. Which identifies batsmen clustering and factor analysis. Where 4 features are used with 1 target column for bowler skill predictions.

The below table shows the vivid Ml techniques applied for knowledge pattern identifications over cricket player data.

TABLE 1: ML techniques in player assessment

10.48047/jocaaa.2024.33.08.341

ML Technique	*Application*
Linear Regression	Batting Skill Assessment
	Bowling Assessment
	Wickets Variance
	Pitch difficulty assessment
	SR Assessment
	Consistency
Logistic Regression	Predicting batsman future score
	Predicting Bowler wickets
	Predicting player performance
XG-Boost	Batsmen Performance
	Innings Performance
	Wicket probability
	Pressure Indexes
	Match Impact Factors
	Economy
Random Forest	Categorizing Bowlers
	Domestic Performance
	SR rating
	Classification
	Pitch Resistance
Outlier Analysis	Bowler Inconsistencies
	Game changing events
	Discovering new talent
	True and false events
	Triggering zones

V. RESULTS AND DISCUSSIONS

Fig. 2 evaluates the batsman strike rate and consistency of runs for the future matches. Bowler consistency prediction finds the overall skill rate probability in future matches.

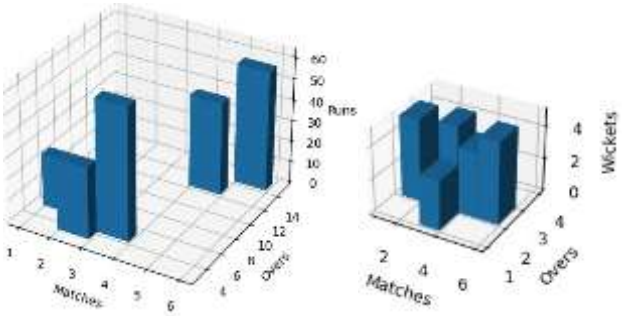

Fig. 2. Batsman, Bowler performance prediction

Fig. 3 shows the result analysis of batsman pressure rate during game play influencing his strike rate. This visualization shows the precision scale of variation or peak stages of pressure application by batsman at specific time stamps.

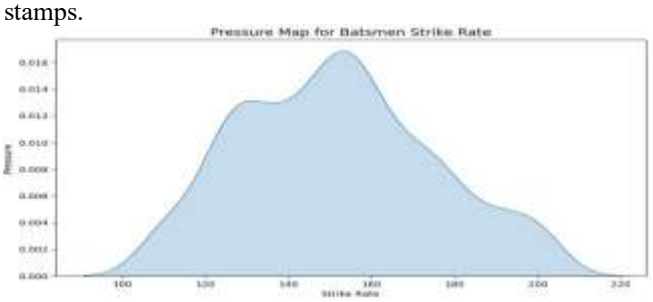

Fig. 3. Pressure Vs Strike rate density mapping

In fig. 4 a heatmap visualization of ball landing positions are analyzed to identify wicket taking probaility zones. A heatmap used to identify high and weak performance zones with relationship strengths among variables. We can find correlation rate among datapoints in distribution.alues refers to the values that are not available.

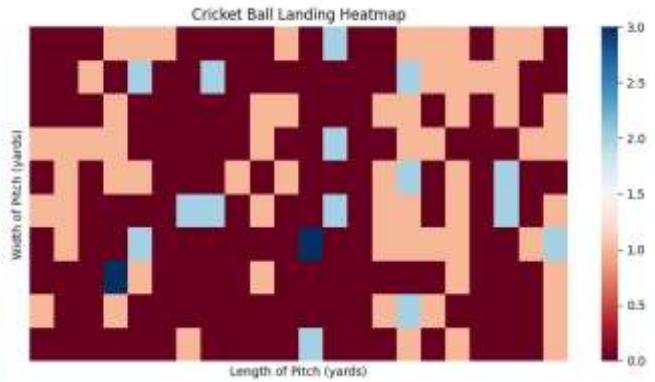

Fig. 4. HeatMap for Ball landing positions on pitch in T20

The PCA offers dimensionality reduction and transforms many correlated variables into small data components. The major components considered in our analysis are Batting Component, Bowling Control Component, Match-Impact Component and Consistency component. each component can be aligned to an axis in PCA for distribution of data assessment. Mainly PCA supports player ranking, Clustering, Role classification and variance estimation.

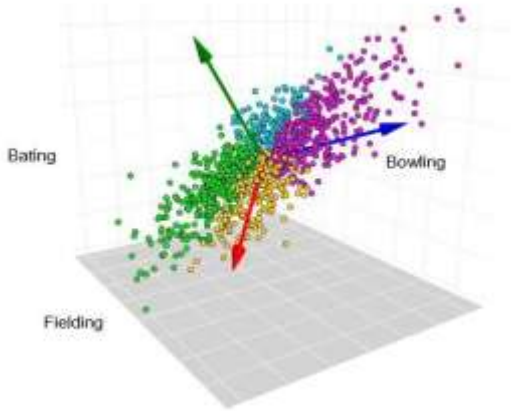

Fig. 5. PCA of Cricket Player roles in T20 matches.

The PCA(Principle Component Analysis) reduces the high dimensional data into smaller sets of uncorrelated subsets. In this work we applied on T20 players roles in matches to investigate the denisty of players skills during gameplay like fielding rate, bowling rate and bating rate. The figure 5 represents the variance in player roles over certain time periods. This analysis identify the time series oriented key roles played by players with the rate. The fielding and bowling distributions identifies the rate of boundaries done by players during a match. Also the directions of major boundary placements.

The application of clustering over batsman performance gives a clear picture of over based performance and match based performance groups of players having similar skill

10.48047/jocaaa.2024.33.08.341

sets. Fig. 6 shows the T20 batsman performance based clustering based on run score per over ratio metric.

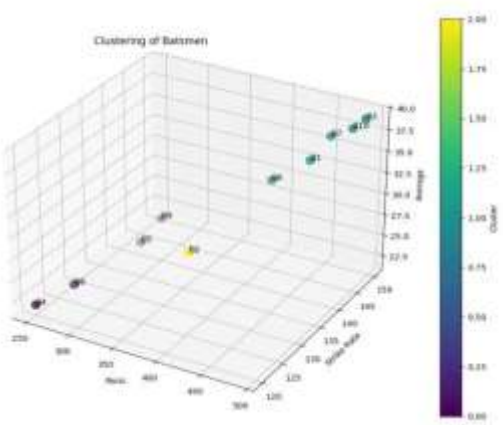

Fig. 6. Clustering of T20 match batsman performance

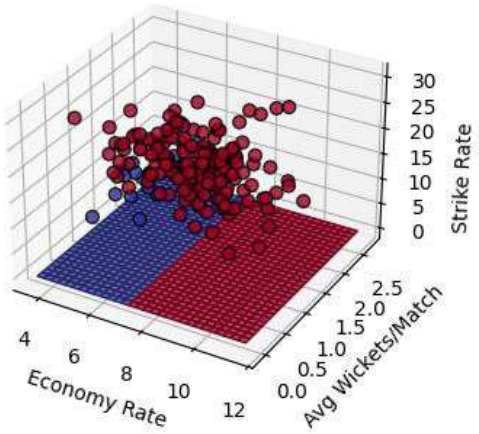

Fig. 7. Clustering of T20 match batsman performance

The clustering is a unsupervised learning technique which enables to group players based on some statisticcal metrics. In this work batting features are clustered over various batsman during T20 match. Based on strike rate and economy rate we can group batsman having similar skills playing in different teams. This knowledge helps to identify the clone skill players among T20 cricket teams. Figures 6 & 7 depicts the batsman groups having similarity among strike rate and economy rate. This provides coaches to identify best batsmen in a match.

The Linear regression applied over player match statistics shown in Fig. 8 gives an insight into batsman performance in facing balls in an over. The R^2 explains the variance in distribution xtremely useful to prediction of success, also useful to compare multiple models and underfitting conditions. On the other hand Logistic regression can be applied over batsman dynamics considering more precission scaled values to identify scientifc precision predictions. Some of the uses like wind and ball speed effects over strike rate. Bowler momentum impact over ball landing position. Parabolic projection to boundary placement accuracy. Also helpful to predict pitch reports and players power play performance. We can apply decision tree regression model for handling non linear data relationships.

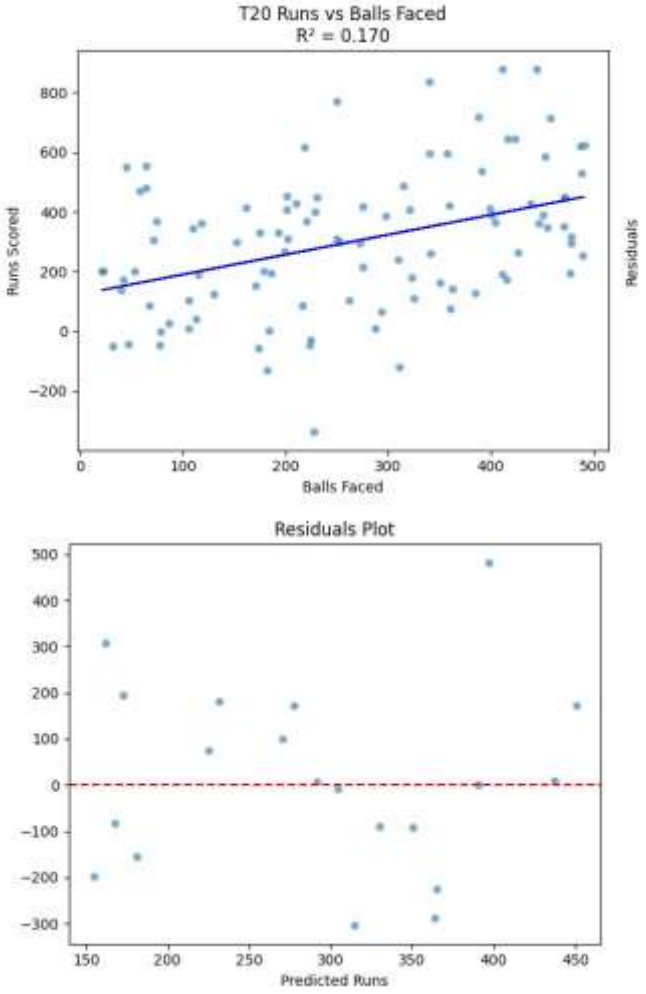

Fig. 8. Linear Regression for Runs Prediction.

The Linear regression applied over player match statistics shown in Fig. 9 gives an insight into batsman performance in facing balls in an over. The R^2 explains the variance in distribution xtremely useful to prediction of success, also useful to compare multiple models and underfitting conditions.

VI. Conclusion And Future Scope

The proposed machine learning model for cricket player skill assessment. By leveraging algorithms like Linear Regression, PCA, Clustering, XG-Boost, Classification and Outlier Analysis over player metrics overcomes the drawbacks of traditional methods. The feature engineering with normalized data provide robust and accurate predictions for player skills. The approach helps players performance as well as coach selection criteria optimization. Cricket data needs richer data analytics for improved decision making system. The fusion of AI algorithms to generate the new knowledge from training data analysis is under research. Supporting Video analytics and computer vision in future our work to be carried out for better solutions for optimal player assessment.

References

[1] Vishal Shaha, Jyotshna Dongardive, "Cricket Betting Through Machine Learning: Enhancing Decision-Making", ICPCCDA -

10.48047/jocaaa.2024.33.08.341

Conference, Springer Nature, ISBN: 978-981-96-6846-5, pp: 271 - 285, 2025.

[2] Ashish Raghav, Neeta Singh, Naresh Kumar, "Sports Analytics in Cricket Predicting Scores in the Indian Premier League Using Deep Learning Techniques", IJDSML, ISSN: 2583-9292, pp: 808 -815, 2025.

[3] Shubham Vora, "Cricket Data Analytics: Building the Best XI Players from T20 World Cup", SSRN, DOI:10.2139/ssrn.5180123,2025, pp: 26 - 32, 2025.

[4] R Suguna, Y Praveen Kumar, "Utilizing Machine Learning for Sport Data Analytics in Cricket: Score Prediction and Player Categorization", 3rd Mysore Sub Section International Conference,IEEE-Explore, DOI:10.1109/MysuruCon59703.2023.10396955, ISBN: 979-8-3503-4036-5, 2024.

[5] Param Dalal, Hirak Shah, "Cricket Match Analytics and Prediction using Machine Learning", IJCA, ISSN: 0975-8887, Vol.1, Issue-6, pp: 27 - 33, 2024.

[6] Probodh Narayan Gour, Mohd. Faheem Khan, "Utilizing Machine Learning for Comprehensive Analysis and Predictive Modelling of IPL-T20 Cricket Matches", IJST, ISSN:0974 – 6846, Vol.17, Issue-7, pp:592 – 597, 2024.

[7] Savita Sangam, S. Aba Aldar, K. Mane, "Cricket Score Prediction Using Machine Learning", JETIR, ISSN:2349 - 5162, Vol.11, Issue-6, pp:88-93, 2024.

[8] Hafeez Ur Rehman Siddiqui, Faizan Younas, Furqan Rustam,"Enhancing Cricket Performance Analysis with Human Pose Estimation and Machine Learning", Sensors Journal, MDPI, Vol.23, Issue-15, pp:234 - 242, 2023.

[9] Dr. Chaitanya Kishore Reddy, Sk. Arshiya Mobeen, "Quantifying and Analyzing the Performance of Cricket Player using Machine Learning", IJISRT, Vol.8, Issue-5, ISSN: 2456-2165, pp: 372 - 378, 2023.

[10] Kumash Kapadia, Hussein Abdel-Jaber, Fadi Thabtah, Wael Hadi,"Sport analytics for cricket game results using machine learning: An experimental study", Applied Computing and Informatics, Vol.18, Issue-3, ISSN: 2634-1964, pp: 256-266, 2022.

[11] Indika Wickramsinghe, "Applications of Machine Learning in cricket: A systematic review", Journal of Machine Learning with Applications, Elsevier, ISSN: 2666-8270, Vol.10, Issue-15, 2022.

[12] Patel Divya Shantilal, Shete Rutuja Arun, "Review on Cricket Analysis and Prediction using Machine Learning Approach", IJCRT, Vol.10, Issue-4, ISSN:2320-2882, pp: 892-896, 2022.

[13] Manoj Ishi, Jayantrao Patil, "An efficient team prediction for one day international matches using a hybrid approach of CS-PSO and machine learning algorithms", Array-Journal, Elsevier, Vol.14, Issue-6, ISSN:2590-0056, pp: 214- 222, 2022.

[14] Prerana Deshmukh, Dr. Sanjay Kumar, "Enhancing Accuracy in Cricket Bat Performance Measurement through AI and Machine Learning", ISSN: 4157 - 4171, Vol.20, Issue-13, pp: 4157-4171, 2022.

[15] Dr B.V.RamaKrishna, B. Sushma, "Machine Learning Approach for Identification of Peer Quality Factors among Sportsman", IJRTE, ISSN: 2277 - 3878, Vol.8, Issue-1, pp: 243-246, 2019.

[16] C Neeraj Pathak, Hardik Wadhwa,"Applications of Modern Classification Techniques to Predict the Outcome of ODI Cricket", Procedia Computer Science, Elsevier, ISSN: 1877-0509, pp: 56-60, 2016.C5.