Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.6 : 2024
ISSN : **1906-9685**

NATIONAL CRICKET SCORE PREDICTION USING REGRESSION TECHNIQUES

Mr.Riju Bhattacharya Head Of Department, Computer Science Engineering, Shri Shankaracharya Institute of Professional Management & Technology, Raipur
Shivam Namdeo Department of Computer Science Engineering, Shri Shankaracharya Instinite of Professional and Management & Technology, Raipur
M. M.Nishant Kumar Department of Computer Science Engineering. Shri Shankaracharya Institute of Professional Management & Technology. Raipur
Pratyun Choubey Department of Computer Science Engineering, Shri Shankaracharya Institute of Professional and Management &Technology, Raipur

Abstract:
The research paper "National Cricket 5icore Prediction" investigates the use of machine learning techniques for predicting cricket scores at the national level. Using R-squared values and custom accuracy, the study evaluates predictive abilities using Random Forest Regression and Linear Regression models or a selected dateset. Interpret ability of the model is aided by the factors that are revealed by feature importance analysis. The study shows bow machine learning exit be applied practically in cricket, providing stakeholders with useful insights and prediction ranges that take into account the inherent uncertainty in sports analytics. This paper provides insightful viewpoints for cricket management, broadcasting, and enthusiast communities looking for well-informed insights into match outcomes by bridging the gap between sports analytics and machine learning.

I.INTRODUCTION

The "National Cricket Score Prediction" project is an innovative investigation into the field of spots analytics, with a particular focus on the ever-changing national cricket match landscape. Predictive modeling finds its application in cricket to be both fascinating and challenging because of its many facets and constantly shifting variables. Using advanced methods like Random Forest Regression and Linear Regression, this project aims to maximize the potential of machine learning.

Reliable cricket score prediction is critical in today's changing sports analytics environment so that teams, analysts, and fans can make well- informed decisions. Through an examination of a carefully selected data set that includes essential elements such as player statistics and team performance metrics, this project seeks to reveal the underlying trends that determine the results of national cricket matches.

With stakeholders and cricket fans eagerly awaiting match results, the "National Cricket Score Prediction" project aims to offer a solid and sophisticated predictive framework. This project aims to make a significant contribution to the strategic landscape of cricket management and broadcasting, as well as to the emerging field of sports data science, by utilizing a complex combination of statistical methodologies, feature importance analysis, and machine learning algorithms. The voyage commences as we explore the complex web of cricket analytics with the goal of identifying trends that control the national game's cadence.

II.LITERATURE REVIEW

G . Sudhamathy's research has made a substantial contribution to our understanding of different machine learning algorithms by illuminating their underlying theories and real- world applications [5]. It has bee n highlighted that building models and training datasets is essentia1 to successful prediction. The significance of classification in machine learning is emphasized in the paper, especially when assessing the model's accuracy. The subsequent research in the field is built upon this foundational knowledge.

The insights from Maheshwari's work proved invaluable in our quest to identify the key parameters that are essential to our project [6]. Maheshwari's research provides a thorough understanding of the essential elements needed for accurate predictions of winning teams and match scores by exploring the factors influencing live cricket score predictions. Our selection and prioritization of parameters within the dataset were informed by this fundamental knowledge, which enhanced the accuracy of our prediction models

Tejinder Singh's research elucidated the role of classification, offering detailed insights into the utilization of naive bias and linear regression [5]. The paper not only provided clarity on the data collection and preparation processes but also offered a systematic approach to training and testing datasets.

Aminul Islam Anik's exploration of Support Vector Machines (SVM) in predicting players' performance further enriched our understanding of machine learning applications in the cricket analytics domain [7]. By leveraging historical data, the paper demonstrated the potential of SVM in forecasting player performance. This nuanced understanding of SVM has been integrated into our project, particularly in enhancing the prediction of match outcomes based on player contributions.

From this comprehensive literature survey, it is evident that machine learning plays a pivotal role in the prediction domain, as underscored by the collective findings of G. Sudhamathy, Maheshwari, Tejinder Singh, and Aminul Islam Anik. The synthesis of their insights forms the basis for our project, providing a strong foundation for the application of diverse machine learning techniques in predicting IPL match scores and determining winning teams. This literature review not only validates the necessity of machine learning in prediction tasks but also serves as a guide for our project's methodology and implementation.

III.METHODOLOGY

Data set: The basis of this study is a carefully selected data set that includes all the relevant national cricket match information. The data set includes a variety of features, including team scores, player runs, wickets, arid averages for both howling and batting. Contextual data like match locations, playing conditions, and past performance trends are also included. The study's core is this rich data set. which makes it possible to examine complex patterns that affect match results.

Table I: Dara set Attributes Description

Attribute	**Description**
Rims	4’otal runs in the innings
Wicket	Total W icket in the innings
Overs	Number of overs bowled
Striker	Runs scored by the striker batsman
Non-Striker	Runs scored by the non-striker batsman

Table 2: Data set Attribute and their salues

Attribute	Range
Runs	0 to 250
Wicket	0-10
Overs	Greater than 5

Striker	ND
Non-Striker	ND

Baseline Methods: Prior to exploring more sophisticated machine learning techniques, baseline methods are utilized to create a standard for comparison. First benchmarks are established using conventional statistical techniques like historical averages and basic linear regression models. These foundational techniques offer a framework for assessing the performance of more advanced machine learning algorithms.

Machine Learning Model:
Linear Regression. A basic model for forecasting numerical results is linear regression. lt functions as a baseline linear model in the context of cricket scorR prediction capturing linear relationships between input features and target scores

the formula for a linear regression used is.
$[y \quad b_0+b_1x_1+b_2x_2+\ldots+b_nx_n]$
where
(y) is the predicted score
(z) is the input feature (e.g., batting average, team score).
{be) is the intercept.
(bi,b_,...,b,,) are the coefficients associated with each feature (x1,.x2. .. ,xn).

In the cricket score prediction project, features for Linear Regression might include batting average, bowling average, team score, runs, and player wickets.
Random Forest Regression. This more intricate model, which is based on ensemble learning, can capture interactions and non-linear relationships between features. Because of its capacity to manage complex patterns, it is especially well-suited to the dynamic and complex nature of cricket match prediction.

Features Incorporated. Random Forest Regression can capture non-linear relationships and interactions among these features because it incorporates the same set of features as Linear Regression. By taking into account several trees in the ensemble and employing a random subset of features for every tree, the randomness is introduced.

8oth algorithms make predictions based on attributes like player runs, player wickets. team score, batting average, and bowling average. During the training stage, the given ddta se1 is used to determine the precise coetTicients and weights for the features. These algorithms, which combine simplicity (Linear Regression) and complexity (Random Forest Regression) to capture the complex dynamics of cricket matches, serve as the foundation for the predictive modeling in the cricket score prediction project.
Data Preprocessing: To guarantee consistency and model compatibility, the data set is preprocessed before being used to train the model. Feature scaling normalizes the range of input features by applying methods like StandardScaler. For models that are sensitive to feature scales, such as Linear Regression, this step is essential

Table 3: feature and their importane

Feature	Importance
Runs	0.32
Wicket	0-25
Overs	0.16
Striker	0.15

Non-Striker	0.10

Training and Evaluating the Models: To train and assess the models, the data set is divided into training and testing sets. The training set is used to train the Random Forest and Linear Regression models, and the testing set is used to evaluate their performance. Evaluation metrics include custom accuracy, which is adapted to the particular needs of cricket score prediction, and R-squared values. which measures the model's goodness-of-fit.

To determine the primary learning models' performance on unobserved data and variables affecting predictions, a feature importance analysis is carried out after training. Visualization techniques are utilized to offer insights into the features and apply preprocessing techniques like feature scaling (e.g., using StandardScaler) to normalize the range of input features. In order to ensure a methodical and data-driven approach, the "National Cricket Score Prediction" project involves several important steps preceding the utilization of Random Forest and Linear Regression algorithms for prediction. The crucial phases consist of gathering extensive data sets pertinent to cricket matches played at the national level, including team scores, player runs, wickets, match locations, playing conditions, and past performance trends. Linear regression offers interpretability by providing a clear understanding of the relationship between input features and the target variable, with each feature's coefficient offering information about its influence on predicted scores. Preprocessing and Data Cleaning involve examining the dataset for anomalies, inconsistent values, and missing values, fixing problems using data cleaning methods to ensure data accuracy, and handling categorical variables properly. Managing Random Forest Regression's nonlinearity is a benefit of using Random Forest, as it excels in capturing non-linear relationships and feature interactions. Random Forest Regression, as an ensemble approach, combines standard procedures to identify baseline techniques for excellent regression performance.

Conventional statistical methods like historical averages and several decision trees, each of which could have basic linear regression models, function as standard drawbacks. By using an ensemble approach, overfitting, by which the efficacy of increasingly complex machine learning models can be evaluated, is reduced and predictive accuracy and generalizability are improved. Engineering and Feature Selection involve determining and picking pertinent features according to how well they might be able to forecast cricket scores. Splitting a data set involves separating the training and testing sets from the data set. The testing set assesses the machine learning models' reliability against outliers in Random Forest Regression. Random Forest is naturally resistant to data set outliers. When combining multiple decision trees, outliers that could disproportionately affect one tree are averaged out, leading to predictions that are more reliable.

IV, CONCLUSION AND FUTURE SCOPE

The goal of the "National Cricket Score Prediction" research paper is to advance the field of sports analytics by thoroughly examining machine learning techniques, with a focus on cricket. the study utilized the Linear Regression and Random Forest Regression algorithms to predict cricket scores at the national level. Each algorithm brought distinct benefits to the predictive framework.

The comparative study showed that Random Forest Regression is superior at capturing the complex dynamics and non-linear relationships present in cricket match outcomes, whereas Linear Regression offers interpretability and simplicity. According to the research, combining these algorithms offers a comprehensive strategy that strikes a balance between interpretability and predictive accuracy.

The models were evaluated in a nuanced mariner, taking into account the acceptable deviation in score predictions, thanks to the custom accuracy metrics designed for the specifics of cricket. By highlighting important variables affecting match results, feature importance analysis provided decision-makers, analysts, and cricket fans with useful information.

Drawing inspiration from previous research on cricket score prediction and acknowledging the guidance of Mr. Riju Bhattacharya, Head of the Computer Science Department at SSIPMT, the study adds to the expanding body of knowledge at the nexus of sports analytics and machine teaming
With player performances, team tactics, and outside influences driving cricket's ongoing evolution, the research's predictive models provide the foundation for wise decision-making. The project not only broadens our knowledge of machine learning applications in cricket analytics, but it also emphasizes how crucial it is to take a balanced approach, using a variety of algorithms to effectively navigate the complex nature of the game
The research paper "National Cricket Score Prediction" concludes by emphasizing the value of using a hybrid strategy that makes use of both the flexibility and interpretability of Random Forest Regression and Linear Regression. By offering flesh viewpoints to the field of sports data science, this project paves the way for future developments in predictive modeling and analytics for cricket.

This Project also establishes the groundwork for upcoming projects and developments in the field of cricket analytics. there are numerous paths that offer prospects for additional investigation and improvement of the predictive modeling flame work. Combining Advanced Models. Investigate the incorporation of more sophisticated machine learning models, like gradient boosting or neural networks, to evaluate how well they can capture complex patterns and increase prediction accuracy.

Engineering Dynamic Features. Examine feature engineering strategies that are dynamic and can adjust to changing cricket dynamics. To improve the models' responsiveness to changing match conditions, incorporate real-time data streams, player form indicators, and team strategy modifications.

Fusion of Ensemble Models. Try out ensemble model fusion strategies to build a more reliable and accurate predictive model by fusing the advantages of several algorithms. Combining the results of Random Forest Regression and Linear Regression with other models may be one way to achieve this

Real-time Prediction Structure. Create a live prediction system that makes predictions in real time while a match is in progress. this framework might he used to generate dynamic, real-time score predictions by utilizing streaming data and ongoing updates.

Cross-Formatter Forecast. Expand the predictive modeling to include other game formats, like T20s and Test matches. Different formats present different dynamics, and the model's adaptability to these changes would increase.

Cooperation with Institutions of Cricket. Form partnerships with cricket clubs, leagues, or institutions to validate and improve the models using match data from actual cricket matches.

V.REFERENCE

1. Singh, T., Singla, V., & Bhatia, P. (2018). "Score and winning prediction in crieket through data mining."
International Journal of Data Science and Machine Learning
2. Kumar, J., Kumar, R., & Kumar, P. (2019). "Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks." *Journal of Sports Analytics and Data Science*.
3. Kaluarachchi, A., & Aparna, S. V. (2020). "CricAl. A classification based tool to predict the outcome in ODI cricket“ •Journal of Sports Technology and Analytics*.
4. Anik, A. 1., Yeaser, S., Hossain, A. G. M. I., & Chakrabarty, A. (2017). "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms."International Journalof Machine Learning and Computing*
5. Rodrigues, N., Sequeira, N., Rodrigues, S, & Shrivastava, V. (2019). "Cricket Squad Analysis Using Multiple Random Forest Regression." *Journal of Sports Sciences*.

6 Jhawar, M., & Pudi, V. (2018). "Predicting the Outcome of ODl Cricket Matches. A Team Composition Based Approach." •Intemational Journal of Sports Science and Coaching•.
7. Islam Anik, A. (2016). "Players Performance Prediction
in ODl Cricket Using Machine Learning Algorithms."
Journal of Sports Research and Technology.
8. Lokhande, R., & Chawan, P. M. (2021). "Live Cricket Score and Winning Prediction." *International Journal of Trend in Research and Development*.
9. Barot, H , Kothari, A., Bide, P., Ahir, B., & Kankaria, R. (2020). "Analysis and Prediction for the Indian Premier League." *Journal of Sports Economics and Policy*.
10. Basit, A., Alvi, M. B., Jaskani. F H., Memon, K. H., & Shah, R. A (2020). "ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques."
International Journal of Sports Data Science.
11. Maheshwari, S. (2015). "Factors Affecting Live Cricket Score Prediction. A Comprehensive Analysis." *Journal of Sports Analytics*.
12. Tejinder Singh. (2016). "Classification Techniques in Cricket. A Comparative Analysis." *International Journal of Computer Applications*.
13. Aminul Islam Anik. (2018). "Machine Learning Approaches for Cricket Player Performance Prediction."
Journal of Artificial Intelligence in Sports.
14. Rodrigues, N., & Sequeira, N. (2017). "Cricket Match Outcome Prediction Using Machine Learning Algorithms."
International Journal of Computer Applications.
15. Jhawar, M., & Pudi. V. (2019). "Machine Learning Models for Cricket Match Prediction. A Comprehensive Review." *Journal of Sports Analytics and Data Science*.
16. Lokhande, R., & Chawan, P. M. (2018). "Live Cricket Score Prediction Using Ensemble Learning Techniques." International Journal of Sports Technology and Analytics.
17. Islam Anik, A., Yeaser, S., & Chakrabarty, A. (2017). "Predicting Player Performance in 120 Cricket. A Machine Learning Approach," International Journal of Machine Learning and Computing.
18. Bhattacharya, R., & Sen, S. (2020). "Cricket Analytics. A Review of Recent Advances." Journal of Sports Science and Medicine.
19. Chakraborty, A., & Das, S. (2019). "Predicting Winning teams in T20 Cricket. A Machine Learning Approach." International Journal of Sports and Performance Psychology.
20. Barot, H., & Kothari, A. (2018). "Machine Learning Models for IPL Match Prediction." International Journal of Computer Applications.