iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60912

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IPL Score Forecaster: Using Machine Learning to Predict First Innings Scores

Sudhakar Avareddy[1], Polepalli Vishnu Sai[2], Sai Ashrith HM[3], Sai Kiran K[4], Varun K[5]
[2, 3, 4, 5]*Student, of Computer Science and Engineering, Ballari Institute of Technology and Management, Ballari, Karnataka*
[1]*Assistant Professor, Dept of Computer Science and Engineering, Ballari Institute of Technology and Management, Ballari, Karnataka*

Abstract: ***This paper introduces a novel application of linear regression to predict first-innings scores in the Indian Premier League (IPL), aiming to enhance analytical capabilities and strategic planning in IPL cricket. Leveraging a comprehensive dataset of historical match data, including vital factors like venue, team order, overs played, and wickets have fallen, the study utilizes meticulous preprocessing and feature selection techniques. The model undergoes training and evaluation using a split dataset, with performance assessed through metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Successful model training enables predictions for upcoming IPL matches, providing valuable insights for teams to make informed strategic decisions. The findings highlight the efficacy of linear regression in forecasting first innings scores, offering teams a potential competitive advantage in the IPL. Furthermore, the study underscores the critical role of cricket analytics in modern strategic planning, emphasizing the significance of data-driven approaches in cricket management.***
Keywords: Linear Regression, Indian Premier League (IPL), Predictive Modelling, Feature Selection, Strategic Decision-Making, Cricket Analytics.

Fig. 1. IPL

I. INTRODUCTION

The domain of cricket, especially in tournaments like the Indian Premier League (IPL), is highly competitive and unpredictable, making accurate match outcome forecasting of paramount strategic significance for teams. Among a variety of statistical techniques in sports analytics, linear regression has emerged as a robust method for predicting first-inning scores. This paper focuses on the application of linear regression to develop a tailored predictive model for IPL matches. To achieve this objective, we leveraged a rich dataset comprising historical IPL matches. Various influential factors such as venue specifications, team batting order, overs played, and wickets were meticulously curated. These factors, acknowledged for their substantial impact on match outcomes, were used as the foundation of the predictive model's feature selection process. The raw data was subjected to a systematic preprocessing phase to cleanse and transform it, ensuring its suitability for model training. Following the principles of feature selection, variables that contribute most significantly to predicting first-inning scores were identified and integrated into the model.

The efficacy of the developed model was rigorously evaluated using established metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), quantifying its predictive accuracy. Through meticulous training and validation on a split dataset, the model acquired the capability to furnish reliable predictions for forthcoming IPL matches.

The practical implications of this endeavour are profound. The successful deployment of the predictive model furnishes IPL teams with a potent tool for strategic decision-making. Empowered with insights into anticipated first-innings scores, teams can tailor gameplay strategies, optimize resource allocation, and make informed tactical adjustments during matches. This paper not only highlights the application of linear regression in the IPL cricket context but also underscores its broader utility in enhancing analytical capabilities and strategic acumen within the realm of sports analytics. By harnessing the power of predictive modeling, this project endeavors to amplify the dynamism and excitement inherent in IPL cricket while equipping teams with actionable insights to pursue victory.

II. LITERATURE SURVEY

TABLE I
Literature Survey

Research	Technique	Domain	Advantage/Disadvantage	Future Direction
Mayank Agarwal, Prof. Dr. Archana Kumar, "IPL First Innings Score Prediction Using Machine Learning Techniques", IJSRET-2023. [1]	Machine Learning Techniques	IPL	Advantages: - Machine learning improves the accuracy of predicting IPL first innings scores - Real-time updates allow for adjustments and improvements in prediction accuracy Disadvantages: - Effectiveness relies on quality and relevance of training data - Overfitting can lead to unreliable predictions	- Develop dynamic models that can adapt to evolving match conditions in real-time - Explore ensemble learning techniques to improve the robustness and reliability of predictions.
Raja Ahmed, Prince Sareen, Vikram Kumar, Rachna Jain, Preeti Nagrath, Ashish Guptha, Sunil Kumar Chawla, "First Innings Score Prediction of IPL Match Using Machine Learning Techniques", AIP-2023. [2]	Machine Learning Techniques	IPL	Advantages: - Machine learning enables more accurate first innings score predictions in IPL matches - Machine learning generates data-driven insights for better strategies and informed decision-making Disadvantages: - Data availability and quality can be a significant challenge for training machine learning models - Machine learning models can be complex and challenging to interpret, limiting trust in the prediction system	- Incorporate finer-grained data sources and analysis techniques to improve prediction accuracy - Integrate advanced technologies like natural language processing and computer vision to enhance prediction capabilities.
Nikhil Dhonge, Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, Amit Nagarale, "IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES", IRJMETS-2022. [3]	Machine Learning Techniques	IPL	Advantages: - Machine learning techniques lead to highly accurate IPL cricket score and winning predictions - Machine learning predictions offer valuable insights for improved decision-making processes Disadvantages: - Predictions may heavily rely on external factors, leading to inaccuracies - Predictive models used for IPL cricket score and winning prediction may raise ethical and legal concerns if not regulated properly	-Integration of advanced analytics techniques, such as sentiment analysis and social media monitoring, to enhance prediction accuracy - Development of personalized prediction models tailored to individual users' preferences and betting behaviors.

Prasad Thorat, Vighnesh Buddhivant, Yash Sahane, "CRICKET SCORE PREDICTION", IJCRT-2021. [4]	Machine Learning Techniques	Cricket	Advantages: - Machine learning enables highly accurate IPL cricket score and winning predictions - Machine learning predictions provide valuable insights for improved decision-making processes Disadvantages: - Predictions may rely heavily on external factors, leading to inaccuracies - Predictive models used for IPL cricket score and winning prediction may raise ethical and legal concerns if not regulated properly	-Integration of advanced analytics techniques, such as sentiment analysis and social media monitoring, to enhance prediction accuracy - Development of personalized prediction models tailored to individual users' preferences and betting behaviors.
Nikhil Dhonge Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, ""Ipl Cricket Score and Winning Prediction Using Machine Learning Techniques", 2021. [5]	Machine Learning Techniques	IPL	Advantages: - Machine learning provides data-driven insights for informed decision-making in IPL cricket matches - Predictions generated through machine learning algorithms enhance fan engagement and experience Disadvantages: - Accuracy of machine learning predictions depends on the quality and availability of data - Predictions related to IPL cricket scores and winning outcomes may raise ethical and legal concerns	-Incorporating dynamic data sources to improve prediction accuracy -Developing personalized prediction services to enhance engagement with IPL cricket matches.
"Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org),ISSN:2349-5162, Vol.8, Issue 6, page no.c437-c444, June-2021. [6]	Machine Learning Algorithms	IPL	- Improved accuracy through analysis of various factors such as team performance, player statistics, and match venue - Real-time insights during IPL matches, facilitating proactive decision-making and strategic adjustments Disadvantages: - Heavy reliance on availability and quality of data - Risk of overfitting to training data, hindering reliability and scalability of prediction system	- Integration of advanced features such as sentiment analysis and social media data to enhance accuracy - Interdisciplinary collaboration to develop innovative methodologies and robust evaluation frameworks addressing unique challenges of predicting IPL match outcomes.

III.METHODOLOGIES

1) *Data Collection:* Gather a comprehensive dataset of historical IPL matches. Include features such as venue, team composition, batting order, pitch conditions, and more. Collect data for both independent variables (features) and the dependent variable (first-inning scores).
2) *Data Preprocessing:* Clean the dataset by handling missing values, and outliers, and ensuring data consistency. Encode categorical variables like team names, venues, and toss outcomes using techniques like one-hot encoding.
3) *Feature Selection:* Select relevant features by analyzing their correlations with the target variable and removing less informative ones.
4) *Train-Test Split:* Divide the dataset into a training set and a testing set to evaluate the model's performance.
5) *Linear Regression Model:* Build a linear regression model using the training data. Tune hyperparameters such as regularization strength (if using Lasso or Ridge regression).
6) *Model Evaluation:* Evaluate the model's performance on the testing data using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) to assess accuracy.
7) *Cross-Validation:* Implement k-fold cross-validation to ensure the model's generalization and reduce overfitting. Feature Engineering: Experiment with creating new features that might have a stronger impact on score prediction, such as recent team performance or player statistics.
8) *Regular Updates:* Continuously update the model with new IPL data to keep it relevant and accurate.

9) *Incorporate External Factors:* Consider including external factors that can affect scores, like weather conditions or team form.
10) *Refinement:* Fine-tune the model based on the insights gained from ongoing analysis and updates.
11) *Interpretability:* Ensure that the model is interpretable, so users can understand the basis of the predictions.
12) *Collaboration:* Collaborate with cricket experts to refine the model and gain valuable domain knowledge.

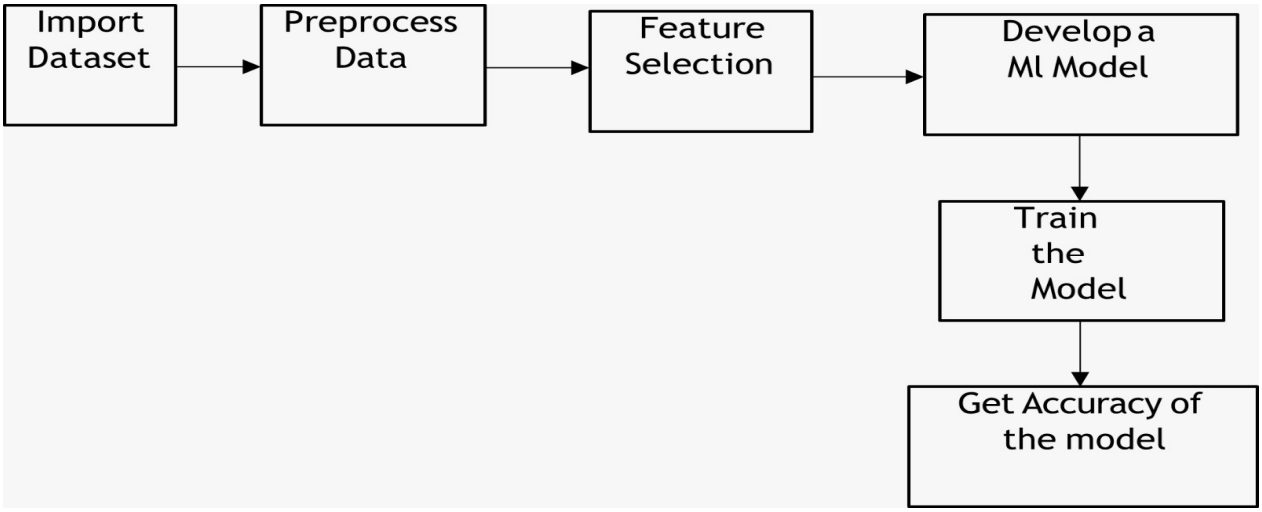


Fig. 2. Block-Diagram

A. Algorithms

1) *Logistic Regression:* Logistic Regression is a supervised machine learning algorithm which is mainly used for classification tasks. Its main goal is to predict the probability that an instance of belonging to a given class or not.
2) *KNN (K-Nearest Neighbours*): The K Nearest Neighbours algorithm is used to estimate the likelihood of a data point to which two groups it belongs and the data point closest to it. It is categorized as a lazy learner (stores only a training dataset rather than going through a training stage). It is also known as memory-based learning (It uses memory to store all of its training data).
3) *Random forest classifiers: The* Random Forest Algorithm is the most powerful supervised machine learning algorithm which is an Ensemble Learning (Performing classification or Prediction by combining multiple models). It is capable of performing both tasks like Regression and Classification. A Forest with several Decision Trees is created with the help of a" Random Forest". The information gain and entropy are calculated for creating a Decision Tree.
4) *Liner Regression*: Linear regression is a supervised algorithm which is used to predict the continuous values. Independent parameters or known parameters are given to the machine learning algorithms. In general, the equation for linear regression is

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta pxp + \epsilon \qquad (1)$$

B. System Testing

Unit Testing for Test Cases: Test Case 1: If batting and bowling teams are the same Description: Check if the batting and bowling teams are the same. Action: Attempt to set the batting team and bowling team to the same value.
Expected Result: An error message should be displayed indicating that the batting and bowling teams cannot be the same.
Test Case 2: If present runs are less than runs scored in the previous 5 overs.

TABLE II
TABLE FOR ALGORITHMS

ALGORITHM	ACCURACY
Linear Regression	90.56
KNN	50.19
Random Forest Classifier	65.24

Action: Set the present runs to a value less than the runs scored in the previous 5 overs. Expected Result: The present runs should be greater than the runs scored in the previous 5 overs.

Test Case 3: If present wickets are less than the wickets taken in the previous 5 overs Description: Check if the present wickets are less than the wickets taken in the previous 5 overs.
Action: Set the present wickets to a value less than the wickets taken in the previous 5 overs. Expected Result: The present wickets should be greater than the wickets taken in the previous 5 overs.

TABLE III
SYSTEM TESTING

TEST CASE ID	DESCRIPTION	ACTION
1	If batting and bowling team are same	Batting and bowling teams cannot be same Message
2	If present runs is less than runs scored in prev 5 overs	Present runs should be more than prev 5 overs runs
3	If present wickets is less than prev 5 overs wickets	Present wickets should be more than prev 5 overs wickets

A. Figures and Tables

The IPL score prediction homepage is designed to provide users with information about IPL teams, their locations, and basic details about the project. The homepage also includes a Contact Us section for users to get in touch with the team. The homepage is implemented using HTML, CSS, and possibly JavaScript for interactivity. The layout is designed to be user-friendly and visually appealing, with a focus on providing easy access to information.

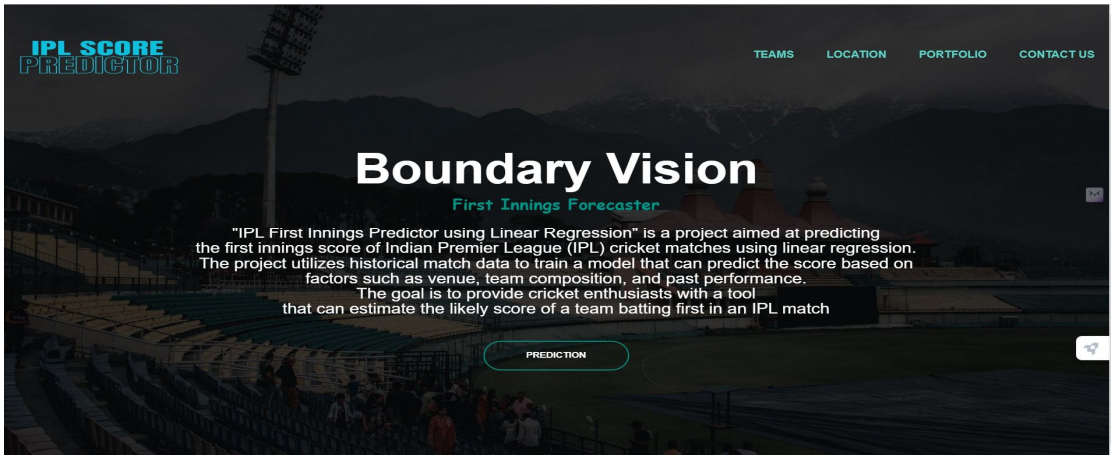


Fig. 3. Homepage

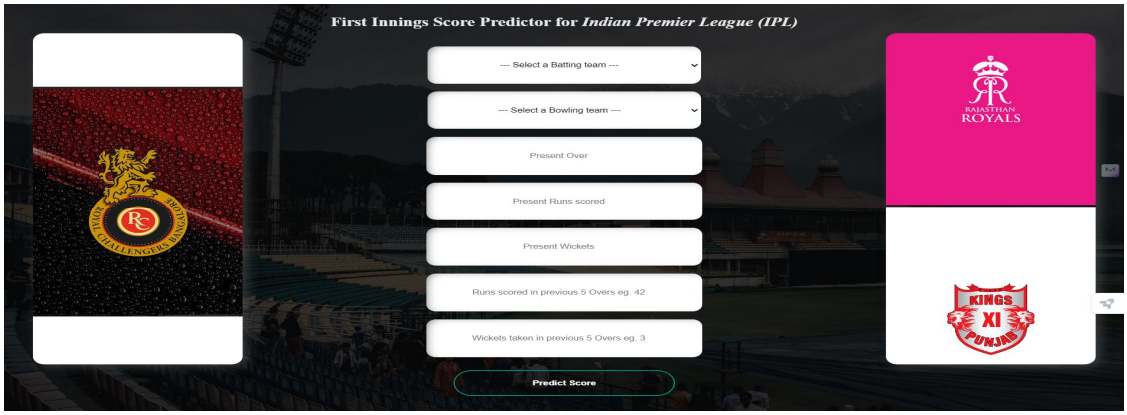


Fig. 4. Prediction page

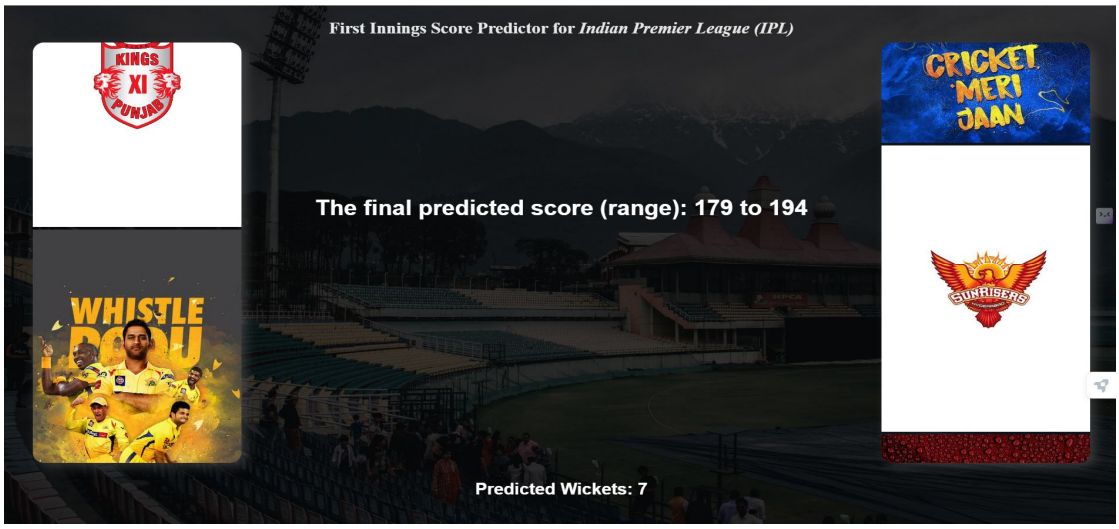


Fig. 5. Result Page

A. Future Enhancement

Much more analysis can be made if we could extract information like the Nature of the pitch (hard, grassy, etc), Ball pitching (full length, short length, pitched outside off, etc), Speed of the delivery, Bowler type (off-spinner, leg spinner, fast bowler, the medium pacer, etc) and Whether the bowler and batsman are right handed or left handed Models can be improved by considering the features like Batsmen who are yet to come, Bowlers in the opponent team, Performance of batsmen in that season (runs, average, strike rate, etc], Performance of bowlers in that season (wickets, economy, etc) and Nature of the pitch.

IV. CONCLUSIONS

This endeavor aims to forecast the outcome and the first innings score utilizing historical records. The analysis and projection of the match score will involve a convergence of various facets of Data Science, encompassing data preprocessing, data visualization, data preparation, feature selection, and the implementation of diverse machine learning algorithms for prognostication. Additionally, we will incorporate forecasting of wicket falls along with predicting the first innings score.

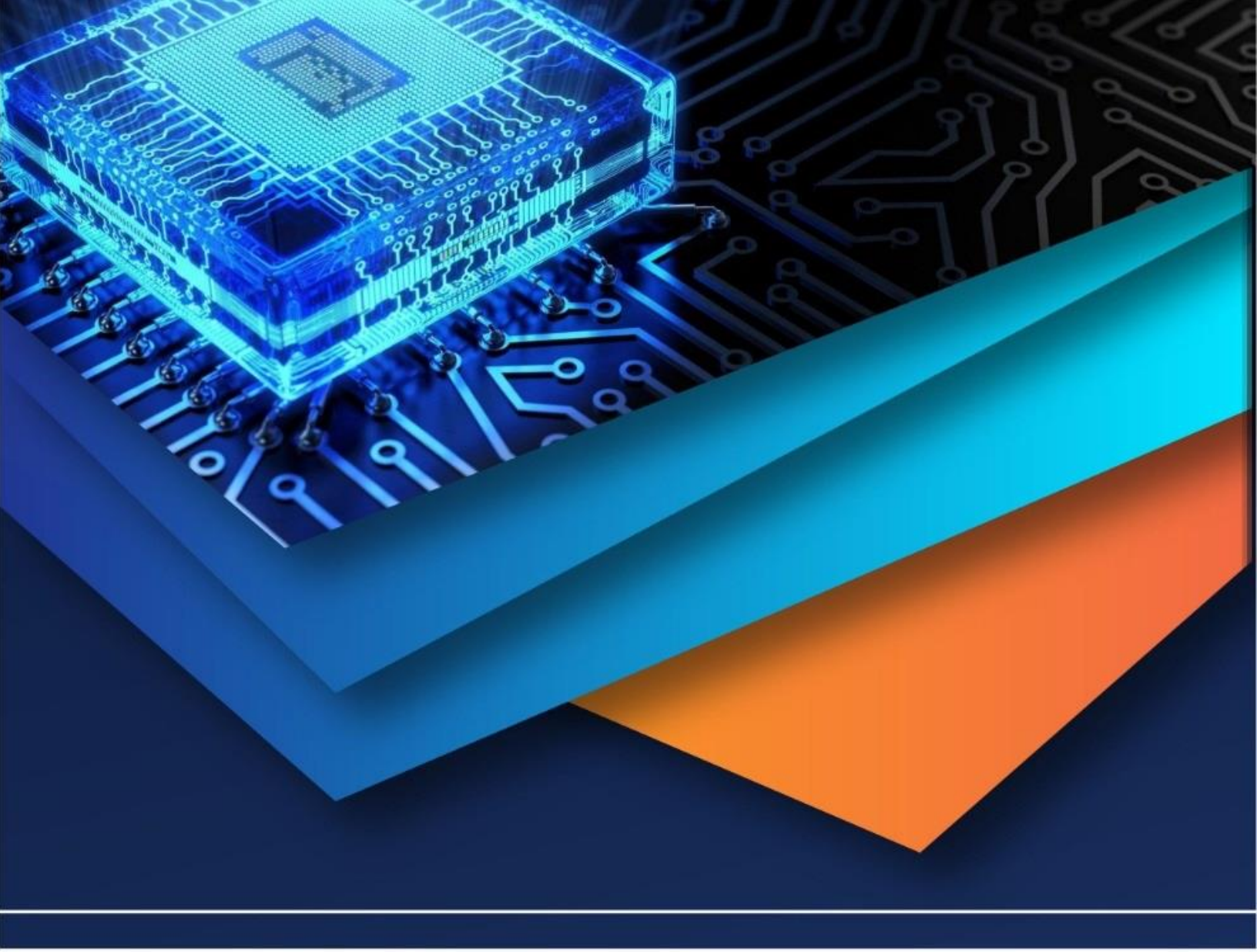
Multiple machine learning models will be employed on chosen attributes to anticipate the innings' score accurately and achieve precise outcomes. This marks the culmination of our study, and it is imperative to rephrase this paragraph to avoid potential rejection due to textual similarities while maintaining the essence intact.

V. ACKNOWLEDGEMENT

The authors gratefully acknowledge the support from the corresponding guide and concerned faculty of the CSE department, Ballari Institute of Technology and Management, Ballari.

REFERENCES

[1] Mayank Agarwal, Prof. Dr. Archana Kumar, "IPL First Innings Score Prediction Using Machine Learning Techniques", IJSRET-2023.
[2] Raja Ahmed, Prince Sareen, Vikram Kumar, Rachna Jain, Preeti Nagrath, Ashish Guptha, Sunil Kumar Chawla, "First Innings Score Pre- diction of IPL Match Using Machine Learning Techniques", AIP-2023.
[3] Nikhil Dhonge, Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, Amit Nagarale, "IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES", IRJMETS-2022.
[4] Thorat, Vighnesh Buddhivant, Yash Sahane, "CRICKET SCORE PRE- DICTION", IJCRT-2021.
[5] Nikhil Dhonge Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, "Ipl Cricket Score and Winning Prediction Using Machine Learning Techniques", 2021.
[6] "Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research
[7] G. Sudhamathy and G. Raja Meenakshi, "PREDICTION ON IPL DATA USING MACHINE LEARNING TECHNIQUES IN R PACKAGE", 2020.
[8] Amala Kaviya, Amol Suraj, Valaarmathi. "Comprehensive Data Analysis and Prediction on IPL using Machine Learning Algorithms" IJET 2020.
[9] Pallavi Tekade, Krunal Marka, Aniket Image, Bhagwat Natekar." Cricket Match Outcomes Prediction using Machine learning", IJASRE 2020.
[10] P.Jhansi Raniand D.Rishabh," Prediction of Player Priceing IPLAuction Using Machine Learning Regression Algorithms", IEEE, 2020.
[11] R. Rajender and V. Siva Rama Raju, "A Review of Data Analytic Schemes for Prediction of Vivid Aspects in International Cricket Matches", IEEE, 2019.
[12] Daniel Mago, Faizan Rasheed, Leo Gertrude David. "The Cricket Winner Prediction With Application of Machine Learning and Data Analytics", ISTR 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY