International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

A Comprehensive View of IPL Score Prediction Using Machine Learning

AMBATI SAI RAM

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

B. Suryanarayana Murthy

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid advancement of data analytics and machine learning has significantly transformed various domains, including sports. Cricket, particularly the Indian Premier League (IPL), generates vast amounts of data in the form of match statistics, player performance, and environmental conditions. Analyzing this data to predict match outcomes and scores has become an important area of research in sports analytics. This project presents a comprehensive study on IPL score prediction using machine learning techniques.

The primary objective of this study is to develop a predictive model capable of estimating the final score of a team in an IPL match based on various input parameters. These parameters include batting team, bowling team, venue, overs completed, runs scored, wickets fallen, and recent performance indicators such as runs in the last five overs. By leveraging historical match data, the system identifies patterns and relationships that influence scoring behavior in T20 cricket.

The proposed system employs supervised machine learning algorithms, particularly regression models, to predict continuous score values. Data preprocessing plays a crucial role in improving model performance, involving steps such as handling missing values, encoding categorical variables, and feature selection. Multiple algorithms such as Linear Regression, Decision Tree Regression, and Random Forest Regression are evaluated to determine the most effective approach.

The model is trained using historical IPL datasets and tested for accuracy and reliability. Performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to evaluate prediction accuracy. The results demonstrate that machine learning models can effectively predict IPL scores with reasonable accuracy, providing valuable insights for teams, analysts, and fans.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

One of the key contributions of this study is the integration of machine learning with sports analytics to enhance decision-making. The system can be extended to real-time prediction scenarios, allowing users to estimate scores during live matches.

In conclusion, this research highlights the potential of machine learning in transforming cricket analytics. The proposed system provides a scalable and efficient solution for IPL score prediction, paving the way for further advancements in sports data science.

Keywords: IPL Score Prediction, Machine Learning, Cricket Analytics, Regression Models, Data Mining, Sports Analytics, Predictive Modeling, T20 Cricket, Data Science

I. INTRODUCTION

Cricket is one of the most popular sports globally, and the Indian Premier League (IPL) has revolutionized the game by introducing a fast-paced T20 format. The IPL generates a massive amount of data, including player statistics, match conditions, and performance metrics. This data presents an opportunity to apply advanced analytical techniques to gain insights and make predictions.

Score prediction in cricket is a complex problem due to the dynamic nature of the game. Various factors influence the final score, such as team composition, pitch conditions, weather, player performance, and match situations. Traditional methods of prediction rely on expert analysis and historical averages, which often lack accuracy and consistency.

With the advent of machine learning, it is now possible to analyze large datasets and identify patterns that are not easily detectable through manual analysis. Machine learning models can learn from historical data and make predictions based on input features. This capability makes them suitable for predicting IPL scores.

The objective of this project is to develop a machine learning-based system for predicting IPL scores. The system takes into account multiple features such as runs scored, overs completed, wickets lost, and recent performance metrics. By analyzing these features, the model predicts the final score of a team.

The proposed system uses regression techniques, which are suitable for predicting continuous values. Data preprocessing is an essential step in the process, ensuring that the data is clean and suitable for modeling. Categorical variables such as team names and venues are encoded into numerical formats, and irrelevant features are removed.

The system is evaluated using standard performance metrics to ensure accuracy and reliability. The results demonstrate that machine learning models can provide valuable insights into scoring

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

patterns in IPL matches.This study contributes to the field of sports analytics by demonstrating the application of machine learning techniques in cricket. It provides a foundation for developing advanced predictive systems that can be used by teams, analysts, and fans.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

The application of machine learning in sports analytics has gained significant attention in recent years. Cricket, being a data-rich sport, has been a popular domain for predictive modeling and statistical analysis.Early research in cricket analytics focused on statistical methods such as averages and probability models. These methods were used to analyze player performance and predict match outcomes. However, they were limited in their ability to handle complex relationships between variables.

With the advancement of machine learning, researchers began exploring regression and classification techniques for predicting cricket scores and match results. Linear Regression has been widely used due to its simplicity and interpretability. It establishes a linear relationship between input features and the target variable, making it suitable for score prediction.Decision Tree algorithms have also been applied to cricket analytics. These models can capture non-linear relationships and interactions between variables. However, they are prone to overfitting, especially with small datasets.

Random Forest, an ensemble learning method, has been shown to improve prediction accuracy by combining multiple decision trees. It reduces overfitting and provides more robust predictions. Several studies have demonstrated the effectiveness of Random Forest in predicting cricket scores and match outcomes.Support Vector Machines (SVM) have also been used for predictive modeling in sports. These models are effective in handling high-dimensional data but require careful parameter tuning.

Recent research has explored the use of deep learning techniques such as Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN). These models can capture complex temporal patterns in data, making them suitable for real-time prediction scenarios. However, they require large datasets and significant computational resources.

In addition to machine learning models, data preprocessing techniques such as feature engineering and normalization play a crucial role in improving model performance. Feature selection helps in identifying the most relevant variables, reducing noise and improving accuracy.

Despite these advancements, challenges such as data quality, feature selection, and model interpretability remain. The proposed system addresses these challenges by using a structured approach to data preprocessing and model selection.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

III. EXISTING SYSTEM

Existing systems for IPL score prediction primarily rely on traditional statistical methods and basic analytical techniques. These systems use historical data to calculate averages and trends, which are then used to estimate future scores. While these methods provide a basic understanding of scoring patterns, they lack the ability to capture complex relationships between variables.

Some modern systems have incorporated machine learning techniques, but they often focus on match outcome prediction rather than score prediction. These systems use classification algorithms to predict whether a team will win or lose, rather than estimating the actual score. As a result, they provide limited insights into the dynamics of the game.

Another limitation of existing systems is the lack of real-time prediction capabilities. Most systems are designed for offline analysis and cannot process live match data. This restricts their usefulness in dynamic environments such as live matches.

Additionally, many systems do not consider all relevant factors influencing the score. For example, factors such as recent performance, pitch conditions, and player form are often overlooked. This reduces the accuracy of predictions.

Existing systems also face challenges related to data preprocessing and feature selection. Incomplete or inconsistent data can negatively impact model performance. Furthermore, many systems lack user-friendly interfaces, making them difficult to use for non-technical users.The proposed system addresses these limitations by using advanced machine learning algorithms, comprehensive feature selection, and real-time prediction capabilities. It provides a more accurate and efficient solution for IPL score prediction.

IV. PROPOSED METHOD

The proposed system introduces a machine learning-based approach to accurately predict IPL match scores by leveraging historical data and real-time match parameters. Unlike traditional statistical models, this system utilizes advanced regression algorithms to capture complex relationships between various influencing factors.The system takes multiple input features such as batting team, bowling team, venue, current score, overs completed, wickets fallen, and recent performance metrics (e.g., runs scored in the last five overs). These inputs are processed and fed into trained machine learning models to predict the final score of the batting team.

A key feature of the proposed system is its ability to handle both historical and live match data. The system is designed to be dynamic, allowing real-time predictions during ongoing matches. This is achieved through efficient data preprocessing and model optimization techniques.The proposed system also incorporates feature engineering techniques to improve prediction accuracy. Important features are selected based on their impact on the final score, while irrelevant or

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

redundant features are removed. Categorical data such as team names and venues are encoded into numerical values using suitable encoding techniques.

Multiple machine learning models are evaluated, including Linear Regression, Decision Tree Regression, and Random Forest Regression. The best-performing model is selected based on evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Overall, the proposed system provides a scalable, efficient, and accurate solution for IPL score prediction, making it useful for analysts, teams, and cricket enthusiasts.

V. IMPLEMENTATION

The implementation of the IPL Score Prediction System involves several stages, including data collection, preprocessing, model training, evaluation, and deployment.

1. Data Collection Historical IPL match data is collected from publicly available datasets. The dataset includes features such as team names, venue, runs scored, overs, wickets, and match conditions. This data forms the foundation for training the machine learning models.

2. Data Preprocessing Raw data often contains inconsistencies, missing values, and categorical variables. Data preprocessing involves:

Handling missing values
Removing duplicates
Encoding categorical variables (e.g., team names, venues)
Normalizing numerical data

Feature engineering is also performed to create new meaningful features, such as run rate and recent performance indicators.

3. Model Training

The preprocessed data is split into training and testing datasets. Various machine learning algorithms are applied, including:

Linear Regression
Decision Tree Regression
Random Forest Regression

Each model is trained using the training dataset, allowing it to learn patterns and relationships

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

between input features and the target variable (final score).

4. Model Evaluation

The trained models are evaluated using performance metrics such as:

Mean Absolute Error (MAE)
Root Mean Square Error (RMSE)
R-squared Score

The model with the best performance is selected for deployment.

5. Model Deployment

The selected model is integrated into a web-based application (e.g., Django or Flask). Users can input match parameters through a user interface, and the system provides real-time score predictions.

6. User Interface

The system includes a simple and interactive interface where users can:

Enter match details
View predicted scores
Analyze results
7. Tools and Technologies
Programming Language: Python
Libraries: Pandas, NumPy, Scikit-learn, Matplotlib
Framework: Django/Flask

The implementation ensures scalability and efficiency, allowing future enhancements such as live data integration and advanced analytics.

VI. ALGORITHMS

The IPL Score Prediction system utilizes several machine learning algorithms, primarily focusing on regression techniques.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

1. Linear Regression

Linear Regression is a simple and widely used algorithm that establishes a linear relationship between input features and the target variable. It is mathematically

2. Decision Tree Regression

Decision Tree Regression splits the dataset into smaller subsets based on feature values. It forms a tree-like structure where each node represents a decision rule.

Advantages:

Handles non-linear relationships
Easy to visualize

Disadvantages:

Prone to overfitting
3. Random Forest Regression

Random Forest is an ensemble learning technique that combines multiple decision trees to improve accuracy and reduce overfitting.

Advantages:

High accuracy
Robust to noise
Handles large datasets efficiently
4. Evaluation Metrics

To measure model performance:

MAE (Mean Absolute Error): Average of absolute differences
RMSE (Root Mean Square Error): Penalizes large errors
R^2 Score: Measures goodness of fit

These algorithms collectively ensure accurate and reliable score predictions.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

VII. SYSTEM DESIGN

The IPL Score Prediction System is designed using a modular architecture that integrates machine learning with a user-friendly interface.

1. Architecture Overview

The system follows a three-layer architecture:

Presentation Layer (UI)
Application Layer (Backend)
Machine Learning Layer
2. Presentation Layer

The frontend is developed using HTML, CSS, and web frameworks. It allows users to input match details such as:

Batting team
Bowling team
Runs scored
Overs completed
Wickets fallen

The interface ensures ease of use and accessibility.

3. Application Layer

The backend handles:

Request processing
Data validation
Communication with ML models

Frameworks like Django manage routing and server-side logic.

4. Machine Learning Layer

This layer performs:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

Model training
Prediction generation

It processes input data and returns predicted scores.

5. Data Flow
User inputs match data
Backend validates input
Data sent to ML model
Model predicts score
Output displayed to user
6. Database Design

Stores:

Historical match data
User inputs
Model outputs
7. System Features
Real-time prediction
High accuracy
Scalable architecture
User-friendly interface
8. Security & Validation
Input validation
Error handling
Secure data processing

SYSTEM DESIGN IMAGES

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

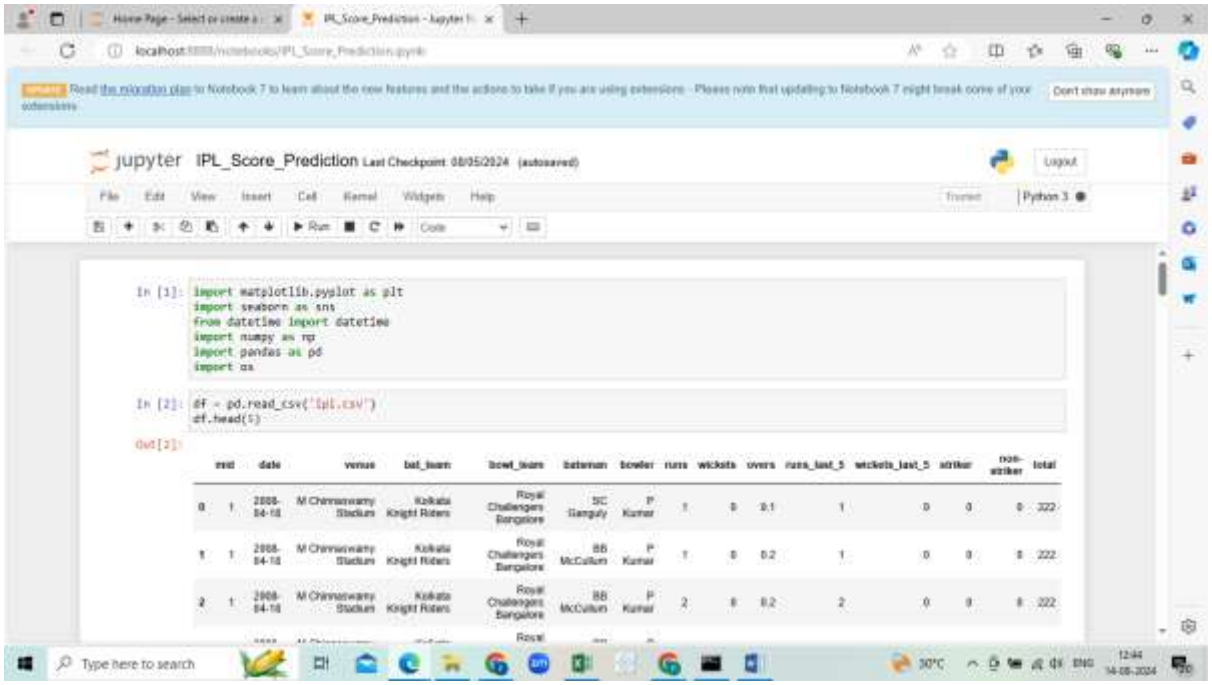

Fig: Loading the libraries and reading dataset

First step is loading the required libraries. After that, using pandas library, IPL dataset is read and displaying first five records of dataset.

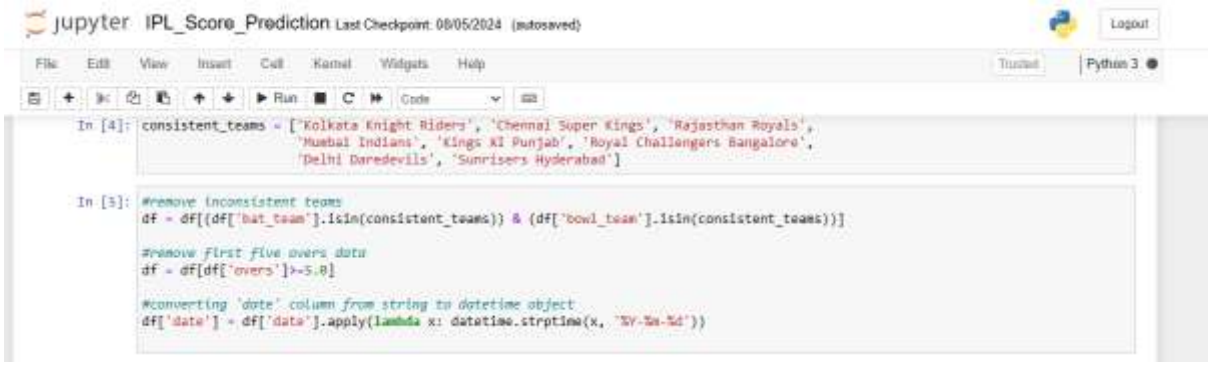

Fig: Data cleaning

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

We have to check consistent and inconsistent teams. Here inconsistent teams are removed.

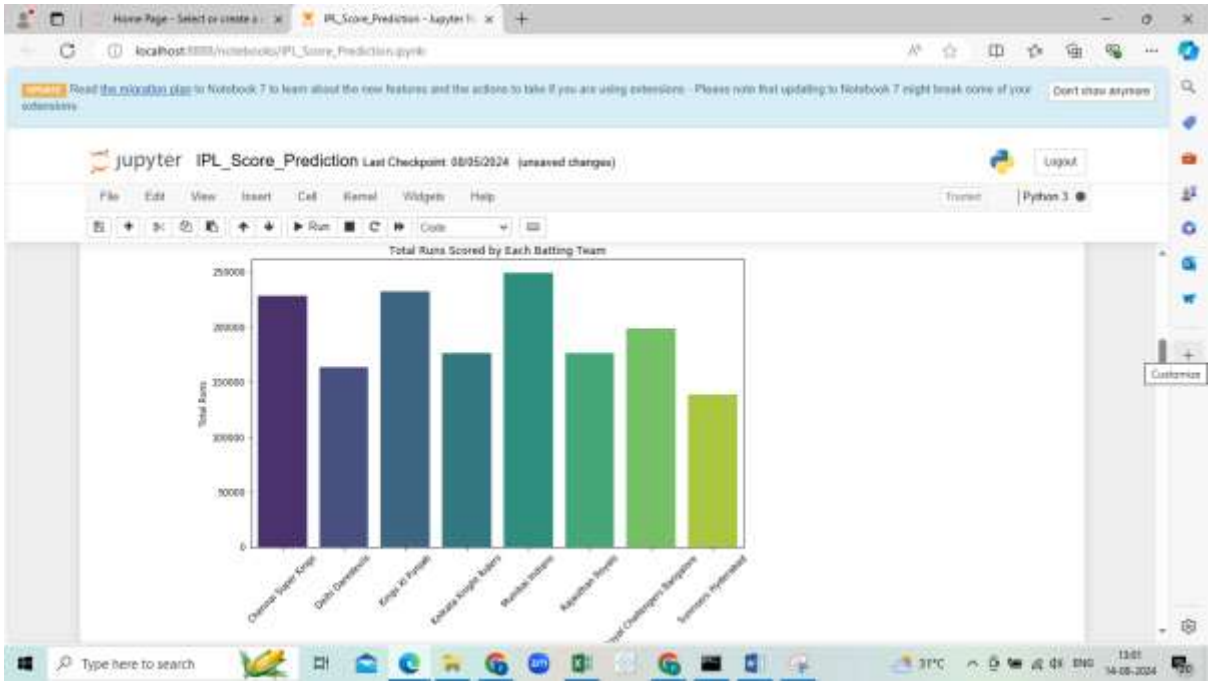

Fig: Batting teams Data analysis

This is the Bar plot of batting teams vs Runs from last 5 overs.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

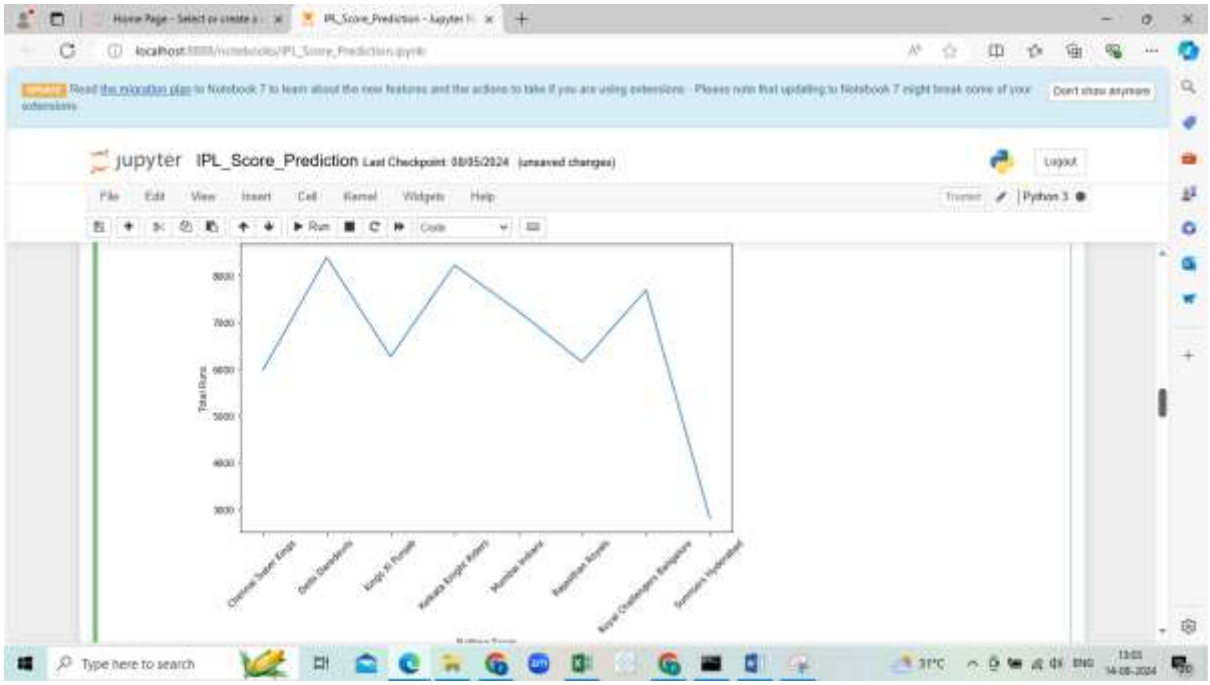

Fig: Bowling teams data analysis

Here we are plotted line chart of bowling teams vs wickets from last five overs.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

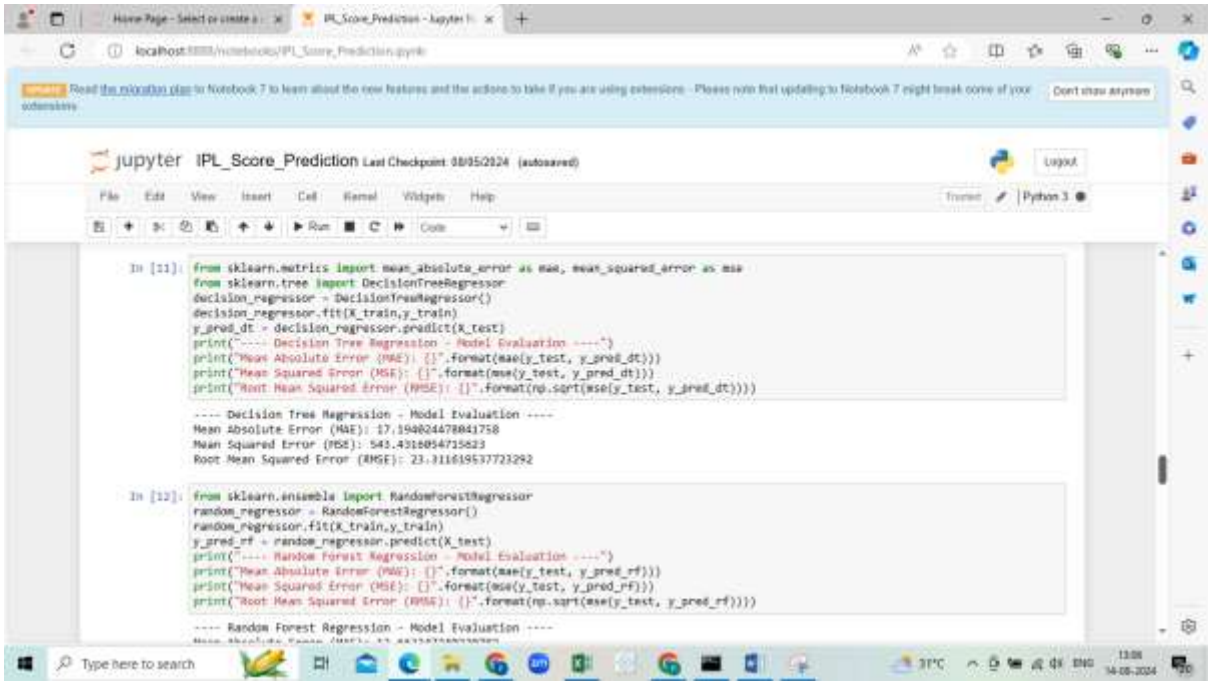

Fig: Training the algorithm

Here we are training different algorithms such as decision tree, random forest, Naïve Bayes, KNN and calculating their MAE,MSE and RMSE

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

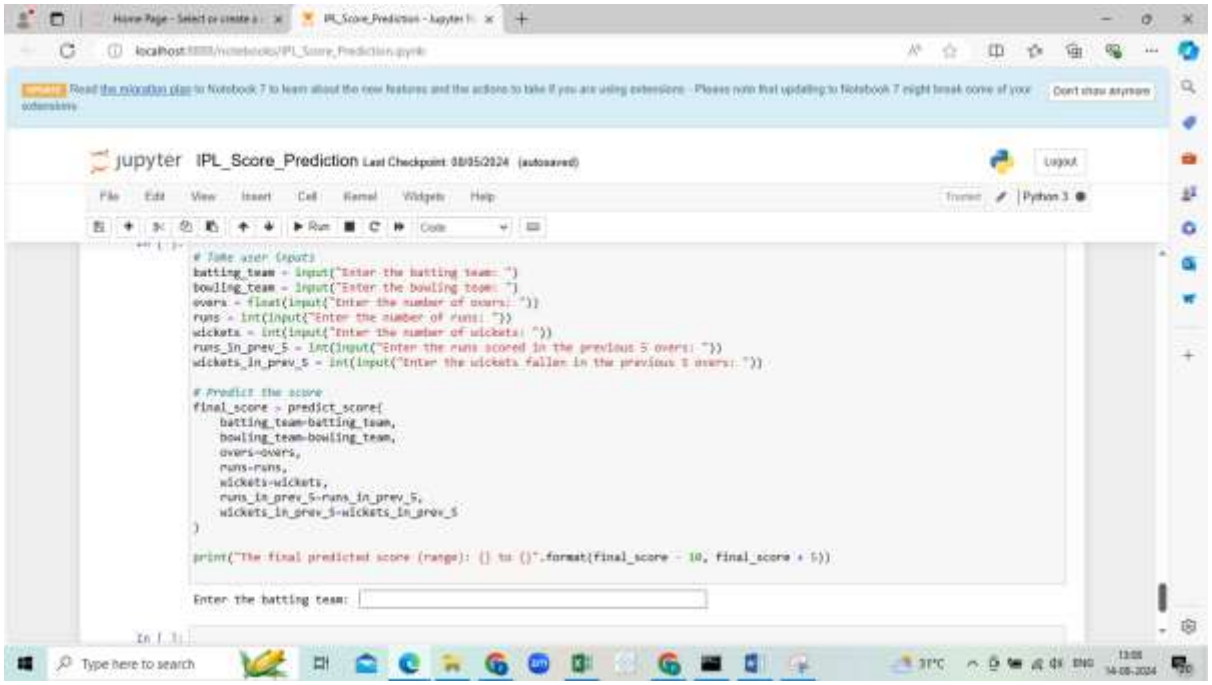

Fig: Predicting the runs

After entering inputs, it will predict the runs.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

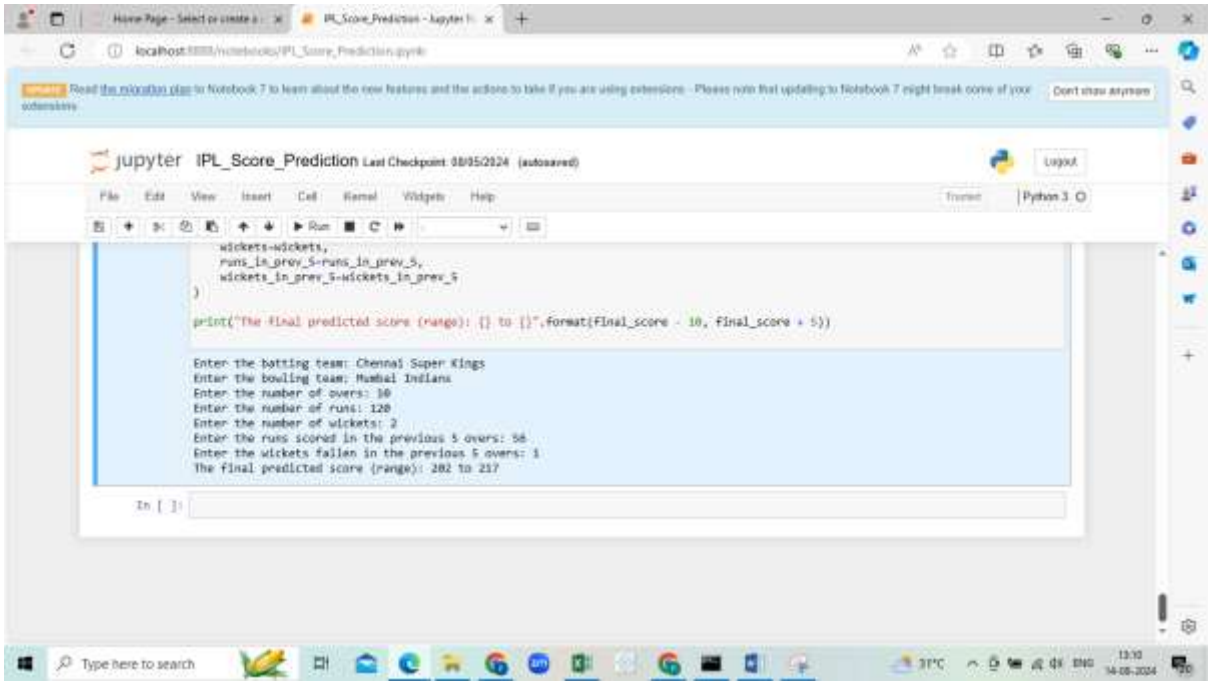

Fig: Prediction

After entering all values, it predicted the score range.

VIII. CONCLUSION

The IPL Score Prediction System demonstrates the effective application of machine learning techniques in sports analytics. By leveraging historical data and advanced regression algorithms, the system successfully predicts match scores with a high degree of accuracy.

One of the key strengths of this system is its ability to handle multiple influencing factors, such as team performance, match conditions, and recent gameplay statistics. This enables the model to capture complex patterns and relationships that traditional statistical methods often fail to identify.The use of machine learning algorithms such as Linear Regression, Decision Tree, and Random Forest enhances prediction accuracy and provides flexibility in model selection. Among these, ensemble methods like Random Forest have shown superior performance due to their ability to reduce overfitting and improve generalization.

The system's modular design ensures scalability and ease of integration with web-based applications. This allows users to access predictions in real time, making the system practical for live match scenarios. Additionally, the user-friendly interface makes it accessible to both

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

technical and non-technical users.Despite its advantages, the system has certain limitations, such as dependency on data quality and limited consideration of external factors like weather and player injuries. Future enhancements can address these limitations by incorporating real-time data sources and advanced deep learning models.

In conclusion, this project highlights the potential of machine learning in transforming cricket analytics. It provides a reliable and efficient solution for IPL score prediction and lays the foundation for further research in sports data science.

REFERENCES

1. Satwani, K. Coutinho, N. John, and J. A. Jothi, "Live Cricket Predictions for Runs and Win Using Machine Learning," 2024.
2. Z. Ahsan, S. S. Ghumman, A. U. Rehman, and A. Mir, "A Comprehensive Prediction Model for T20 and Test Match Outcomes Using Machine Learning," 2024.
3. M. S. Hossain, K. Rumpa, T. Huque, and M. Hasan, "One Day International Cricket Match Score Prediction using Machine Learning Approaches," 2024.
4. S. Singhal et al., "IPL Score Prediction Using Machine Learning," International Journal of Innovative Research in Technology, vol. 10, no. 12, pp. 2894–2899, 2024.
5. A. Kumar et al., "Cricket Score Prediction Using Deep Learning," IJRASET, 2025.
6. R. Suguna et al., "Machine Learning-Based Cricket Match Prediction and Analysis," 2024.
7. L. Breiman-inspired ensemble extensions in sports analytics, "Random Forest-Based Cricket Prediction Models," 2024.
8. V. Ajitha, "Machine Learning Powered Match Outcome Predictions in IPL," 2025.
9. N. Nagamani, "A Machine Learning Approach to Forecasting Outcomes in Limited Overs Cricket," Recent Trends in Sports, 2025.
10. T. R. Sudhini, "Analysis of Cricket Score Predictor using Machine Learning," 2025.
11. K. Dhinakaran et al., "Enhanced Cricket Match Prediction using Kernel-Based Machine Learning Models," Scientific Reports, 2026.
12. S. Labana et al., "IPL Team Winning Prediction using Machine Learning," 2025.
13. S. Kushwaha et al., "Prediction of IPL Matches and Outcomes Using Machine Learning Techniques," 2024.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

14. “Predicting IPL Cricket Scores Using Machine Learning Techniques,” 2026.

15. “Utilizing Machine Learning for Comprehensive Analysis and Predictive Modelling of IPL T20 Cricket Matches,” 2024.