

**PROJECT SYNOPSIS**  
**ON**  
**“IPL Match Prediction Using Machine Learning”**

This synopsis is submitting to  
**G.H. Rasoni College of Engineering**  
In partial fulfilment of the requirement for the award of Degree of  
**Bachelor of Technology in Information Technology**



**SUBMITTED BY**  
**SHIVAM NARAD (A-62)**  
**SHRAWIL SARVE (A-63)**  
**CHETAN KUMBHARE (B-49)**  
**HIMANSHU DONGARE (B-59)**

**GUIDED BY**  
**PROF. RITU PAWAR**

**2023-24**

**Department of Information Technology**  
**G.H. Rasoni College of Engineering (Autonomous Institute Affiliated to**  
**Rashtrasant Tukdoji Maharaj Nagpur University, Nagpur)**

# **INDEX**

	Page No
ABSTRACT.....	1
1. INTRODUCTION.....	2
1.1 BACKGROUND.....	2
1.2 AIM & OBJECTIVE.....	2
1.3 SCOPE OF PROBLEM.....	3
2. LITERATURE SURVEY.....	3
3. PROPOSED SYSTEM.....	4
3.1 PROPOSED APPROACH.....	4
3.2 PROPOSED ARCHITECHTURE.....	4-7
4. PLAN OF RESEARCH WORK.....	8
5. REFERENCES.....	9

## **ABSTRACT :**

The Indian Premier League (IPL) has emerged as one of the most popular and competitive T20 cricket leagues globally, with a vast following and enormous betting interest. Predicting the outcome of IPL matches accurately is a challenging and intriguing problem, with significant implications for fans, teams, and the betting industry. This research project explores the application of machine learning techniques to predict the winners of IPL matches.

We start by collecting comprehensive historical data on IPL matches, including player statistics, team performance, venue conditions, and various contextual factors. Feature engineering is employed to extract relevant information, and data preprocessing techniques are applied to ensure data quality.

Several machine learning algorithms are evaluated, including decision trees, random forests, support vector machines, and deep neural networks. The models are trained on historical match data and validated using cross-validation techniques to ensure robustness and generalization. Feature importance analysis is conducted to identify the most influential factors in match outcomes.

This research not only contributes to the growing field of sports analytics but also offers valuable insights for cricket enthusiasts, team management, and sports betting professionals. Accurate match win predictions can enhance the overall IPL experience and provide stakeholders with a competitive advantage in this highly dynamic and exciting cricket tournament.

## **1. INTRODUCTION:**

### **1.1 BACKGROUND :**

The Indian Premier League (IPL) has undergone a remarkable transformation since its inception in 2008. What started as a cricketing tournament quickly evolved into a global sporting phenomenon, blending world-class cricketing talent with entertainment, glamour, and a fanatical fan base. The IPL's unique format, featuring franchise-based teams representing different cities and regions of India, has captivated audiences worldwide and made it one of the most-watched cricket leagues on the planet.

While the IPL's popularity is unquestionable, it also presents a captivating challenge for cricket enthusiasts, analysts, and sports betting aficionados: predicting match outcomes accurately. The dynamic nature of T20 cricket, with its short game duration, emphasis on power-hitting, and the influence of diverse factors such as pitch conditions, team compositions, and player form, makes match outcome prediction a complex task. However, it is a challenge that holds immense appeal due to the potential for innovation through data-driven approaches and machine learning techniques.

Machine learning and data analytics have revolutionized the sports industry, offering a new dimension to understanding player performance, team strategies, and match dynamics. These technologies have been successfully applied in various sports, including soccer, basketball, and baseball, to predict game outcomes and gain a competitive edge. With the IPL's increasing reliance on data analytics for player auctions, in-game strategies, and fan engagement, the application of machine learning to predict match winners has become an intriguing area of research.

### **1.2 AIM AND OBJECTIVE :**

The objective of this project is to develop an efficient system for predicting the outcome of T20 cricket matches. Several factors, both in-game and pre-game, such as the venue, past team performance, and the outcome of the toss, play a significant role in determining the result of a match. This project also places a strong emphasis on conducting in-depth exploratory data analysis, building predictive models, and creating data visualizations related to the Indian Premier League (IPL). To achieve the best possible match outcome prediction, a variety of methods will be employed, including supervised machine learning using a Random Forest Classifier and statistical techniques.

Additionally, to ensure convenient access and usability of the predictions, the project intends to deploy a user-friendly web application that is compatible with various web browsers.

### **1.3 SCOPE OF PROBLEM :**

The scope of the problem is to develop a machine learning model for predicting the outcomes of Indian Premier League (IPL) cricket matches. The project will involve the implementation of a pipeline that includes data pre-processing techniques such as one-hot encoding for categorical features and feature selection. Random Forest will be employed for the initial model development, and Logistic Regression will be used as an alternate classifier for comparison. The goal is to leverage historical IPL match data to build a predictive model that can assist cricket enthusiasts and betting enthusiasts in making informed decisions regarding match outcomes.

## **2. LITERATURE SURVEY:**

Apurva Lawate, Nomesh Katare, Salil Hoskeri, Santosh Takle, Prof. Supriya. B. Jadhav (2021). [2] This paper is about a model that can predict the projected score of 1st inning as well as the winner in a IPL cricket match. The performance of model depends on various features like wickets taken in last 5 overs, runs scored in last 5 overs, overs, overall score and wickets at current ball. The proposed model contains data from IPL matches played between in 8 years. This paper will give us step by step insights on how one can predict projected score of 1st inning while the match is still in progress. Linear Regression algorithm is used to predict the score. This model explains about 75.226% of data.. Using this model, we can get good insight during the match on how much score will the current batting team obtain.

Omkar Mozar, Soham More, Shubham Nagare and Prof. Nileema Pathak (2022). “Cricket Score and Winning Prediction” [4] As we all know cricket is the most played game. There are so many series in cricket which are played in our country, one of them is the Indian Premier League (IPL). Now it is conducted among 8 teams. Our proposed system consists a model that has two parts the first one is prediction of score and the second one is team winning prediction. In this the score prediction is done with the help of Lasso Regression algorithm whereas in winning prediction SVM classifier, decision tree classifier and random forest classifier are used. The model uses the supervised machine learning algorithm to predict the winning. Random Forest Classifier is used for good accuracy and stable accuracy so that desired predicted output is accurate.

Dhonge, N., Dhole, S., Wavre, N., Pardakhe,M., & Nagarale [7] (2021) As cricket is the mostly played game. There are so many series are played in country one of them is Indian Premier League (IPL). Now it is conducted among 8 teams. In these papers the model has been proposed that has two methods the first one is prediction of score and the second one is team winning prediction. In these the score prediction includes linear regression, lasso regression and ridge regression whereas in winning prediction SVC classifier, decision tree classifier and random forest classifier are used. The model used the supervised machine learning algorithm to predict the winning. Random Forest Classifier used for good accuracy and the stable accuracy so that desired predicted output is accurate.

### **3. PROPOSED SYSTEM:**

#### **3.1 PROPOSED APPROACH:**

By implementing a Random Forest model and enhancing the feature engineering process, we have potentially build a more accurate and robust predictor for IPL match outcomes.

#### **3.2 PROPOSED ARCHITECHTURE:**

##### **1) Data Loading:**

- It begins by importing the necessary libraries: numpy and pandas.
- It loads two CSV files, 'matches.csv' and 'deliveries.csv', into Pandas DataFrames named match and delivery.

##### **2) Data Preprocessing:**

- It extracts the total runs scored by each team in each inning and stores this information in the total\_score\_df DataFrame.
- It filters the data to retain only the first inning in the total\_score\_df.
- It merges the match DataFrame with the total\_score\_df to create a new DataFrame match\_df containing information about matches and their total runs.

##### **3) Team Name Standardization:**

- The code standardizes some team names to ensure consistency. For example, it replaces 'Delhi Daredevils' with 'Delhi Capitals' and 'Deccan Chargers' with 'Sunrisers Hyderabad'.

##### **4) Filtering Matches:**

- The code further filters matches based on certain conditions, like eliminating matches with the Duckworth-Lewis method applied (dl\_applied == 0) and retaining matches involving specific teams in the teams list.

#### 5) Data Merging:

- It merges the filtered match data with delivery data to create the `delivery_df` DataFrame. This DataFrame contains information about deliveries in each match.

#### 6) Calculations and Feature Engineering:

- It calculates and adds several columns to `delivery_df`, including `current_score`, `runs_left`, `balls_left`, and `wickets`.
- It defines a function `result` to determine whether the batting team won the match and applies it to create a 'result' column in `delivery_df`.
- It calculates the current run rate (`crr`) and required run rate (`rrr`) for each delivery.

#### 7) Splitting Data for Machine Learning:

- It prepares the data for machine learning by creating the `final_df` DataFrame with relevant columns.
- It shuffles the rows in `final_df` using the `sample` method.
- It drops rows with missing values and rows where there are no balls left (`balls_left != 0`).
- It splits the data into features (`X`) and the target variable (`y`).
- It imports additional libraries for machine learning: `ColumnTransformer`, `OneHotEncoder`, `LogisticRegression`, `RandomForestClassifier`, and `Pipeline`.

#### 8) Machine Learning Pipeline:

- It sets up a machine learning pipeline that includes one-hot encoding for categorical features and a logistic regression model.

#### 9) Model Training and Evaluation:

- It fits the pipeline on the training data (`X_train`, `y_train`) and makes predictions on the test data (`X_test`).
- It calculates the accuracy of the model using `accuracy_score`.

#### 10) Match Summary and Progression:

- It defines two functions: `match_summary` to print match summary information, and `match_progression` to track the progress of a match and make predictions.
- The `match_progression` function takes a match ID, extracts relevant data for the match, and predicts the probability of the batting team winning at each over.
- It also calculates and prints other match-related information

#### 11) Visualization:

- The code plots a graph to visualize the match's progress, including wickets, win probability, and lose probability.

#### 12) Model Serialization:

- It imports pickle to serialize (save) the trained model pipeline as 'pipe.pkl'.

#### PROPOSED SYSTEM ADVANTAGES :

- Complexity is less compared to previous process.
- Ability to learn and extract complex features.
- Accuracy is good.
- With its simplicity and fast processing time, the proposed algorithm gives better execution time.
- machine learning algorithm (Random forest) is performed to predict the value effectively.
- Prediction is accurate.



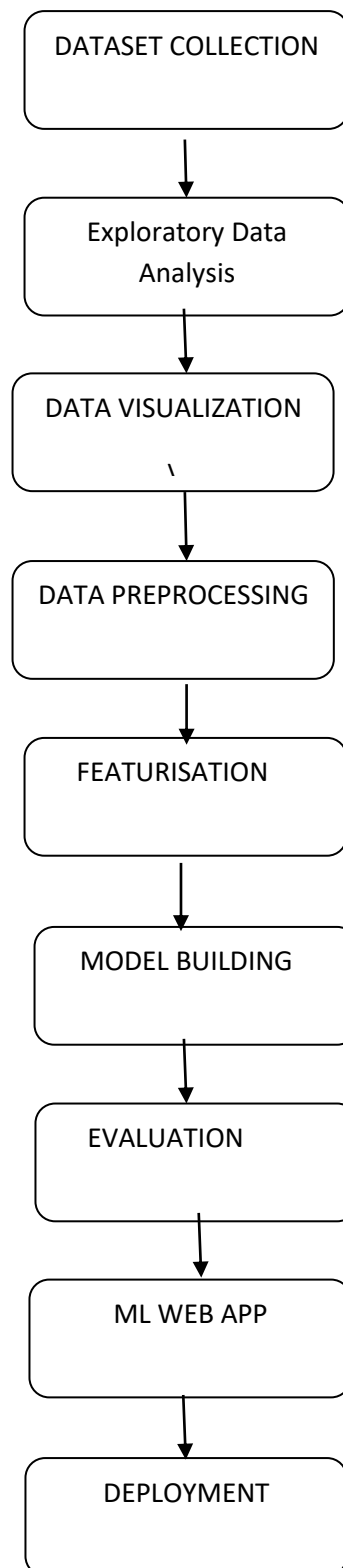


FIG: System Architecture

#### **4 PLAN OF RESEARCH WORK:**

<b>Sr No</b>	<b>Month</b>	<b>Activity Planned</b>
1	July	Literature survey.
2	August	Data collection and paper publication on literature review.
3	August	Design phase – Data flow analysis.
4	September	Implementation.
5	September	Comparative study and result analysis and thesis writing.
6	October	Thesis writing and paper publication based on implementation.

## 5 REFERENCES:

- [1] Ann Thomas, Dr. Jibrael Jos , “A COMPARATIVE STUDY OF MACHINE LEARNING MODELS IN IPL SCORE PREDICTION”, International Research Journal of Modernization in Engineering Technology and Science, Volume:04,Issue:04,April-2022.
- [2] Apurva Lawate, Nomesh Katore, Salil Hoskeri, Santosh Takle, Prof. Supriya. B. Jadhav (2021). “Cricket Analysis and Prediction of projected Score and Winner using Machine Learning”.
- [3] Kiran Gawande, Prof. Sumit Harale, Simran Pakhare “PREDICTIVE ANALYSIS OF AN IPL MATCH USING MACHINE LEARNING” IJCRT Volume 10, Issue 4 April 2022.
- [4] Omkar Mozar, Soham More, Shubham Nagare and Prof. Nileema Pathak (2022). “Cricket Score and Winning Prediction” Volume:09,Issue:04,April-2022.
- [5] Srinivasa Reddy gajjala, Deepika Gonnabattula, Manideep kanakam, Sai Swaroop Maddipudi, “Prediction of IPL Match Score and Winner Using Machine Learning Algorithms”, Journal of Emerging Technologies and Innovative Research Volume 8 Issue 6,June-2021.
- [6] Rajesh Goel “Dynamic Cricket match outcome prediction” Journal of Sports Analytics 2021.
- [7] Dhonge, N., Dhole, S., Wavre, N., Pardakhe,M., & Nagarale, A (2021). “IPL Cricket Scoreand Winning Prediction Using Machine Learning Techniques”
- [8] Srikantaiah, Aryan Khetan1, Baibhav Kumar, Divy Tolani, Harshal Patel “Prediction of IPL Match Outcome Using Machine Learning Techniques.” Intelligent Computing Communication & Security (ICIC 2021)

### Guide

Prof. Ritu Pawar  
Assistant / Associate Professor  
Department of Information Technology  
G H R C E, Nagpur

### HOD

**Dr. Mahendra Gaikwad**  
Head of Department  
Information Technology  
GHRCE, Nagpur