**Predicting Rising Stars in Indian Cricket**

**MINI PROJECT**

By

**Dhruv Talati 60004180022**

**Naitik Rathod 60004180054**

**Nishit Mistry 60004180066**

**Manan Parikh 60004180049**

Guide:

**Aniket Kore**

Asst. Professor

|  |  |  |
| --- | --- | --- |
|  |  |  |

University of Mumbai

2020-2021

**CERTIFICATE**

This is to certify that the mini project entitled **“Predicting Rising Stars in Indian Cricket”** is a bonafide work of **“Dhruv Talati(60004180022), Naitik Rathod(60004180054), Nishit Mistry(60004180066), Manan Parikh(60004180049)”** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of B.E. in Computer Engineering.

**(Name and sign)**

**Guide**

**Prof. Aniket Kore**

**(Name and sign) (Name and sign)**

**Head of Department Principal**

**Mini Project Report Approval.**

This mini project report entitled **Predicting Rising Stars in Indian Cricket** by **Dhruv Talati, Naitik Rathod, Nishit Mistry, Manan Parikh** is approved for the partial fulfillment of the degree of ***B.E. in Computer Engineering.***

Examiners

1.---------------------------------------------

2.---------------------------------------------

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

-----------------------------------------

(Signature)

-----------------------------------------

Dhruv Talati     60004180022

Naitik Rathod   60004180054

Nishit Mistry     60004180066

Manan Parikh        60004180049

Date:

**Abstract**

There are immense number of predictions and speculations about cricket ongoing on the internet all the time. These are usually based on how good the player plays in the previous matches. A system which predicts the future scope of a player is exigent. Authors propose a system based on machine learning techniques where we will predict the rising stars in cricket for the bowling, batting, and all-rounders. Data will be collected from various sources and our own database will be created for the project. The preprocessing and analysis based on considerations of different features will be done. Generative machine learning algorithms will be used and various factors will be taken for making the predictions about the rising stars. Multiple factors will be considered for the predictions as in, number of runs, wickets , 50s , 100s, etc.

The system will predict top 10 Rising Stars in each domain i.e., batting, bowling and all-rounders. The system will display which current player will be replaced by the future star. There will be a cap on players age and threshold will be set for the rising stars by considering the lowest score from analysis of the current players.

For proof of applicability, we will be using the older data and showing the predictions. The older data will be taken from espncricinfo datasets. We will be considering the data from different trophies going on in India, IPL data, and then considering the parameters for the predictions.

Multiple machine learning algorithms like SVM, CART and hybrids of the models will be employed to check for better accuracies among all. This system will be very useful for selectors who need to predict the player performance and pick the players for a team. Selection process will become easier and all the data and stats will be available at one place.

**Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Contents** | **Page No.** |
| **1** | **INTRODUCTION** |  |
|  | **1.1 Description** |  |
|  | **1.2 Problem Formulation** |  |
|  | **1.3 Motivation** |  |
|  | **1.4 Proposed Solution** |  |
|  | **1.5 Scope of the project** |  |
| **2** | **REVIEW OF LITERATURE** |  |
|  | **2.1 Previous work** |  |
|  | **2.2 Research Gap** |  |
| **3** | **SYSTEM ANALYSIS** |  |
|  | **3.1 Functional Requirements** |  |
|  | **3.2 Non-Functional Requirements** |  |
|  | **3.3 Specific Requirements** |  |
|  | **3.4 Use-Case Diagrams and description** |  |
| **4** | **ANALYSIS MODELING** |  |
|  | **4.1 Data Modeling** |  |
|  | **4.2 Activity Diagrams / Class Diagram /sequence /collaboration /state** |  |
|  |  |  |
| **5** | **DESIGN** |  |
|  | **5.1 Architectural Design for proposed system** |  |
|  |  |  |
| **6** | **IMPLEMENTATION** |  |
|  | **6.1 Algorithms / Methods Used** |  |
|  | **6.2 Working of the project** |  |
| **7** | **TESTING(white box test cases )** |  |
| **8** | **RESULTS AND DISCUSSIONS** |  |
| **9** | **CONCLUSIONS & FUTURE SCOPE** |  |

Appendix

Literature Cited

Acknowledgements

ii

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Caption** | **Page No.** |
| 3.1 | Use-case diagram of the proposed system |  |
| 4.1 | Data Flow Diagram of the proposed system |  |
| 4.2 | Activity Diagram of the proposed system |  |
| 4.3 | Class Diagram of the proposed system |  |
| 4.4 | Sequence Diagram of the proposed system |  |
| 4.5 | Collaboration Diagram of the proposed system |  |
| 4.6 | State Diagram of the proposed system |  |
| 5.1 | Architectural Design of the proposed system |  |
| 6.1 | Home Page - i |  |
| 6.2 | Home Page - ii |  |
| 6.3 | Home Page - iii |  |
| 6.4 | Home Page - iv |  |
| 6.5 | Login Page |  |
| 6.6 | Batting Stats |  |
| 6.7 | Bowling Stats |  |

iii

**List of Tables**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Table Title** | **Page No.** |
| 7.1 | White Box Testing Results |  |

iv

**List of Abbreviations**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Abbreviation** | **Expanded form** |
| i |  |  |
| ii | RSs | Rising Stars |
| iii |  |  |
| iv |  |  |
| v |  |  |
| vi |  |  |
| vii |  |  |

v

**Chapter 1**

**Introduction**

* 1. **Description**

Cricket is no longer a sport only played by the elites or the maharajas in India as it was a century ago. Today, in India the sport is played by the masses and is loved by all. No other sport has enjoyed the love and affection that cricket has in India. The complexity of cricket has also increased many folds. Earlier it was just the test matches today there are Test matches, ODIs, Twenty20s and the matches are played internationally, domestically and by various age groups like U-19 etc. In India currently there are many tournaments being played like Indian Premier League, Ranji Trophies, Syed Mushtaq Ali and many more. The data and analysis that is available is mainly on the capped players with very little data analysis done on the uncapped players. This makes selectors task of different teams a cumbersome one. We therefore propose a solution to this predicament. We will be using machine learning algorithms and classification algorithms for prediction of rising stars of cricket. There will also be a classification of the rising stars based on their role in the team as a batsman, bowler, all-rounder or a wicket keeper. This will enable teams to have a data on the players who have performed consistently using the rising star prediction algorithm. The teams can make better informed decisions during the selection process and make valuable addition to their teams accordingly. The players will also be benefitted as they can be assured that a one bad game during which the selectors have come to watch their performance will not be a roadblock in their career because selectors have data on their consistency, their performance with respect to co-players etc.

* 1. **Problem Formulation**

The current models and systems available are not designed to analyze the data of different tournaments and make accurate predictions of the Rising Stars. Authors here propose to create a system that is capable of handling the live data from different formats so that the predictions made can be close to perfect.

* 1. **Motivation**

In India, one could ask anyone what they dreamt of becoming in childhood and most often than not the answer would be “in my childhood I always dreamt of becoming a cricketer”. With the 2nd largest population in the world and sheer amount of people playing cricket it is always a tough task for one to predict the rising stars that would become one of the legends of the game in the future.

We propose a method for predicting the rising stars of India’s favorite sport cricket. With thousands of cricket matches being played every day in a country of India’s size it is necessary for franchises and BCCI to have access to data which shows them the rising stars on whom they could keep an eye. There are large number of tournaments like Vijay Hazare, Ranji Trophy, Syed Mushtaq Ali, under-19 matches etc., being played and this data is residing in different locations making it difficult for teams to identify players suited for the slots they are looking to fill in their teams.

* 1. **Proposed Solution**

We propose a solution using a machine learning based model for predicting the rising stars of the game with high accuracy to help the selectors build a better team. The system will be based on the live data being fetched and analysis being done on the spot. Everytime the system is called; data will be checked for any updates and then the further processing will begin. The RSs of all domains will be displayed, along with the current player which can be replaced by the upcoming star. Stats of all the players will also be visible along with the RSs list so that all the information could be fetched in one place.

* 1. **Scope of the Project**

With the current pace at which the cricket matches are being played, there are numerous amounts of changes that are being done in the statistics and the data available on the player. Having a system which is based on live data is necessary and holds credibility over the models which are based on relatively older data. Due to the ease of use which it will provide having the RSs and the stats in one place, it could be a very informative project that can be of great use to the people involved over the selection process of the players.

2

**Chapter 2**

**Review of the Literature**

* 1. **Previous Work**

Authors in [1] suggest that finding the rising stars(RSs) within the domains is of great importance as the organizations can put efforts in betterment and expertise of the RS. Rising star is an emerging player who could become a star in future based on the consistent performance.

The authors have put forward the concept of co-players who play with the RS as it is an essential factor for the rising star prediction in cricket. Three features are defined namely, co-batsmen, team, opposite team for batsmen and similarly co-bowlers, team, opposite teams for bowlers. Total 9 features for the RS in batsmen and 11 features for RS in bowling domain have been considered. Data used is taken from espncricinfo and processed into two datasets.

Multiple Machine learning algorithms have been used and four most appropriate have been used for binary classification. Generative classifiers outperform the others. The RSP is made with high accuracies and rankings are compared with the ICC rankings of 2013-2016 players. Basic terminologies and concepts of the game of crickets are also elaborated. Among the generative models, Bayesian Network(BN) and Naïve Bayesian (NB) are used. Support Vector Machines (SVM) and Classification and Regression Tree (CART) are used in the discriminative models. For the feature evaluation, authors have used the state-of-the-art evaluators information gain, gain ratio and chi-squared statistics. The weighted average of the batsman and bowler is calculated and then their performance is evaluated. Statistical analysis of each feature is shown for their data. The authors used 10 fold cross validation for training and to validate the classifiers for using their datasets of each domain.

In batting domain, highest accuracy achieved was 87.5%, 87.3%, 84%, 78% using BN, NB, CART, SVM on the first dataset. Even the second dataset shows the highest accuracies of 89%, 88%, 80%, 73% fir CART, BN, NB, SVM for the same feature.

In bowling domain, the highest accuracies are 80%, 78.8%, 78.5% and 78.5% by applying SVM, NB, CART and BN using their first dataset and 77.8%, 75%, 74% and 72% accuracies

for BN, CART, SVM and NB classifiers using the other dataset. Category wise and model wise analysis is done and the rankings of ICC are compared with the predictions.

In [2], two types of classification models are considered to learn the desired predictive function F^RS (.) and two algorithms are chosen for each model category.

It is the first attempt that uses supervised machine learning methods for prediction of rising stars. Four famous algorithms are chosen for binary classification of rising stars, although other ML algorithms may also be used. In this work, famous algorithms are selected from wide collection, based on efficient performance and classification accuracy. A set of eleven features are designed on the basis of content and graph information. This feature combination was not considered for prediction of rising stars in previous research. The performance of recommended algorithms is critically analyzed in terms of evaluation metrics and MEMM classifier demonstrates best performance. This novel idea is implemented for rising stars prediction in database domain. It can be implemented for other domains and may be utilized for rising paper prediction.

In this work, the performance of applied classifiers is analyzed by Precision, Recall and F1 evaluation metrics. They have mainly used F1 score to examine the effects of different features for rising star classification accuracy and prediction.

The authors in [3] suggest a solution to the problem of choosing a good batsman in twenty20 using a two-stage method for measuring and ranking batting parameters in cricket using ordered weighted averaging (OWA) operator and regression.

Authors have collected the raw data from the cricinfo website.

The first stage measures the performance of players taking five various capabilities such as Highest Score (HS), Average (AVG), Strike Rate (SR), 4s, 6s using OWA operator and then establishes the ranking of parameters using a regression model.

The authors of this paper try to provide a scientific basis to prove that strike rate is an important measure for good batsmen in shorter version of cricket using OWA and regression methods.

It is shown that for different generated OWA weights, corresponding to various uncertainty levels, the ranking obtained for batting parameters is not sensitive to the change of these weights. In OWA operator there are two important measures, the dispersion (or entropy) and the orness.

In this section, authors propose a two-stage OWA-regression method to prioritize batting parameters including, HS, Avg, S/R, 4s, and 6s for 40 players chosen from IPL 4. In the first stage they use the OWA method for finding the performance of batsmen.

The second stage  they compute the estimated parameters of the model. The corresponding result shows that the most important batting parameters are S/R, HS, Avg, 4s and 5s in the decreasing order of importance.

For different OWA weights, the ranking of batting parameters does not change. This

indicates that the changes of OWA weights, obtained from different orness levels, does not affect the ranking of batting parameters.

On the basis of the mathematical results they have shown that S/R is the most important parameter to choose a batsman for Twenty20 format.

This research paper can help selectors better identify the good batsman further increasing

their chances of winning in the T20 format.

Authors in [4] explore the application of Social Network Analysis (SNA) to rate the players in a team performance. They generate a directed and weighted network of batsmen–bowlers using the player-vs-player information available for Test cricket and ODI cricket. Additionally they generate a network of batsmen and bowlers based on the dismissal record of batsmen in the history of cricket—Test (1877–2011) and ODI (1971–2011).

They obtain data from the cricinfo website. The website contains the information of proceedings of all Test matches played since 1877 and all ODI matches from 1971 onwards. These include the runs scored by batsmen, wickets taken by bowlers, outcome of a game and also the information of the mode of dismissal of a batsman. They collect the data of player-vs-player for Test cricket (2001–2011), and ODI cricket (1999–2011) from the cricinfo website. They have also collected the batting and bowling average of the player

The performance of a batsman is judged by the ‘quality’ of runs scored and not the number of runs scored. Hence, runs scored against a bowler with a lower bowling average carry more credit than runs scored against a bowler of less importance. Authors introduce a performance index of a batsman (PIB) against a bowler given by the following equation

PIB =ABa/CBo

where ABa is the batting average of the batsman against the bowler he faced and CBo refers to the career bowling average of the bowler.

In this paper [5] authors have used a concept of co-players is introduced for predicting the rising stars for better team selection. Analysis for selecting the best features is done using machine learning techniques, the actual prediction of top 10 rising stars is done.

 In this work, authors have defined a concept of co-players, i.e., a player is evaluated considering not only his own performance but also of other players he has played in a common time span. Team performances and opposite team performances are also taken into account. The authors have also introduced the concept of allrounders which will have its own set of features.

The data is taken from ESPNcricinfo for years 2006–2018 for all three domains, i.e. batsman, bowler and all-rounder.

Authors have used the Support vector machine (SVM) which is a discriminative classifier trained for classification and analysis of data based on standard data sets and it is said to be a supervised learning model, where data is plotted in space as a point.

The authors have used features like co-batsmen runs, co-batsmen strike rate and other.

Features for the batting domain. Similarly, they have defined the features for bowling domain and allrounder domain.

To evaluate a player as a rising star or not a rising star, the player who had played at least 20 matches is considered. SVM model is trained, and further applicable rising stars are noted. And the final list of rising stars is derived, according to their RS score. The result obtained based on computation on features states player ability depends not only on his performance but also on performance of other players and team which ensures a good evaluation system. The list was compared by the authors with ICC and accuracy obtained as: 60% for batting domain, bowling accuracy: 70%, all-rounder accuracy:40%.

The authors in [6] have discussed the existing methods to find the rising stars and the pros and cons of the methods are discussed. The datasets and the evaluation of the performances are described too. Open challenges and the future scope is discussed towards the end.

Methods for finding Rising stars in Bayesian Networks are sub categorized into four: Ranking methods, Prediction Methods, Clustering Methods, Analysis Methods. Earlier published papers using these methods are discussed and their improvements done later as discussed. Findings and Limitations of multiple publications and their dataset are stated and the scope of improvement among those is described. Basically the authors state that collaborations with renowned researchers can also lead to a rising future. Similarly, all the methods in Bayesian networks are discussed, their earlier works, best works and future scopes are mentioned. Applications of finding rising stars in other domains like community question answering networks, sports networks, telecommunication networks are elaborated with relevant works in those fields. Datasets of academic networks and miscellaneous networks are specified with the main features of all the datasets.

Challenges like falsely predicted rising stars, using multiple data sources, applications in multiple domains, long term prediction impacts and many other challenges and future directions are discussed in brief.

The authors in [7] obtained all the data from www.cricinfo.com using scraping tools, parsehub and import.io. For batting, they considered matches played from January 14, 2005 to July 10, 2017. The senior most player during this span was SR Tendulkar, so they collected innings by innings list of the performance of all the batsmen from December 18, 1989 when he played his first ODI match. For bowling, they considered matches played from January 2, 2000 to July 10, 2017. The senior most player during this span was PA de Silva, so they collected innings by innings list of the performance of all the batsmen from March 31, 1984 when he played his first ODI match. Since the past stats of the players such as average, strike rate etc. are not available directly online for each match they played, they calculated from the innings by innings list for each match. They imported all the data in MySQL tables and used php to manipulate them.

For predictive analytics, they used Weka and Dataiku. Both these tools are a collection of machine learning algorithms for data mining and also provide some preprocessing functionalities.

They have made various attributes i.e for Batting attributes:

- No of innings: The number of innings the player has played till the day of that match, which shows the experience of the player

- Batting Avg: This attribute depicts the run scoring capability of that player

- Strike Rate: This attribute depicts how fast the player can score runs.

- Other attributes included the centuries, fifties, zeros and the highest score of that player

Bowling attributes:

- No of innings: the total number of matches the player has played till the day of the match

- Overs: Total number of overs bowled by the bowler, for experience

- Bowling Avg: Bowling average is the number of runs conceded by a bowler per wicket taken.

- Bowling Strike Rate: Bowling strike rate is the number of balls bowled per wicket taken.

- Four/Five Wicket Haul: Number of innings in which the bowler has taken more than four wickets.

Derived attributes:

- Consistency: This attribute describes how experienced the player is and how consistent he has been throughout his career.

- Form: Form of a player describes his performance over last one year. All the traditional attributes used in this formula are calculated over the matches played by the player in last 12 months from the day of the match.

- Opposition: Opposition describes a player’s performance against a particular team. All the traditional attributes

used in this formula are calculated over all the matches played by the player against the opposition team in his entire career till the day of the match.

- Venue: Venue describes a player’s performance at a particular venue. All the traditional attributes used in this formula are calculated over all the matches played by the player at the venue in his entire career till the day of the match.

This paper [8] is centered on the implementation of machine learning to foretell the winner of an IPL match.

The historical dataset was obtained from various sources like ESPN Cricinfo and iplt20. Feature engineering techniques were applied by authors to derive more insights about the current dataset. Most of the Machine Learning algorithms work better with numerical values than the string values. Hence all the string formats in the dataset were converted to the numerical formats utilizing the Label Encoding by the authors.

To produce accurate results, all the unnecessary features from the dataset were eliminated eg- Umpire Name, StadiumName etc.

To rule out the class imbalance authors have designed the model to predict the winner based on the essential features instead of the Team names, declaring either Team 1 or Team 2 as a winner. Authors have considered the data of only 11 players for a team based on the highest number of matches they have played during the IPL was considered.

Authors have referred to the DREAM 11 points table like total score of bowler etc to derive the formula.

Authors have weighted the features according to their relative importance over other measures

eg-For Batting Analytic Hierarchy Process (AHP) the attributes were arranged by the authors in their decreasing order of importance:-

Batting Average > Innings > Strike Rate > 50’s > 100’s > 0’s

Similarly Bowling AHP and the yearly ranks of each team based on the win ratios was noted by the authors and the ranks were derived using AHP.

Teams’ ranking was done according to the teams’ points, and past performance features were fed to the model for predicting the results.

The consistency of a team adds more weightage to its current performance than the overall performance and hence authors have assigned 80 percent weightage to the current performance of a team and 20 percent weightage to their overall performance.

A machine learning model is asymmetric in nature and is neither capable of identifying the symmetry of features nor has a way to input the information about the symmetry of features. Hence, this information was entered to the model by generating a symmetric duplicate for every row in the training set by the authors.

Several machine learning models were applied by the authors  to the

selected features to predict the IPL match results. The best results were concluded using the tree-based classifiers. The highest accuracy of 60.043% with Random Forest with a standard deviation of 6.3% and an ambiguity of 1.4% was observed by them.

The research by authors focused on predicting the winner for an IPL match using machine learning and utilized the available historical data of IPL from season 2008-2019.

The authors in [9] made initial step involved in developing the system to pre-

process the raw data. The raw data pertaining to players of countries are fetched from espncricinfo portal. Also, the statistical information pertaining to the players are fetched

from espncricinfo statsguru repository. In this work, the data mining algorithms used are k-means clustering, decision trees, random forest and support vector machines. The software used to perform data mining operations over the selected dataset is WEKA

The data fetched from espncricinfo portal is populated into an excel file and the CSV loader is used to feed the data into the knowledge system. The dataset is then passed on to the class assigner where the typical classes are provided to the data. The class labelled data is then passed to the cross fold validation to create training and testing data. The test set and training set is then fed to the classifier algorithms and the results are observed. The classifier block is changed according to the type of classification algorithm that is to be used.

The precision values for batsman, bowler and all-rounder are 0.94, 0.95 and 0.81 respectively. The recall values for batsman, bowler and all-rounder are 0.93, 0.92 and 0.81 respectively. The prediction accuracy of the classifier is 91.87% when decision tree classification is used, 93.46% when SVM is used and 95.78% when random forest used.

The authors in [10] have used the cricket players dataset collected over the past 10 years from a website and the main outcome of the system is to predict if the team will win the match or not. Random forest model is used as it gave the best results for the authors as compared to the other algorithms. Various metrics were used for performance testing of the players like weather conditions, previous scores, maiden overs ,etc.

* 1. **Research Gap**

All the previous work done in this domain has been done for the rising stars in the international players only. There is no study where the upcoming players/ uncapped players of any country are analyzed for the RSs. The authors here propose to predict the RSs considering the players performances in the Vijay Hazare trophy, IPL, Ranji Games. Upcoming players capable of being a rising star will be predicted and can be used for the selection purposes.

1. **System Analysis**
   1. **Functional Requirements**

Get Cricket Data - download, filter, and store the required data in the local database. Structured data from apis like cricapi and espncricinfo will be used. System should be able to save the records of players and store it in the local database for analysis.

Analysis Strategy - specify new equations that will consider the players performance in all different formats played. The equation will calculate the feature score for all the players. The minimum score among the capped players will be set as the threshold for predicting the rising stars among the upcoming players.

Displaying RSs and player stats: For each player, obtain data from various reliable sources. Analyze the data and display the RSs and their stats. This will make it easier for selectors as all the data will be available at one place.

* 1. **Non-Functional Requirements**

Performance - is the amount of work accomplished by a system compared to the time and resources required to do that work. Our system would compile the data and analyze the data from the various sources to achieve higher accuracies with faster performance.

Interoperability - is the ease with which the system is able to operate with a diverse range of data. Data from different formats like Tests, ODIs, Twenty20s will be processed under same system for the rising stars prediction. Different tournaments can be taken under consideration for the model.

Reliability - is the accuracy and dependability of a system to produce correct outputs for the inputs it receives. Data is being fetched from reliable sources such as ESPNcricinfo and cricapi. This will have lesser errors and accurate predictions will be made.

* 1. **Specific Requirements**

Player data in local database used for training the models should be secured and no manipulation should be done to the data after the model is trained as it will lead to faulty results.

Data obtained from various sources should be stored in similar manner for faster training and prediction purposes.

* 1. **Use Case Diagram**

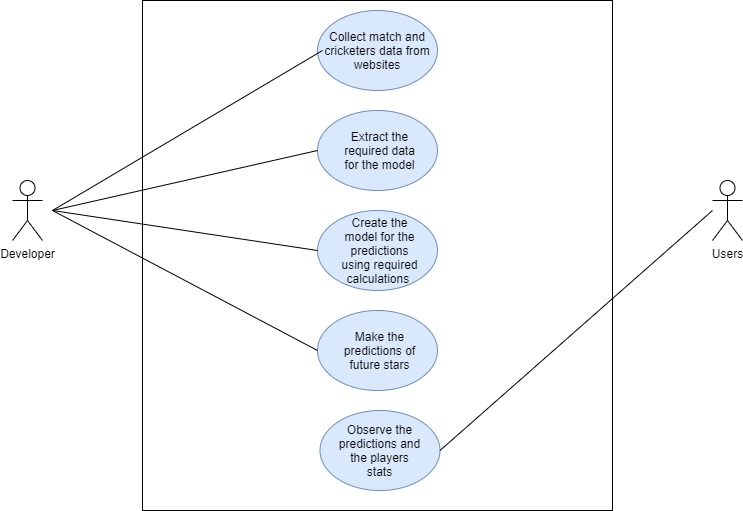
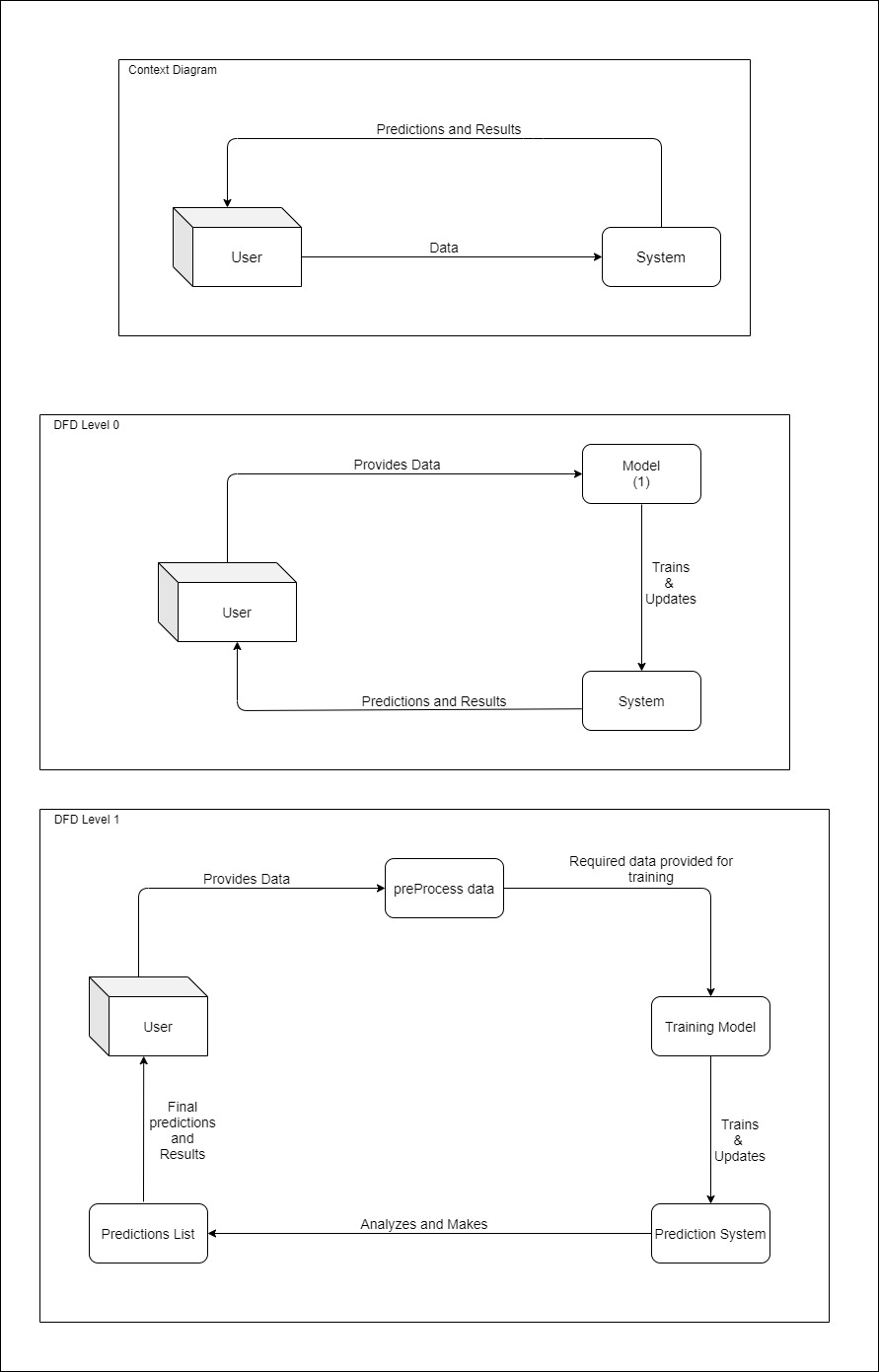


Fig 3.1: Use Case Diagram of the proposed system

1. **Analysis Modelling**
   1. **Data Modelling**



* 1. **Activity Diagrams / Class Diagram /sequence /collaboration /state**

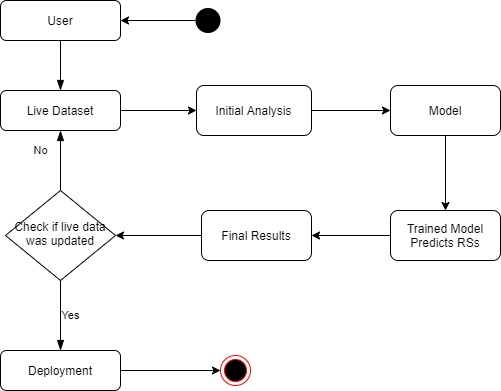


Fig 4.2 Activity Diagram of the proposed system

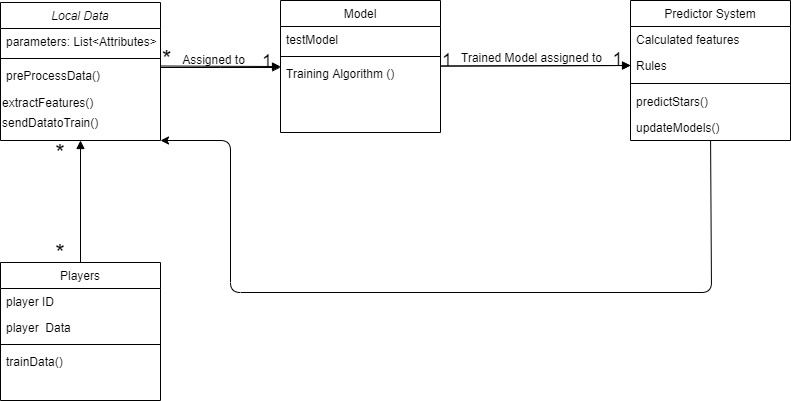


Fig 4.3 Class Diagram of the proposed system

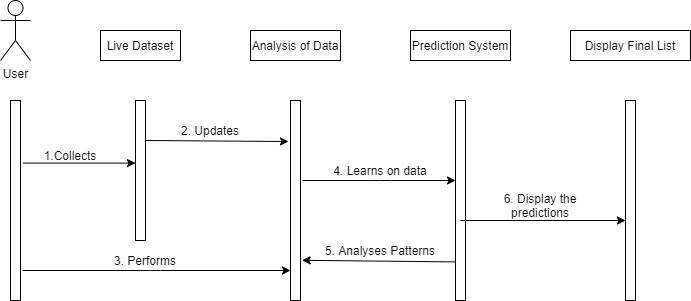


Fig 4.4 Sequence Diagram of the proposed system

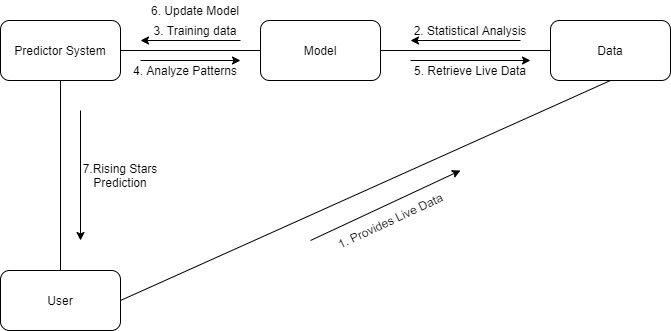


Fig 4.5 Collaboration Diagram of the proposed system

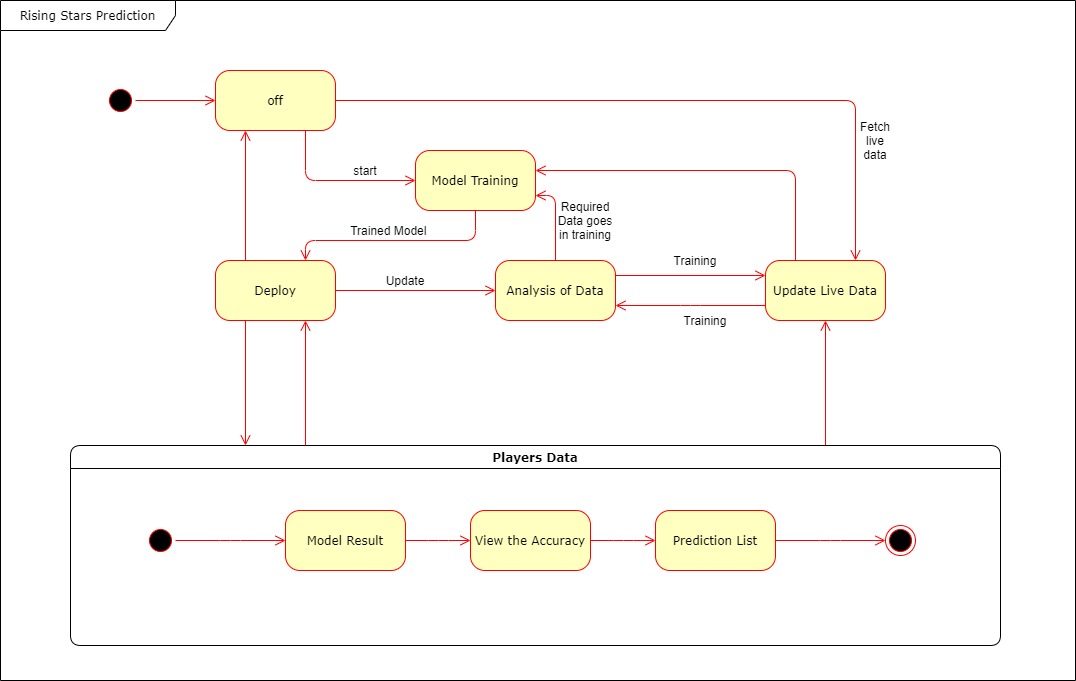


Fig 4.6 State Diagram of the proposed system

1. **Design**
   1. **Architectural Design for the Proposed System**

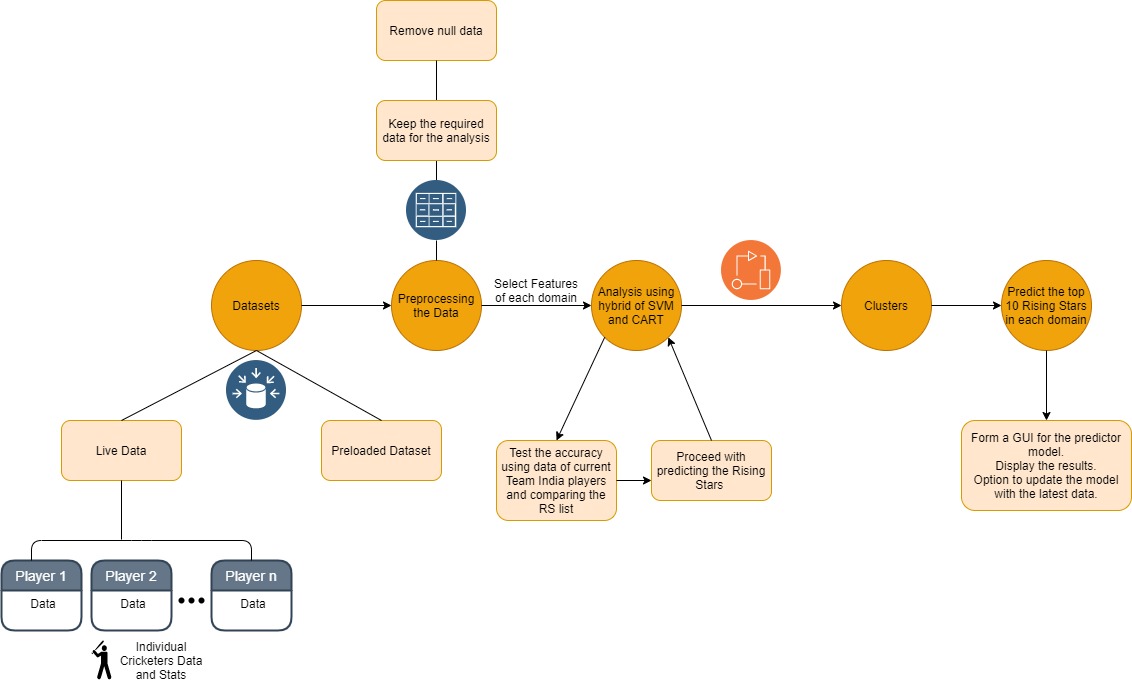


Fig 5.1: Architectural Design of the proposed system

1. **Implementation**
   1. **Algorithms/Methods used**

Currently, no algorithms were used in the implementation. The Player’s data is being fetched from the cricapi and being displayed in the webpage.

The player’s stats including the batting and bowling in all the formats is fetched and displayed currently. The data analysis and the analysis based on different features will be done further, with the model training and testing and the prediction part.

* 1. **Working of the project**

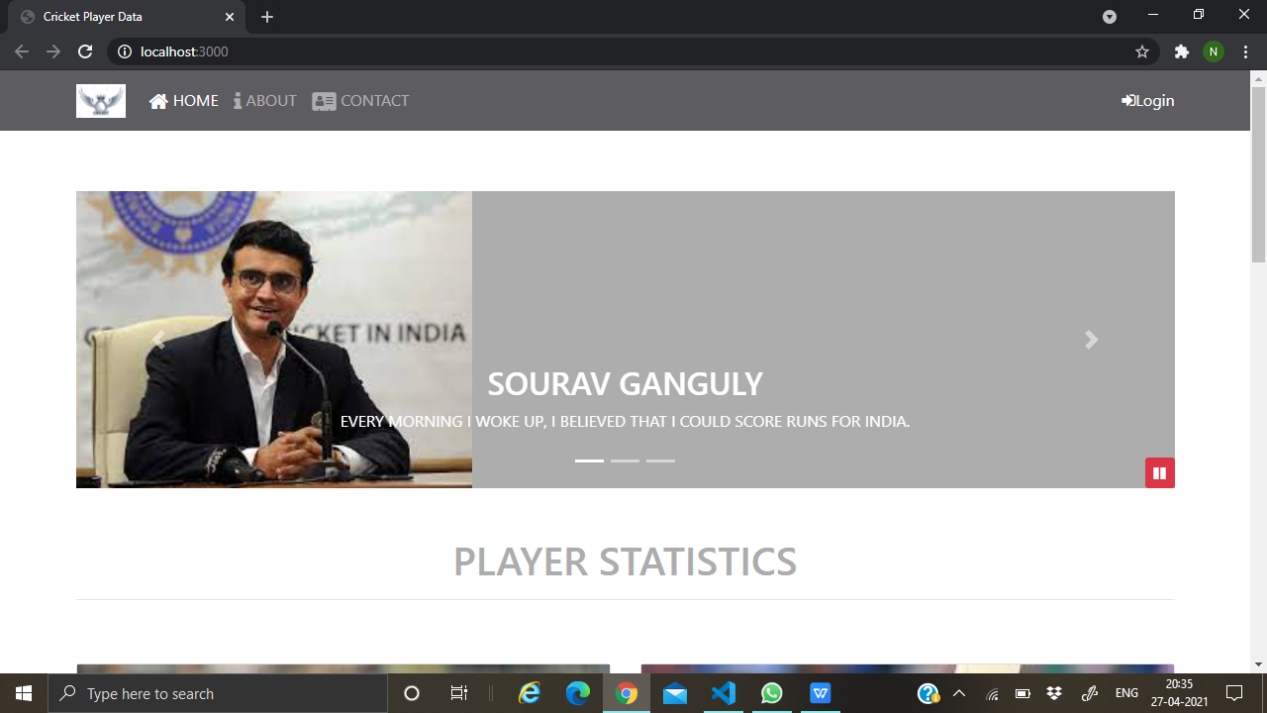


Fig 6.1 Home Page-i

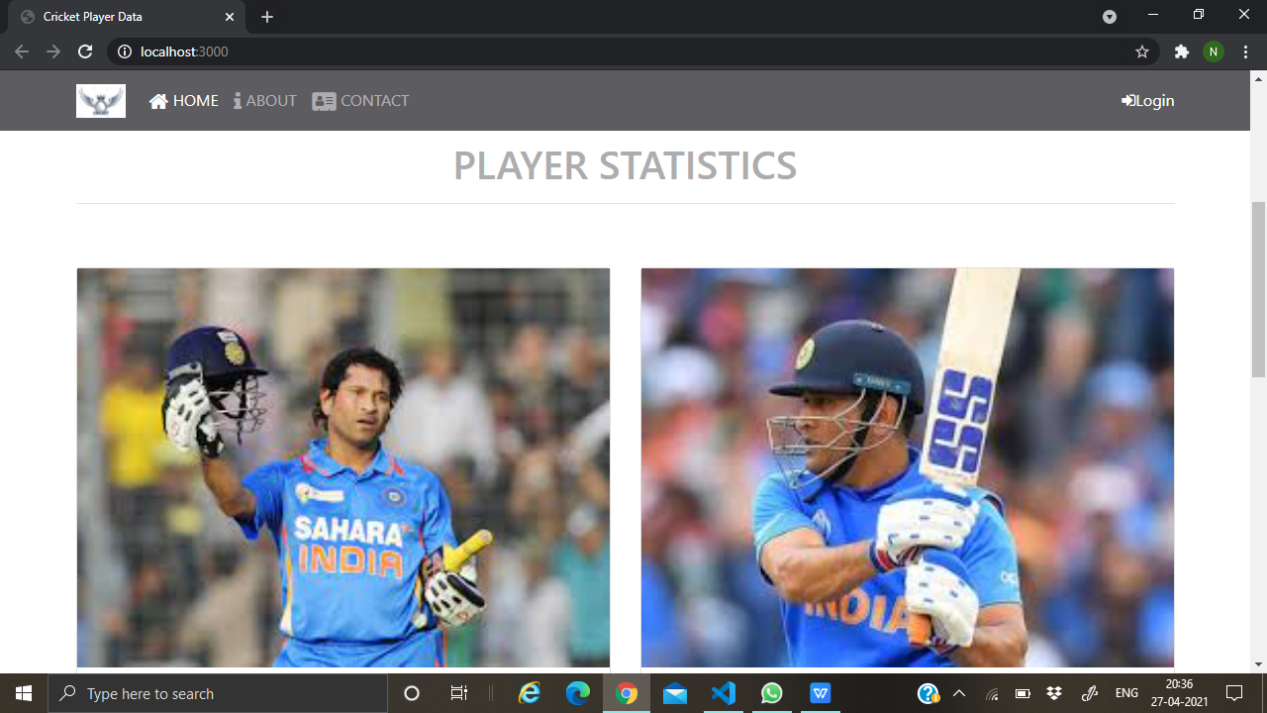


Fig 6.2 Home Page-ii

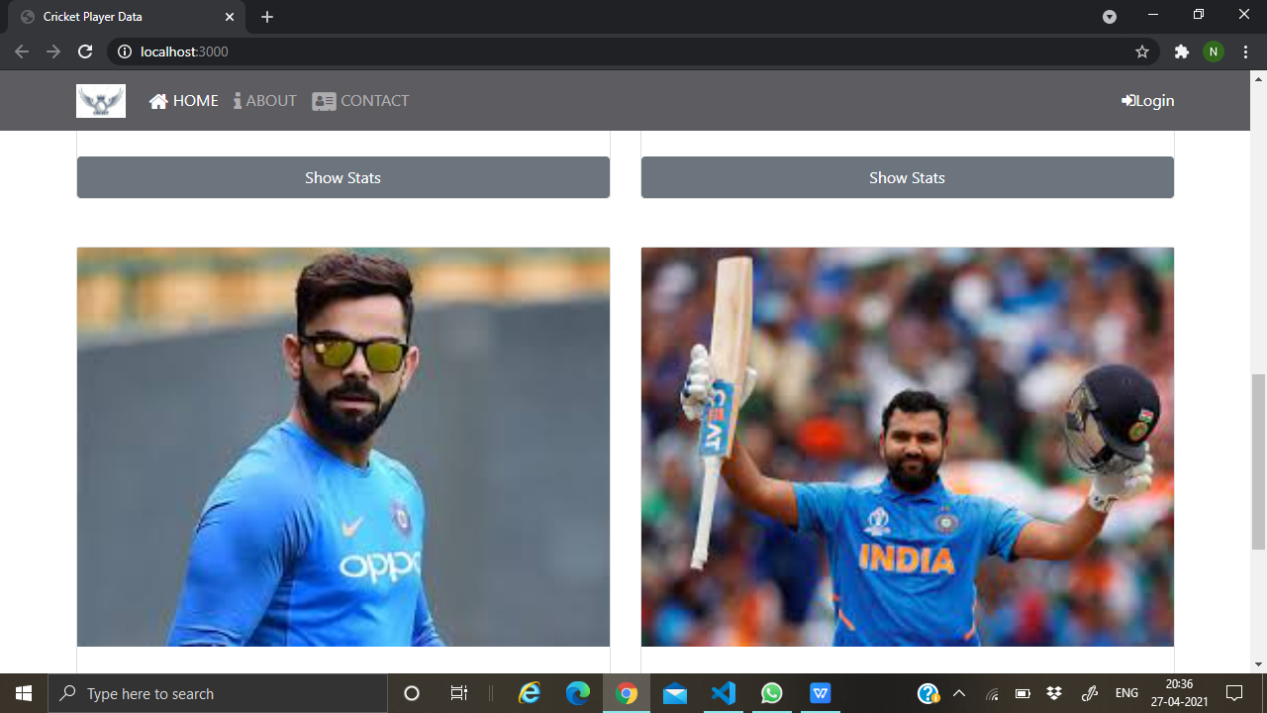


Fig 6.3 Home Page-iii

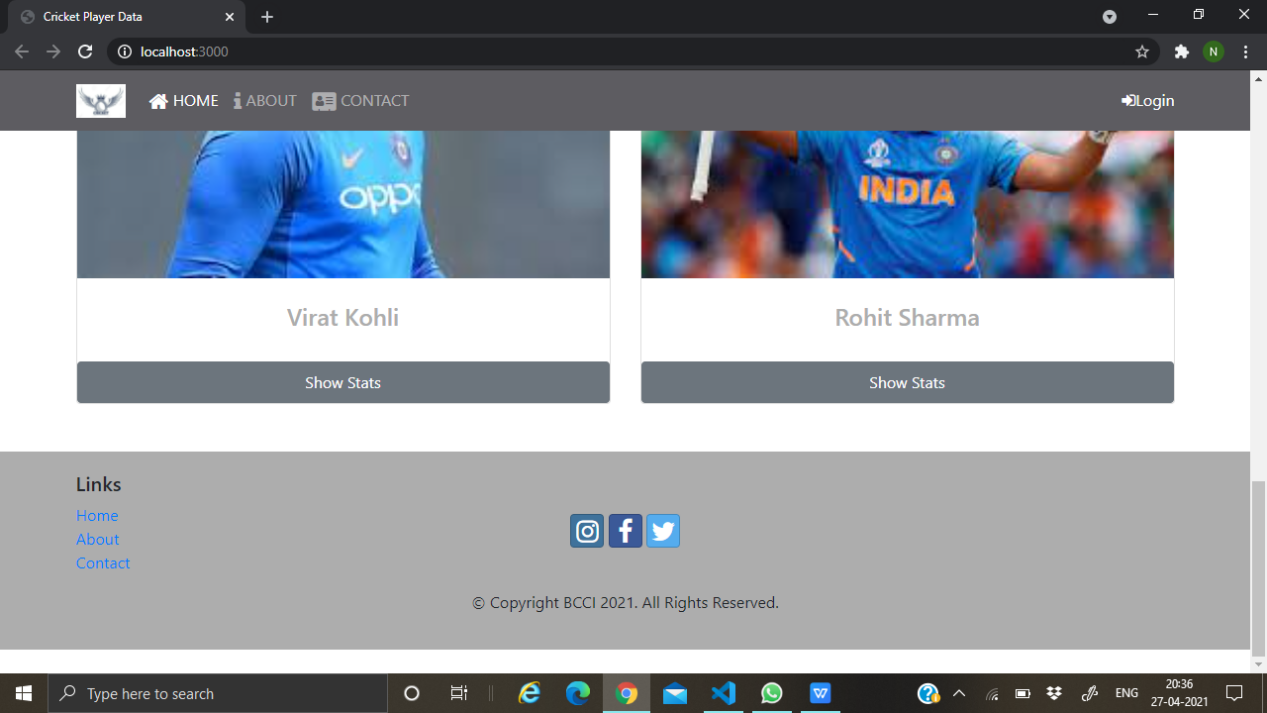


Fig 6.4 Home Page-iv

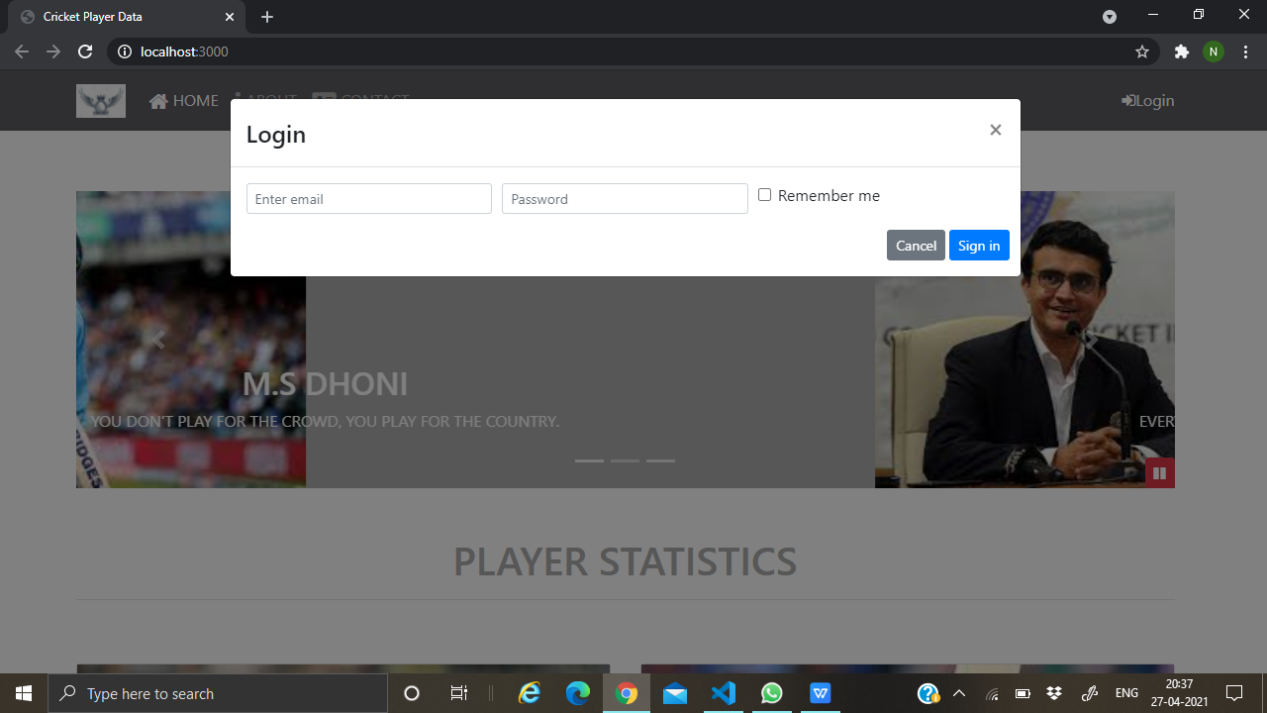


Fig 6.5 Login Page

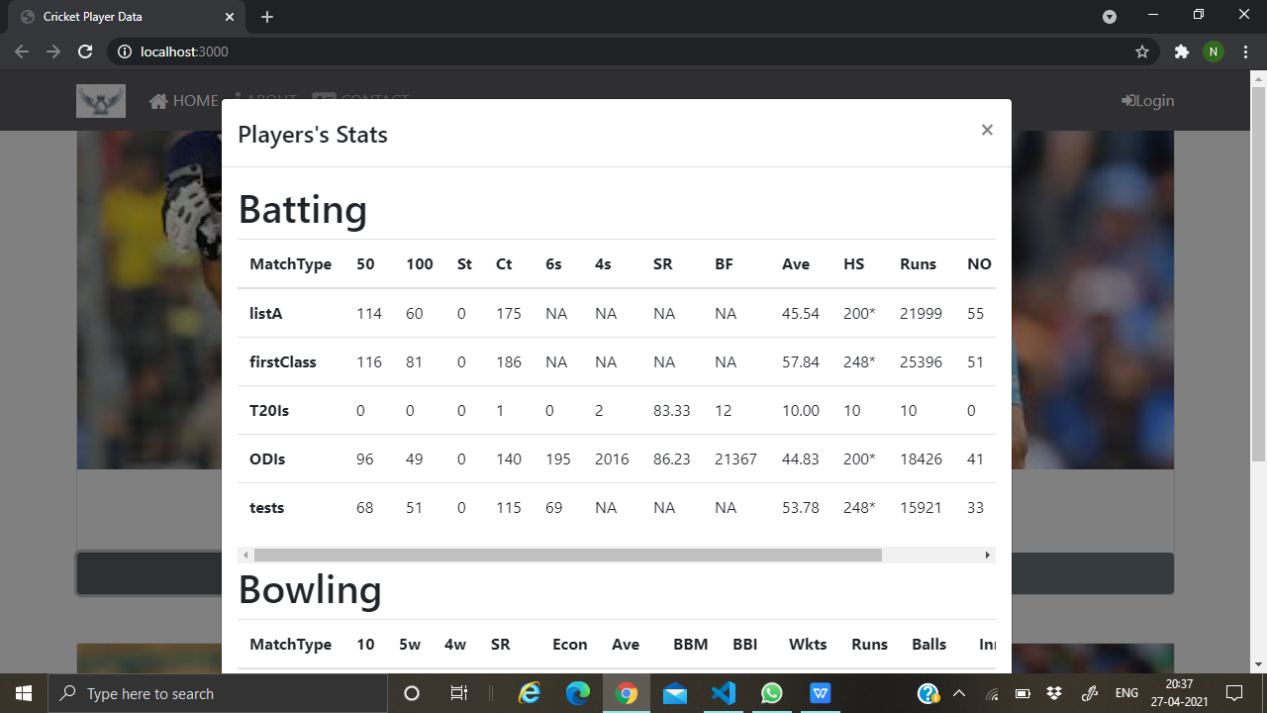


Fig 6.6 Batting Stats

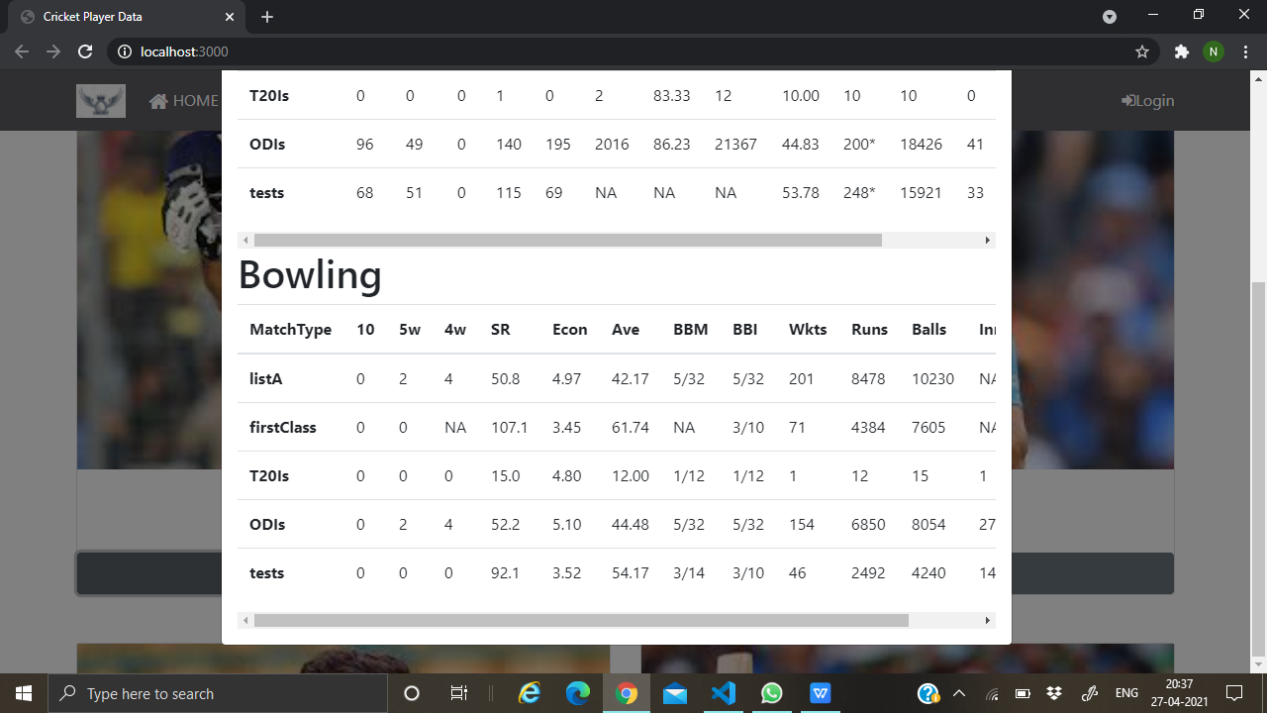


Fig 6.7 Bowling Stats

1. **Testing(White Box Testing)**
2. **Results and Discussions**

Currently the GUI is implemented where the data is fetched and displayed. This whole data will be used in the analysis and the training of the predictor models and for the prediction system altogether which will be further implemented. The model will be based on our own formula to consider different tournaments and different parameters altogether. The Rising Stars of the Indian Cricket will be displayed in all the domains i.e., batting, balling and all-rounders.

1. **Conclusion and Future Scope**

The analysis and the prediction part will be done in further development. Accuracies will be tested and predictions will be made. Similar systems can be implemented by different countries for their upcoming stars using data from IPL, Big Bash League etc.

**Literature Cited:**

1. Prediction of Rising Stars in the Game of Cricket

<https://ieeexplore.ieee.org/document/7878604>

1. Using machine learning techniques for rising star prediction in co-author network

<https://link.springer.com/article/10.1007/s11192-014-1455-8>

1. Measuring batting parameters in cricket: A two-stage regression-OWA method

<https://www.sciencedirect.com/science/article/abs/pii/S026322411400133X?via%3Dihub>

1. Quantifying individual performance in Cricket — A network analysis of batsmen and bowlers

<https://www.sciencedirect.com/science/article/abs/pii/S0378437113008819?via%3Dihub>

1. Rising Star Evaluation Using Statistical Analysis in Cricket

<https://link.springer.com/chapter/10.1007%2F978-981-15-4218-3_31>

1. Finding rising stars in bibliometric networks

<https://link.springer.com/article/10.1007%2Fs11192-020-03466-w>

1. PREDICTING PLAYERS' PERFORMANCE IN ONE DAY INTERNATIONAL CRICKET MATCHES USING MACHINE LEARNING

<https://www.researchgate.net/publication/323611656_Predicting_Players'_Performance_in_One_Day_International_Cricket_Matches_Using_Machine_Learning>

1. Prediction of IPL matches using Machine Learning while tackling ambiguity in results

<https://indjst.org/articles/prediction-of-ipl-matches-using-machine-learning-while-tackling-ambiguity-in-results#:~:text=Findings%3A%20As%20per%20the%20investigation,of%201.4%25%2C%20was%20observed>.

1. Data mining based Classification of Players in Game of Cricket

<https://ieeexplore.ieee.org/document/9215413>

1. Prediction of Cricket Players using Machine Learning

<https://link.springer.com/chapter/10.1007/978-981-15-2407-3_20>

[1] H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dawood and Y. Yang, "Prediction of Rising Stars in the Game of Cricket," in IEEE Access, vol. 5, pp. 4104-4124, 2017, doi: 10.1109/ACCESS.2017.2682162.

[2] Daud, A., Ahmad, M., Malik, M.S.I. *et al.* Using machine learning techniques for rising star prediction in co-author network. *Scientometrics* **102,**1687–1711 (2015). <https://doi.org/10.1007/s11192-014-1455-8>

[3] Gholam R. Amin, Sujeet Kumar Sharma, Measuring batting parameters in cricket: A two-stage regression-OWA method, Measurement, Volume 53, 2014, Pages 56-61, ISSN 0263-2241,

<https://doi.org/10.1016/j.measurement.2014.03.029>.

[4] Satyam Mukherjee, Quantifying individual performance in Cricket — A network analysis of batsmen and bowlers, Physica A: Statistical Mechanics and its Applications, Volume 393, 2014, Pages 624-637, ISSN 0378-4371,

<https://doi.org/10.1016/j.physa.2013.09.027>

[5] Khot A., Shinde A., Magdum A. (2021) Rising Star Evaluation Using Statistical Analysis in Cricket. In: Tripathy A., Sarkar M., Sahoo J., Li KC., Chinara S. (eds) Advances in Distributed Computing and Machine Learning. Lecture Notes in Networks and Systems, vol 127. Springer, Singapore. <https://doi.org/10.1007/978-981-15-4218-3_31>

[6] Daud, A., Song, M., Hayat, M.K. *et al.* Finding rising stars in bibliometric networks. *Scientometrics* **124,**633–661 (2020). <https://doi.org/10.1007/s11192-020-03466-w>

[7] Passi, Kalpdrum & Pandey, Niravkumar. (2018). Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning. 111-126. 10.5121/csit.2018.80310.

[8] Tripathi A, Islam R, Khandor V, Murugan V. (2020). Prediction of IPL matches using Machine Learning while tackling ambiguity in results. Indian Journal of Science and Technology. 13(38): 4013-4035. <https://doi.org/10.17485/IJST/v13i38.1649>

[9] A. Balasundaram, S. Ashokkumar, D. Jayashree and S. Magesh Kumar, "Data mining based Classification of Players in Game of Cricket," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 271-275, doi: 10.1109/ICOSEC49089.2020.9215413.

[10] Aleemulla Khan P., Thirupathi Rao N., Bhattacharyya D. (2020) Prediction of Cricket Players Performance Using Machine Learning. In: Fiaidhi J., Bhattacharyya D., Rao N. (eds) Smart Technologies in Data Science and Communication. Lecture Notes in Networks and Systems, vol 105. Springer, Singapore. https://doi.org/10.1007/978-981-15-2407-3\_20