A Mini Project Synopsis on

Cricket Score Prediction

T.E. - I.T Engineering

Submitted By

Pratik Dhumal	**19104031**
Kushal Todi	**19104047**
Ruchita Raut	**19104033**

Under The Guidance Of

Prof. Geetanjali Kalme

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year: 2021-22

CERTIFICATE

This is to certify that the Mini Project report on **Cricket Score Prediction** has been submitted by Pratik Dhumal (19104031), Ruchita Raut (19104033), and Kushal Todi (19104047) who are Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as partial fulfillment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in a satisfactory manner as per the curriculum laid down by the University of Mumbai.

Ms.Geetanjali Kalme
Guide

Prof. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place: A.P.Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide Prof. Geetanjali Kalme. Expressing gratitude towards our HoD, **Prof. Kiran Deshpande,** and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teachers Geetanjali Mam and Nahid Mam who gave us their valuable suggestions and ideas when we needed them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

CHAPTER 1

1. Introduction

- Cricket score prediction is an area where the first innings score of a cricket match is predicted using some techniques.
- Lots of people like watching cricket and they also like to predict the final score. Our Project focuses on an accurate prediction of cricket scores for matches considering the previous dataset available and also considers the various factors that play an important role in the score prediction.

1.1 Purpose

- Cricket prediction is a project in which we attempt to predict the final score of the team batting first and find the winning probability of the chasing team. Cricket is a sport that creates a huge amount of data in every game. This dataset is used as an input to make accurate predictions. Many strategies in the game are dependent on the final score, and our project will help captains and coaches to make correct decisions based on the predictions. We used powerful machine learning techniques such as random forest regression and logistic regression to make accurate predictions.
- We will be predicting the score of a match by considering various factors like runs scored, overs bowled, wickets were taken, etc. The reason behind selecting these features is that we need to build a model that can understand the dynamicity of the cricket game. Therefore, we are considering the factors which will be focusing on the dynamic nature of the cricket game.
- Lots of people like watching cricket and they also like to predict the final score. Our Project focuses on an accurate prediction of cricket scores for matches considering the previous dataset available and also considers the various factors that play an important role in the score prediction.

1.2 Objectives

1. To improve the efficiency of players by analyzing the data the predict future outcomes, which will help in effective strategic planning.
2. We can get the news related to sports.
3. The main aim is to predict the match outcomes using various parameters, the performance of each player based on the historical data. To achieve reliable accuracy, we need to analyze a large amount of data.

1.3 Scope

1. Can be applied in national and district level cricket leagues.
2. Can be applied in strategic planning of the team.
3. Can be used for formulating teams for specific matches.
4. Can be used to provide hidden insights, such as batting partnerships and who is the best batting partner.

1.4 Literature Review:

We studied papers based on the area of our research i.e., Cricket prediction. We studied 8 IEEE papers and drew some conclusions from the study [1].

A comparison of all 8 IEEE papers has been made. The work proposed [2] deals with the score prediction of the first innings and also predicts the outcome of the match after the second innings. Lasso Regression algorithm is used to predict the first innings score and outcome prediction is done by using the Naive Bayes Classifier.

In, [3] the research aims at predicting the result of an ongoing cricket match on an over-by-over basis based on the information and data that is available from each over. The author tests the datasets on various machine learning models. It has been found that the Random Forest algorithm has the highest accuracy.

The work proposed [4] deals with the sentiments. The author predicts the outcome and man of the match by using twitter-based positive and negative sentiment analysis. Naive Bayes classifier, SVM, Random Forest algorithm, and Logistic Regression,are some of the models used for prediction. In, cricket squad analysis is done. This paper provides a mathematical approach to selecting the players. RMSE value of Multiple Random Forest Regression is greater than LR, SVR, and Decision Tree.

The work proposed in, [5] deals with the player's performance prediction in ODI matches. This paper proposes the model which consists of statistical data of Bangladesh players. The main aim of this research is to predict the performance of players based on the records using SVM with linear kernel and SVM with the polynomial kernel.

Chapter 2

2. Problem Definition

There is the need for a system that provides for estimation or prediction of total runs which can be made using the current match statistics, for example, overs done, wickets gone, or runs made in the last 5 overs. In this we have designed a system that can provide the Score and Winning Prediction in Cricket match, the system can analyze multiple parameters like wickets fallen before and after 5 over, runs scored, etc. When declaring a time for a particular championship it is very important to select the best team so that the chances of the team winning the championship become easy. This problem had to be solved to generate the best players from both teams for the best battle. To solve this problem, we have collected the historical data of some teams and using prediction algorithms like Lasso regression and the Random Forest algorithm we are predicting the Final Score that will be scored by the batting team.

CHAPTER 3

3. Proposed System

We aim to build a model that can predict the score of a cricket match efficiently. We are looking to build a model that can consider various parameters that contribute to the score prediction.

a) **Data Collection:** We will be taking the dataset. The dataset will be taken in the CSV format. The data collected from the website will be cleaned in the next step.

b) **Data Cleaning:** In the data cleaning step, we want to remove unwanted columns like match id, venue, batsman name, bowler name, a score of the striker, and score of the non-striker. These columns will not be required during prediction hence we will be dropping those columns. In the IPL dataset, some teams are not playing in the IPL anymore. Teams like Deccan Chargers, Kochi Tuskers Kerala, Pune Warriors India, Gujarat Lions, Rising Pune Supergiant, etc. are not part of IPL. So, we need to eliminate those teams from the dataset and we only need to consider the consistent teams. We will be considering the data after 5 overs. The date column in the dataset is present in the string format but we want to apply some operations on the date column for that we will need to convert the string to a date-time object.

c) **Data Preprocessing:** After cleaning the data, we will need our data to be preprocessed. In the data preprocessing step, we will be performing one-hot encoding. One hot encoding is explained in detail in the implementation section. We will need to rearrange the columns of our dataset in the data preprocessing step. The purpose of rearranging columns is that we need our columns to be properly arranged in some sequence.

d) **Data Splitting:** After data preprocessing, we will be splitting our data in such a way that IPL matches played before 2017 will be considered for the training of the model and IPL matches played after 2017 will be considered for test data.

e) **Model Generation:** We will be using the Random Forest Regression and Lasso Regression model for the prediction. The model with the highest accuracy will be selected

for the prediction. The model which we will be using for the prediction is explained in the implementation section.

f) **Final Prediction:** Finally, the data will be passed through the model and then the user inputs will be taken. After getting the user inputs and matching them with the historical data we will be predicting a range of the score i.e. from the lower bound to the upper bound.

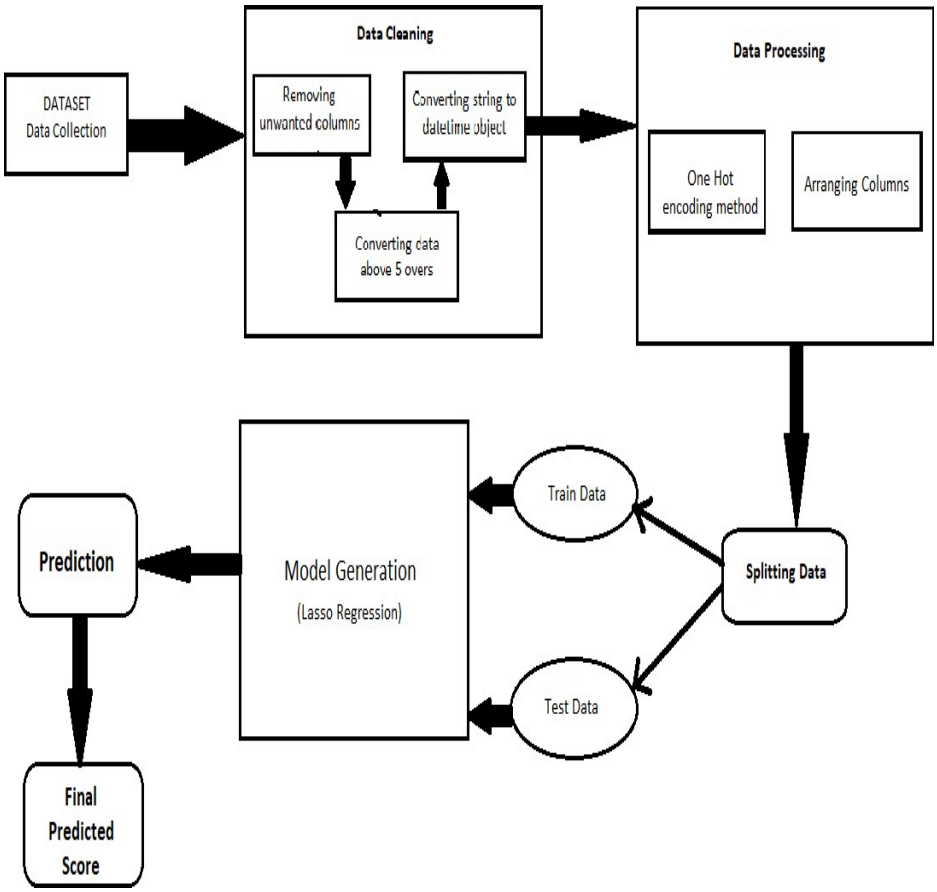

Fig. 3.1 Model Architecture

So in our ML-based application, two algorithms will be compared first is Random Forest Algorithm and the second is Lasso Regression.

1. **Random Forest Regressor:** Random Forest is an ensemble technique that is used to perform regression and classification tasks. Ensemble techniques combine results of various machine learning models and give the best accurate prediction of any individual model.

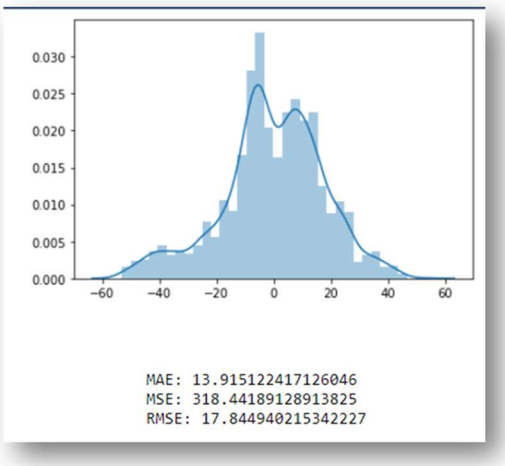

Fig. 3.2 Prediction using Random Forest

2. **Lasso Regression:** Lasso Regression is a regularization technique. The "LASSO" stands for Least Absolute Shrinkage and Selection Operator. It is used for more accurate predictions. Lasso Regression is derived from Linear Regression.

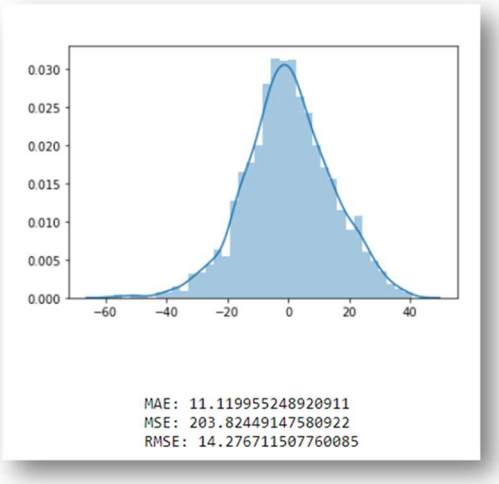

Fig. 3.3 Prediction using Lasso Regression

3.1 Features and Functionality

1. Login/Register:

The User has to enter the credentials to get signed in, to the website, or else register to get the access.

2. Home page:

On this page, the user will get to see the news related to all the matches.

There will be the prediction button, by clicking on it the user will be redirected to the score prediction tab.

The user will get to know about the upcoming match schedules and the live-going matches data.

3. Prediction Module:

This page is the main motto of our project cricket score prediction. In this the user has to input various parameters like the team's name, overs bowled, the input of how many runs have been scored till now, and wickets fallen.

4. Prediction Result Page:

After clicking on predict button the predicted score range (result) will be shown on this page.

CHAPTER 4

4. Project Outcomes

The user first will login into the application, then the user will be redirected to the Homepage where he will get to see the news and upcoming match schedules. When the user will click on the Make Prediction button the user will be redirected to the Prediction module page and the user has to input various parameters to make the prediction. After giving inputs the user will be again redirected to the Final Score prediction page.

The algorithms used are Random Forest and the lasso regression so in this, we have seen that Lasso regression gives the most accurate score. The final output will be given using lasso regression.

Lasso Regression accuracy is:

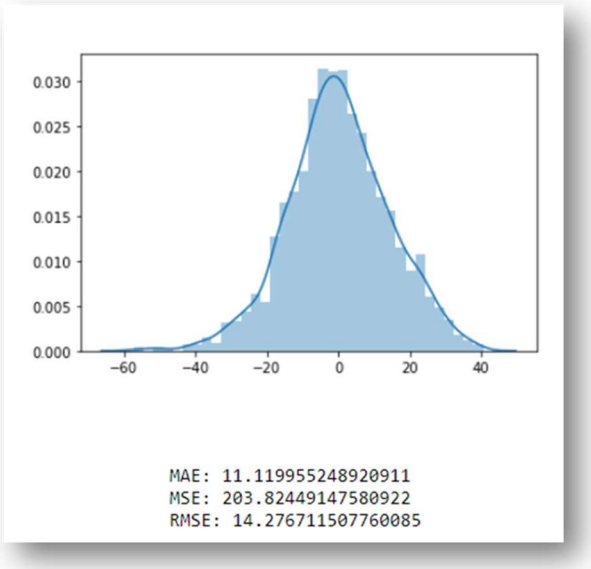

Fig. 4.1 Lasso Regression

CHAPTER 5

Technological Stack

Hardware Requirements

- Operating System – Windows, Android, IOS
- Processor – Pentium Intel or above
- RAM – minimum 256MB
- Storage – 16MB

Software Requirements

- **Frontend :**

- **HTML**: HTML stands for HyperText Markup Language. It is the standard markup language for creating Web pages. It describes the structure of a Web page. It consists of a series of elements. Its elements tell the browser how to display the content
- **CSS**: CSS stands for Cascading Style Sheets. It describes how HTML elements are to be displayed on the screen, on paper, or in other media. It can control the layout of multiple web pages all at once.
- **REACT**: React.js is an open-source JavaScript library that is used for building user interfaces specifically for single-page applications. It's used for handling the view layer for web and mobile apps. React also allows us to create reusable UI components.

- **Backend :**

- **Python:** Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems

- **Machine Learning Algorithm:** Machine Algorithm used is Lasso Regression.

- **Flask:** Flask is used for developing web applications using python, implemented on Werkzeug and Jinja2. Advantages of using the Flask framework are: There is a built-in development server and a fast debugger provided.

CHAPTER 6

Project Design

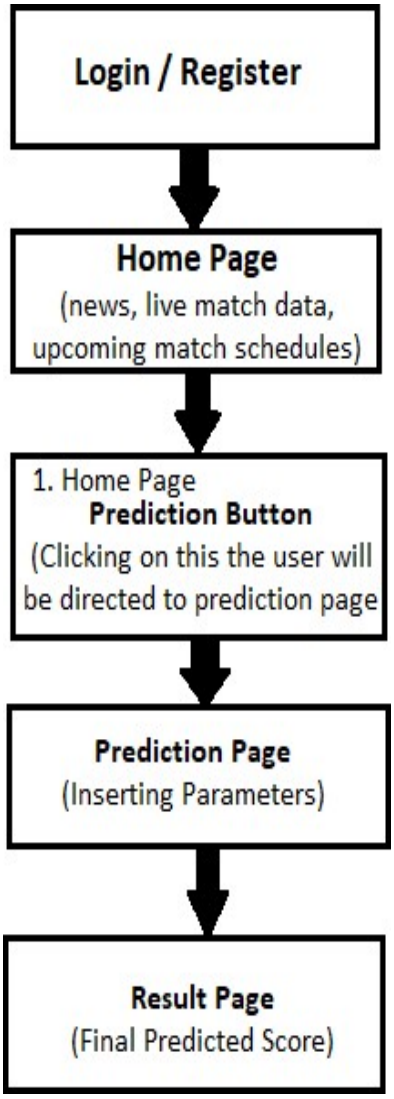

Fig. 6.1 UI Flow

Log in/Register page:

This is the Login/Register page user the user has to enter the required credentials to get redirected to the homepage. The user which is logged In, their email will be displayed as "Currently logged in as: test@gmail.com".

Homepage:

This is the homepage of our app which includes the Latest News, Authentication, and score prediction button. We have used React bootstrap for styling our project. There will be the prediction button, by clicking on it the user will be redirected to the score prediction tab.

Prediction Page:

This is the Prediction page where the user will input the parameters required and then he/shehas to click on the predict score tab to get the predicted score.

Final Result Page:

This is the final prediction page where the user will get the predicted score range i.e lower bound to upper bound.

We will be using the Flask framework for the development of this project. In the Flask framework, the routing of the pages is done based on the URLs. On the home page, we will be having our main prediction tab. In that tab, we will be taking the user inputs. The user inputs include, Name of the batting team, name of the bowling team, number of runs scored, number of overs bowled, number of wickets taken, number of runs scored in the previous 5 overs, and number of wickets taken in the previous 5 overs. Once the user hits predict score button then the form is submitted. After submission of the form, our model comes into the picture in the backend. The inputs are compared with the historical data and then a score is predicted. The user is redirected to the result page where the user can see the actual predicted score. On the prediction page, the user will be getting the output in the form of a range i.e. from the lower bound to the upper bound.

CHAPTER 7

Project Scheduling:

Group Members	Time	Work to be done
Everyone	2nd week of February	Deciding Project Topic and Research on ML
Ruchita Raut	4th week of February	Research on ProjectTopic read a researchpaper, and choose an algorithm.
Pratik Dhumal	2nd week of March	Designing the UI and Creating the ML model
Kushal Todi	4th week of March	Creating the Frontend and Implementing ML model
Everyone	2nd week of April	Connecting the Frontend and the ML Model using Flask
Everyone	4th week of April	Testing and fixing errors and making required changes. Making Report and Representation.

CHAPTER 8

Screenshots of Application:

Log in/Register page:

This is the Login/Register page user the user has to enter the required credentials to get redirected to the homepage. The user which is logged In, their email will be displayed as “Currently logged in as: test@gmail.com”.

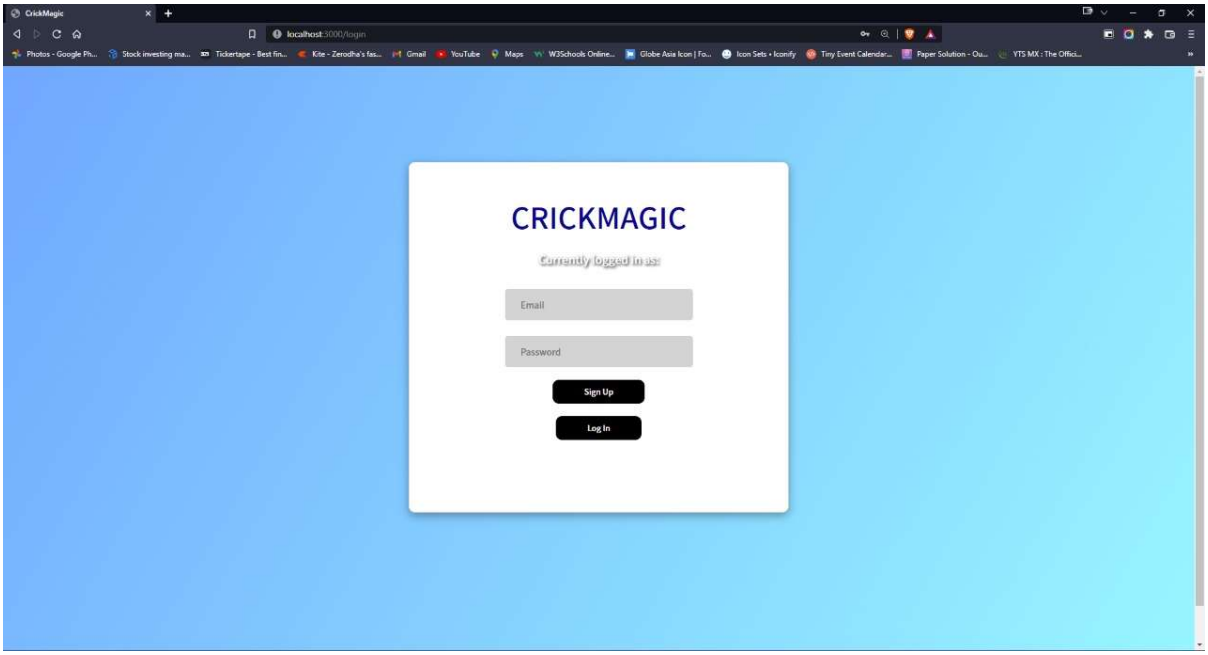

Fig. 8.1 Login/ Register Page

Homepage:

This is the homepage of our app which includes the Latest News, Authentication, and score prediction button. We have used React bootstrap for styling our project. There will be the prediction button, by clicking on it the user will be redirected to the score prediction tab.

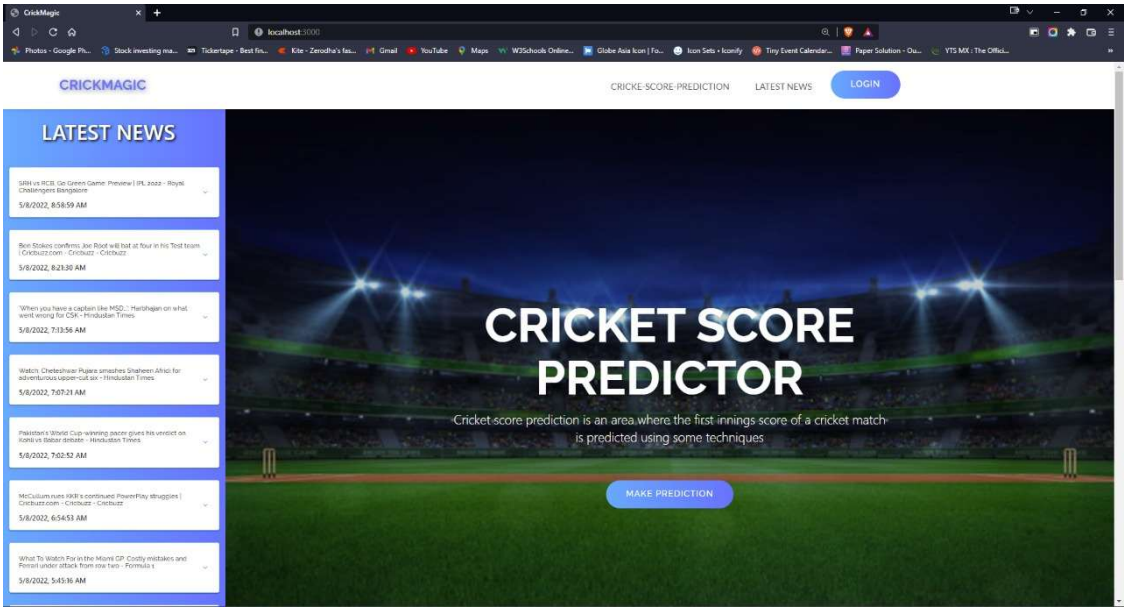

Fig. 8.2 Home Page

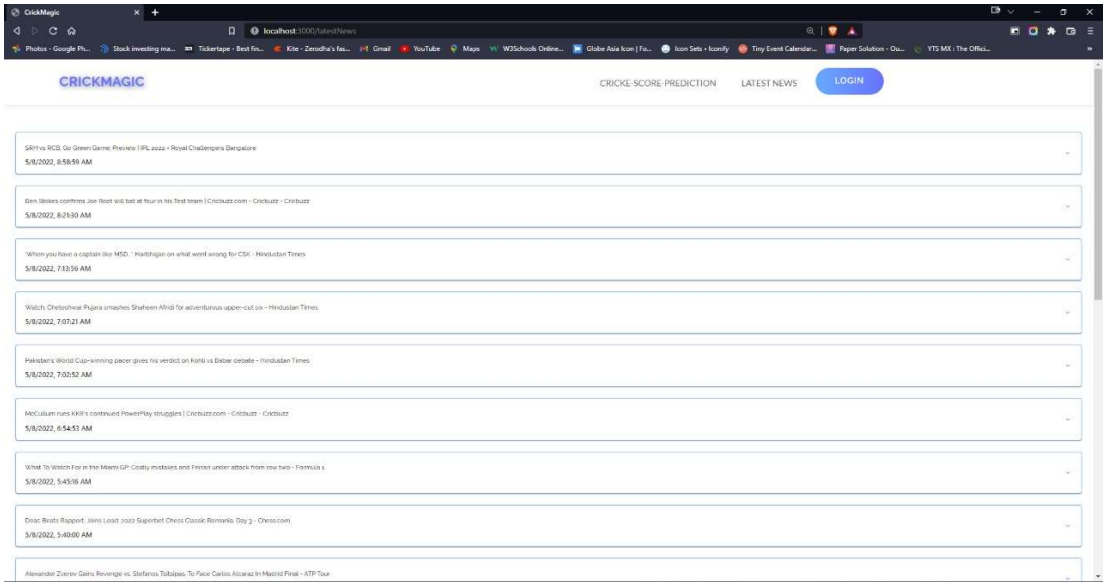

Fig. 8.3 Latest News Section

When the user clicks on any of the news the user will be redirected to the news page where he

or she can read the whole news.

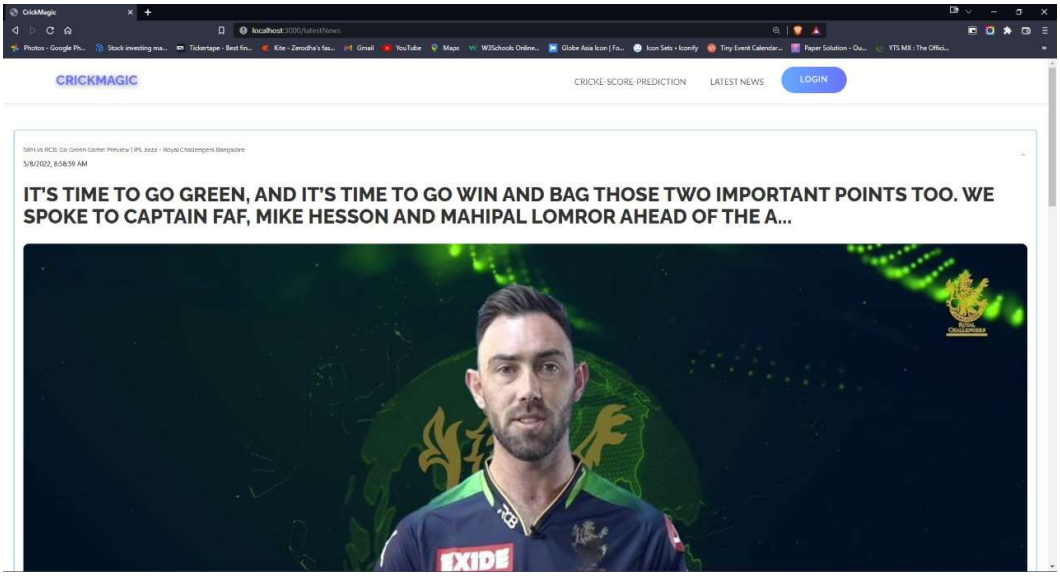

Fig 8.4 News whole page

Prediction Page:

This is the Prediction page where the user will input the parameters required and then he/shehas to click on the predict score tab to get the predicted score.

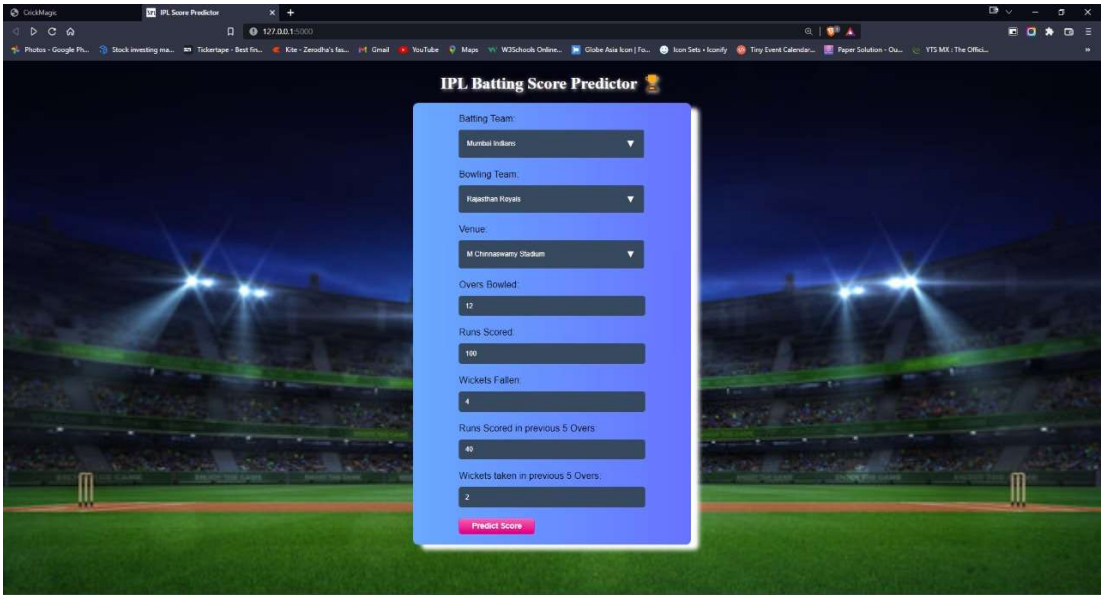

fig. 8.5 Prediction Page (with Inputs)

Final Result Page:

This is the final prediction page where the user will get the predicted score range i.e lower bound to upper bound.

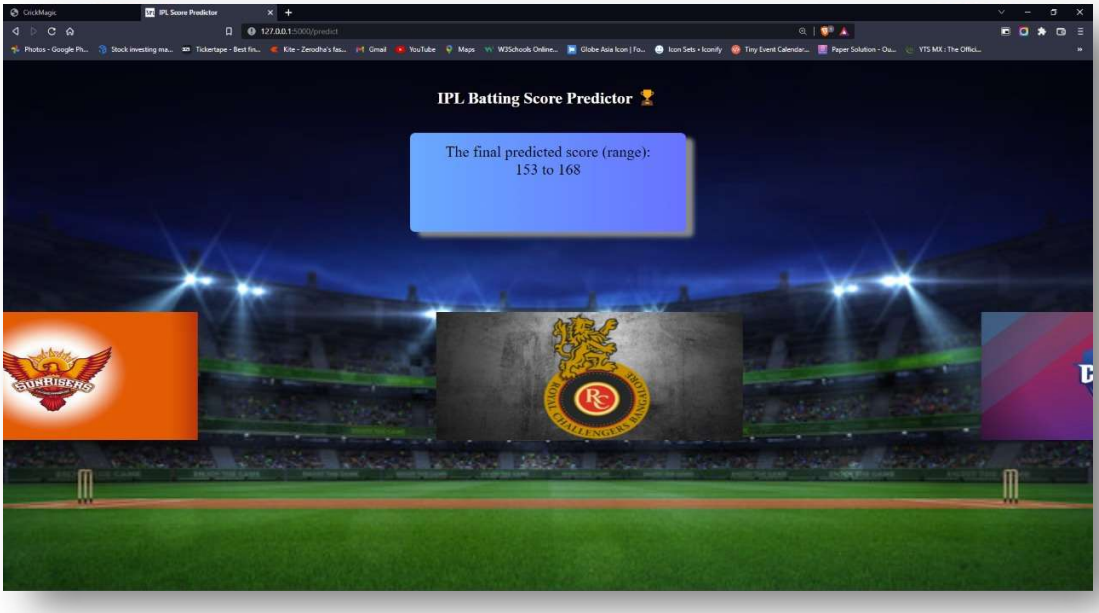

Fig. 8.6 Final Prediction Page

Chapter 9

Conclusion:

From the results, we can conclude that the Lasso Regression algorithm has the highest accuracy of the prediction. So, we are using the Lasso Regression model for prediction purposes.

In the future, we can implement a model for predicting the chasing probability. We can work on improving the accuracy of the model used in this project. Factors like venue, pitch, and the opponent team can be considered for the prediction.

References:

1) [1] Prasad Thorat, Vighnesh Buddhivant, Yash Sahane; Review Paper on Cricket Score Prediction; April 2021

2) [2] Tejinder Singh, Vishal Singla, Parteek Bhatia; - Score and Winning Prediction in Cricket through Data Mining; Oct 8-10, 2015

3) [3] D. Thenmozhi, P. Mirunalini, S. M. Jaisakthi, Srivatsan Vasudevan, Veeramani Kannan V, SagubarSadiq S; Moneyball - Data Mining on Cricket Dataset; 2019

4) [4] A.N.Wickramasinghe, Roshan D.Yapa; Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data; 2018

5) https://www.geeksforgeeks.org/implementation-of-lasso-regression-from-scratch-using-python/

6) https://www.geeksforgeeks.org/implementation-of-lasso-regression-from-scratch-using-python/

7) https://towardsdatascience.com/which-evaluation-metric-should-you-use-in-machine-learning-regression-problems-20cdaef258e