

(ISO 9001:2015 Certified), Accredited with ‘A’ Grade by NAAC

: 08258 - 281039 – 281263, Fax: 08258 – 281265

### Department of Computer Science and Engineering

B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

Report on Mini Project

IPL SCORE PEDICTION

#### Course Code : 18CS601 Course Name : Machine Learning

##### Semester: 6 SEM Section: A

**Submitted To:**

## Ms.Divya Jennifer D’Souza Assistant Professor

## Department of Computer Science and Engineering

**Submitted By**

Blen Joswin Saldanha-4NM18CS039

Dhanush Raj-4NM18CS048

**Date of submission:**

#### Signature of Course Instructor

CERTIFICATE

“IPL Cricket Score Prediction ” is a simple work carried out by Blen Joswin Saldana (4NM18CS039) and Dhanush Raj (4NM18CS048) in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering prescribed by N.M.A.M Institute Of Technology,Nitte during the year 2020-2021.

It is certified that all corrections/suggestions indicated for Internal

Assessment have been incorporated in the report. The Mini project report has

been approved as it satisfies the academic requirements in respect of the

project work prescribed for the Bachelor of Engineering Degree.

ACKNOWLEDGEMENT

We believe that our project will be complete only after we thank the people

who have contributed to make this project successful.

First and foremost, our sincere thanks to our beloved principal, Dr. Niranjan

N. Chiplunkar for giving us an opportunity to carry out our project work at our

college and providing us with all the needed facilities.

We sincerely thank Dr. K.R. Udaya Kumar Reddy, Head of Department of

Computer Science and Engineering, Nitte Mahalinga Adyantaya Memorial

Institute of Technology, Nitte.

We express our deep sense of gratitude and indebtedness to our guide Mrs.

Divya Jennifer D’Souza, Assistant Professor GD I, Department of Computer Science

and Engineering, for her inspiring guidance, constant encouragement, support

and suggestions for improvement during the course of our project.

We thank all the teaching and non-teaching staff members of the Computer

Science and Engineering Department and our parents and friends for their honest

opinions and suggestions throughout the course of our project.

Finally, we thank all those who have supported us directly or indirectly

throughout the project and making it a grand success.

Blen Joswin Saldana

(4NM18CS039)

Dhanush Raj

(4NM18CS048)

# ABSTRACT

The **Indian Premier League** (**IPL**) is a professional [Twenty20 cricket](https://en.wikipedia.org/wiki/Twenty20_cricket" \o "Twenty20 cricket) league, contested by eight teams based out of eight different Indian cities. The league was founded by the [Board of Control for Cricket in India](https://en.wikipedia.org/wiki/Board_of_Control_for_Cricket_in_India" \o "Board of Control for Cricket in India) (BCCI) in 2007. It is usually held between March and May of every year and has an exclusive window in the [ICC Future Tours Programme](https://en.wikipedia.org/wiki/ICC_Future_Tours_Programme" \o "ICC Future Tours Programme).

The IPL is the most-attended cricket league in the world and in 2014 was ranked sixth by average attendance among all sports leagues. In 2010, the IPL became the first sporting event in the world to be broadcast live on [YouTube](https://en.wikipedia.org/wiki/YouTube" \o "YouTube). The brand value of the IPL in 2019 was ₹475 billion (US$6.7 billion), according to [Duff & Phelps](https://en.wikipedia.org/wiki/Duff_&_Phelps" \o "Duff & Phelps). According to BCCI, the 2015 IPL season contributed ₹11.5 billion (US$160 million) to the [GDP](https://en.wikipedia.org/wiki/GDP" \o "GDP) of the [Indian economy](https://en.wikipedia.org/wiki/Economy_of_India" \o "Economy of India).

There have been [thirteen seasons](https://en.wikipedia.org/wiki/List_of_Indian_Premier_League_seasons_and_results" \o ") of the IPL tournament. The current IPL title holders are the [Mumbai Indians](https://en.wikipedia.org/wiki/Mumbai_Indians" \o "Mumbai Indians), who won the [2020 season](https://en.wikipedia.org/wiki/2020_Indian_Premier_League" \o "2020 Indian Premier League).[[10]](https://en.wikipedia.org/wiki/Indian_Premier_League" \l "cite_note-10) The venue for the [2020 season](https://en.wikipedia.org/wiki/2020_Indian_Premier_League" \o "2020 Indian Premier League) was moved due to the [COVID-19 pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India" \o "COVID-19 pandemic in India) and games were played in the [United Arab Emirates](https://en.wikipedia.org/wiki/United_Arab_Emirates" \o "United Arab Emirates).

Table Of Contents

**S.No | Title | Page No.**

**----------------------------------------------------------**

1 Introduction ---------------------- 6

2 Software requirements ------------ 7

3 Hardware requirements ----------- 8

4 Design and Implementation ------------ 9-12

5 Result ----------------------------- 12-15

6 Conclusion ------------------------ 16

7 References ------------------------- 16

# INTRODUCTION

# 

Since the dawn of the IPL in 2008, it has attracted viewers all around the globe. High level of uncertainty and last moment nail biters has urged fans to watch the matches. Within a short period, IPL has become the highest revenue generating league of cricket. Data Analytics has been a part of sports entertainment for a long time. In a cricket match, we might have seen the scoreline showing the probability of the team winning based on the current match situation.

Being a cricket fan, visualizing the statistics of cricket is mesmerizing. While I was jumping blogs on Medium and kernels (Let’s say code playbook) on Kaggle, I was fascinated by the analysis done. Hence, I decided to get my first hands-on experience by building a classifier to predict the winning team.

In Machine Learning, the problems are categorized into 2 groups mainly: Regression Problem and Classification problem. The Regression problem deals with the kind of problems having continuous values as output while in the Classification problem the outputs are categorical values. Since the output of winner prediction is a categorical value, the problem which we are trying to solve is a Classification problem.

Software Requirements

**➢** Operating System : Windows

➢ Google Colaboratory:

# What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser,with

* Zero configuration required
* Free access to GPUs
* Easy sharing

Machine Learning Using Colab

With Colab you can import an image dataset, train an image classifier on it, and evaluate the model, all in just [a few lines of code](https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/quickstart/beginner.ipynb" \t "https://colab.research.google.com/notebooks/_blank). Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including [GPUs and TPUs](https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index" \l "using-accelerated-hardware), regardless of the power of your machine. All you need is a browser.

Colab is used extensively in the machine learning community with applications including:

* Getting started with TensorFlow
* Developing and training neural networks
* Experimenting with TPUs
* Disseminating AI research
* Creating tutorials

Hardware Requirements

➢ **RAM :** Minimum 4GB (8GB recommended)

➢ **Processor :** Intel i3 or above

➢ **HDD :** 20 GB Disk space

➢ **INTERNET CONNECTION**: Internet Connection is required for google colab.

# DESIGN AND IMPLEMENTATION

## Step 1 — Understanding The Dataset

Our Dataset contains data of every ball played by batsman from the year 2006 to 2017 in every match.

## Step 2 — Load The Required Libraries And Dataset

After understanding the datasret,then we can load the dataset and required libraries.

Code:

# Importing essential libraries

import pandas as pd

# Loading the dataset

df = pd.read\_csv('ipl.csv')

## Step 3 — Remove Unwanted Data Columns From The Dataset

The dataset may contain unwanted data columns.So it’s we remove unwanted data columns from dataset.

Code:

# --- Data Cleaning ---

# Removing unwanted columns

columns\_to\_remove = ['mid', 'venue', 'batsman', 'bowler', 'striker', 'non-striker']

df.drop(labels=columns\_to\_remove, axis=1, inplace=True)

## Step 4 — Removing Unconsistent Teams In IPL

In IPL sometimes there are 10 teams which play the league,sometimes there will be 8 teams.

So it is important to consider only consistent teams.

Code:

consistent\_teams = ['Kolkata Knight Riders', 'Chennai Super Kings', 'Rajasthan Royals' 'Mumbai Indians', 'Kings XI Punjab', 'Royal Challengers Bangalore',

'Delhi Daredevils', 'Sunrisers Hyderabad']

df=df[(df['bat\_team'].isin(consistent\_teams))&(df['bowl\_team'].isin(consistent\_teams))]

## Step 5 — Removing The First 5 overs score and update the new one and do required operations

We remove the first five overs score for the prediction purpose and we update the new Score.

And then we convert the column 'date' from string into datetime object.

Then we convert the column 'date' from string into datetime object

Code:

# Removing the first 5 overs data in every match

df = df[df['overs']>=5.0]

# Converting the column 'date' from string into datetime object

from datetime import datetime

df['date'] = df['date'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d'))

# --- Data Preprocessing ---

# Converting categorical features using OneHotEncoding method

encoded\_df = pd.get\_dummies(data=df, columns=['bat\_team', 'bowl\_team'])

## Step 6 — Rearranging the dataset columns

## We rearrange the dataset columns for our requirements.

Code:

# Rearranging the columns

encoded\_df = encoded\_df[['date', 'bat\_team\_Chennai Super Kings', 'bat\_team\_Delhi Daredevils', 'bat\_team\_Kings XI Punjab',

'bat\_team\_Kolkata Knight Riders', 'bat\_team\_Mumbai Indians', 'bat\_team\_Rajasthan Royals',

'bat\_team\_Royal Challengers Bangalore', 'bat\_team\_Sunrisers Hyderabad',

'bowl\_team\_Chennai Super Kings', 'bowl\_team\_Delhi Daredevils', 'bowl\_team\_Kings XI Punjab',

'bowl\_team\_Kolkata Knight Riders', 'bowl\_team\_Mumbai Indians', 'bowl\_team\_Rajasthan Royals',

'bowl\_team\_Royal Challengers Bangalore', 'bowl\_team\_Sunrisers Hyderabad',

'overs', 'runs', 'wickets', 'runs\_last\_5', 'wickets\_last\_5', 'total']]

## Step 7 — Splitting The Data into Train And Test

We must split the dataset into train and test part.Later remove the remove the 'date' column.

Code:

# Splitting the data into train and test set

X\_test = encoded\_df.drop(labels='total', axis=1)[encoded\_df['date'].dt.year >= 2017]

y\_train = encoded\_df[encoded\_df['date'].dt.year <= 2016]['total'].values

y\_test = encoded\_df[encoded\_df['date'].dt.year >= 2017]['total'].values

# Removing the 'date' column

X\_train.drop(labels='date', axis=True, inplace=True)

X\_test.drop(labels='date', axis=True, inplace=True)

## Step 8— Select The Best Reggression Model

For the prediction purpose we use algorithms of regression.There are different types of regressions.But we use regression model which has highest accuracy level.

Code:

# --- Model Building ---

# Linear Regression Model

from sklearn.linear\_model import LinearRegression

regressor = LinearRegression()

regressor.fit(X\_train,y\_train)

r2score=regressor.score(X\_test,y\_test)

print("Accuracy With Linear Regression= ",r2score)

# Lasso Regression Model

from sklearn.linear\_model import Lasso

lasso=Lasso()

lasso.fit(X\_train,y\_train)

l2score=lasso.score(X\_train,y\_train)

print("Accuraccy With Lasso Regression= ",l2score)

# Ridge Regression Model

from sklearn.linear\_model import Ridge

ridge=Ridge()

ridge.fit(X\_train,y\_train)

r2score=ridge.score(X\_train,y\_train)

print("Accuraccy With Ridge Regression= ",r2score)

# Ridge Regression Model

from sklearn.linear\_model import BayesianRidge

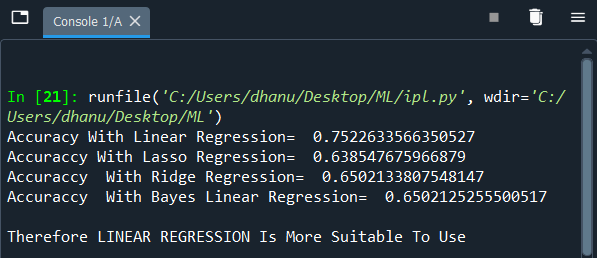
bayes=BayesianRidge()

bayes.fit(X\_train,y\_train)

b2score=bayes.score(X\_train,y\_train)

print("Accuraccy With Bayes Linear Regression= ",b2score)

print("\nTherefore LINEAR REGRESSION Is More Suitable To Use")

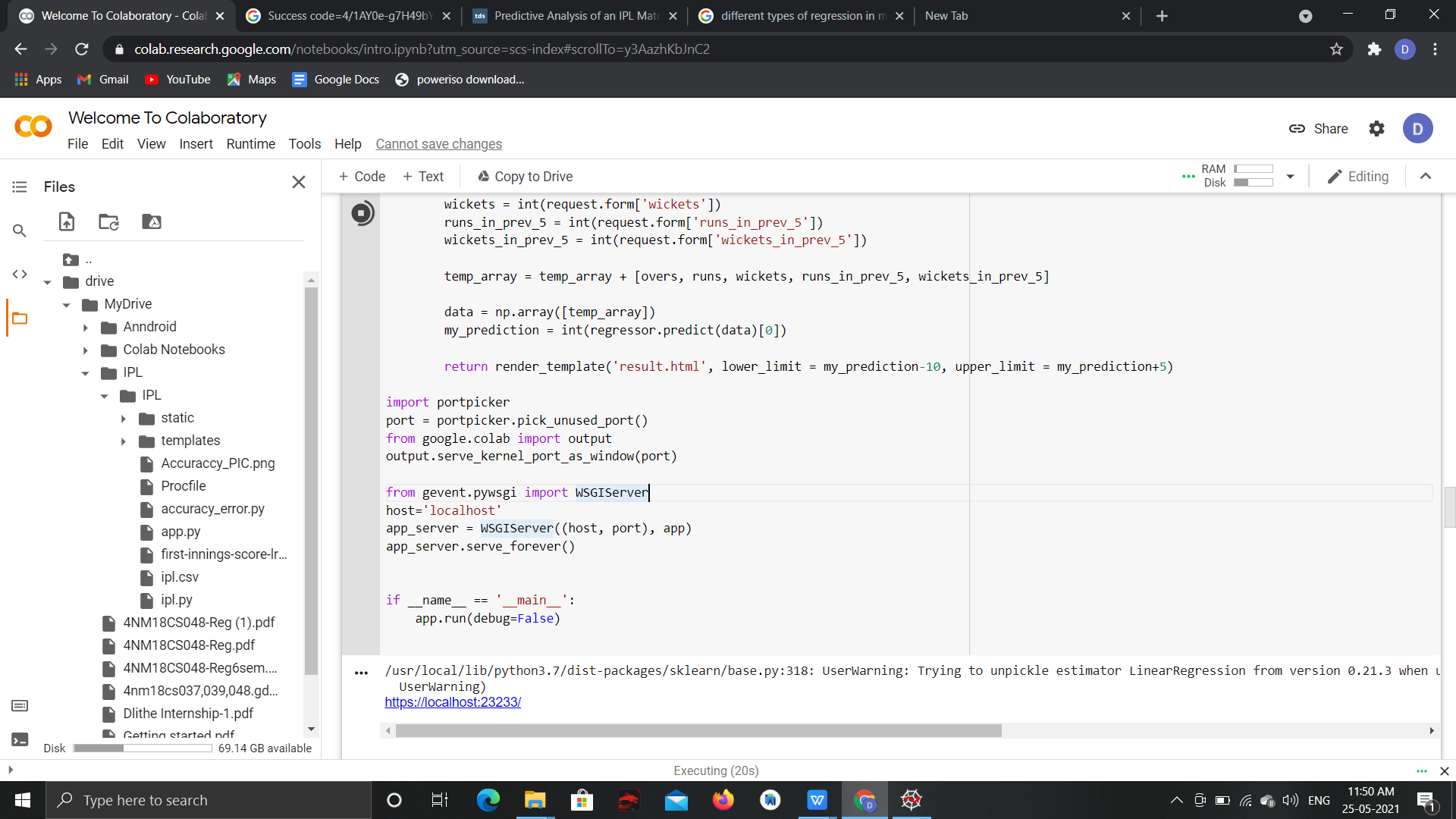


We select Linear Regression Model for regression.

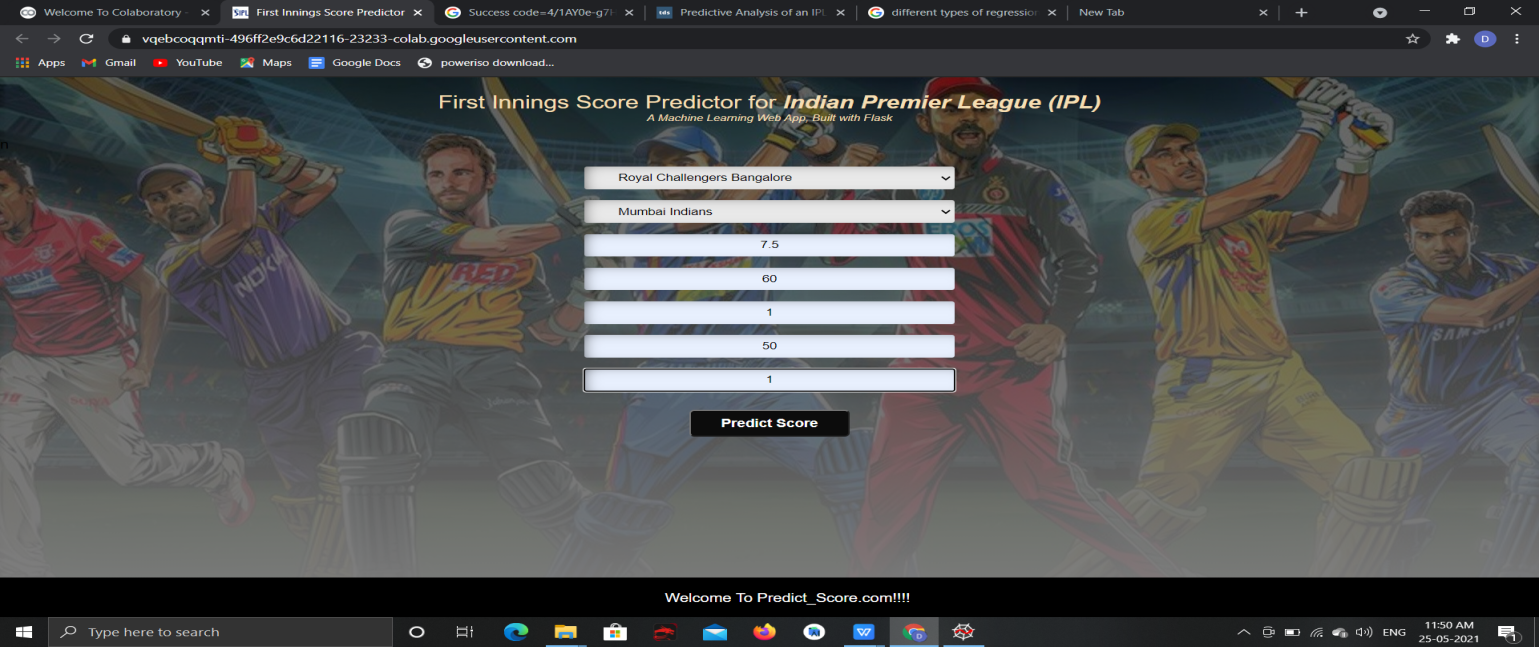
# RESULT

1. We use python library known as “flask”.Flask is web app framework in python.

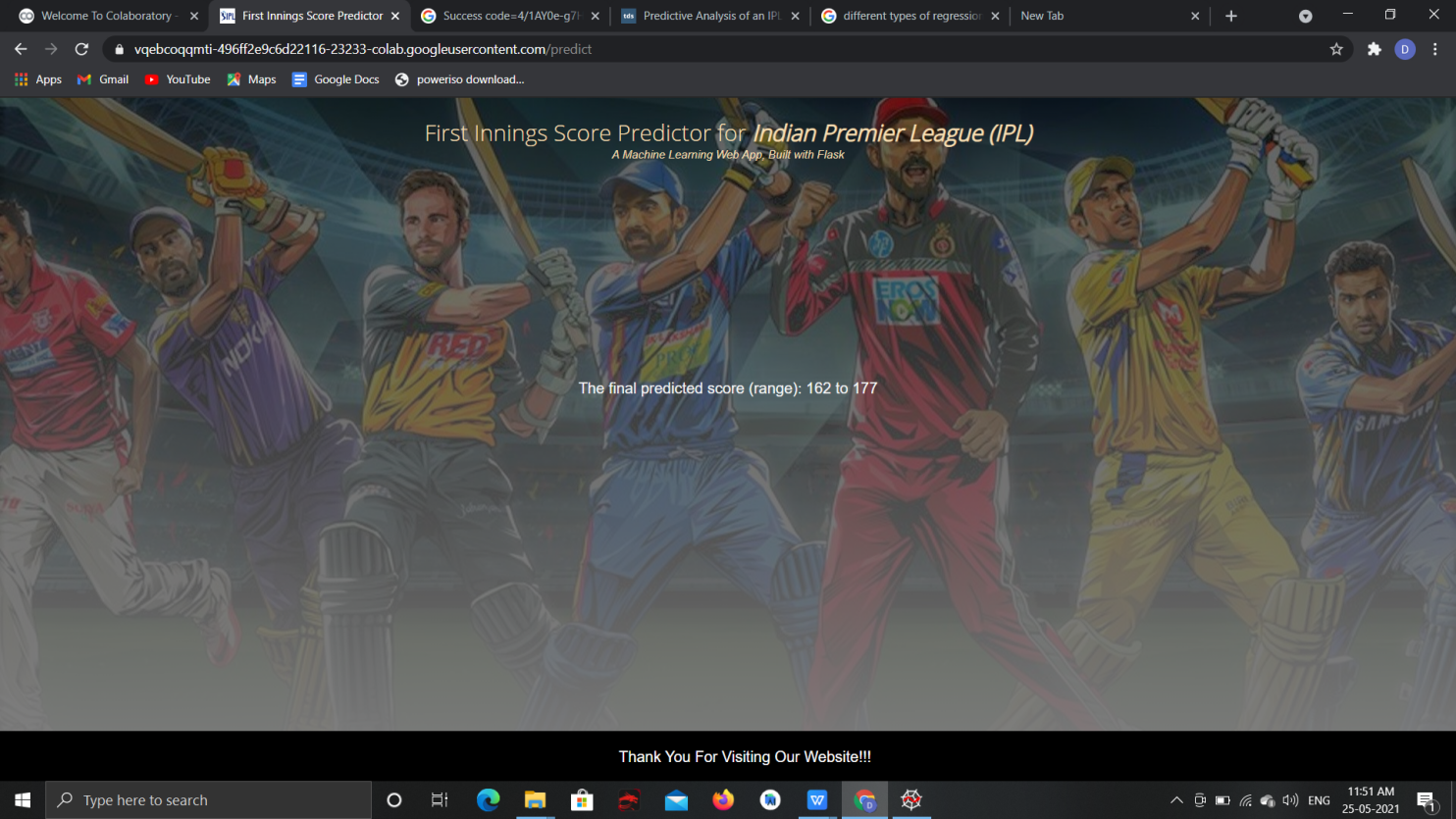
When we run the testing model(app.py) we localhost.By clicking it we will get the web page.



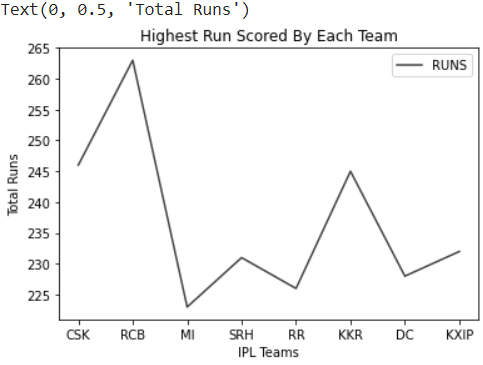
1. After clicking the localhost we get the webpage ,where we enter the appropriate inputs.

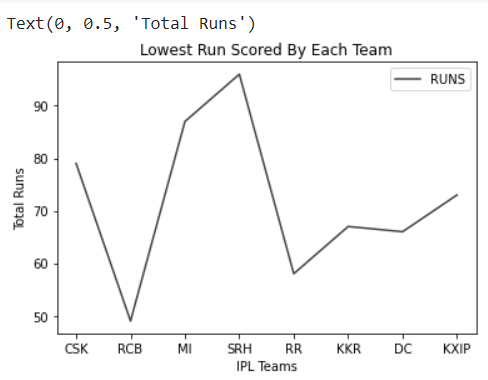


1. After entering inputs we click submit button.We get output will in the form of range.



1. Graphs from CSV File





Conclusion

To conclude, IPL/T20 is a tournament which can be unpredictable to decide a winner. Even, Machine Learning model prediction can go wrong sometimes. A match result can be achieved in many ways like using Players points, team combinations etc. Here, for this study, I have chosen how powerplay can be used to predict the match-winner. Some teams score quick runs during powerplay so that they can get a good score on the board. On the other hand, some teams play slowly in powerplay but in the middle overs, they start hitting. So, it all depends on the day of the match, the opposition they are playing with, pitch type, etc. In the cricketing field, to achieve the full convergence into the data science world, it would require a lot of additional data to meet the full picture of analysis.

References

1. Youtube

<https://www.youtube.com/watch?v=4CtyDxfhoN8>

1. Google

<https://towardsdatascience.com/predicting-ipl-match-winner-fc9e89f583ce>