CAPSTONE PROJECT WORK REPORT

PHASE – II

PREDICTION OF IPL USING ML

SUMITRA.T

A report submitted in part fulfilled of the degree of
BSc. Computer Science with Data Analytics

Supervisor: Mrs. Jayapriya, M.C.A, M.E (Ph.d)
Associate Professor
Department of Computer Science with Data Analytics

Department of Computer Science with Data Analytics

KPR College of Arts Science and Research

(Affiliated to Bharathiar University, Coimbatore) Avinashi Road, Arasur, Coimbatore – 641 407

November 2022

CAPSTONE PROJECT WORK REPORT

PHASE - II

PREDICTION OF IPL USING ML

Bonafede Work Done by

SUMITRA.T

REG.NO: 2028B0057

Dissertation submitted in partial fulfilment of the requirements for the award of Bharathiar University, Coimbatore-46.

Signature of the Guide **Signature of the HOD**

[Mrs. Jayapriya.P]

Submitted for the Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

TABLE OF CONTENTS **PAGE NO**

ACKNOWLEDGEMENT

In the accomplishment of completion of my Capstone Project Work Phase – II on **Analyze IPL using python** I would like to convey my special gratitude to **Dr. S. Balusamy, Principal of KPR College of Arts Science and Research** and **Mrs. P. Jayapriya, Associate Professor and Head, of Department of Computer Science with Data Analytics**. Your valuable guidance and suggestions helped me in phase - I of the completion of this project. I will always be thankful to you in this regard. I am ensuring that this project was finished by me and not copied.

Student Signature

Place:

Date:

ORGANIZATION PROFILE

KPR COLLEGE OF ARTS SCIENCE AND RESEARCH

(Affiliated to Bharathiar University, Coimbatore)

Avinashi Road, Arasur, Coimbatore – 641 407)

ABOUT THE COLLEGE

KPR College of Arts Science and Research is the latest addition to the KPR fleet. The College is located in a picturesque campus of about 11. Acres. The College is run by KPR charities under the leadership of our Chairman Dr. K.P. Ramasamy. The KPR Group is one of the largest industrial conglomerate in the country with interest in Textiles, Sugar, Wind Turbines, Automobiles and Education. The College was established in the year 2019 with a vision of providing top class education and life skills to students and thereby serve the nation and beyond. KPRCAS today offers 12 UG programmes in Management, Commerce and Computer Science streams. The Students of KPRCAS undergo intense training not only in the syllabus and curriculum of the affiliating University but are also trained in various areas. So that they emerge as industry ready graduates to meet the varying demands of the competing industries. Character building and Leadership qualities are inculcated into the students to make them responsible citizens focusing on the development of society and nation. A plethora of Clubs and Events encouraged the students to take part in sports and other cultural activities. KPRCAS offers three years undergraduate courses, which are exclusively for Business, Commerce and Computer Science Stream. The students are equipped with skills and knowledge needed to take up various leadership positions and to develop the society. Beyond Book Teaching help them to be professionals. KPRCAS emphasis on making the students academically brilliant, and also prepare them for the real corporate world. The learning curve begins here for the students of KPRCAS.

ABOUT THE DEPARTMENT

Bachelor of Computer Science with Data Analytics (B.Sc. (CS with DA)) was established in the year 2020. Data Analytics helps to raise the quality of data in the entire business system. The goal of data analytics is to construct the means for extracting business-focused insights from data This requires an understanding of how value and information flows in a business, and the ability to use that understanding to identify business opportunities. The primary aim of a data analyst is to increase efficiency and improve performance by discovering patterns in data. Data analysts exist at the intersection of information technology, statistics and business. They combine these fields in order to help businesses and organizations succeed. The students get exposed to Big Data, Business Intelligence, Data Mining, Data Visualization, Advanced Excel, Predictive Analytics and R Programming.

SYNOPSIS

A professional Twenty20 (T20) cricket competition for India was founded in 2008 and is known as the Indian Premier League (IPL). There are teams in significant Indian cities in the league, which uses a round-robin group and knockout system. Founded with the intention of developing cricket in India, the Twenty-20 cricket tournament league known as the Indian Premier League (IPL)**.** Twenty20 (T20) cricket round robin Indian, developing young and gifted players. Teams from various Indian cities compete against one another in the competition each year.The first recorded game of cricket was played on Thursday, March 10, 1300, between Westminster and Newenden by King Edward 2. According to some theories, "CREAG" was CRICKET's old English name. Children who lived in the weald, a region of dense woodlands and clearings in south-east England that spans Kent and Sussex, possibly invented cricket during the SAXON or NORMAN eras. The first definitive is dated January 1597, month 17, first. The first time adults played cricket was in 1611. In 1611, two men in Sussex were charged with violating the law by playing cricket on Sunday instead of attending church. This was the first time that cricket was mentioned as an adult activity. A BOYS GAME of dictionary-defined cricket is held in the same year. Cricket attracted the first gamblers in the 1660s. In the 17th century, cricket was brought to North America by way of the English colonies. In the early half of the 20th century, colonists brought it to the West Indies and east India company Mariners brought it to India. It arrived in Australia almost immediately after the start of colonisation in 1788. The main crisis for cricket came in the middle of the 1780s. According to the ICC rankings, the various cricket formats have continued to be a major competitive sport in the majority of the former British Empire countries, noting particularly the Indian subcontinent and most recently the Netherlands. The International Cricket Council (ICC) introduced the test championship table in June 2001 and a one-day international championship table in October 2002. Twelve countries largely saw first-class cricket for the first time in 2002 thanks to the ICC Intercontinental Cup. The first ICC Twenty-20 world cup competition took place in 2007.

CHAPTER – 1

1.1 INTRODUCTION

The Twenty20 cricket tournament in India is played in the Indian Premier League (IPL), a professional league. It was created by the Mumbai-based Board of Control for Cricket in India (BCCI), under the direction of Rajeev Shukla, a vice president of the BCCI who also acts as the league's chairman and commissioner. Currently, nine teams with international players are competing in it. After a confrontation between the BCCI and the Indian Cricket League, it was established. In collaboration with India Times, IPL was the first sporting event to ever be streamed live on YouTube in 2010. In its fifth season, its brand value is anticipated to be over $2.99 billion USD. The franchise-based system used by the IPL is modelled after how Americans hire and. The rights to own the team that represented each city were put up for auction, with the highest bidder taking home the prize. On January 24, 2008, the same subject matter was the subject of an auction with a $400 million base price. The final price of the sale was $723 59 million. The most expensive franchise was the Mumbai location owned by Mukesh Ambani's Reliance Industries Limited (RIL), which brought in $111.9 million. Vijay Mallya's United Breweries paid $111.6 million for the Bangalore location. For $107 million, media company Deccan Chronicle got the IPL franchise for Hyderabad, while India Cements paid $91 million for Chennai. Bollywood also made a mark by having two of its biggest performers secure the ownership of their respective teams, with Preity Zinta and her boyfriend Ness Wadia purchasing the Mohali team for $76 million and Red Chillies Entertainment of Shah Rukh Khan and Juhi Chawla purchasing Kolkata for $75.09 per team. The Indian Premier League (ICL) was founded with a billion Indian Rupee corpus and was intended to consist of six teams playing Twenty20 cricket, with ambitions to expand to sixteen teams within three years and eventually transition to 50-over contests. In India, the ICL would now be the richest professional league if these ambitions had come to fruition.

1.2 LITERATURE REVIEW

Predicted the emerging players from batsman as well as from the bowlers using machine learning techniques. predicted estimation of the location of a moving ball based on the value of the cricket sensor network. predicted ranking system which is based on the social network factors and their evaluation in the form of composite distributed framework using Hadoop framework and Map Reduce programming model is used for processing the data. predicted the outcome of IPL-2020 based on the 2008-2019 IPL datasets using Data Mining Algorithms with an accuracy of 82.73%. predicted player evaluation in IPL based on the 2008-2019 datasets using Data Mining Technique. Data mining algorithms are used which gives evaluation using player statistics assessing a player's performance and determining his base price. They predicted about how to select a player in the IPL, based on every player's performance history using algorithms like decision tree, Naïve Bayes and Multilayer perceptron (MLP). MLP outperforms better than other algorithms used Support Vector Machine (SVM), CTree, and Naïve Baiyes classifiers with accuracies of 95.96%, 97.97% and 98.98% respectively, to predict the probability of the winner of the matches. predicted the match outcome based on the toss and venue. predicted match outcome using home ground, time of the match, match type, winning the toss and then batting first by using Naïve Bayes classifier. Passi *et al.* [9], predicted the performance of players based on the runs and the number of wickets. Both the type of problems is treated as classification problems where the list of runs, and list of wickets are classified in different ranges based on machine learning algorithms. The Random Forest algorithm outperforms better than other algorithms.predicted the value of the traits of the batsmen and the bowlers in the current match. This would help in selecting the players for the upcoming matches by using past performances of a player against a specific opposition team by using Multiple Random Forest Regression. predicted the possible fixture for a cricket match based on the various venue, teams, number of holidays between each match in a fair and efficient manner.

CHAPTER - 2

2.1 SYSTEM SPECIFICATION

2.1.1 HARDWARE CONFIGURATION

Operating System	Self-Hosted Technical Requirement	Cloud Technical Requirement
Windows	Windows 8.1+	Mac OS 10.14+
Linux	Ubuntu LTS releases 18.04 or late	Ubuntu LTS releases 18.04 or later
RAM	8 GB	
HDD	1 TB	
Processor	64-bit, four-core, 2.5 GHz minimum per core (If your dataset size is significantly larger than the medium dataset, we recommend 8 cores.)	
Mouse	Dell MS116 1000DPI USB Wired Optical Mouse	
Keyboard	Dell KB522 Business Keyboard-Black	
Monitor	Dell 24 Monitor-S2421HN in-Plane Switching (IPS)	

2.2 Software Specification

IDE	Anaconda
Language Support	Python
Platform	Jupyter Notebook
Browser	Google Chrome Version 101.0.4951.67
Database	MySQL 8.0.29

CHAPTER – 3

3.1 SYSTEM DESIGN

3.1.1 EXISTING SYSTEM:

In the existing system there is a formula to calculate the projected score and a win predictor based on the win percentage of a team and polls. These techniques won't give accurate results because they are based on perceptions and predictions based on a particular instant. Consider calculating projected score, the formula.
(Projected score = current run rate * overs in an innings).
The average accuracy obtained by following the above technique is very less.

3.1.2 PROBLEM STATEMENT:

Given IPL datasets of past 9 years, the main objective of this paper is to predict the outcome of an IPL match between two teams based on the analysis of previously stored data using Machine Learning algorithms. The information will be analysed and pre-processed. After pre-processing the data will be used to train different models in order to give the outcomes. We will analyse the various datasets and use key variables such as strike rate, bowler economy, etc. and feed it as input to an algorithm will help us get the probable outcome of a match.

3.1.3 PROPOSED SYSTEM:

The proposed system uses Random Forest algorithm for Score Prediction of an innings and Logistic Regression for win prediction. The system scrapes data of all IPL matches from 2008 to 2020 from espncricinfo.com. And then we perform Data Cleaning followed by Data Preprocessing. Data Analysis and Visualization is done to make better understanding of the results and exploring various valuable insights. A model is deployed using one of the salesforce platforms, Heroku.

CHAPTER - 4

4.1SYSTEM DESIGN

4.1.1 INPUT DESIGN

LIBRARIES:

```
import pandas as pd
import plotly. express
as px

import plotly. graph_objects
as gg data = pd. read_csv
("IPL 2022.csv") print (data.
Head ())
```

NUMBER OF MATCHES WON:

```
figure = px.bar(data, x=data["match_winner"],
title="Number of Matches Won in IPL 2022")
figure. show ()
```

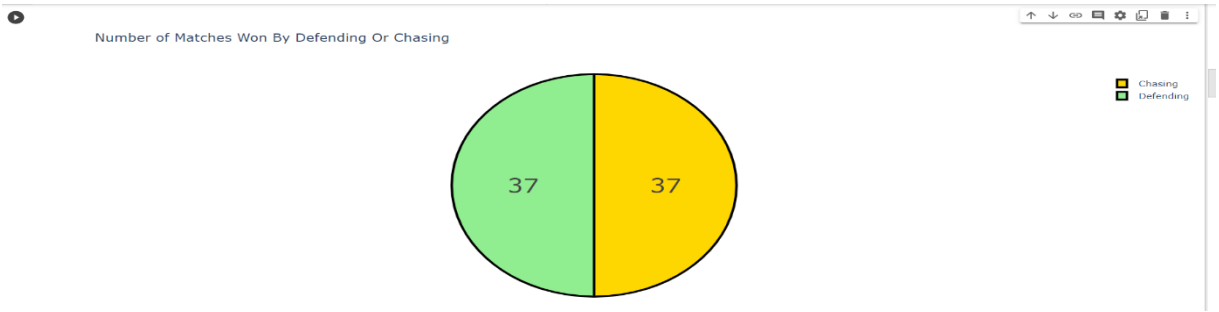
With eight victories, Gujrat now has the most victories in the competition. Gujrat's success in the IPL as a new team is impressive. Let's now examine how most of the teams succeed. Here, we'll examine whether most teams win by chasing (batting second) or protecting (batting)

WIN BY DEFEAT OR CHASING:

INPUT DESIGN:

```
Data[“won_by”] = data[“won_by”].map({“Wickets”: “Chasing”,  “Runs”: “Defending”})

Won_by = data[“won_by”].value_counts()

Label = won_by.index

Counts = won_by.values

Colors = [‘gold’,’lightgreen’]

Fig = go. Figure (data= [go. Pie (labels=label, values=counts)])

Fig.update_layout(title_text=’Number of Matches Won By Defending Or Chasing’)

Fig.update_traces(hoverinfo=’label+percent’, textinfo=’value’, Textfont_size=30,
```

OUTPUT:

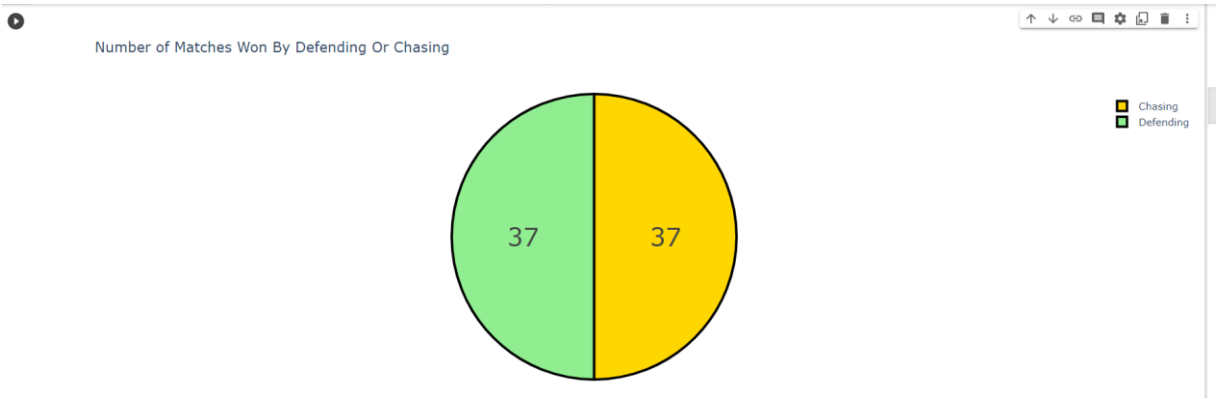


WINNING:

INPUT DESIGN:

```
toss =
data["toss_decision"].value_cou
nts () label = toss. Index counts
= toss.values colors =
['skyblue','yellow']
fig = go.Figure(data=[go.Pie(labels=label,
values=counts)])
fig.update_layout(title_text='Toss Decision')
fig.update_traces(hoverinfo='label+percent',
textinfo='value', textfont_size=30,
marker=dict(colors=colors,
line=dict(color='black', width=3))) fig.show()
```

Thus, after winning the toss, the majority of captains decide to field. So far, 43 games have had captains opt to field first, and only three have had captains

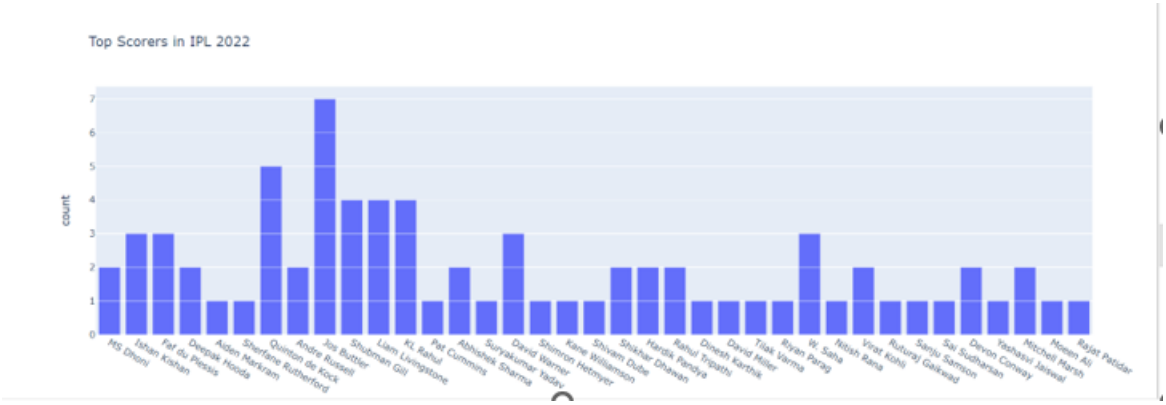


TOP SCORES:

INPUT DESIGN:

figure = px.bar(data, x=data["top_scorer"], title="Top Scorers in IPL 2022")
figure.show()

Jos Buttler has currently scored the most goals in 5 games.



RUN SCORES:

INPUT DESIGN:

Let’s analyze it deeply by including the runs scored by the top scorers:

figure = px.bar(data,

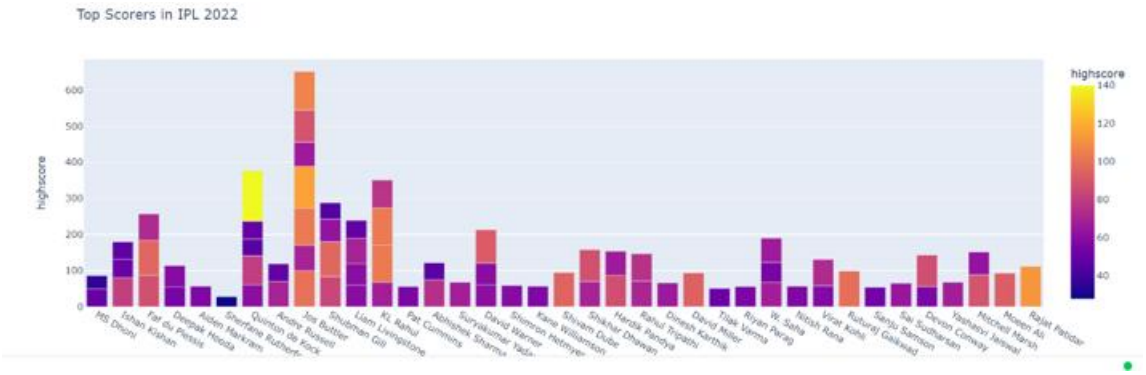
x=data["top_scorer"],

y = data["highscore"],

color = data["highscore"],

title="Top Scorers in IPL 2022")

figure.show()



Jos Buttler has currently scored the most goals in 5 games. He appears

So till now, Jos Buttler has scored three centuries, and KL Rahul has scored two centuries.

MATCH AWARDS:

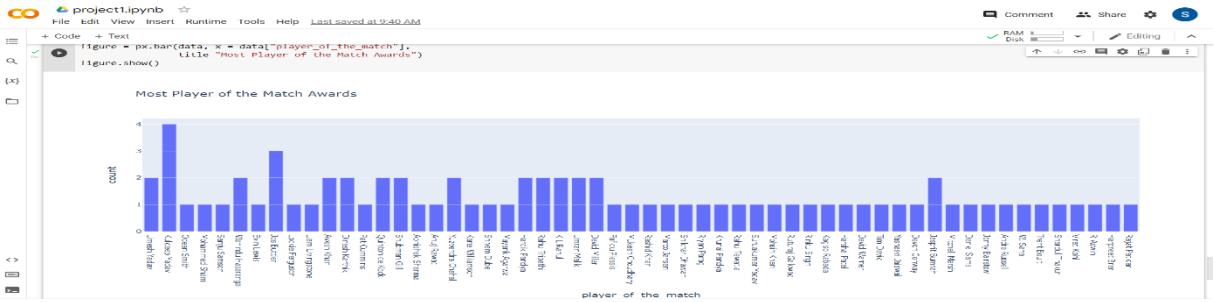
INPUT DESIGN:

Now let's have a look at the most player of the match awards till now in IPL 2022:

figure = px.bar(data, x = data["player_of_the_match"],

title="Most Player of the Match Awards")

figure.show()



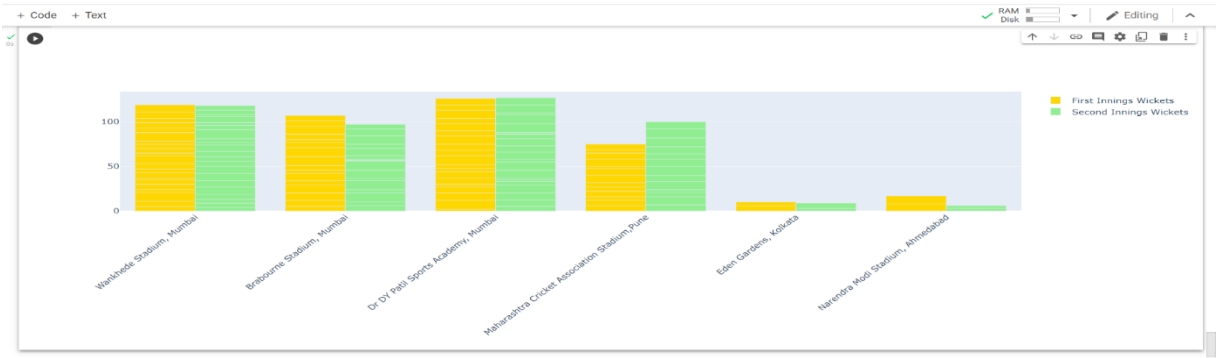
Yuzvendra Chahal has the top bowling statistics in four games, as can be shown.

Yuzvendra Chahal will therefore have a terrific.

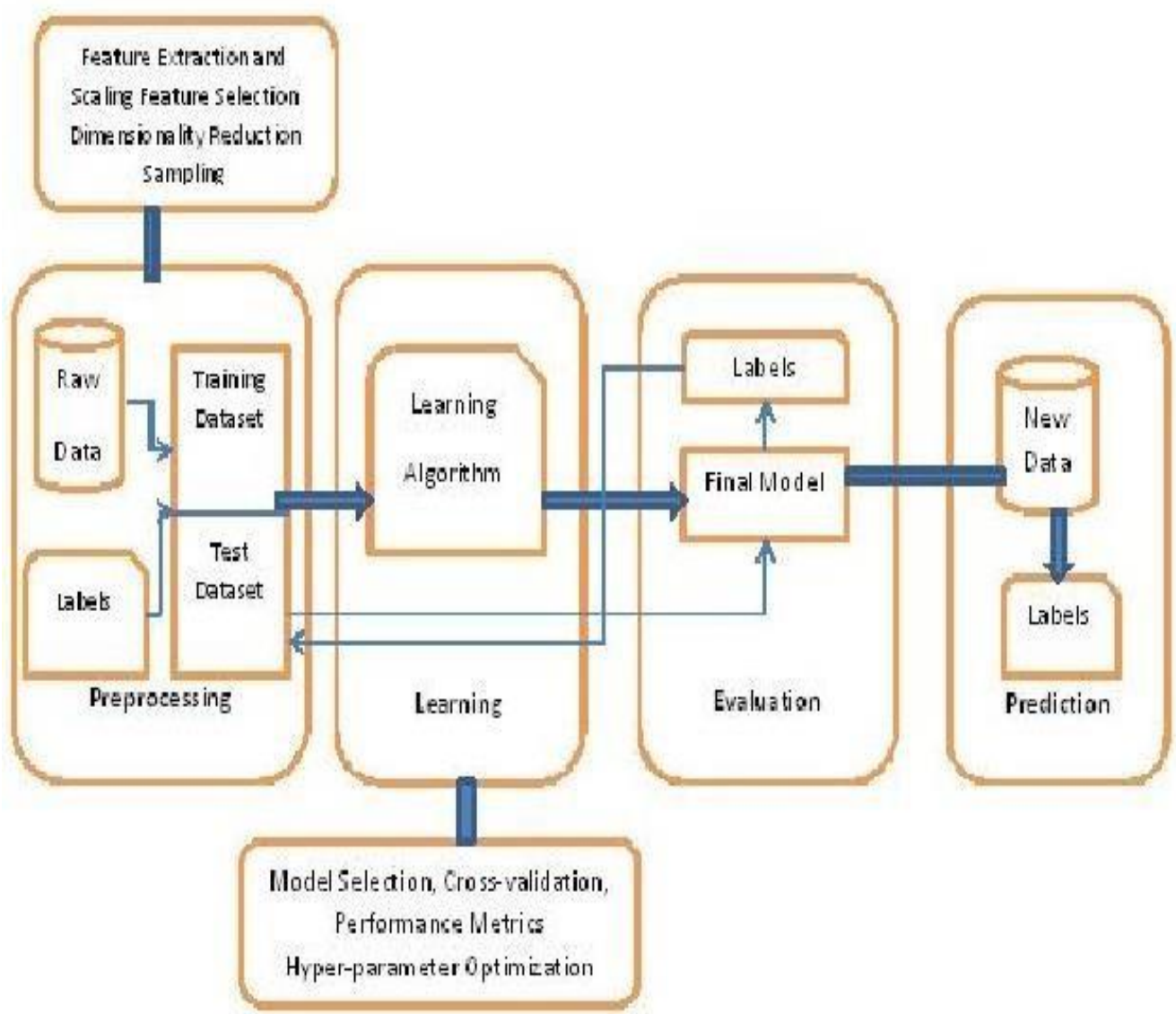
WICKETS:

INPUT DESIGN:

```
 figure
=go.Figure()
figure.add_trace(g
o.Bar(
x=data["venue"],
y=data["first_ings
_wkts"],
name='First
Innings  Wickets',
marker_color='go
ld'
))
figure.add_trace(go.Bar(
x=data["venue"],
y=data["second_ings_wkts"],
name='Second      Innings      Wickets',
marker_color='lightgreen'
))
figure.update_layout(barmode='group',              xaxis_tickangle=-45)
figure.show()
```



SYSTEM ARCHITECTURE:



CHAPTER – 5

CONCLUSION:

Though **IPL Data Analysis Python Project** is complete for beginners still it has a few limitations that advanced pandas users can overcome very easily. Matplotlib has been used for graph plotting that needs all the data in a refined format. Seaburn python module has built-in capabilities for better graph plotting. In the IPL we observe that only the current run-rate is used to predict the final score which is not an efficient way, since there are many other factors which can affect the projected score, we use some tools and libraries of python like 'Requests','Beautifulsoup' and analyze the conditions to predict the total score. With this project we can able to analyze and predict the score and winning probability of a team. Predicting a winner in a sport such as cricket is especially challenging and involves very complex processes. But with the introduction of machine learning, this can be made much easier and simpler. In this paper, various factors have been identified that contribute to the results of the Indian Premier League matches. Factors that have a major impact on the outcome of an IPL match include the teams playing, the venue, the city, the toss winner and the toss decision. We have analyzed IPL data sets and predicted game results based on player performance. The methods used in the work to obtain the final test are Logistic regression, Support Vector Machine (SVM), Decision tree, Random Forest classifier and K-nearest neighborhood. Random Forest classification (RFC) outperforms the other algorithm.

BIBLIOGRAPHY:

1. Bertland, Linda (2002).

"Searching the Internet: Search Engines and Subject Indexes." School Libraries on the Web [Online], available at http://www.sldirectory.com/search.html.

2. Bopp, Richard E. and Linda C. Smith (1995).

Reference and Information Services: An Introduction. 2nd ed. Englewood, Colorado: Libraries Unlimited.

3. Henley, Jennifer-Jordan (1996).

"Writing a General Research Paper." Roane State Community College Online Writing Lab [Online], available at http://www.rscc.cc.tn.us/owl/Research.html.

4. Irwin, Ken (2002).

"Criteria for Web Site Evalutation." University of Michigan Shapiro Undergraduate Library [Online], available at http://guides.lib.umich.edu/content.php?pid=30524.

5. Kuhlthau, Carol Collier (1993).

Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex Publishing.

6. Kuhlthau, Carol Collier (1994).

Teaching the library research process (2d ed.). Metuchen, NJ: The Scarecrow Press.

7. Lamm, Kathryn (1995).

10,000 Ideas for Term Papers, Projects, Reports & Speeches. New York: Macmillan.

8. Williams, Sharon and Laura Reidy.

"Introductions and Thesis Statements." Hamilton College, Nesbitt-Johnson Writing
Center [Online], available at
http://www.hamilton.edu/academics/Resource/WC/Intro_Thesis.
html.

SOURCE CODE:

LIBRARIES:

```
import pandas as pd
import plotly. express
as px

import plotly. graph_objects
as gg data = pd. read_csv
("IPL 2022.csv") print (data.
Head ())
```

NUMBER OF MATCHES WON:

```
figure = px.bar(data, x=data["match_winner"],
title="Number of Matches Won in IPL 2022")
figure. show ()
```

WIN BY DEFEAT OR CHASING:

```
Data[“won_by”] = data[“won_by”].map({“Wickets”: “Chasing”,  “Runs”: “Defending”})

 Won_by = data[“won_by”].value_counts()

 Label = won_by.index

Counts = won_by.values

Colors = [‘gold’,’lightgreen’]

Fig = go. Figure (data= [go. Pie (labels=label, values=counts)])

Fig.update_layout(title_text=’Number of Matches Won By Defending Or Chasing’)

Fig.update_traces(hoverinfo=’label+percent’, textinfo=’value’, Textfont_size=30,
```

WINNING:

```
toss =
        data["toss_decision"].value_cou
        nts () label = toss. Index counts
        = toss.values colors =
        ['skyblue','yellow']
         fig = go.Figure(data=[go.Pie(labels=label, values=counts)])
fig.update_layout(title_text='Toss Decision')
fig.update_traces(hoverinfo='label+percent', textinfo='value',
textfont_size=30, marker=dict(colors=colors, line=dict(color='black',
width=3))) fig.show()
```

TOP SCORE :

```
figure = px.bar(data, x=data["top_scorer"], title="Top Scorers in IPL 2022")
figure.show()
```

RUN SCORES:

Let’s analyze it deeply by including the runs scored by the top scorers:

```
figure = px.bar(data,
x=data["top_scorer"],
y = data["highscore"],
color = data["highscore"],
title="Top Scorers in IPL 2022")
figure.show()
```

MATCH AWARDS:

```
figure = px.bar(data, x = data["player_of_the_match"],
title="Most Player of the Match Awards")
figure.show()
```

WICKETS:

```
figure
=go.Figure()
figure.add_trace(g
o.Bar(
x=data["venue"],
y=data["first_ings
_wkts"],
name='First
Innings  Wickets',
marker_color='go
ld'
))
figure.add_trace(go.Bar(
x=data["venue"],
y=data["second_ings_wkts"],
name='Second  Innings  Wickets',
marker_color='lightgreen'
))
figure.update_layout(barmode='group',          xaxis_tickangle=-45)
figure.show()
```

*****