# 1. INTRODUCTION:

1.1 Overview:

BCCI has been organizing the IPL T20 cricket tournament every year. The use of analytical methods in various aspects of cricket including results prediction and toss decision for winning is very important.There is a huge demand for the algorithm that best predicts the result of cricket because of its popularity and huge amount of money involved in the game. Thus the analysis of IPL results becomes more important. Prediction of outcome of a match and toss decision using dashboard , plots and predictive model(with in the error limits) based on Machine learning is an important aspect in cricket. Records of the past performance of players,wininng, venue,etc and other related data can be analysed to create models that predicts the winning team. This model is created using IBM Cognos Analytics,IBM Watson Studio,NoDe Red,IBM Cloud their results can be compared based on the Evaluation Measures as accuracy, precision, recall, sensitivity and error rate.

1.2 Purpose:

The objective of the solution is to create dashboard that visualizes IPL data for certain set of stated problems. It will also be able to predict the team who will win the next match based on input provided by user (Climate Condition, toss result, field/batting selection etc) using Machine learning algorithm (classification- as output is categorical). For analysis of current data set dashboard will be used (bar graph, scatter plot etc) and for prediction Node Red (to provide GUI).
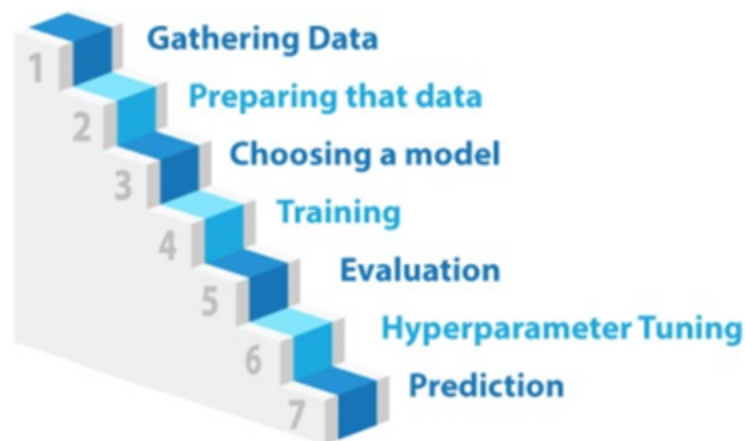
Literature Survey

The dataset has **18 columns**. Let's get acquainted with the columns.

- **id**: The IPL match id.

- **season**: The IPL season

- **city**: The city where the IPL match was held.

- **date**: The date on which the match was held.

- **team1**: One of the teams of the IPL match

- **team2**: The other team of the IPL match

- **toss_winner**: The team that won the toss

- **toss_decision**: The decision taken by the team that won the toss to 'bat' or 'field'

- **result**: The result('normal', 'tie', 'no result') of the match.

- **dl_applied**: (1 or 0)indicates whether the Duckworth-Lewis rule was applied or not.

- **winner**: The winner of the match.

- **win_by_runs**: Provides the runs by which the team batting first won

- **win_by_runs**: Provides the number of wickets by which the team batting second won.

- **player_of_match**: The outstanding player of the match.

- **venue**: The venue where the match was hosted.

- **umpire1**: One of the two on-field umpires who officiate the match.

- **umpire2**: One of the two on-field umpires who officiate the match.

- **umpire3**: The off-field umpire who officiates the match

2.1 Problem Description

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 drawn the fans to watch the matches in large numbers. Within a short period, IPL has become the highest revenue generating league of cricket. With all this, the amount of data being generated in terms of matches revenue scores etc has also become huge. Analysing such vast amounts of data would give great insights in forecasting match results ,top scores and wicket takers etc.

The objective of this solution is to create dashboard that visualizes following capabilities and also forecast the future results
1. To find the team that won the most number of matches in the entire IPL.
2. To find the team that lost the most number of matches in the entire IPL.

3. Does winning a toss increase the chances of victory.

4. To find the player with the most player of the match awards.

5. To find the city that hosted the maximum number of IPL matches.

6. To find the most winning team for each season.

7. To find the on-field umpire with the maximum number of IPL matches.

8. To find the biggest victories in IPL while defending a total and while chasing a total.

9. Which team won the most matches while batting first.

10. Which team won the most matches while batting second.

11. List of teams which have won matches by most runs cumulatively

2.2 Proposed Solution

Proposed Model By collecting the previous season data of each player and team, we perform training on the data by developing data analyzation. The information makes use of to calculate points of each player and also calculate the overall strength of each team based on past performance. The output is displayed with the help of various types of graphs where user can predict the match result or take decision in selecting the players for a particular match.
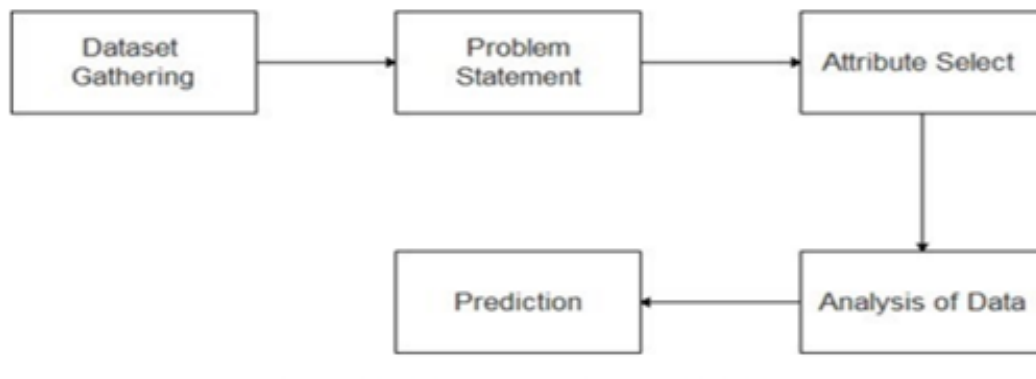

Fig. 1. Block Diagram for Proposed Model Gathering Dataset:

 The first step is to collect dataset of past matches which is enough to perform the operations. Understanding Problem Statement: Problem statement consist of the data to be displayed like, in how many matches team opted batting first after winning toss and won the match and how many choose fielding and won the match.

 Attribute Selection: Base on the problem statement the data will be fetched from the.csv File particular data is fetched.

Analyzing: The raw data is converted into useful information by performing certain operations. This is done to provide information in appropriate manner.

Prediction: The future is predicated whenever that situation occurs with the help of the information the

user can predict the result of the match.

1. **General Description**
   The objective of the solution is to create dashboard that visualizes IPL data for certain set of stated problems. It will also be able to predict the team who will win the next match based on input provided by user (Climate Condition, toss result, field/batting selection etc) using Machine learning algorithm (classification- as output is categorical). For analysis of current data set dashboard will be used (bar graph, scatter plot etc) and for prediction Node Red (to provide GUI).

2. **Novelty / Uniqueness:**
   Prediction model based on Machine Learning is the novelty added by us to the project that would be able to predict winning percentages of teams playing, runs to be scored in match by teams, Selection of batting and bowling after toss, toss winning percentages of team,etc predictions can be made.

1. **Business / Social Impact:**
   Proposed solution impacts the fantasy league application business as the user of these applications will be able to predict more precisely with the help of dashboard , plots and predictive model(with in the error limits) based on Machine learning. So more and more customers will be attracted in order to gain more profits.
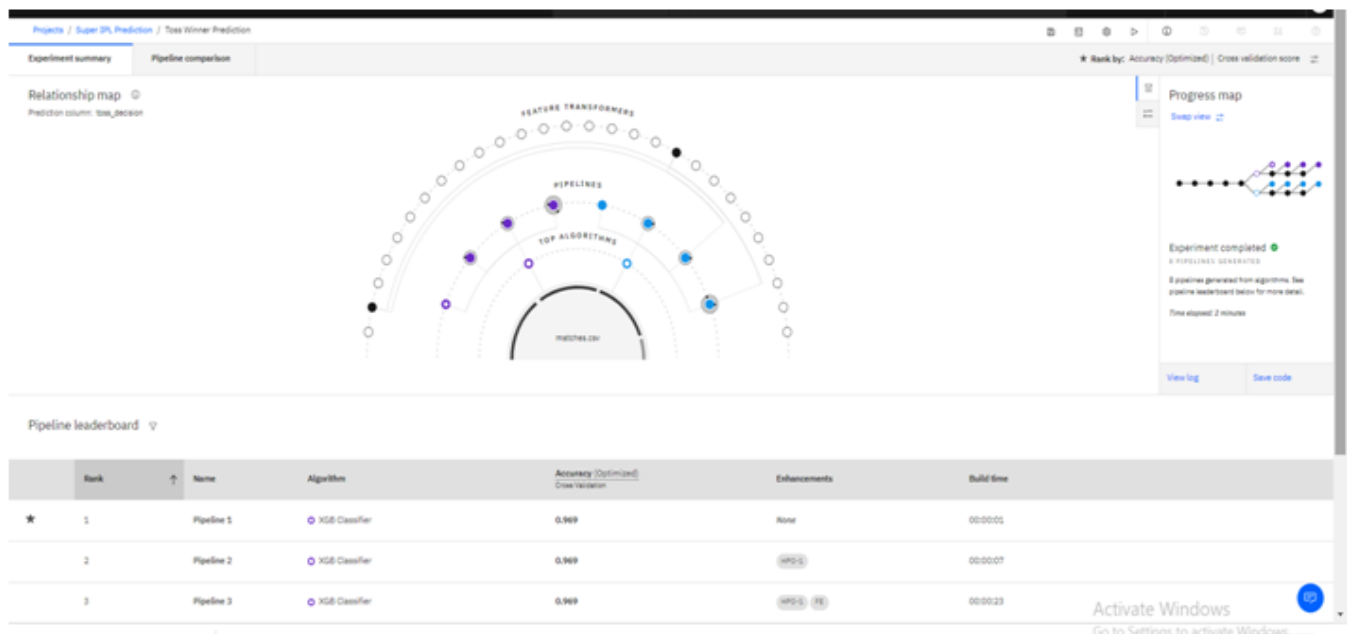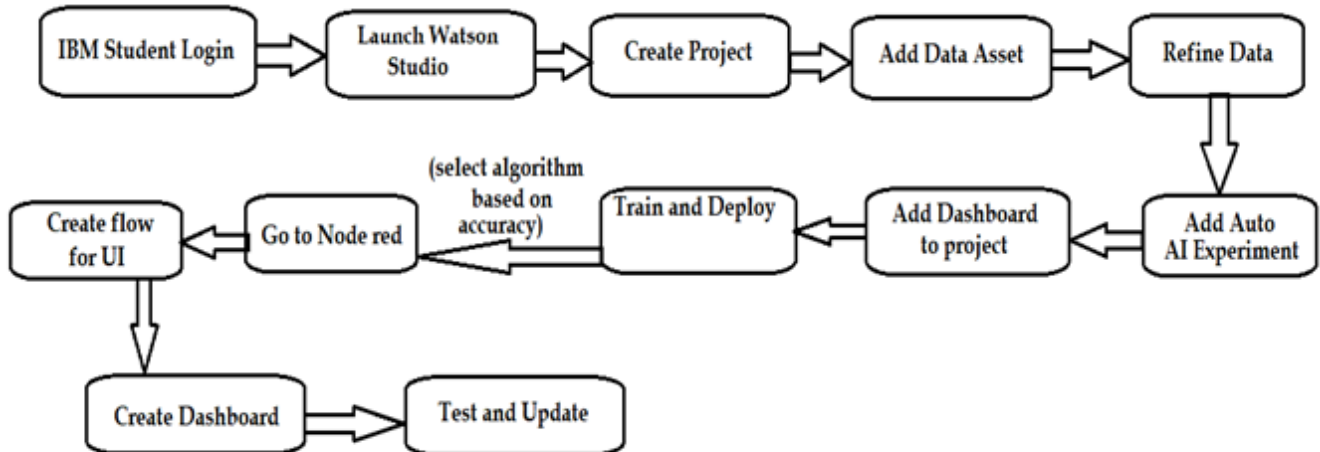

Fig 2 : Training Model on the basis of data set given.

Theoretical Analysis

3.1 Block Diagram

3.2 Software Used

1. IBM Cognos Analytics
2. IBM Cloud
3. IBM Watson Studio.
4. Machine Learning
5. NoDe Red

Experimental Investigation:

- **Team Winning Count**

	team	wins
0	Mumbai Indians	109
1	Chennai Super Kings	100

2	Kolkata Knight Riders	92
3	Royal Challengers Bangalore	84
4	Kings XI Punjab	82
5	Rajasthan Royals	75
6	Delhi Daredevils	67
7	Sunrisers Hyderabad	58
8	Deccan Chargers	29
9	Gujarat Lions	13
10	Pune Warriors	12
11	Delhi Capitals	10
12	Rising Pune Supergiant	10
13	Kochi Tuskers Kerala	6
14	Rising Pune Supergiants	5

- **Player of the match count**

	player	wins
0	CH Gayle	21
1	AB de Villiers	20
2	RG Sharma	17

3	MS Dhoni	17
4	DA Warner	17
5	YK Pathan	16
6	SR Watson	15
7	SK Raina	14
8	G Gambhir	13
9	V Kohli	12

- **Toss Win Count:**

	team	wins
0	Mumbai Indians	98
1	Kolkata Knight Riders	92
2	Chennai Super Kings	89
3	Kings XI Punjab	81
4	Royal Challengers Bangalore	81
5	Delhi Daredevils	80
6	Rajasthan Royals	80
7	Sunrisers Hyderabad	46
8	Deccan Chargers	43

9	Pune Warriors	20
10	Gujarat Lions	15
11	Delhi Capitals	10
12	Kochi Tuskers Kerala	8
13	Rising Pune Supergiants	7
14	Rising Pune Supergiant	6

- **Player of the Match Count :**

	player	wins
0	CH Gayle	21
1	AB de Villiers	20
2	RG Sharma	17
3	MS Dhoni	17
4	DA Warner	17
5	YK Pathan	16
6	SR Watson	15
7	SK Raina	14

8	G Gambhir	13
9	V Kohli	12

■ On Field Umpire Count

	umpire1	umpire2
HDPK Dharmasena	73.0	14.0
Asad Rauf	51.0	NaN
S Ravi	49.0	57.0
AK Chaudhary	43.0	15.0
Aleem Dar	38.0	NaN
...	...	...
K Srinivasan	NaN	3.0
KN Anantapadmanabhan	NaN	3.0
SD Ranade	NaN	2.0
Nand Kishore	NaN	1.0
Subroto Das	NaN	1

- **Winner Per Season Count**

	year	team	wins
0	2008	Rajasthan Royals	13
0	2009	Delhi Daredevils	10
0	2010	Mumbai Indians	11
0	2011	Chennai Super Kings	11
0	2012	Kolkata Knight Riders	12
0	2013	Mumbai Indians	13
0	2014	Kings XI Punjab	12
0	2015	Chennai Super Kings	10
0	2016	Sunrisers Hyderabad	11
0	2017	Mumbai Indians	12
0	2018	Chennai Super Kings	11
0	2019	Mumbai Indians	11

Flow Chart

RESULT

- Mumbai Indians is the most successful team in IPL.

- Mumbai Indians has won the most number of toss.

- There were more matches won by chasing the total(419 matches) than defending(350 matches).

- When defending a total, the biggest victory was by 146 runs(Mumbai Indians defeated Delhi Daredevils by 146 runs on 06 May 2017 at Feroz Shah Kotla stadium, Delhi).

- When chasing a target, the biggest victory was by 10 wickets(without losing any wickets) and there were 11 such instances.

- The Mumbai city has hosted the most number of IPL matches.

- Chris Gayle has won the maximum number of player of the match title.

- Winning toss gives a slight edge(52% probability of winning) against the opponents.

- Five Indian players have figured in the top ten IPL players list.

- **S. Ravi(Sundaram Ravi) has officiated the most number of IPL matches on-field.**

- **Eden Gardens has hosted the maximum number of IPL matches.**

- **Till 2019, 40 venues have hosted 756 IPL matches.**

Advantages and disadvantages

7.1 Advantages

- Predictive models help in predicting for future results.

- Easy analysis of large data set.

- Compatible model for future data as well.

- Graphical User Interface for better understanding.

- User friendly.

- Reduce Human effort for making conclusions.

- Can be developed/used by user with zero coding skill.

7.2 Disadvantage

- It requires large computation power.

- It requires large data set for training the model.

- Not 100% accurate.

8. Applications
- For predicting result beforehand and increasing chances of winning in gambling business.
- Presence of player which increases chances on winning the match.

9. Conclusion

The IPL data analysis is the most famous technology in today's world. With the help of the data analysis we can find out the pros and cons of the player and work on the performance of the individual player and also the whole, it will help in taking decision. The application can be used for the selection commission to select the best player including bowler, batsman and even the fielder for the team and to perform well on the field during match. It is helpful for all type of game to work on the performance and predict the future performance of the player and team. The following methodology helps in other T20 leagues around the world like PKL,LPL,CBL,BBL,SuperSmash,T20 Blast, MSL, BPL, APL, and World League CLT20.

10. **Future Scope**

The solution is not limited to current data set as prediction will be enhanced with more data entries in future. It will not only help in prediction of winning team but also will suggest whether the toss winner must go for fielding or batting based on past records. This will enhance gambling business in World by providing more accurate predictions. It can be made more accurate by integrating more data into it and using better designed algorithm.

11. **Bibliography**

- https://www.kaggle.com/nowke9/ipldata?select=matches.csv
- https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=stories-get-started-dashboards
- cloud.ibm.com
- https://www.import.io/post/business-data-analysis-what-how-why/

: