

IPL Match Prediction using Machine Learning

Sunil Bhutada¹, Subhani Shaik², Aishwarya Laxmi Nethi³, Pabba Adeeshwar⁴,
Sairam Parshi⁵

¹*Professor, IT Department Sreenidhi Institute of Science and Technology
Yamnammpet, Hyd-bad.*

²*Associate Professor, IT Department Sreenidhi Institute of Science and
Technology Yamnammpet, Hyd-bad.*

³⁻⁵*B. Tech IV year, IT Department Sreenidhi Institute of Science and Technology,
Yamnammpet, Hyd-bad*

Abstract

Cricket, the mainstream and widely played sport across India which has the most noteworthy fan base. Indian Premier League follows 20-20 format which is very unpredictable. IPL match predictor is a ML based prediction approach where the data sets and previous stats are trained in all dimensions covering all important factors such as: Toss, Home Ground, Captains, Favourite Players, Opposition Battle, Previous Stats etc, with each factor having different strength with the help of KNIME Tool and with the added intelligence of Naive Bayes network and Eulers strength calculation formula.

Keywords: *Naïve Bayes Classification, Eulers Strength Formula, Cricket Prediction, Supervised Learning, KNIME Tool.*

1. Introduction

With technology growing abundantly in the last few decades, an inside and out obtaining of information has gotten moderately simple. Subsequently, Machine Learning is turning out to be a significant pattern in sports examination in light of the accessibility of live just as chronicled information. Analytics of sports would be procedure for gathering previous game information and investigating it extricate basic information from it, from an expectation which encourages powerful and dynamic judgement. It could be whether to buy a player or not in auction, else whom to set on the field in coming match, using more competitive task like, preparing the strategies to matches in future depending on the prediction being made using various factors from past matches.

In this context, the information of last 5 years IPL has been collected and is placed in .csv file format where the delimiters used is comma (,) where this format is widely used in computer science. As we have gathered a lot of data from last few years IPL, there is a huge chance for the data to be unstructured, hence we performed data cleaning and removed unnecessary information.

2. Literature survey

Different ML logics are used and they have proved its efficiency. Connection between ML, games goes previously to the underlying long periods for man-made consciousness. Man-made consciousness contemplated that machine learning draws near utilizing the game of checkers [1]. In cricket, to foresee a match's result, essential duty is separating basic highlights that influence after effect of a match. Fascinating works have been done in order to foreseeing result. This literature study wraps up most produced works which anticipated an aftereffect in a match coordinate earlier for 50 overs cricket. Bandalasiri

investigated elements such as home ground, victory hurl, impact of DL technique for 50-50 [2]

3. Problem Statement

The general Match result is if the team won the match or no. But just focusing on winning or losing does not give accurate prediction. We should also consider other factors such as home grounds etc. Considering other factors so forth would help in deciding the match prediction result along with the strength which supports the decision which was earlier predicted.

Basically a T20 match has a lot of aspects which influence the game result, here in this project we have focused on all these aspects which have a probability of becoming the decision making factor of the match, hence by including such aspects we have increased the strength to our analysis.

If there are two team's P and Q then the result won't just be either P or Q won the match, but this analysis will give us the predicted winner along with some confidence which is the strength which we have obtained.

4. Process Flow

The process stream shows how the entire procedure of the project is done and the result is seen. The following flow chart shows the procedure:

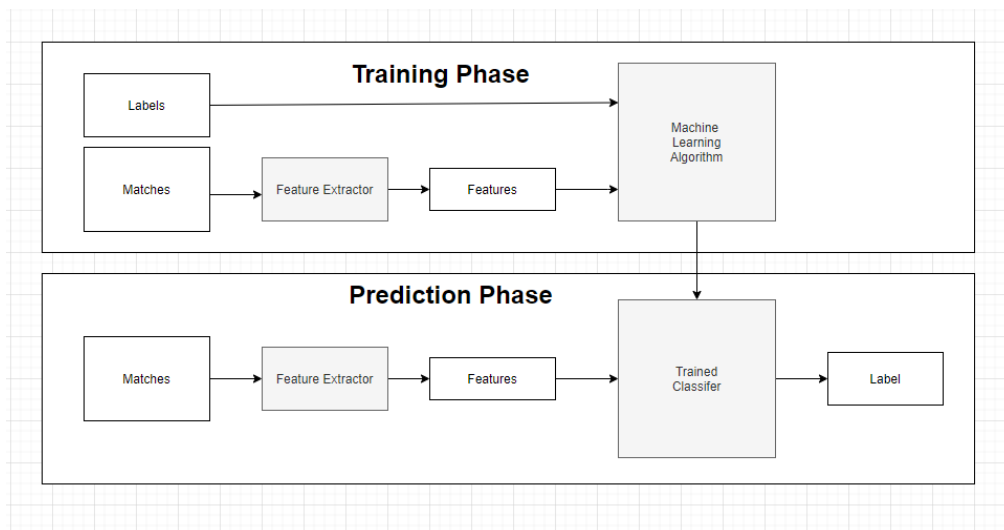


Figure 1. Process Flow

5. Methodology

5.1 Understanding Data

The data understanding is the assemblage or gathering of the match data. The raw data was in an unstructured format with various fields. We have taken the data and converted it into a structured .csv file format. This dataset is composed of all aspects. Based on these aspects we have done our prediction.

5.2 Data preparation

Data pre-processing a crucial role in the process. This step deals with modifying the data and then converting it into a usable format. Cleaning the data consists of numerous steps which comprises of eliminating null values, joining similar attributes. It plays a significant role and is very continuous and complicated stage.

5.3 Data Visualization:

The following is the data depicting the various statistics related to the project which impacts the outcome of the winning team.

5.3.1 Team Winning Percentage

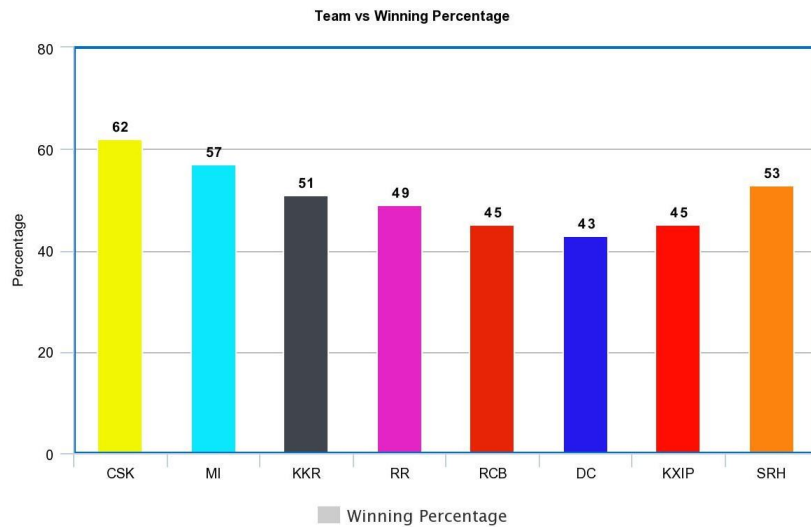


Figure 2. Team Winning Percentage

5.3.2 Home Ground Advantage

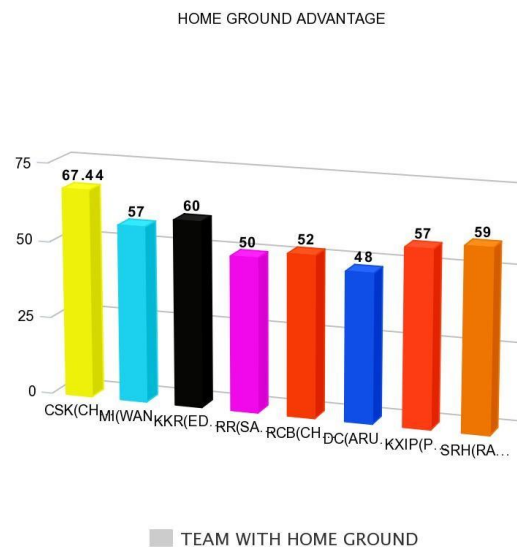


Figure 3. Home Ground Advantage

5.3.3 Teams with their IPL Trophies

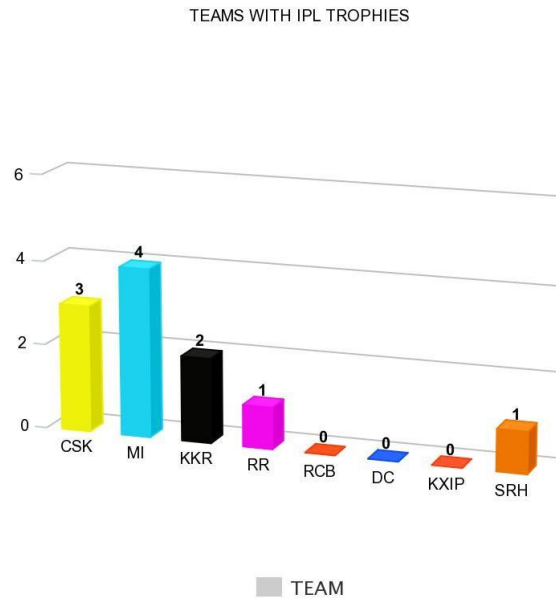


Figure 4. Teams with their IPL Trophies

5.3.4 Teams Qualifying for Playoffs

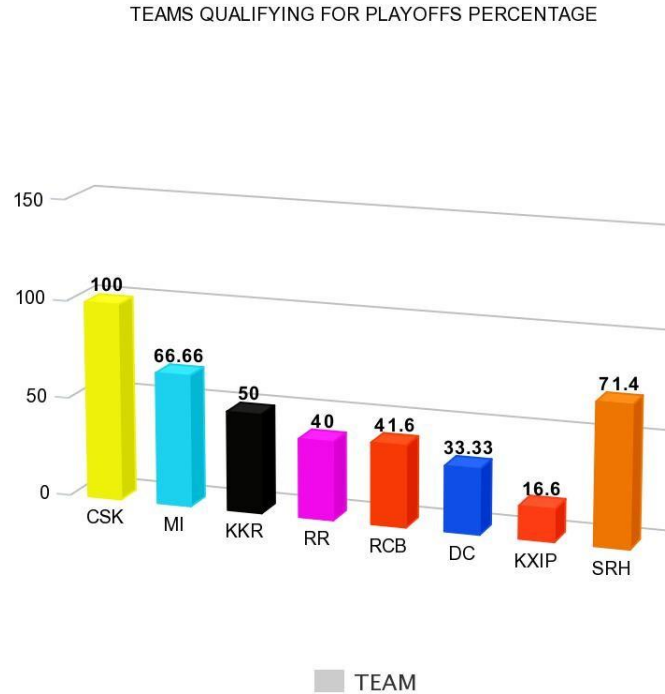


Figure 5. Teams Qualifying for Playoffs

5.3.5 Individual Team Performance

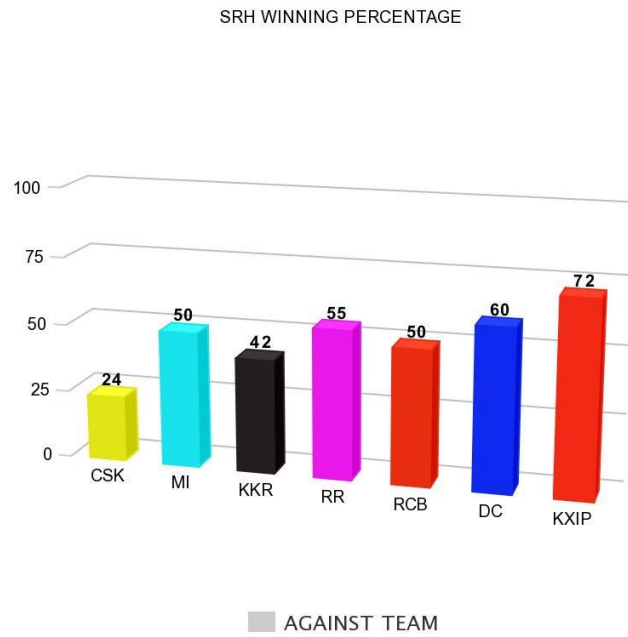


Figure 6. SRH

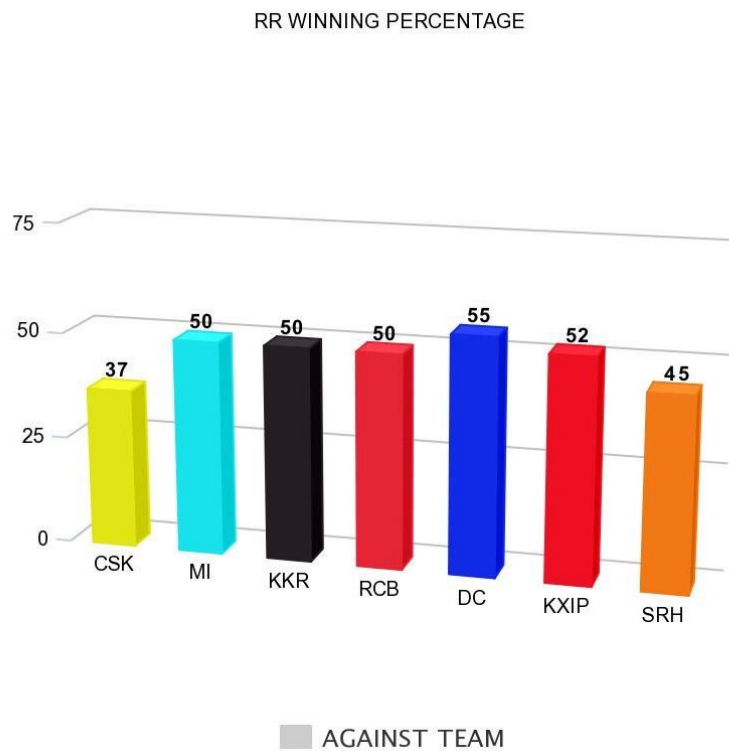


Figure 7. RR

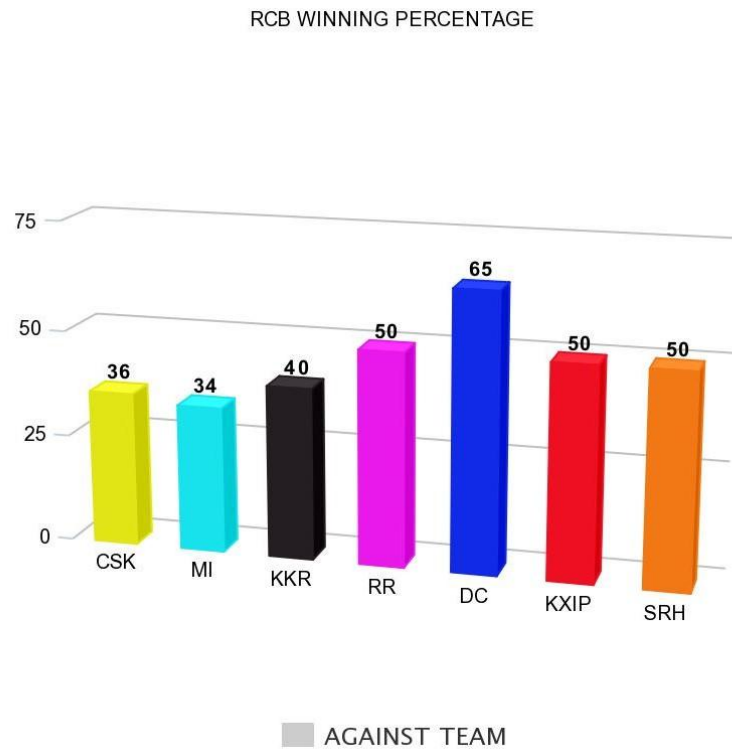


Figure 8. RCB

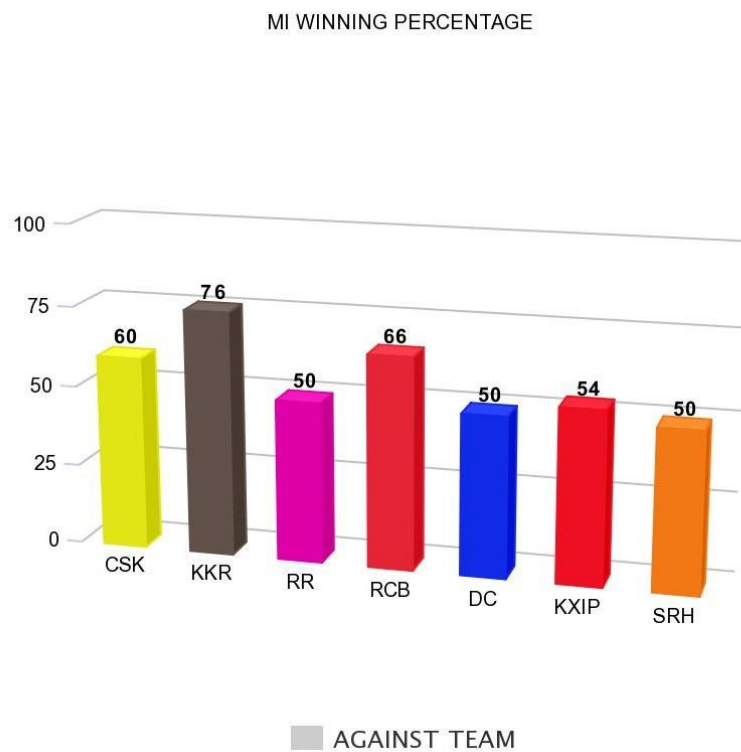


Figure 9. MI

KXIP WINNING PERCENTAGE

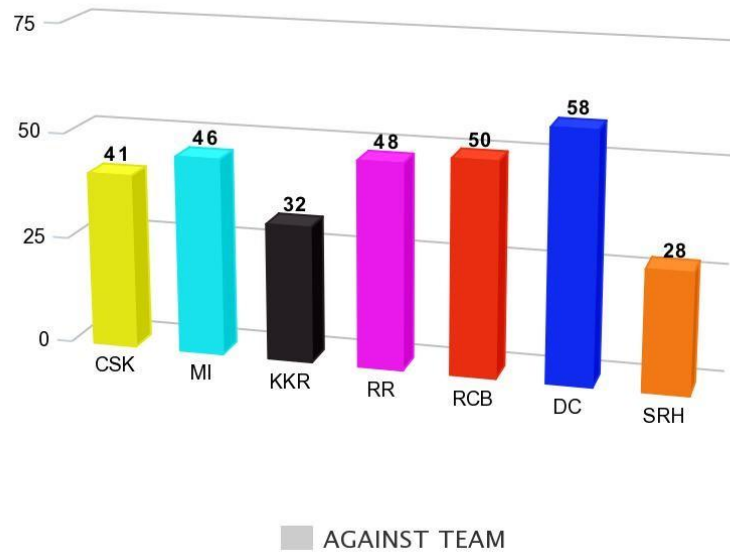


Figure 10. KXIP

KKR WINNING PERCENTAGE

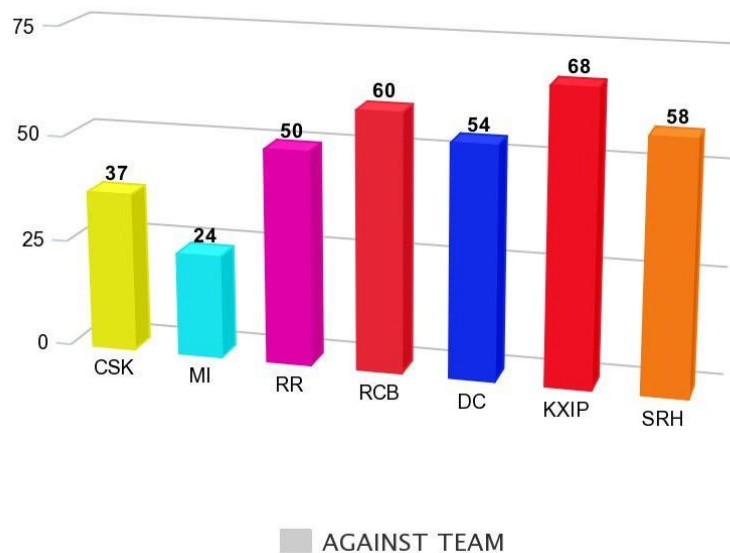


Figure 11. KKR

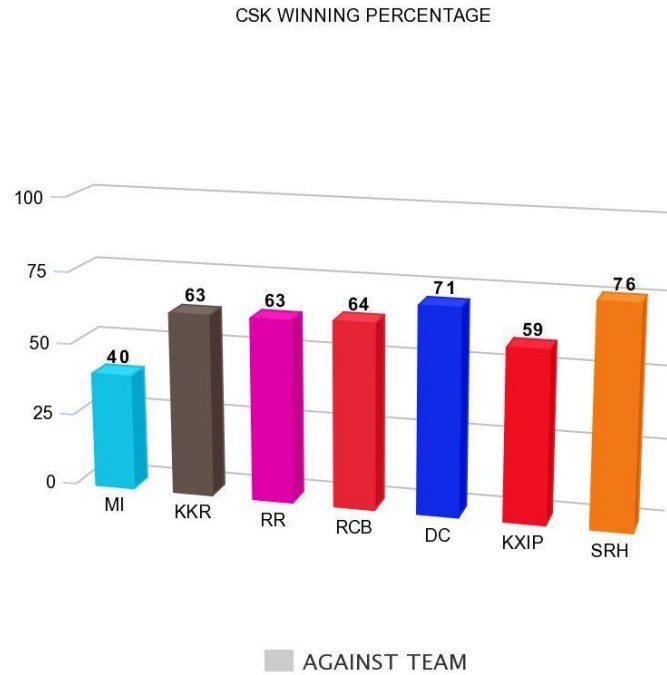


Figure 12. CSK

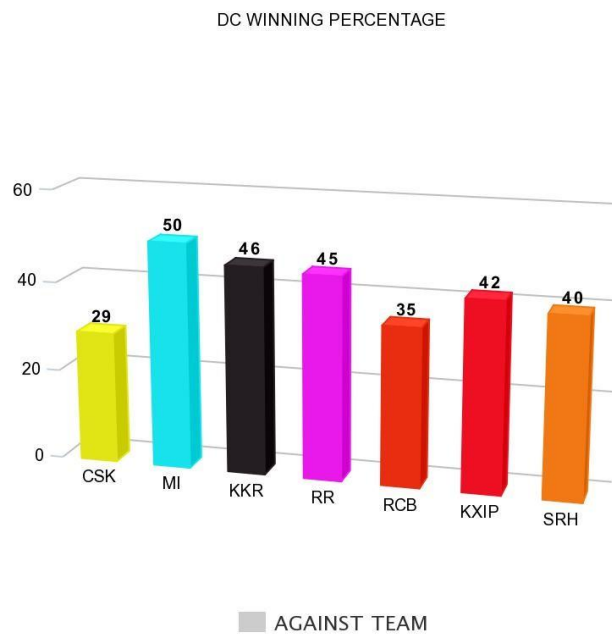


Figure 13. DC

5.4 Modelling:

5.4.1 Supervised learning: There are two sorts of learning ML methodologies, Supervised & Unsupervised Learning, we utilised Supervised model as it has a predefined label, hence there won't be vagueness in the handling of the information, Here the name is foreseeing the triumphant group by thinking about different factors and highlights.

5.4.2 Eulers Strength Calculation Formula: When there are various factors which decides the outcomes of the match, one needs some intelligent strength given to each factor in various instances one of such formulas is Eulers Strength Calculation formula, the formula is:

$$N - E + F = K + 1$$

N= relative strength

E= degree

F= no of factors

K= no. of components

5.4.3 Naive Bayes Theorem: In Discrete Mathematics Naive Bayes Theorem is an important theorem which uses conditional probability and be applied when the outcome is already given. In our system we first consider both of the matches individually and finds the reverse probability using Naive Bayes Classification and predicts the team which ever had more probability.

$$P(A|B) = \frac{\{P(B|A) P(A)\}}{\{P(B)\}}$$

A and B are two different teams.

5.4.4 KNIME tool: Knime is a straightforward yet wealthy in GUI apparatus for examination and displaying . It has highlights of Matlab, R and other reenactment modules for explicit spaces (like concoction/pharmaceutical...etc). It has a mix of the two information examination, displaying and representation capabilities(with existing yield modules associated with the model you structure). The yields can be spared in PMML/XML group which can be later devoured by some other expository apparatuses. Contributions across different sources (DB associate/XML/.CSV/.xls document groups) can be associated with the model.

6 Existing System

Existing system purely depends on how the team deals with the very important factors that influence the outcome there are pretty intelligent systems out there like Dream 11, which use Analytics of the IPL and other cricket data for predicting the outcome.

7 Proposed System

In our system we aimed to prioritise the important factors which influence the match by giving them the balanced strength using intelligent formulas like Eulers Strength calculation formula. We researched on beauty behind T20 cricket which is very sensitive as even one over can change the match and hence by showing the importance of Machine Learning in the prediction by giving the support or the confidence to the Winner that is predicted. The Eager Learning which we have taught also helps the DBA for maintaining the database. Our system is even more flexible for training the further more data at user level as well as server level.

8 Future Scope

Though it is a fun machine learning project, it can be extended to corporate level also, where Sporting channels would like to show the metrics which can be used to increase the audience pulse, which shows the factors which can change even a losing game inclined to winning. It uses for creating online polls, Sports Business. IPL teams do their own analysis on each and every player as they spend millions on each of them, they use various analysis like Player Vs Ground Battle, before bidding the player.

9 Results

After performing the ML algorithms, we obtained the following results based on different factors:

```
IPL MATCH PREDICTOR

Enter Home Team
1.CSK  2.MI  3.KKR  4.RR
5.RCB  6.DC  7.KXIP  8.SRH

Input:1

Enter Away Team
1.CSK  2.MI  3.KKR  4.RR
5.RCB  6.DC  7.KXIP  8.SRH

Input: 8

Time Session
1.Afternoon
2.Evening
2

Want to predict before or after the toss being Tossed?

1.Before Toss
2.After Toss

Note: Tosses are considered important in cricket, as captains decide whether to bat or bowl first depending on the condition of the pitch.
2
Who won the Toss?
1.CSK
2.SRH
2
Toss Decision
SRH choose to bat first
SRH choose to bowl first
1
-----please wait while we predict the match-----

Prediction:
CSK wins the match with confidence of 58.3
```

Figure 14. CSK vs SRH (with toss)

```
IPL MATCH PREDICTOR

Enter Home Team
1.CSK  2.MI  3.KKR  4.RR
5.RCB  6.DC  7.KXIP  8.SRH

Input:1

Enter Away Team
1.CSK  2.MI  3.KKR  4.RR
5.RCB  6.DC  7.KXIP  8.SRH

Input: 8

Time Session
1.Afternoon
2.Evening
2

Want to predict before or after the toss being Tossed?

1.Before Toss
2.After Toss

Note: Tosses are considered important in cricket, as captains decide whether to bat or bowl first depending on the condition of the pitch.
1
-----please wait while we predict the match-----

Prediction:
CSK wins the match with confidence of 64.4
```

Figure 15. CSK vs SRH (without toss)

Tosses are considered important in cricket, as captain makes the decision if the team wants to bowl or bat on the basis of pitch. from above results we can see how the toss is influencing the match outcomes.

```
IPL MATCH PREDICTOR

Enter Home Team
1.CSK  2.MI   3.KKR  4.RR
5.RCB  6.DC   7.KXIP 8.SRH

Input:2

Enter Away Team
1.CSK  2.MI   3.KKR  4.RR
5.RCB  6.DC   7.KXIP 8.SRH

Input: 3

Time Session
1.Afternoon
2.Evening
2

Want to predict before or after the toss being Tossed?
1.Before Toss
2.After Toss
2

Note: Tosses are considered important in cricket, as captains decide whether to bat or bowl first depending on the condition of the pitch.
2
Who won the Toss?
1.MI
2.KKR
2
Toss Decision
KKR choose to bat first
KKR choose to bowl first
1
-----please wait while we predict the match-----

Prediction:
KKR wins the match with confidence of 57.099999999999994
```

Figure 16. MI vs KKR (with toss)

10 Conclusion

From the study there are numerous elements which impact result of any IPL match is observed. Main factors that fundamentally impact any IPL match could be their host group, non-home group, arena, winner of toss and many more. This relatively helped in the calculation of strength. Different ML techniques were handed down for IPL data set which contributed to this study. The data set consists of all the IPL matches that were held from the past 6 years that is from 2014 to 2019. The prepared models were utilized to foresee the result of IPL matches. The T20 cricket has a scope for changeability, because even few balls can totally change the game. IPL was started 12 years back, there were very less number of games played compared to 50-50 and test games. Thus, structuring ML for anticipating game result with a precession of 75% is exceptionally good at this stage.

References

- [1] A. L. Samuel, "Some studies in machine learning using the game of checkers. iirecent progress," in Computer Games I, pp. 366–400, Springer, 1988.
- [2] A. Bandulasiri, "Predicting the winner in one day international cricket," Journal of Mathematical Sciences & Mathematics Education, vol. 3, no. 1, pp. 6–17, 2008.
- [3] Indian Premier League Official Website
- [4] P. Langley, W. Iba, K. Thompson, *et al.*, "An analysis of bayesian classifiers," in *Aaai*, vol. 90, pp. 223–228, 1992.
- [5] S. Kampakis and W. Thomas, "Using machine learning to predict the outcome of English county twenty over cricket matches," *arXiv preprint arXiv:1511.05837*, 2015.
- [6] L. Passfield and J. G. Hopker, "A mine of information: can sports analytics provide wisdom from your data?," *International journal of sports physiology and performance*, vol. 12, no. 7, pp. 851–855, 2017.
- [7] T. H. Davenport, "What businesses can learn from sports analytics," *MIT Sloan Man-agement Review*, vol. 55, no. 4, p. 10, 2014.