

# Predicting the winner of an IPL Match, and Best Players for Intelligent Team Selection

Sai Pranav M  
Computer Science and  
Engineering  
PES University  
Bangalore, India  
saipranavmadineni@gmail.com  
m

Rohan Bennur  
Computer Science and  
Engineering  
PES University  
Bangalore, India  
rohan.bennur@gmail.com

**Abstract**—All real-life sports contain a vast amount of statistical information regarding individual players, team, games and seasons. Cricket is one of the most popular sports played in the world. Each team has a desire for winning and choosing the best playing eleven has been the most challenging task. Indian Premier League (IPL) is being played for the past 10 years. Abundant data is being gathered and mining over this data can lead to some hidden patterns in understanding the likelihood of a team winning and in also selecting the best playing eleven of a team. In addition, years of experience, current form, the impact of the players, batsmen available and the number of runs to be chased can have an influence on the outcome; hence these factors are likewise expected and compared. Hence, a model can be developed which can predict the outcome of the match at various stages of the game. State-of-the-art data mining and Machine Learning techniques will be utilized in the making of the model. The performances of data mining techniques such as Artificial Neural Networks like CNN's and DNN's will be implemented using various different metrics at different stages of the game and accuracy of the models can be studied. These can be extensively useful in predicting individual Player's Performance.

**Keywords:** Cricket, IPL, Machine Learning, Artificial Neural Networks, CNN's, DNN's

## I. INTRODUCTION

Cricket is an outdoor sports game played between two teams and the team that scores maximum runs wins the game. This game is played at domestic and international levels and is played across three different formats: one day international (50-over match), Test match (5-day match) and T20 (20-20) format respectively. Cricket is followed and loved by more than a billion people all around the world. Hence there is a need for a data mining model to make prediction to predict the winner of the fixture.

The recent T20 format has gained huge recognition through-out the globe. India winning the inaugural T20 World Cup in 2007, led to a massive foundation for the introductory edition of Indian Premier League in 2008. This league was followed by the immense population due to the fast-paced fixtures. Out of all domestic leagues around the world, IPL is leading in terms of money, entertainment, footfall, popularity, number of views, etc. After the enormous success of IPL, various other countries also started a similar type of franchise league. In IPL, players from various nationalities feature in a different franchise. Selection of the players depends upon the auction which is carried prior to each season. The franchise with the highest bidder for a particular player gets the ownership to feature that player. There is no upper limit for bid price but the overall budget is limited.

In this project, prediction of a match is considered before the start of the game depending upon the impact-factor of each player, playing in that respective team. The winning prediction would be done after every 5-overs of the second inning utilising various algorithms of the Artificial Neural Networks. Player impact factor can be formulated by considering their previous performances. This information comprises of factors such as batting strike rate, runs scored, 30+ scores, bowling economy, wickets taken and many others. The overall impact factor of that team can then be formulated by taking a mean of all the impact factors of the featuring 11 players.

Use of Artificial Intelligence, machine learning, deep learning and data science makes life easier in every aspect. Use of machine learning and predicting the outcomes before the match actually played will allow the players as well as the coaches to analyse the improvement areas. Machine learning is booming and firmly identified with (and frequently covers with) computational insights, which also focuses on prediction-making through the use of technology. It has solid connections to numerical improvement; which hypothesis conveys strategies and application areas to the field. Machine learning is some of the time conflated with data mining where the latter subfield concentrates more on exploratory information analysis and is known as supervised learning. Evaluating the performances of players is not a straight-forward task. Manually analysis of all the past record of each player is practically impossible. So intelligent system to predict the performance of the players based on their past record can be helpful for team management and team selectors.

## II. CRICKET STATISTICS

Statistics mainly fall in two categories: Batting and Bowling. These two statistics have most influence on final outcome of match.

### A. Batting Statistics

All Statistics related to batting are:

1. Innings: The number of innings player played.
2. Not Outs: The number of times the batsman remains not out.
3. Runs: The number of runs scored by batsman.
4. Highest Score: The highest score ever made by the batsman in the past.
5. Batting Average: The batting average of batsman.

6. Centuries: The number of times batsman score 100 runs.
7. Half-Centuries: The number of times batsman score 100 runs.
8. Balls Faced: The total number of balls faced by batsman.
9. Strike Rate:  $\text{Strike Rate} = (100 * \text{Runs}) / \text{Ball Faced}$
10. Run Rate: The average number of runs scores by team per 6 balls.

#### B. *Bowling Statistics*

Statistics related to bowlers are:

1. Over: Six balls bowled by bowler is an over.
2. Maiden Overs: The number of maiden overs (In which the bowler conceded zero runs in an over) bowled.
3. Wickets: The number of wickets taken by bowler.
4. No-Balls: The number of no-balls bowled by bowler.
5. Wide: The number of wide balls bowled by bowler.
6. Bowling Average: The average number of runs conceded per wicket.
7. Strike Rate: The average number of balls bowled per wicket taken.
8. Economy Rate: The average number of runs conceded per over.
9. Five Wickets in an Innings (5w): The number of innings in which the bowler took five wickets or more.
10. Ten Wickets in a Match (10w): The number of matches in which the bowler took ten wickets.

Apart from batting and bowling there are other statistics as well which affect the final outcome of cricket match. Other important statistics are: Toss Win, Weather, Partnership, Format of the match, Venue, Pitch condition, Age of the player, right-handed or left-handed, Bowling Speed, Play in Pressure, Opponent team, Importance of match (e.g. Final Match is important than other match), One day or Day night match, Injury.

### III. LITERATURE SURVEY

#### A. *Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach [1]*

In this paper, the researchers have used three features i.e. Toss, Venue, Strength A/B (relative strength of team A wrt to team B) to detect a match winner.

They use 4 models here for prediction of a match winner. The final weights of the models are chosen such that top 6 performing batsmen and bowlers match with predicted top 6 players.

Here they've distributed the training dataset to be all the matches played between 2010 and testing dataset to be all the matches played after 2014.

SVM, Random Forests, Logistic Regression, Decision Trees and kNN have been used for binary classification.

They claim that the kNN algorithm yields better results as compared to other classifiers.

Takeaway-This paper conveys relying completely on the historical data is not only insufficient, but also fallacious since it does not portray the current competence of a team.

#### B. *Prediction of Live Cricket Score and Winning [2]*

In this paper a model has been proposed that has two methods, first predicts the score of first innings not only on the basis of current run rate but also considers number of wickets fallen ,venue of the match and batting team. The second method predicts the outcome of the matching the second innings considering the same attributes as of the former method along with the target given to the batting team. These two methods have been implemented using Linear Regression Classifier or Q-Learning base decision tree approach and Naïve Bayes Classifier for first innings and second innings respectively. In both methods, 5 over intervals have been made and at each interval above mentioned attributes have been updated. The dataset consists of all the matches played between 2002 and 2014 of every team independently.

Takeaways: The dataset used is dynamic , it is updated after every 5 over interval.

Using prediction like Naive Bayesian algorithm, they've also predicted the best 11 players from each team which can be used on online platforms like fantasy leagues for getting maximum points.

#### C. *The Use of Data Mining for Basketball Matches Outcomes Prediction [3]*

This research paper has been chosen since it uses binary classification algorithms that is required for sport result prediction.

The problem intended to solve is the outcome of the match and spread.

Spread represents and advantage in the number of points given to one of the teams in order to equalize their chances for victory.

For each match two groups of attributes have been used . The first group consists of standard basketball statistics which are 141 in number and the second group of attributes consists of information about league standings. Fresh information about the games that were played previous day, as well as new statistical data for all teams were downloaded and using this model calculates the prediction for future games.

For development of classification and regression model RapidMiner is selected as a tool.

The Naive Bayes algorithm and multivariate linear regression is used in predicting the outcome and spread for a team respectively.

Takeaway-This model used 141 attributes in the first group of attributes which led to increased epochs for the model in reaching 67% accuracy.

#### IV. PROBLEM STATEMENT

Cricket is one of the famous outdoor sports that contain a large set of statistical data in real world. With millions of fans following the IPL matches, the prediction of outcome of a match and the best player of a match before it takes place is a real word problem. The predicted data can then be used to create teams on platforms like Dream11 or MPL.

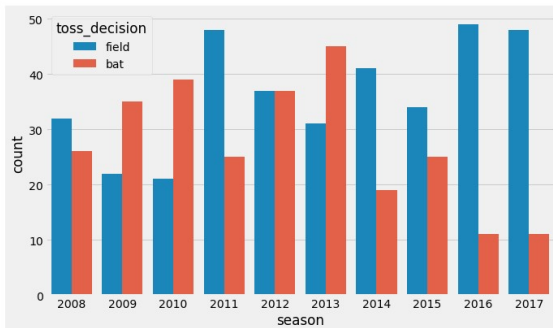
#### V. OBSERVATIONS

The dataset [4] consists of two files, matches.csv and deliveries.csv.

- No missing values were found in both the datasets.
- There were few outliers, which were ignored.

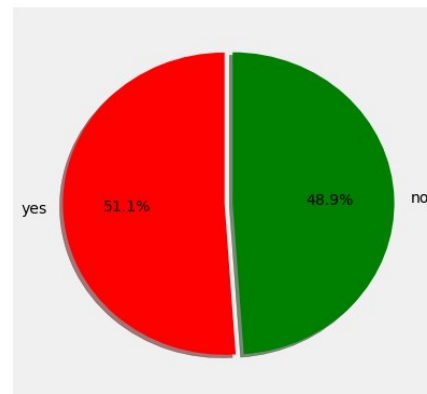
##### A. Toss Decisions across Seasons

The decision for batting or fielding varies largely across the seasons. In some seasons, the probability that toss winners opt for batting is high, while it is not the case in other seasons. In 2016 though, the majority of toss winners opted for fielding. However in the current season of IPL 2020, majority of teams have opted to bat on winning the toss as the venues in Dubai have favored the teams batting first.

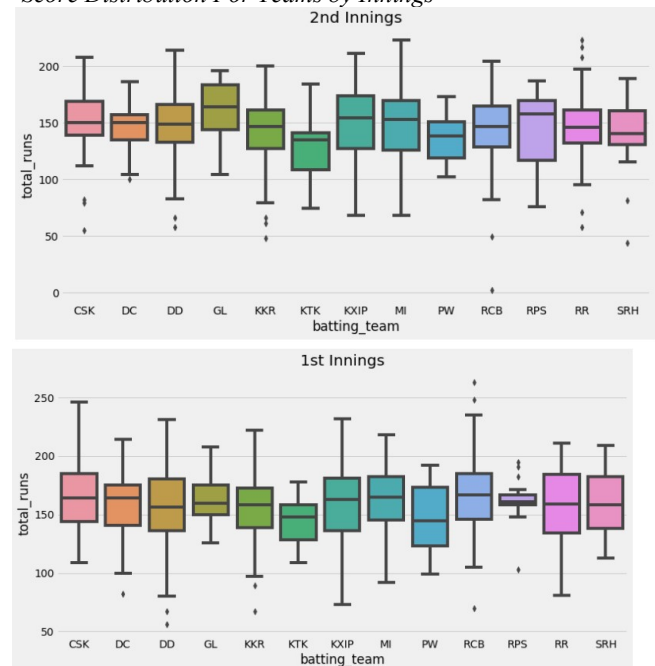


##### B. Toss winner- the match winners?

The toss winner is not necessarily the match winner. The match winning probability for toss winning team is about 50%



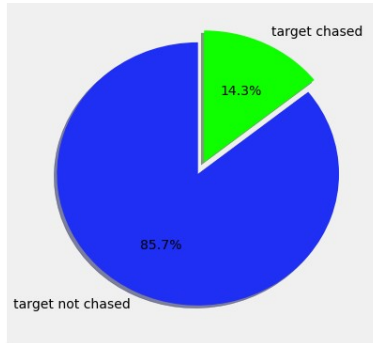
##### C. Score Distribution For Teams by Innings



According to Graph 1 the batting by CSK in innings 1 looks to be the best. Graph 2 also conveys the same story.

In Graph 2 we see a point near 0 which may seem to be outlier. But it is so because the match was disrupted due to rain or other reason.

#### D. Chances of chasing 200+



The above graph shows that batting first and putting 200+ scores on the board has greater chance of winning.

#### VI. PROPOSED SOLUTION

Most of the existing data use only match related statistics and have omitted the venue as a feature. Venue can have a major impact on the result of a match, as the home team will have

an immense support from their fans which will in turn boost their confidence and they'll end up being a better side. The existing studies use only the machine learning algorithms, but have never used Artificial Neural Networks like DNN or CNN. We intend to further improve performance using the above mentioned ANN's.

#### REFERENCES

- [1] Madan Gopal Jhanwar, and Vikram Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach" *MLSA@PKDD/ECML*, 2016
- [2] Rameshwari A. Lokhande and Pramila M. Chawan, "Prediction of Live Cricket Score and Winning", *International Journal of Trend in Research and Development*, Volume 5(4), July 2018
- [3] Dragan Miljković, Ljubiša Gajić, Aleksandar Kovačević, Zora Konjović, "The Use of Data Mining for Basketball Matches Outcomes Prediction", *Intelligent Systems and Informatics (SISY)*, 2010 8th International Symposium, September 2010
- [4] Dataset: [kaggle.com/manasarg/ipl](https://kaggle.com/manasarg/ipl)