# PREDICTION OF MATCH WINNERS OF IPL USING MACHINE LEARNING ALGORITHMS

Dev Karan Singh, Sarthak Agarwal, Sanjeev Gupta, Manisha Singh, Utkarsh Saxena

Department of Computer Science and Engineering, Moradabad Institute of Technology, Moradabad, U.P., India

**Abstract:** Cricket is a popular sport not only in India but also around the world. Specifically T-20 format of this game is become more popular in last few years. Today one of the tournament named Indian Premier league(IPL) related to this format has grown very rapidly. But cricket is always said the game of uncertainty. Prediction of winner of tournament or match has also a concern area of fans or followers.

On the other hand technology is evolving at high rate. Machine learning algorithms are always the first choice of researchers to predict something after training a model. So in this paper we are predicting the winners of Indian Premier League matches using different supervised learning techniques.

Index Terms: IPL, Classification algorithm, Machine learning, Prediction, Support Vector Machine

# I. INTRODUCTION ABOUT IPL

A new style of game has taken the cricketing world by squall. It is an enhancement of the one day game also known as 50-50 over game, to make the cricket fast, furious and instant. It was done to bring new audience of into cricket. It is full of action and appealing to the younger crowd.

The game is played with traditional one-day rules on normal cricket grounds, with some exceptions. The first international game of 20-20(T 20) played in Australia, between Australia A and Pakistan on Jan 13th 2005 at the Adelaide Oval. The game has changed a lot since then and griping the world tightly.

In a short time, T20 became the most popular and lucrative form of cricket, particularly in India. On the same pattern India has launched Indian Premier League (IPL) where huge crowd attend matches and millions more watch on television.

The teams for IPL are selected by means of an auction. Selection of a team from a pool of available players is done by means of auctioning of players. It also involves huge amount of money because of its popularity in India. So it becomes important to form a best combination of players to win the tournament. Out of the thirteen teams that have played in the Indian Premier League since its inception, one team has won the competition four times, one team has won the competition thrice, one team has won the competition twice and three other teams have won it once. Mumbai Indians are the most successful team in league's history in terms of the number of titles won.

Team +
Delhi Capitals
Gujarat Lions
Kings XI Punjab
Kolkata Knight Riders
Mumbai Indians
Rising Pune Supergiant
<b>Royal Challengers Bangalore</b>
Sunrisers Hyderabad
Chennai Super Kings
Rajasthan Royals
Deccan Chargers
Pune Warriors India
Kochi Tuskers Kerala

Fig 1: IPL Teams [1]



Fig 2:Teams Position in IPL[2]

# **II. INTRODUCTION ABOUT MACHINE LEARNING**

Machine learning is a branch of artificial intelligence (AI). It provides systems the ability to automatically learn and improve by experiencing different examples.

SAS [3] defines Machine Learning as a method of data analysis that automates analytical model building. ML is based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Primarily it aims at eliminating the human intervention or assistance by allowing the machine learn and behave automatically and accordingly.

Since 1943 when the first paper came into existence on neural network till date machine learning has changed a lot. Now we are having GPU(Graphics Processing Unit) which enables machines to process high amount of data rapidly and accurately. This advancement in computing in the last few years has made it easy to acquire in many domains. [4].

Traditionally ML can be classified into three categories Supervised learning, Unsupervised and Reinforcement learning.



Fig 3: Types of Machine Learning

In Supervised Learning what has been learned in the past is used to

predict new knowledge using labeled examples. Support Vector Machine, Naive Bayes Classifier, Neural Networks, Nearest neighbors are some of the popular approaches related to it and are in focus of this paper.

When the information used to train is neither classified nor labeled then unsupervised machine learning algorithms are used. Hierarchical, Partitioning, Density and Grid based approaches are popular of clustering which works in unsupervised way.

Reinforcement machine learning algorithm is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Deep Adversarial Networks, Q-Learning, Temporal Difference are few of the popular one.

## **III. MACHINE LEARNING IN CRICKET**

This paper aims at different features of cricket match which can affect the result of a match. In cricket certainly there exist some features which can be focused by a team to win the game. There are various research paper which explore the use of machine learning in sports especially in cricket like Bandulasiri [5],the effect of Duckworth Lewis method [6], Rabindra Lamsal[7] has identified seven factors, Bailey and Clarke work [8] etc. Similarly in different research scholar has shown the results by applying various algorithm like Naive Bayesian model, Artificial Neural Network model etc. All these paper has shown that key features are important to identify for raising the chance of winning.

#### **IV. METHODOLOGY**

We utilize supervised learning algorithms on the grounds that Supervised learning is the place you have input factors (i) and a output factor (j) and you use a calculation to take in the planning capacity from the contribution to the yield. The goal is to deduce the planning limit so well that when you have new info information (i) that you can foresee the output factor(j) for that information. We don't utilize unsupervised learning algorithm in this paper on the grounds that Unsupervised learning is the spot you simply have input data (i) and no contrasting yield factor(j) The target of unsupervised learning is to show the hidden structure or dissemination in the data to take in additional edge about the information.

# A. Data Collection

The Indian Premier League's legitimate website [9] is the essential premise of information for this task. The information was web rejected from the site and kept in the proper configuration utilizing a python library called beautiful soup. The dataset has the sections with respect to coordinate number, IPL season year, where match has been held and the arena name, the match victor subtleties, taking part groups, the edge of winning and the umpire subtleties, player of the match. Indian Premier League was just 11 years of age, which is the reason, after the prepreparing, just 634 matches were accessible. Here, a portion of the segments may contain invalid qualities and a portion of the characteristics may not be required for coordinate champ expectation which is examined in information preprocessing.

#### **B.** Data Preprocessing

No quality data, no quality results. This is the fundamental line behind the success of any algorithm. Because before applying any algorithm we must pre-process the data if we expect best results out of it. There are many problems which may exist in data like incorrect values, Inconsistencies, Incompleteness etc.There are many pre-processing steps to deal with these problems like data cleaning, Data integration, data transformation, Data reduction, data discretization etc. Here, in this progression we have attempted to investigate more in the dataset to discover any oddities present. Each dataset may have certain deformities which must be managed to make it a standard structure for performing counts.

## (i) Data cleaning

There are some invalid qualities in the dataset in the sections, for example, victor, city, setting and so on. Because of the nearness of these invalid

qualities, the grouping is impossible precisely. Along these lines, we attempted to supplant the invalid qualities in various sections with sham qualities.

## (ii) Choosing Required Attributes

This progression is the primary part where we can dispose of certain sections of the dataset that are not valuable for the estimation of game dominating group. This is evaluated utilizing highlight significance. The considered qualities have the accompanying component significance.

## C. Train and test splitting

To make an important preparing set, the issue should be grasped for which it is being settled for. The chief set which is being utilized is the preparation set, the greatest of the two. Running a preparation set through an AI framework tells the net the best way to weigh differing features, transforming them coefficients according to their likelihood of constraining botches in the results. Those coefficients, in any case called boundaries, will be contained in tensors and together they are known as the model, since it encodes a model of the data on which it is being trained. They are the most fundamental takeaways which are procured from setting up an AI framework. The subsequent set is the test set. It fills in as a seal of underwriting, and isn't used until the end. After it is being readied and data is set, the neural net can be tried against this last subjective assessment. The results it produces should support that the net accurately sees pictures, or recollects that them at any rate [x] level of them. In case exact figures are not met, come back to the preparation set and investigate the errors made. Taking the privilege dataset would not make any sort of issue and the framework will work easily.

#### V. EVALUATION MODELS

## A. NAIVE BAYES CLASSIFICATION

Naive Bayes classifiers[10] are a group of basic "**probabilistic classifiers**" in view of applying Bayes' hypothesis with groundbreaking (Naive) autonomous presumptions between the highlights. Study has shown that a simple Bayesian classifier, naïve Bayesian classifier, has comparable performance with decision tree and selected neural network classifiers. Naive Bayes classifiers are profoundly adaptable, requiring various boundaries straight in the quantity of factors (highlights/indicators) in a learning issue. Most extreme probability preparing should be possible by assessing a shut structure articulation, which takes linear time, instead of by costly iterative guess as utilized for some different kinds of classifiers. Another perspective about Naïve Bayes classifiers is that this technique utilizes the "naive" idea that all highlights are disconnected.

#### **B.** SUPPORT VECTOR MACHINE

A support vector machine (SVM), which can be utilized reciprocally with a support vector arrange (SVN), is additionally viewed as a regulated learning calculation. SVMs work by being prepared with explicit information previously sorted out into two distinct classifications.

Thus, the model is built after it has just been prepared. Moreover, the objective of the SVM strategy is to separate which classification any new information falls under. The supreme objective is that the SVM will discover a hyperplane that separates the dataset into two groups [11].

Additionally, the benefits of utilizing the SVM technique are that it will, in general, be exact and performs amazingly well on datasets that are littler and more compact. Likewise, this procedure is entirely adaptable since it tends to be utilized to order or even decide numbers. Additionally, bolster vector machines have the capacity to deal with high dimensional spaces and will in general be memory.

# C. RANDOM FOREST CLASSIFIER

Random Forest classifier is a strategy utilized for relapse and characterization procedures. In the Random Forest Classifiers, to order another case, there are number of trees in working arbitrarily in a forest putting input vector down and obligation of each tree is to give a class name or target variable as a decision in favor of the class. Furthermore, which node has most noteworthy votes will be picked by Random Forest Classifier. To expand the precision anticipated and to control the overfitting. [12]

#### VI. RESULT

According to the comparative study of different algorithms like naive bayes with 61% efficency, linear SVM with 64% etc., algorithms provides us the best result with an efficiency of 61% to 75% it means almost 61 to 75 out of 100 prediction made by the system on the test set are correct. Rest all the other algorithms used in comparative study are discarded because their score on the text is less than 50% after analyzing the data a graph is also plotted which show the winning prediction of each team in a particular season.



Fig 1.4: Algorithms Efficiency

#### VII. CONCLUSION

Keeping in mind the popularity of Cricket, specially IPL in India, and effectiveness of different machine learning algorithms in prediction of different aspects we have to identify which team will win the match. For that purpose a comparative study of different algorithms on different factors are generated with a number of features of a team like which team win the toss, which team choose to bat first and which team bowl first etc. On these factors different algorithms model will be trained but only three algorithms will provide the satisfactory results.

VIII. FUTURE DEVELOPMENTS

Cricket is always said a game of Uncertainty. The system can extended with new features like player recommendation for strategic lead after analyzing the opponent team on various factors like the player of the opponent team and their abilities, climatic conditions, play grounds etc can be used for player recommendation. One more point which can enhance the efficiency of model is addition of more sample set or new Indian Premier League season data.

## REFERENCES

- [1]. https://en.wikipedia.org/wiki/List\_of\_Indian\_Premier\_Leag ue\_seasons\_and\_results
- [2]. http://statisticstimes.com/sports/indian-premier-leaguewinners.php
- [3]. https://www.sas.com/en\_in/insights/analytics/machinelearning.html
- [4]. https://en.wikipedia.org/wiki/Machine\_learning
- [5]. Bandulasiri, "Predicting the winner in one day international cricket," Journal of Mathematical Sciences & Mathematics Education, vol. 3, no. 1, pp. 6–17, 2008.
- [6]. F.C. Duckworth, A.J. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches," Journal of the Operational Research Society, Vol. 49 (3), pp. 220-227, 1998.
- [7]. Rabindra Lamsal, Ayesha Choudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Classification Based Machine Learning Algorithm".
- [8]. M. Bailey and S. R. Clarke, "Predicting the match outcome in one day international cricket matches, while the game is in progress," Journal of sports science & medicine, vol. 5, no. 4, p. 480, 2006.
- [9]. https://www.iplt20.com (Official website of Indian Premier League).
- [10]. P. Langley, W. Iba, K. Thompson, et al., "An analysis of bayesian classifiers," in Aaai, vol. 90, pp. 223–228, 1992.
- [11]. C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp.273–297, 1995.
- [12]. T. K. Ho, "Random decision forests," in Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278–282, IEEE, 1995.