# IPL Win Prediction System To Improve Team Performance using SVM.

Siddharth Sinha, Prakash Tripathi, Aakash Vishwakarma, Prof. Amruta Sankhe

*Department of Information Technology, Atharva College of Engineer*

*Abstract*

*Machine learning and data analytics are trending field currently with rise in huge amount of data generated and sophisticated algorithms developed. One of the fields of machine learning is predictive analytics, where probability of a particular outcome in the future is predicted, based on their historical data. Cricket is one of the most popular sports globally, especially in India. Huge amount of resources, effort, and time is spent to organize cricket matches and leagues, such as IPL (Indian Premier League). Thus, with large stakes involved, it creates enormous pressure on players, coaches and team management to perform well. Hence, in this paper, we worked on creating a win prediction system, which predicts the probability of a team winning in a particular match, through various parameters (features), based on their historical matches. The input parameters are changed, so that the highest probability of winning can be obtained, in that particular match. Thus, it helps team captain, coaches, and management to choose those parameters (players), in that match, to increase their win probability. Also, the strengths and weaknesses of the order of the bowling and batting of the team are obtained to improve team performance.*

## 1. INTRODUCTION

Cricket is one of the most viewed sports in entire world. Many natural factors affecting the sport, enormous media scope, and a huge betting market have given strong reason to model and train the sport from various aspects. Although the complex and different rules determine the sport, the power of players and their performances on a given day, and many other parameters play an important role in influencing the best result of a match. Because the technology is improving at a faster speed and also the large market in betting and huge demand for cricket has influenced the common people to use machine learning calculations to predict the results of cricket matches. Use of machine learning and data science makes life easier in every aspect, using machine learning and predicting the outcomes before the match will help the players and coaches to analyze the weak areas. We adopt several machine learning concepts and algorithms so as to predict the winner of the cricket match. And machine learning is booming and machine learning is firmly identified with (and frequently covers with) computational insights, which also focuses on prediction-making through the utilization of technology. Its solid connections to numerical improvement, which conveys strategies, hypothesis and application areas to the sector. Machine learning is a few of the time conflated with data processing where the latter subfield concentrates more on exploratory information analysis and is understood as supervised learning.

With the progression in innovation and additionally in sports, predicting the results of a match has evolved to be so basic. Cricket is one among the most popular and most watched team games in the world. We are predicting the results of a T20 (Twenty twenty) cricket matches using machine learning concepts like supervised learning to predict the champions of the matches. We utilize career statistics and also the team performances like batting and bowling performances so as to train the models. However, the unpredictable rules governing the sport, the capacity of players and different parameters play a necessary part in influencing the results of the match. Therefore we are using supervised learning algorithms to predict the end result of the sport and it will help the coaches of the team to understand and analyze where actually the team is going wrong.

## 2. LITERATURE REVIEW

Daniel et al. [1] have examined many features to predict the match winner before start of the match by training the selected features on machine learning models. To achieve this they have

applied many machine learning algorithms on test and training dataset of different sizes which are Random Forest, SVM, Naive Bayes, Logistic Regression and Decision Tree. This system will help board of cricket like evaluating the team's strength and cricket analysis.

Jhanwar and Vikram [2] have made predicting model for the outcome of a One Day International (ODI) match by using a supervised learning approach. They have used statistics of each active player as well as recent performance of each player. Thus, they have finalised KNN as their preferred algorithm for classification.

Rameshwari Lokhande and P.M.Chawan [3] In this paper, match outcome prediction is done while the match is in progress i.e. live match prediction. Thus, features such as number of wickets fallen, match venue, team ranking, pitch report, home ground advantage, etc, are considered, rather than pre-match features such as past player performance, past team experience, etc. Thus, this paper discusses about various cricket elements gathered from various research papers related to cricket match outcome prediction.

Jaishankar and Rajkumar [4] had predicted the outcome of a One Day (ODI) cricket matches using machine learning concepts such as supervised learning. This will also identify the champs of wrong and area of improvement in the model so as to increase the accuracy of the model. So in these paper we had used various types of machine learning algorithms and comparison has been made in order the identify the best trained model with best accurate algorithms. We had considered the past statistics of matches along with team credibility in terms of performance like batting, bowling, etc. However the unpredictability of the match result will be there at some extent due to capacity of the players in terms of performance, different types of parameters. So in these paper we had trained model with help of four different types of algorithms and also comparing amongst them so as to get best and desired outcome with highest accuracy

Rory P.Bunker and Fadi Thabtah [5] had made a prediction model with numerous features which is very important because higher the quality of features will increase the accuracy of the model. Large amount is invested which made necessary to build more accurate model. Analyzing and made predefined strategies so as to build classification model which can help various owners , selectors by deploying best accurate model has been done . There are various features present but out which gives more accuracy to the prediction model will be considered like past data of historical matches, individual player data, opposition information. This paper gives a analytical inspection of the composition in ML, focusing of application of Artificial Neural Network (ANN) to sports result prediction. While performing so we had noticed the data sources, learning approach utilised, relevant means of model implementation and particular issues of predicting match outcome. This paper will help to propose sport prediction model in which machine learning strategy can be used.

Sandesh et al. [6] in this paper performance of the players in terms of the strength and weaknesses is used to predict the match outcome. This will help the coaches, captains of the team, selectors and managers by considering various statistical measures of the players. They have used supervised machine learning methods such as SVM with multiple kernels such as linear, nonlinear and RBF to predict match outcome against particular opponent . K-means have been used to find 3 to 4 best nearest similar players to a particular player. Comparison has been done for players in each order such as top order , middle order and lower order.

Akhil et al. [7] has made a prediction system for a T20 matches in a particular IPL matches in dynamic nature (i.e. during match is in progress). Currently the score of first innings can be predicted on the basis of current run rate which can be evaluated by the basic formula however it does not include various features like home ground ,venue, toss etc which is too important during prediction because it can change the outcome of the match and also there is no prediction model to predict the outcome of the second innings of the match. In this system they have tried to predict the each inning of the match by taking all the features which is needed by using Multiple Variable Linear Regression along with Logistic regression and finally predicting the outcome of the match using Random Forest algorithm.

Tejinder Singh, Vishal Singla, Parteek Bhatia [8] had built up an expectation model with numerous highlights to build the adequacy of the model . The model has two strategies, first strategy is to predict the score of first inning by taking the current run rate, number of wicket fallen, batting group and scene of the match. The subsequent strategy predicts the result of the match in the subsequent innings considering indistinguishable qualities from of the previous technique. For first inning Linear Regression Classifier is used and for second inning  Naïve Bayes Classifier is used. In the two strategies the 50 over match has been divided into 5 over each and thus each Is has been recorded which has been played between 2002 to 2014.

Stylianos Kampakis, William Thomas [9] point of this examination was to understand how predictable cricket matches are. The objective challenge was the English twenty over area cricket cup. The first highlights close by designed highlights offered data of around 500 team and their player insights. The model has been trained based on two process i.e. firstly considering team features only and then considering both team and player features. The working of the models was tried over individual seasons from 2009 to 2014 having   been prepared over past season information for each situation. The ideal model was a straightforward to predict  the outcome of the match combined with consideration of  complex hierarchical features.

Padma et al. [10] focuses on Indian ODI cricket matches and used Market Basket tools, with attributes for mining various association rules. They focussed on various elements like toss result, teams, home advantage, etc, to apply the rules. They used 10 years of past data of matches and calculated their support and confidence measures to predict cases which lead to team loosing.

## 3.  DESIGNING THE DATASET

The main source of past matches data is official website of Indian Premier League. The data is scrapped from this official site and stored in a Comma Separated Values (CSV). Data of batsmen and bowlers played in previous years of IPL are collected.. Pre-processing of data was performed on both bowling and batting datasets for modelling the system. Then the data was split into 80% training data for the modeling purpose and 20% validation data for the testing purpose.

## 4.  ASPECTS TO FORESTALL CRICKET WINNER

To win a match, it depends on multiple factors such as batting, bowling, fielding, team performances and player performances. Factors considered for batting are shown in table 4.1, and factors for bowling in table 4.2. Here 'B' represents "Binary", 'I' represents "Interval" and 'N' represents "Nominal".

**Table 4.1: Batting features**

| Features | Type | Feature Description |
|----------|------|---------------------|
| All Rounder | B | It defines the player can perform in all three roles i.e. as a batsman, bowler and fielder. Values can be 1 or 0. 1 indicates player is all- rounder and 0 otherwise. |
| Innings | I | How many innings the players has played not matches. |
| Not Out | I | How many times a player has been not-out in his entire career |
| Runs | I | Total how many runs the player has made in his T20 career |
| HS (highest score) | I | Highest runs scored in a particular  inning by the player |
| Average | I | Avg. of a player depends on no. of instances he has been out with respect to |

| | | no. of innings he has played |
|---|---|---|
| Strike Rate | I | It represents an average of how many runs that players has scores for every 100 balls that they have faced. |
| Century | I | How many centuries are made in T20? |
| Half Century | I | How many half centuries are made in T20? |
| 4s | I | How many 4's the player have hit in the entire innings of T20. |
| 6s | I | How many 6's the player have hit in the entire innings of T20. |
| Result | B | It shows does the player is selected to play IPL or not. Here the selected status is indicated by 1 and not selected status is indicated by 0. |

**Table 4.2: Bowling Features**

| Features | Type | Feature Description |
|---|---|---|
| Innings | I | How many times the player has been bowled in terms of innings |
| Balls | I | How many balls bowled |
| Runs | I | Total How many numbers of runs conceded |
| Wickets | I | How many wickets taken |
| Average | I | Average number of runs conceded divided by total no. of wicket he has taken |
| Team | N | Players team |
| Economy | N | Average number of runs conceded for each over bowled |
| SR | I | Total How many balls bowled / wickets taken by bowler. |
| 4w | I | In how many innings the bowler grabbed as a minimum four wickets |
| 5w | I | In how many innings the bowler grabbed as a minimum five wickets |
| Result | B | It shows does the player is selected to play IPL or not. Here the selected status is indicated by 1 and not selected status is indicated by 0. |

## 5. METHODOLOGY

Support vector machine (SVM) is used to predict the match outcome. It is a supervised machine learning classification technique that builds a plane separating the positive and negative data points. Linear and nonlinear SVM kernels are used to predict the match result as a B category label. The prediction classes are the probability of win (W) or loss (L) by a team in a match. SVM creates a feature vector with finite dimensions, and each dimension represents a feature obtained from previous matches. SVM then trains the model on this feature vector generated. Batsman and bowlers are ranked on their various parameters and thus a ranking index is created, on which then SVM is trained. The feature vector consists of six categories, four for batting, and two for bowling. Batting category includes "opening", "top order", "middle order", and "tail enders". Bowling category includes

20

"pacers" and "spin". This feature vector is created for both the teams, and cumulative ranking is created for each category by subtracting the feature vector of a team from the feature vector of opponent team, in same category. Through this project, captain, coaches and other stakeholders select the players of both the team, according to their batting and bowling order and a probability is predicted of their win or loses.

Steps of the project:
1. Scrape the data from epncricinfo website.
2. Process the unstructured data obtained by scraping, into structured format (CSV), and store it into database.
3. Using various statistical measures, player performance is quantified and a ranking index is generated.
4. The model then runs on the players selected in both the teams, and prediction of probability of win or lose is generated.
5. Team structure is then analysed and the strongest and weakest slots are shown.
6. Similar players to a particular player are obtained using KNN and K-means, according to their performance statistics.

## 6. BLOCK DIAGRAM
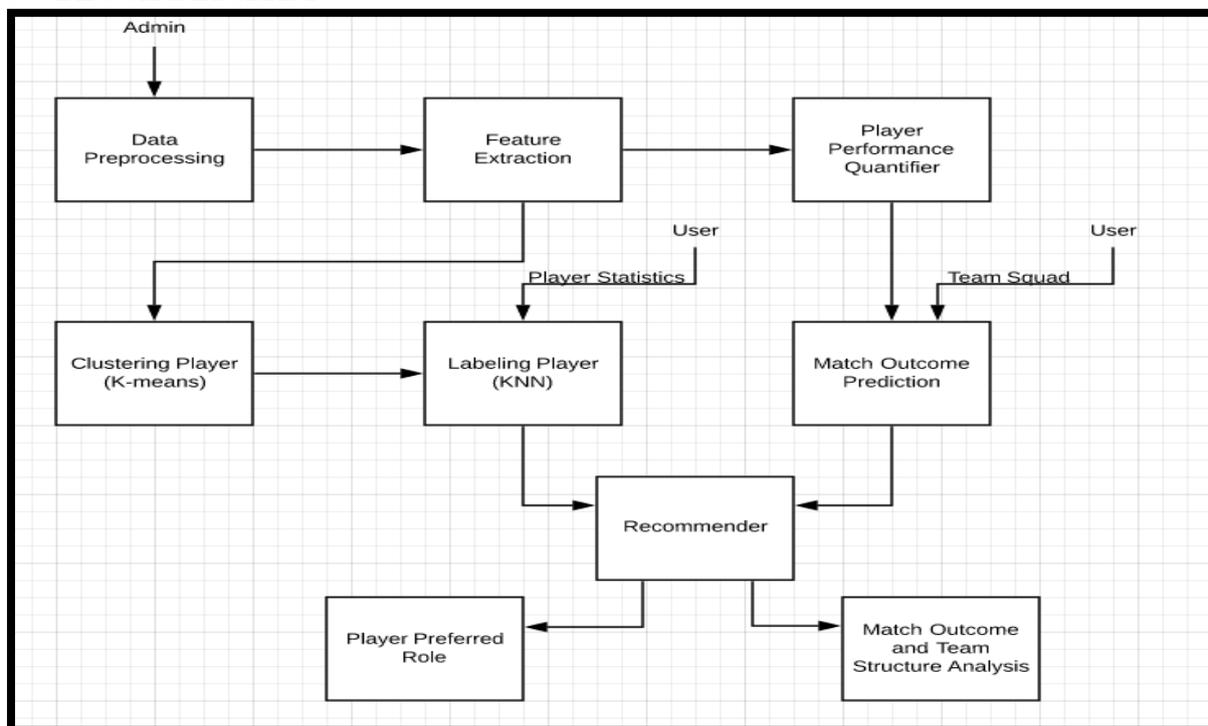


**Fig 6.1: Bowling Features**

The data gathered by scraping is pre-processed, converted in CSV format, and then stored into database. Various feature extraction methods were applied to select only relevant features. Then, player performance metrics were quantified to create a complete database of various player parameters and teams. Match outcome is predicted using SVM (Support Vector Machine) algorithm. The parameters that resulted in highest probability of winning are our recommendation to the team, so that team can improve their chances of winning. Also, K-means clustering algorithm is used to find a most suitable alternative to an injured player. Thus, match outcome and team structure are analysed to understand strengths and weaknesses of the team.

## 7. RESULTS

The accuracy of Naive Bayes classifier is 88.4 %, while that of SVM is 96.3 %, thus we chose SVM as our preferred classifier. The dataset contains 20 percent testing set and 80 percent training set.

**Table 7.2: Model Classification Test Cases**

| Model Classification Test Cases | | | | |
|---|---|---|---|---|
| **Classifier** | **Test Case** | **Precision** | **Recall** | **F-Measure** |
| SVM | Team 1 win | 96.1 | 95.3 | 91.5 |
| | Team 2 win | 96.4 | 94.1 | 93.4 |
| | Match draw | 74.1 | 69.3 | 78.6 |
| Naive Bayes | Team 1 win | 86.3 | 91.3 | 81.7 |
| | Team 2 win | 89.6 | 87.6 | 84.5 |
| | Match draw | 66.4 | 72.5 | 69.7 |
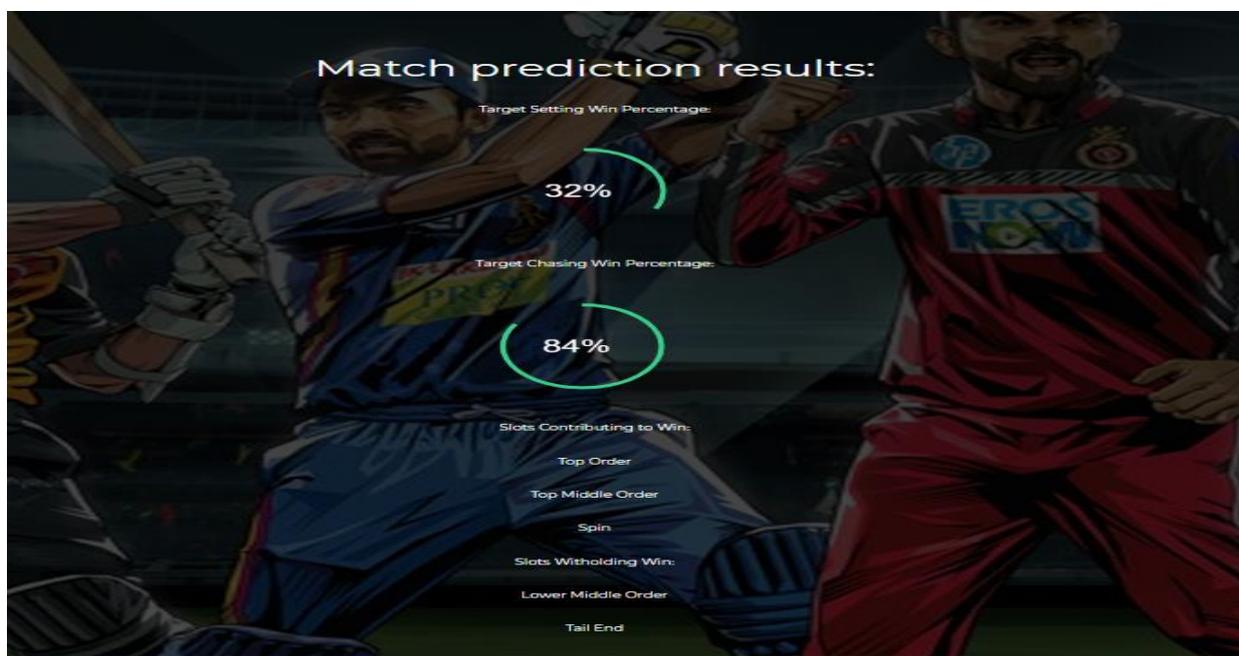


**Fig 7.1: Prediction Output**

## 8. CONCLUSION

This paper has proposed the prediction model. This model is used for predicting the outcome of the match based on historic data. During the extraction of features various features has been involved but most important features has been taken during prediction. We have also made a team structure in terms of slots which defines most important slots contributing to match winning and a ranking system for the players through their performance statistics. We have used K-means to cluster all players according to their performance and KNN (K-nearest neighbour) is used to find interchangeable player to a particular player. SVM model was trained using linear, polynomial and RBF (Radial Basis function). The accuracy, precision and recall rate of SVM and RBF are compared. RBF kernel has performed better compared to linear and polynomial, during testing. This proves that our data set is not separable linearly. Thus, we preferred SVM with RBF kernel for prediction.

**REFERENCES.**

[1] Vistro, Daniel, Leo Gertrude David, "The cricket winner prediction with application of machine learning and data analytics," International Journal of Scientific and Technology Research, Volume 8, Issue 09, 2019.

[2] Jhanwar, Vikram, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," International Institution of Information Technology, Hyderabad, 2016.

[3] Lokhande, Chawan, Pramila, "Prediction of Live Cricket Score and Winning," International Journal of Trend in Research and Development, Volume 5(1), (2018).

[4] Jaishankar, Rajkumar, "A review paper on cricket predictions using various machine learning algorithms and comparisons among them," International Journal for Research in Applied Science and Engineering Technology, 2018.

[5] Rory, Fadi. "A Machine Learning Framework for Sport Result Prediction," Applied Computing and Informatics, Volume 15, Issue 1 2017.

[6] Jayanth, Sandesh Bananki, Akas Anthony, Gududuru Abhilasha, Noorni , Gowri Srinivasa, "A team recommendation system and outcome prediction for the game of cricket," Journal of sports analytics, vol 4, pp. 263-273, 2018.

[7] Akhil, Venkata, Venkatesh, Sai, Chavali, "Cricket score and winning prediction using data mining," International journal of advance research idea and innovations in technology, 2018.

[8] Tejinder, Vishal, Parteek, "Score and Winning Prediction in Cricket through Data Mining," International Conference on Soft Computing Techniques and Implementations, 2015.

[9] Stylianos, William, "Using machine learning to predict the outcome of English county twenty over cricket matches," Cornell university, 2015.

[10] Raj, Padma, "Application of association rule mining: A case study on team India," 2013 International Conference on Computer Communication and Informatics, Coimbatore, India, January 2013.