© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

SUPER PREDICTOR FOR THE INDIAN PREMIER LEAGUE

Rachit Mehul Pathak
rachitmehul.pathak2020@vitstudent.ac.in
VIT University, Chennai Campus, Kelambakkam - Vandalur Rd,
Rajan Nagar, Chennai, Tamil Nadu 600127

Abstract

I decided to work on the Super Predictor for The Indian Premier League and my main objectives was to calculate and tell which factors are majorly crucial and of higher significance for predicting the outcome of a game. Factors such as Most Matches Played In (Stadiums), Most Winning Team in Each Season, Top 5 Teams by win-rate, Net run rate of a team across various seasons, Wins by Batting First / Second, Matches Won After Toss Win, Player with Most Awards and On-Field Umpire with Maximum Matches were considered.

Introduction

T20 cricket was introduced in 2003, after which it gained immense popularity due to its shorter format. BCCI in 2008, initiated the Indian Premier League which attracted the attention of millions of people all over the world. The use of analytical methods in various aspects of cricket is very important.

There is a huge demand for an algorithm that best predicts the results of cricket because of its popularity and huge amount of money involved in the game. Records of the past performance of players and other correlated data can be analyzed to create models that predict the winning team.

The domain of the project is set in the industry of sports and entertainment. The scope of data analytics and machine learning in sports world is very vast. The state-of-the-art machine learning models can outperform old methods and thus, using the algorithms and visual descriptions help in analyzing and formulating outcomes and conclusions.

Sports analysis is done for either the sports teams directly involved in the game or sports betting firms. Also, analysis about the players and their performance statistics can be explained by using data related to any sports or game. This includes the weather conditions, the state of the pitch, the crowd, the toss winner, etc. The main objective is to improve the team performance as well as the moral of the team which in turn leads to better odds of the team winning the game. The competitive nature of the games as well as the high stakes make this a very viable topic to apply data analytics.

Literature Review

[1] This paper reflects upon the changes taking place due to COVID on the IPL and the sponsors. The game was delayed and it had an impact on the national GDP as the season ticket's are a big industry. The sponsors in the post COVID seasons of IPL mostly stayed the same on the pretext of the immense profit it provided. Dream11 was the new sponsor which was taken after COVID. BCCI managed the covid by hosting the games in UAE with the primary source of income as the broadcasting rights and thus it took a hit on the income losing about $2 billion and $1 billion in the salaries of players. Since IPL is a revenue business model, it helped the bigger brands to invest and sponsor the events for better output and opportunity. Similar revenue models are being taken into account for other sport league across the nation. Use of OTT platforms and channels was seen to be increased as a ripple effect.

[2] This paper focuses on the importance of celebrity endorsement on the performance of

© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

teams in various leagues. Indian Premier League contributes to the most of any other league present and the participation and the hype is maximum of any other in India. The celebrity endorsed to a particular team came with the pressure of advertising and the characteristics of VIP. The article conjointly tries to take a gender at young people's recognitions with respect to how this celeb-endorsement works. With a higher advent of much famous celebrity, the team moral was found to be much higher as well as the merchandise of the team as well as fan support also differed. This paper gave us an insight into the fact that just numeric data isn't enough to predict this vast of a case study and many underlying factors are there which have a boost to the team in general.

[3] This paper reflects upon the use of different Machine Learning Algorithms in the Indian Premier League. The models are taken as both prescriptive and descriptive. The visualization of data and the outlines provide with conclusions we can draw from just the descriptive analysis of the data like which team won most games to which batsman scored the most runs. The predictive model takes into account various attributes like the history between the teams and the players, as well as the outcome of the game with respect to scores on a particular ground. This paper gave a key insight into the use of various attributes and relationships which we can take into account and discover the relevance.

[4] This paper determines the attributes and factors that play an important role taking into account the last 3 seasons of the Indian Premier League. The method was the dataset being ran through regression analysis and the t-test. The finalists were supposed to be predicted as the conclusion of the paper. They concluded that there are many underlying factors which are not taken into account in the numerical version of the dataset. With the quantitative and descriptive analysis of the dataset they were able to predict the finals of the Indian Premier League season of 2021.

[5] This paper presents an analysis of the Duckworth-Lewis-Stern (DLS) method for One Day International (ODI) cricket matches. Few classifiers were used to predict the outcome of the game during the second innings of the game and the giant leap of predication was run every 10 overs to check the accuracy of the models. The accuracy of the DLS method is checked in between games versus various algorithms for predicting the winner of the game. The result of a cricket match is predicted during the second inning. A DLS table was created which showed the overs bowled and the predicted outcome being right. The overall prediction's maximum accuracy was found to be higher than 80%. Thus this gives us an insight in how the DLS works and how useful of a concept it is in case of reduced over matches in the Indian Premier League.

[6] This paper gives an account about the various attributes and how they are taken into account while player selection in the Indian Premier League. Cricket, being a very competitive field, having proper account of the players performance around the year and back are very important. Broadly classifying them into Batting and Bowling Skill attributes, we can divide and rank them on the basis of the corresponding performance in the field. After training the model, the features were taken into account for a future year IPL and the results were found to be with an F1 score of 0.8.

[7] This paper refers to the statistical analysis and prediction for the Indian Premier League games. The factors affecting a match have been taken into account with Machine Learning to Predict the outcome of a match. A novel analysis of batting and bowling rating has been proposed and algorithms like SVM, Logistic Regression, Random Forest and Naïve Bayes have been used

© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

and tested on the datasets. Decision Tree was found to have the minimum accuracy and Logistic Regression to have the maximum accuracy. Using the results, we can infer and correlate with our own results and check how models behave differently on different datasets and still predict the outcome of the match at a high rate of accuracy.

[8] This paper provides a model which provides a multi-objective optimized squad selection based on batting and bowling performance. Also, the proposed model tries to formulate a balanced squad by constraining the number of pure batsmen, bowler, and all-rounders in the team. Also, bounds are taken on the star players which provide greater value to the overall squad. The problem was treated as a 0/1 knapsack problem and the algorithms taken into account were combinatorial optimization algorithms. The trade-off squads were formed and the theoretically selected players performed well in IPL 2020 matches.

[9] The following research paper talks about predicting IPL results based on match id, innings, batting team, bowling team, wide runs, byes, leg byes, penalty runs, extra runs, total runs, player dismissed, using machine learning algorithms like Random Forest. The results of the study show that factors such as toss, venues and teams play an important role in deciding the outcome of the game. It has also been shown that right selection of Machine Learning model helps to extend accuracy of prediction.

[10] The tool used presented this paper can be used to predict the performance of players. The tool employs Hbase, an open source distributed non-relational database for storing data. It can be used to successfully predict the outcome of of IPL matches, to predict the performance of a particular team and provide statistical analysis of players based on different parameters. The developed models can help decision makers during IPL matches to evaluate the strength of one team against another.

[11] In the given research paper, it has been proposed that models be created on prediction of the score and a second one is on team winning prediction. The models have been created on the basis that each team consists of 4 International players and 7 Indian players. There are 8 founding franchises. The performance of any team is decided by key player performance and team conditions. The performance of a team's players determines its victory factor. Machine learning models have been used to make the given predictions. Useful applications of the given models are online fantasy games, used by team analyst, which provides stats to cricket lovers and they can also use to access an opponent’s strengths and weakness.

[12] Prediction of outcome of a match using machine learning algorithms is a crucial aspect in cricket. Records of the past performance of players and other related data are often analyzed to make models that predict the winning team. In the given research paper models have been created using algorithms like Decision Tree, Naive Bayes and K Nearest Neighbors. The results are compared using evaluation metrics such as accuracy, prediction, error rate and sensitivity. Random forest is employed for match analysis using factors such as teams playing the given match, toss winners and venues. SVM and RBF kernels have also been used for prediction.

Methodology

The prediction system created in this project was highly accurate. To accomplish this, a dataset with the right columns and very few outliers was chosen. The data was preprocessed using the right techniques by eliminating missing data fields and converting most data fields into numeric data. On the preprocessed data, statistical analysis was conducted using R programming language to find

© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

the trend and possibilities of various factors. Furthermore, Machine Learning algorithms such as Naïve Bayes, Random Forest and K-means clustering were used to create an accurate prediction system. Towards the end of the project, the results were carefully noted down.

Results and Analysis

I. Teams that Won Both Toss and Match

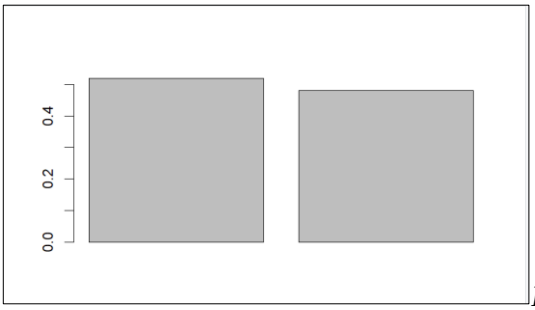

Fig 2: Matches won by teams that won the toss vs teams that lost the toss

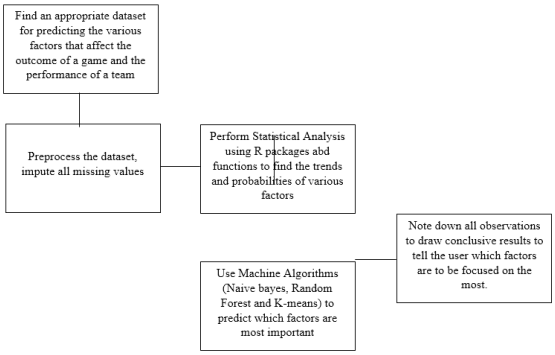

Fig 1: Architecture of Super Predictor System

I. Team with Most Number of Wins

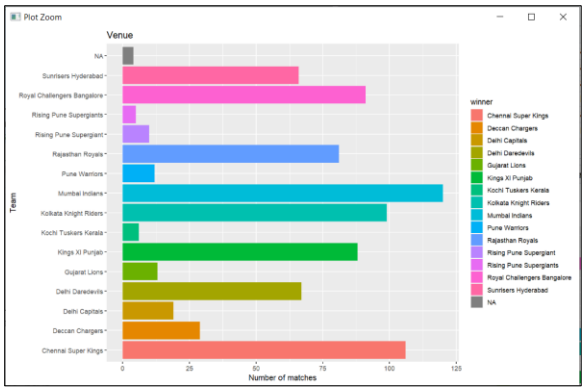

Fig 3: Bar plot for IPL teams with most wins

II. On- field Umpire with Maximum Matches

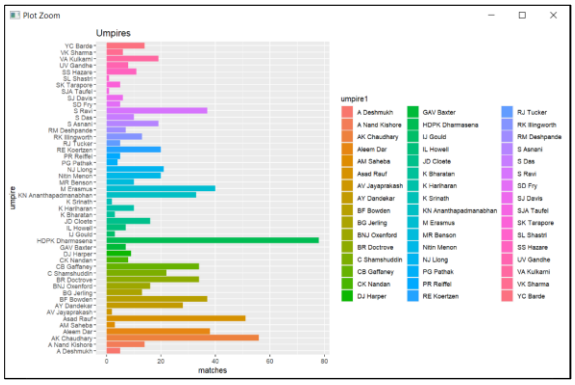

Fig 4: Bar plot for umpires with maximum matches

III. Probability of Winning for a Team that Elects to Bat First vs Team that Elects to Field

P(batting) = 0.392157

P(fielding) = 0.607843

IV. Home Team Advantage

P(win at home) = 0.240530

© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

V. Teams Won by D/L Method

```
> dl
                        winner method
1              Kings XI Punjab    D/L
2            Chennai Super Kings    D/L
3              Delhi Daredevils    D/L
4          Kolkata Knight Riders    D/L
5            Chennai Super Kings    D/L
6           Kochi Tuskers Kerala    D/L
7          Kolkata Knight Riders    D/L
8  Royal Challengers Bangalore    D/L
9            Sunrisers Hyderabad    D/L
10           Sunrisers Hyderabad    D/L
11 Royal Challengers Bangalore    D/L
12     Rising Pune Supergiants    D/L
13         Kolkata Knight Riders    D/L
14     Rising Pune Supergiants    D/L
15 Royal Challengers Bangalore    D/L
16         Kolkata Knight Riders    D/L
17              Rajasthan Royals    D/L
18             Kings XI Punjab    D/L
19             Delhi Daredevils    D/L
> count(dl)
   n
1 19
```

Fig 5: Teams that won by D/L method

VI. K-means for visualising the team that win matches after winning the toss

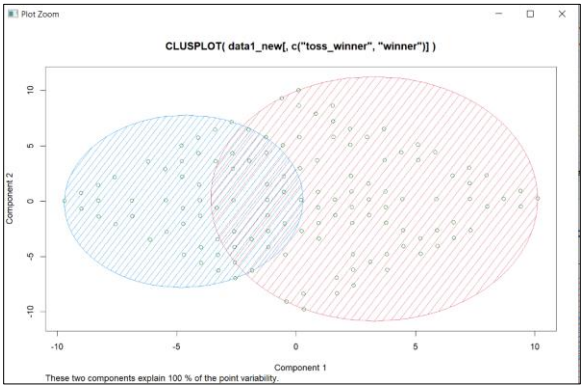

Fig 6: Clustering using K-means

VII. Random Forest

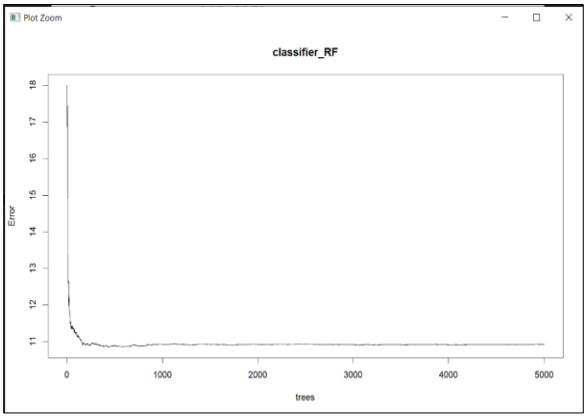

Fig 7:Error vs trees for Random Forest

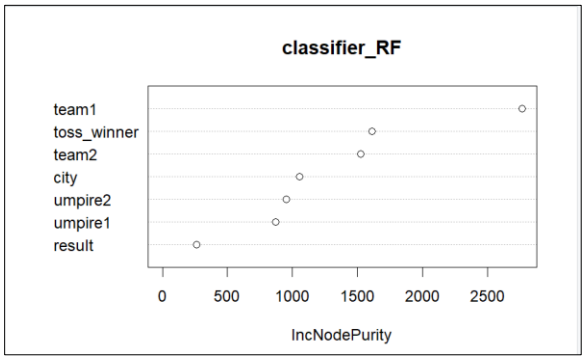

Fig 8: Variable Importance plot

VIII. Naïve Bayes

```
Overall Statistics

               Accuracy : 0.125
                 95% CI : (0.0891, 0.1688)
    No Information Rate : 1
    P-Value [Acc > NIR] : 1

                  Kappa : 0

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
Sensitivity                NA       NA       NA       NA       NA       NA
Specificity            0.8681  0.95833  0.97222  0.90278  0.97917   0.8819
Pos Pred Value             NA       NA       NA       NA       NA       NA
Neg Pred Value             NA       NA       NA       NA       NA       NA
Prevalence             0.0000  0.00000  0.00000  0.00000  0.00000   0.0000
Detection Rate         0.0000  0.00000  0.00000  0.00000  0.00000   0.0000
Detection Prevalence   0.1319  0.04167  0.02778  0.09722  0.02083   0.1181
Balanced Accuracy          NA       NA       NA       NA       NA       NA
                     Class: 7 Class: 8 Class: 9 Class: 10 Class: 11 Class: 12
Sensitivity                NA       NA    0.125        NA        NA        NA
Specificity             0.993056   0.8924       NA   0.96528   0.93403  0.993056
Pos Pred Value             NA       NA       NA       NA        NA        NA
Neg Pred Value             NA       NA       NA       NA        NA        NA
Prevalence             0.000000   0.0000    1.000   0.00000   0.00000  0.000000
Detection Rate         0.000000   0.0000    0.125   0.00000   0.00000  0.000000
Detection Prevalence   0.006944   0.1076    0.125   0.03472   0.06597  0.006944
Balanced Accuracy          NA       NA       NA       NA        NA        NA
                     Class: 13 Class: 14 Class: 15
Sensitivity                 NA        NA       NA
Specificity            0.98958     0.875  0.92014
Pos Pred Value              NA        NA       NA
Neg Pred Value              NA        NA       NA
Prevalence             0.00000     0.000  0.00000
Detection Rate         0.00000     0.000  0.00000
Detection Prevalence   0.01042     0.125  0.07986
Balanced Accuracy           NA        NA       NA
```

Fig 9: Naïve bayes statistics

© INNOVATIVE RESEARCH THOUGHTS
Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

The project was able to distinguish between the major factors that affect the outcome of a given game. We performed statistical and predictive analysis on the given dataset to procure most of the results of the project.

Statistical analysis showed that:

1. teams which won the toss had a higher chance of winning the game.
2. Mumbai Indians was the team that won the greatest number of games.
3. HDPK Dharmasena was the on-field umpire for the maximum number of games.
4. a team had a better chance of winning the game if it elected to field instead of bat.
5. The team playing at its home ground wins atleast 24% of its matches.
6. There was a total of 19 wins through the Duckworth-Lewis method.

Our predictions using machine learning algorithms proved that the factors which affect the game the most include toss decision, city in which the match is played and the opponent team which participates in the match. These predictions were made using the Random forest, Kmeans and Naïve Bayes algorithms.

Conclusion

The main aim was to find which factors play a major role in the outcome of a cricket game, or performance of a team in each Indian Premier League Season. It was also to find which player has a major impact on a season and a given game. It also gave us an elaborate account on which team to focus on in an upcoming game and how this would play a role in the teams being fan favourites or having a better auction season.

With the use of data analysis and machine learning algorithms implemented using R, relationships were able to be determined between the factors and conclude the hierarchy of the factors which affect the outcome of the game in given order.

The factors which immensely affect the outcome of the game are:

1. Toss decision
2. The stadium in which the match is played

References

[1] Dhillon, A. S., & Sharma, R. (2021, January 18). Impact of Covid-19 on Indian Premier League and its Sponsors. Unknown. https://www.researchgate.net/publication/348565977_Impact_of_Covid19_on_Indian_Premier_League_and_its_Sponsors

[2] Dhillon, A. S., & Sharma, R. (2020, October 10). Celebrity Endorsement Effectiveness on Teams of a different Sports League in India. Unknown. https://www.researchgate.net/publication/348352228_Celebrity_Endorsement_Effectiveness_on_Teams_of_a_different_Sports_League_in_India

[3] Akhila, G., Hemachandran, K., & Jaramillo, J. R. (2021, January 1). Indian Premier League using different aspects of machine learning algorithms. Unknown. https://www.researchgate.net/publication/356719085_Indian_Premier_League_Using_Different_Aspects_of_Machine_Learning_Algorithms

[4] Barot, H., Kothari, A., Bide, P., Ahir, B., & Kankaria, R. (2020, June). Analysis and prediction for the Indian Premier League. 2020 International Conference for Emerging Technology (INCET). http://dx.doi.org/10.1109/incet49848.2020.9153972

[5] Abbas, K., & Haider, S. (2019, December). Duckworth-Lewis-Stern method comparison with machine learning approach. 2019 International Conference on Frontiers of

ISSN : 2454 – 308X | Volume : 09 , Issue : 04 | July - September 2023
DOI : https://doi.org/ 10.36676/irt.2023-v9i4-012

Information Technology (FIT). http://dx.doi.org/10.1109/fit47737.2019.00045

[6] Ghosh, A., Sinha, A., Mondal, P., & Saha, P. (2021, January 27). Indian Premier League player selection model based on Indian domestic league performance. Unknown. https://www.researchgate.net/publication/350149401_Indian_Premier_League_Player_Selection_Model_Based_on_Indian_Domestic_League_Performance%5D

[7] Barot, H., Kothari, A., Bide, P., & Kankaria, R. (2020, June 1). Analysis and prediction for the Indian Premier League. Unknown. https://www.researchgate.net/publication/343406867_Analysis_and_Prediction_for_the_Indian_Premier_League

[8] Verma, S., Pandey, V., Pant, M., & Snasel, V. (2022). A balanced squad for Indian Premier League using modified NSGA-II. IEEE Access, 10, 100463–100477. https://doi.org/10.1109/access.2022.3204649

[9] Reddy, K. R., A, P. S., V, C., & Reddy, S. S. (2022). Super Predictor of Indian Premier League (IPL) using Various ML techniques with help of IBM Cloud. International Journal for Research in Applied Science and Engineering Technology, 10(6), 1793–1807. https://doi.org/10.22214/ijraset.2022.43654

[10] Singh, S., & Kaur, P. (2017). IPL visualization and prediction using hbase. Procedia Computer Science, 122, 910–915. https://doi.org/10.1016/j.procs.2017.11.454

[11] Jaswanth, K., Arun K., Deekshith B.D., Ashwik K., Pavithra j (2022). Analyzing and Predicting Outcomes of IPL Cricket Data. International Research Journal of Engineering and Technology (IRJET), 9(3), 1125-1128.

[12] Gawande, K., Harale, S., Prof, & Pakhare, S. (2022). PREDICTIVE ANALYSIS OF AN IPL MATCH USING MACHINE LEARNING. International Journal of Creative Research Thoughts, 10(4).