

ANALYSING AND PREDICTING THE RESULTS OF IPL MATCHES BASED ON VARIOUS FACTORS.

Anusha C
PES1UG19CS074

Computer Science and Engineering
Pes University
Bangalore, India
anusha20c@gmail.com

Thrupthi M S
PES1UG19CS541

Computer Science and Engineering
Pes University
Bangalore, India
thrupthisomashekar@gmail.com

Divya K C
PES1UG19CS147

Computer Science and Engineering
Pes University
Bangalore, India
divyachandrashekhardiv@gmail.com

Abstract — We performs in depth analysis on the matches played during the Indian Premier League (IPL). Cricket, especially the T20 format, has maximum uncertainty, where a single over can change the momentum of the game. IPL Data Analysis is all about the analysing the data that is already present in data set using data science, machine learning and R. This is an application design for the purpose of analysing the data by fetching the attribute from the dataset and predicting the future of the match. EDA will help us to find patterns in data and determining relationships in data and. Visualizing the trends in performance of the team and predicting what can be the results in the next coming matches. This will helps to identify the team that has more chances to identify the team that has more chances to win the upcoming seasons.

In this project we try to evaluate the effects of these major factors on win count:

- 1.LOCATION OF THE MATCH
- 2.TOSS WINNING
- 3.PERFORMANCE IN PREVIOUS MATCHES

I. INTRODUCTION

Cricket is the biggest tournament played in all most all the countries. It is the game between two teams in which each team has 11 players the final result will be either loss or win or at the rare cases points will be shared with both teams which mean no team has lost or won. Sometimes the game is unpredictable because of that game keeps on changing each and every time.

The Indian Premier League (IPL) is a Twenty20 cricket league tournament held in India contested during April and May of every year where top players from all over the world take part. The IPL is the most-attended cricket league in the world and ranks sixth among all sports leagues. The madness of cricket in people is like anything by looking into this, the main objective of our work is

predicting the match result before the game starts based on the past statistic data that is present in the

form of data set. in this the study of Indian Premier League (IPL) is done using that past 12 seasons played till the date

in cricket and IPL, Data Science is used in a somewhat unique and interesting manner. Data analysis is very important in IPL to predict the match result.

You won't believe that IPL teams have started hiring proper companies who are experts in such Data Analysis. **Performance Analytics Companies** that analyze how good players are, and develop strategies for that players. These Data Analysis companies analyze data about players in detail to understand who is good at what aspect.

In one interview, **Virender Sehwag** encapsulated the importance of Data Science very nicely. He said that *"Every game you play, they will record your good performance, your bad performance, you played against which bowler, you scored against which team and which bowler, and the whole data will easily show you that you are good against Pakistan but you're not good performed against Bangladesh, you're good against South Africa but you're not good against England. In 2003 when our computer analytics guy come in and he showed me videos and different kinds of data analysis, I got amazed!!"*

Machine Learning techniques are also used to predict the match results. Different models are created with the help of programming and computers in which, inputs like the position of a player, location of the match, the weather of the day, etc. are all added as variables and on the basis of previous matches, these models predict the future results of the matches. If you provide the data input of the previous matches, such as the venues of the matches as well as teams that played, players that were present as well as the type of players that were present, then in the future it could be predicted the result of the matches presently being played.

Obviously, it will not be 100% accurate but it could be quite useful.

II. LITERATURE REVIEW

In cricket, to predict an outcome of a match, the primary task is to extract out the essentials factors (features) which affect result of a match. Interesting works have been done in the field of predicting outcome in cricket.

Author [1] has analyzed the factors like home field advantage, winning the toss, game plan (first batting or first fielding) and the effect of Duckworth Lewis method [2] for one-day cricket format. Furthermore, Bailey and Clarke mention in their work [3] that in one-day cricket format, home ground advantage, past performances, venue, performance against the specific opposition, current form are statistically significant in predicting total runs and predicting the outcome of a match.

The software used for modelling is Anaconda and Python libraries like pandas, NumPy and IPython to work with the data structure and applying algorithms [3,4]. The main result obtained was based on the impact of toss winner and resultant match winner. The predictive model considered the innings score at regular intervals and the final scores to predict the match result. The model predicted score and run rate projected score were quite near to the final score, in particular the score predicted by the model was more accurate to the actual score. When no feature selection was applied to the dataset the model's accuracy was not satisfactory, i.e. slightly above 50%.

Author [5] experimented with twenty over format cricket matches to predict the outcome using various data mining techniques. The main aim of the study was to combine pre-game and in-game data to predict the outcome. They considered the T20 International match data along with IPL data till 2015 as the training data set. In depth analysis was conducted by segmenting the data on the basis of venue, one team against all other teams, batting first and so on. Decision Tree was applied to predict the match outcome, and produced models with around 78% accuracy for the team that bats first and 75% when it bats second. IG technique was used for feature selection.

Authors [6] performed a comparative analysis of various regression and classification models in prediction of a football game and the results showed that classification-based models outperformed regression-based classification models. In cricket, Author [7, 8] proposed an adoption of h-index and PageRank to rank cricket teams and models to predict best suitable team line-up for a particular game using statistical modeling and network centrality-based approach. Authors [9–10] worked on identifying the role of external factors in the outcome of an ODI. The key features under consideration were home field advantage, winning the toss, game plan (batting first or fielding first), match type (day or day and night), competing team, current form etc. Author [11] used logistic regression technique to explore the statistical

significance of various features and to build a model for result predictions in ODI.

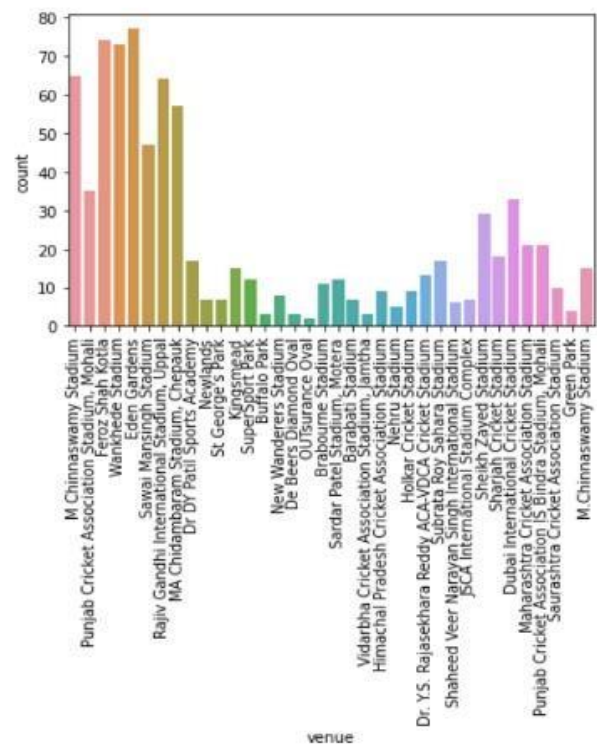
Live Cricket Score and Winning Prediction work [12] describes about the building of the model which predicts the score for the chasing team and will estimate the score of the second innings of match. The proposed work uses the concepts of Linear Regression, Naive Bayes Classifier and Reinforce Learning Algorithm. The factors such as toss result, ranking of the team, home team advantages are considered.

we are building a model for predicting the winning team based on the factors like home match, toss winning, performances in previous matches and team plan (batting or fielding).

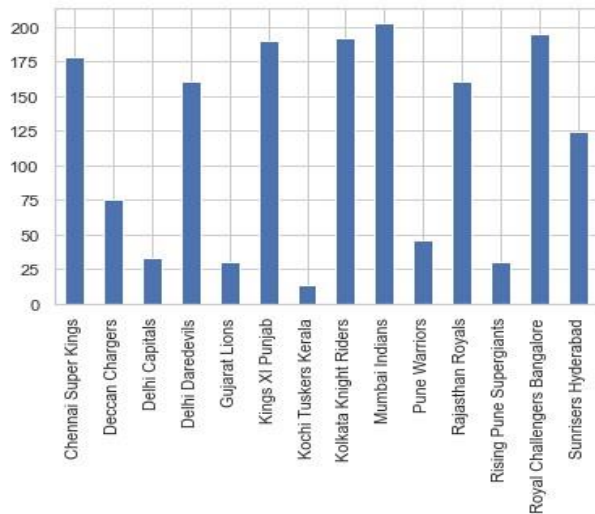
III. INITIAL INSIGHTS /DATA VISUALIZATION

The data of IPL matches collected from 2008 to till 2019. This was analysed and visualized. The dataset had records of matches covering almost 12 seasons. There are two datasets “matches.csv” and “deliveries.csv” are used to analyse and to visualize the data. The source of the data is <https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020>.

We can notice the list of teams that played between the period 2008 to 2020. If you are a pro IPL fan then you will see some old team names on the list which are not playing these days but they contributed some valuable information in the IPL history.

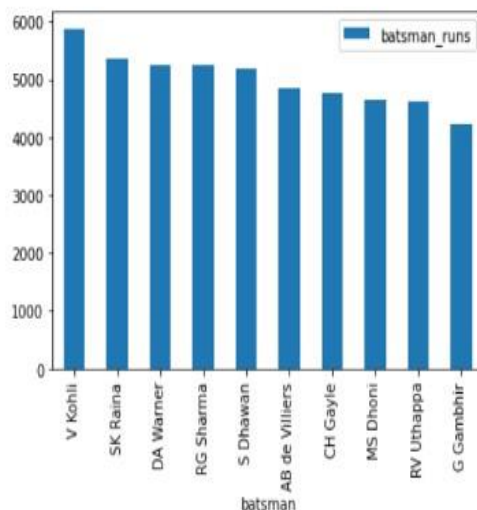


As you see that Eden Garden is the fan-favourite ground of IPL, nearly 80 matches are hosted there.



We count the value of each team playing in column one and add the count value of each team from team two to get the desired output. For example, if CSK played 90 times from team one and 85 times from team 2 then the total of 175 matches are shown in the above graph. As you clearly see that Mumbai Indians played the highest number of matches in the IPL. We grouped the top 10 batsmen from the delivery dataset and sum up their runs. After this plot this information into a graph.

Out[32]: <AxesSubplot:xlabel='batsman'>

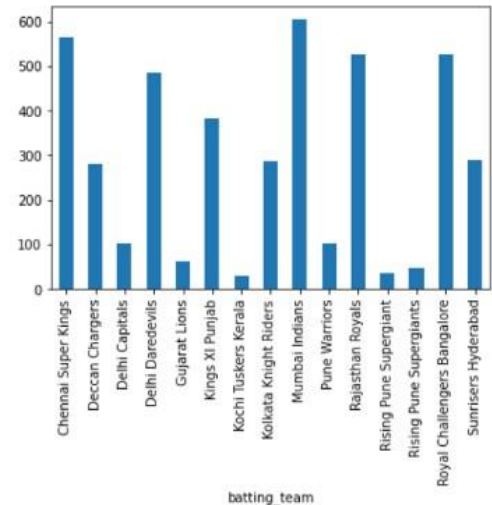


Clearly, King Kohli is at the top followed by Suresh Raina and other batsmen.

The bowler with team-wise performance:

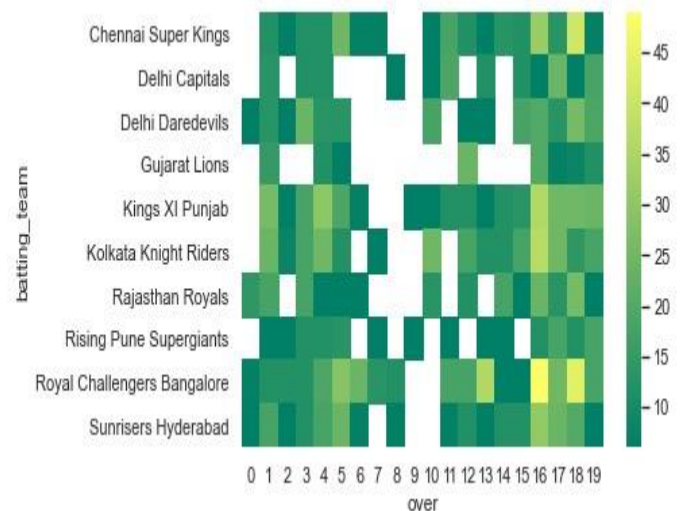
We are taking the example of PP Chawla. This bowler has given the highest number of runs in the IPL history till 2020. We summed up the total runs given by PP Chawla to the opponent team.

Out[36]: <AxesSubplot:xlabel='batting_team'>

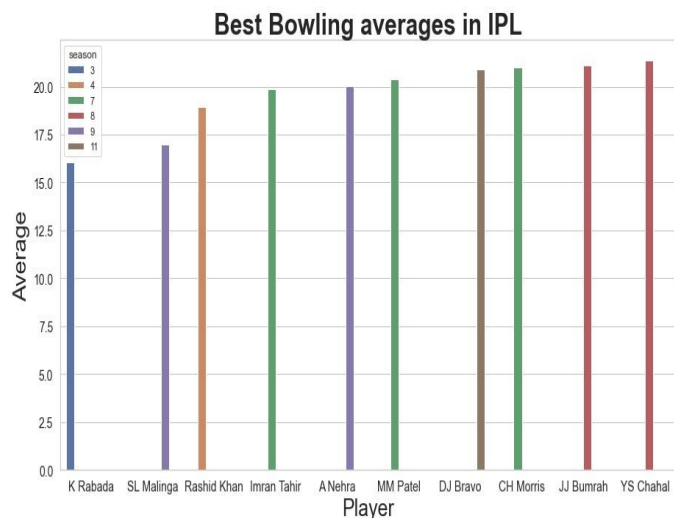


It's clear that if you have PP Chawla in your team then don't let him play against MI, CSK, RCB, RR, and DC.

Over wise batting performance of each team in IPL (2008– 2020)

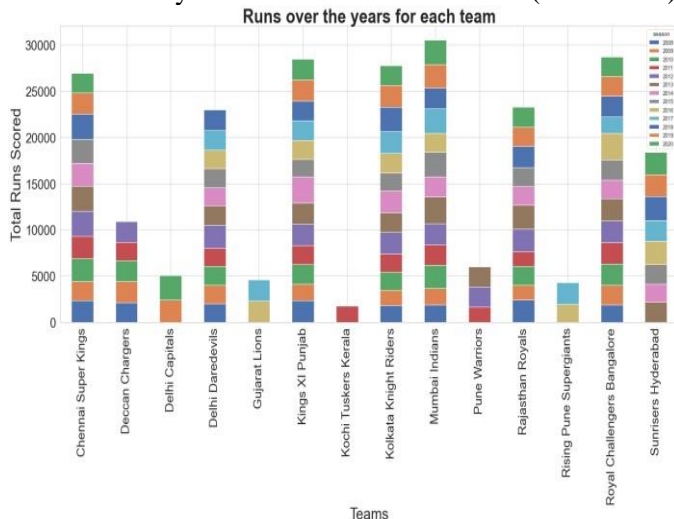


As you clearly see that if you are playing against MI or CSK then you have to play with your best bowling attack line-up from 1st over. MI's batsmen are silent in over 2 and 3 after that they went on rampage mode against their opponent. The same goes for CSK and RCB too. This data is not only helpful from the bowling team's perspective but also the batting team. If you are a team manager and you see using this data that your team is not performing well in the death overs then you probably focus to buy a good finisher in the next auction. As you see from the above heatmap the most of the team is lagging in finish the map, except CSK and MI. Which bowlers have performed well over last few seasons and can be considered further for upcoming auctions?



We can analyse easily from above graph, K Rabada has performed well

Runs scored by all the teams across seasons (2008-2019)



IV PROBLEM STATEMENT:

In this project we try to predict the effect of these factors on win count:

- 1.location of the match (home ground or away ground)
- 2.the toss winning
- 3.performance in previous matches

In this project we try to take into account the different factors like position of a player, location of the match, the weather of the day, etc. are all added as variables and on the basis of previous matches and possible trends in the performance of players individually and their contribution to the performance of a team as a whole. We try to answer the question: We also will try and predict the performance of different teams in the upcoming 2021 IPL tournament by training the model for data available. We can also

answer for the questions like How valuable is the player going to be for the team? And what Stadium and in which weather does a player performs better?

V.REFERENCES

- [1] A. Bandulasiri, "Predicting the winner in one day international cricket," Journal of Mathematical Sciences & Mathematics Education, vol. 3, no. 1, pp. 6–17, 2008.
- [2] F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches," Journal of the Operational Research Society, vol. 49, no. 3, pp. 220–227, 1998
- [3] 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
- [4] W. McKinney, Python for data analysis: Data wrangling with Pandas, NumPy, and IPython, O'Reilly Media, Inc., 2012.
- [5] Munir, F., Hasan, M.K., Ahmed, S., Md Quraish, S., 2015. Predicting a T20 cricket match result while the match is in progress (Doctoral dissertation, BRAC University).
- [6] Delen, D., Cogdell, D., Kasap, N.: A comparative analysis of data mining methods in predicting NCAA bowl outcomes. Int. J. Forecast. 28(2), 543–552 (2012)
- [7] Daud, A., et al.: Ranking cricket teams. Inf. Process. Manag. 51(2), 62–73 (2015)
- [8] Saqlain, S.M., Usmani, R.S.A.: Comment on "ranking cricket teams". Inf. Process. Manag. 53(2), 450–453 (2017)
- [9] Bailey, M., Clarke, S.R.: Predicting the match outcome in one day international cricket matches, while the game is in progress. J. Sports Sci. Med. 5(4), 480 (2006)
- [10] Morley, B., Thomas, D.: An investigation of home advantage and other factors affecting outcomes in English one-day cricket matches. J. Sports Sci. 23(3), 261–268 (2005)
- [11] Bandulasiri, A.: Predicting the winner in one day international cricket. J. Math. Sci. Math. Educ. 3(1), 6 (2006)
- [12] Rameshwari Lokhande and P.M. Chawan, "Live Cricket Score and Winning Prediction", International Journal of Trend in Research and Development, Vol. 5, No. 1, pp. 3032, 2018.