Volume 10, Issue 1, January – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14603646

Prediction and Analysis of Franchise Cricket

Rayini Arun Kumar Reddy[1]; Sahith kumar Reddy[2]; and T Sushmitha[3]
1,2,3 Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology,
Hyderabad, Telangana

Abstract:- In our present world, sports produce a very large amount of statistical data. What makes cricket different from other sports is the number of variables involved in it right from the pitch to conditions playing under, a breeze to the length of the boundary line and likewise many other which makes every game has its prominence maybe that's the reason haven't got bored of the game though it is more than 100 years old. From the day we started using analytics in cricket to today the game evolved massively, it changed the way the players, coaches look at the game and it brought a new dimension to the game. IPL has been a carnival of cricket and showing the potential of cricket to the world and acted as a bridge in carrying the game to a wider range of audiences. In present-day IPL we are using every statistic that's available because of the high competitiveness of the tournament. This project is a sincere effort to find hidden insights in the IPL by using the data of previous seasons.

I. INTRODUCTION

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India that has captured the imagination of cricket fans worldwide. Founded in 2008, the league has since grown in popularity and has become a lucrative platform for both players and franchises. The IPL is not only a source of entertainment but also a goldmine for data analytics. Franchises are leveraging data analysis to make informed decisions on team selection, strategy, and performance analysis. Data analysis has become an integral part of the IPL, and it has helped franchises gain a competitive edge over their rivals. In this paper, we will explore the various uses of data analysis for IPL franchises and how it has impacted the league. The main intention of performing this analysis is to help the franchises in order to build a strong team and help them with the players individual statistics over the years in the IPL.

II. LITERATURE REVIEW

In this section, an overview of a few recent research activities will be presented where sports (cricket) data has been used for knowledge and analysis.

Haghighat M. Rastegari H and Nourafza N. (ACSIJ Advances in Computer Science: an International Journal, 2013) have reviewed data mining techniques for result prediction in sports [1]. They have also evaluated the advantages and disadvantages of the reviewed data mining techniques. The techniques that are reviewed are classification techniques like Artificial Neural Networks (ANN), Support Vector Machines (SVM), Bayesian method, Decision trees, Fuzzy system and Logistic Regression. The datasets that are used were, first 15 weeks of NFL 2003, Rugby League, ACB Basketball League 2008-9 season, NBA League 2005-6 to 2009-10 seasons, last 15 years of Netherland soccer and many other datasets. By evaluating the literature in this, they have detected two major challenges, first is the need for further research in order to obtain better prediction accuracy and second is the lack of general and comprehensive statistics datasets that force researchers to collect data from sports websites. They have suggested a few things to improve the prediction accuracy, they are - using the data mining and machine learning techniques that have yielded good results in other fields, also using hybrid algorithms can boost the accuracy, and including more valid features for prediction.

Several studies have been done on player's compensation in various sports. For example, Estenson (1994) studied player compensation in baseball. Likewise Dobson and Goddard (1998) and Kahn (1992) considered some of the issues in the field of football. There are also studies related to ice-hockey by (Jones and Walsh, 1988) and basketball (Berri, 1997). There are quite a few studies that deal with scheduling cricket matches. But there hasn't been any significant research in the field of player price analysis which says what amount of money a player can be given which is done by "Siddhartha K Rastogi and Satish Y Deodhar". Hedonic price analysis is based on the hypothesis that a good/service can be treated as a collection of attributes that differentiates it from other goods/services.

Pabitra Kumar Dey, Gangotri Chakraborty, Purnendu Ruj, and Suvobrata Sarkar in the paper "A Data Mining Approach on Cluster Analysis of IPL" [3] has made used of MATLAB to produce a clustering algorithm based on fuzzy logic to classify the batting statistics of IPL into a number of clusters. Here they divide the data into four clusters and the goal is to determine the grouping in a set of unlabelled data. The criterion for the classification is run/ball as a parameter. When no. of clusters are 4 the accuracy if classification is 73.48%. They obtained the results for the test set and the accuracy was measured for the particular machine learning model used to predict the winner of the match. They predict based on the different teams of the IPL, and are predicted by analysing every over, so that the winner can be predicted in almost any situation of the match. The accuracy for a selected number of attributes for each team using feature selection was also measured. For every model generated the highest accuracy for a team to win is being predicted. Thus, the graph displays which team has the highest accuracy in each generated model.

Daniel Mago Vistro, Faizan Rasheed, Leo Gertrude David (International Journal Of Scientific & Technology Research, 2019) have proposed a model to predict the winner in cricket match using Machine learning and Data Analytics [2]. The Machine learning algorithms used to train with datasets are SVM, Random forest, Naive Bayes, Decision trees and Logistic Regression. They have also used an XGBoost classifier which is also called gradient boosting and makes very fast calculations by tree algorithms. The dataset taken is IPL data of the year 2008 to 2017. They are using the model's confusion matrix to evaluate the performance. By visualizing attributes of data with the target variable the best features were selected. The accuracy obtained by the models were Decision tree classifier - 76.9%, Random forest classifier - 80.76% and XGBoost classifier - 94.23%. The prediction produced by their model required a lot of domain information and expertise for observations**.**

The paper (A MCDM Approach for Evaluating Bowlers Performance in IPL [6]) is about drawing statistics on the bowler in all of his three forms (criteria for ranking—1.played at least 3 matches. 2. Bowled at least 8 overs 3. Got at least 1 wicket) and performance and ranking is given using AHP-TOPSIS and AHP-COPRAS.The first three attributes are negative attributes as lower the value of the attribute, higher the performance of the bowler. The latter three are positive attributes. If a bowler's performance is good in a match we can't say or analyze his overall ranking in the series so the criterion is that the bowler should at least play 3 matches, bowled for at least 8 overs and took at least a wicket

III. SOFTWARE REQUIREMENTS

Practically every field uses Python for a variety of tasks and activities. It supports a variety of paradigms for programming, including structured (particularly procedural), functional, and object-oriented. The extensive standard library of this language has earned it the moniker "batteries-included" language. Python gives programmers some of the best flexibility and capabilities, which will improve their efficiency and capacities as well as the quality of their code. Python also has a vast library that helps with the heavy workload. Libraries for machine learning methods include NumPy, Pandas, Scikit-Learn, and NLTK.

IV. METHODOLOGY

The IPL data is sourced from the Kaggle website, which acts as the input source for the analysis. The data extraction process involves retrieving the raw data from the source and storing it in a dataset. The raw dataset is then cleaned to remove any inconsistencies, errors, or missing values. Once the data cleaning is complete, the dataset undergoes pre-processing, where it is transformed and normalized to prepare it for analysis. The processed data is then analyzed using various data mining techniques, including analysis. This includes visualizations such as charts, graphs, and tables to present the data in a user-friendly manner.

➢ *Dataset*

The IPL datasets on Kaggle contain data on all IPL matches from 2008 to 2020. The dataset includes information on the match details such as venue, date, and winner, as well as ball-by-ball data, player statistics, and team standings. The ball-by-ball data includes information on the deliveries such as the type of delivery, runs scored, wickets taken, and the player involved. The player statistics include information on batting, bowling, and fielding performance. Cricsheet, on the other hand, provides a more extensive dataset that includes ball-by-ball data for all international cricket matches, including the IPL. The dataset contains information on over 700 T20 matches played in the IPL between 2008 and 2021. It includes detailed information on the deliveries, such as the type of delivery, runs scored, wickets taken, and the fielding events. The dataset also includes information on the players, such as their batting and bowling statistics.

V. RESULT AND EVALUATION

There are numerous aspects which play crucial part in cricket, We considered some of the most important and impactful factors and analyzed them thoroughly.

A. Toss Decision vs Result:

The toss in cricket, often perceived as a simple coin flip, plays a significant role in determining the outcome of a match, particularly in the Indian Premier League (IPL). Environmental factors such as dew heavily influence toss decisions. In evening matches, dew creates challenging conditions for bowlers by reducing their grip on the ball, making it advantageous for teams to bowl first. Similarly, the pitch is another critical factor. Its behavior can vary based on location, weather, and even the duration of the match, adding complexity to the decision-making process.

Beyond environmental considerations, team composition and strategy also impact toss decisions. Teams must assess their strengths and weaknesses, as well as those of their opponents, to determine whether batting or bowling first will provide a strategic advantage. While the toss may appear inconsequential at a glance, its implications for team performance and match dynamics are profound. As such, IPL teams devote considerable effort to analyzing these variables, recognizing the pivotal role the toss plays in shaping the trajectory of the game. This study delves into the underlying factors that make toss a critical element in IPL cricket.

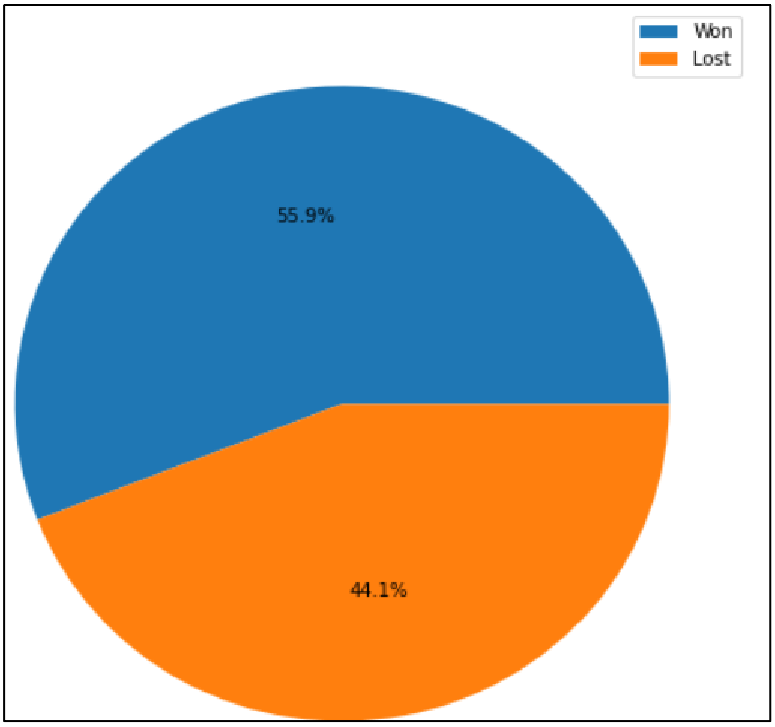


Fig 1:Percentages of Win and Lose after Electing Field First

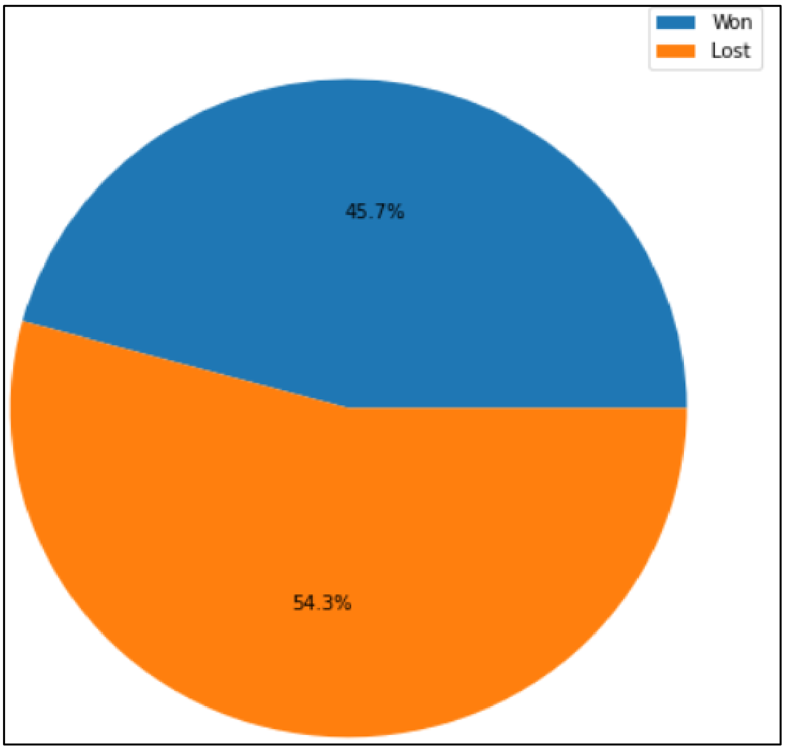


Fig 2:Percentages of Win and Lose after Electing Bat First

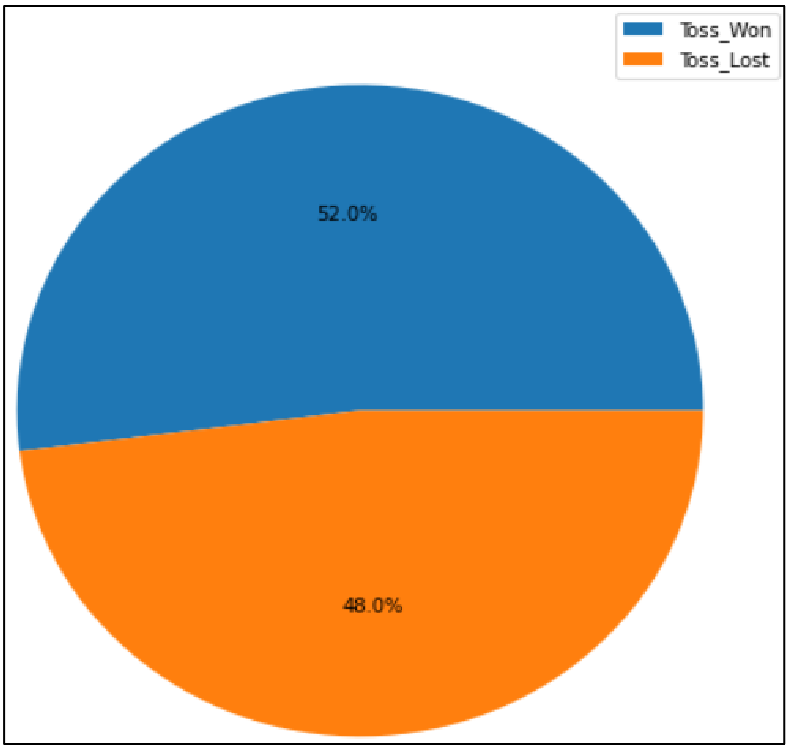


Fig 3:Percentages of Win and Lose after Winning Toss

B. Analysis of Phasewise Runrate for Team Batting First vs Team Batting Second(Year Wise) (Powerplay, First Two, Next Four overs vs Middle overs vs Death overs):

Let's break this analysis down year by year to see how teams' batting strategies have changed over time. When batting first, teams usually approach the innings in phases: the first two overs, the next four, the middle overs, and the death overs. The first two overs are about understanding the conditions and adjusting their play. The next four focus on taking advantage of the powerplay to score quickly. In the middle overs, teams work on building partnerships and keeping the scoreboard ticking. Finally, the death overs are all about going big, with set batsmen aiming to rack up as many runs as possible.This phased approach helps us see how strategies have evolved. The run rate—a key metric in the IPL—is a simple but powerful way to measure batting performance. It's calculated by dividing the total runs by the overs faced. It not only shows how quickly a team scores but also gives franchises insights into their overall performance and areas to improve.

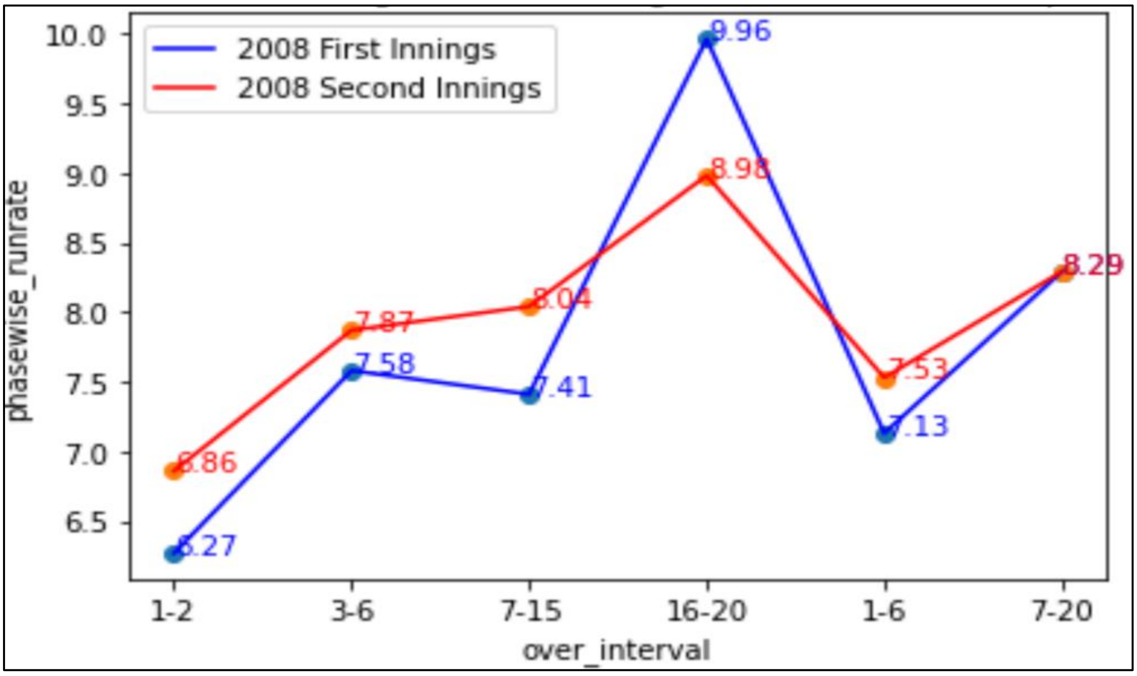


Fig 4:2008 Batting first vs Batting Second Runrate Comparison

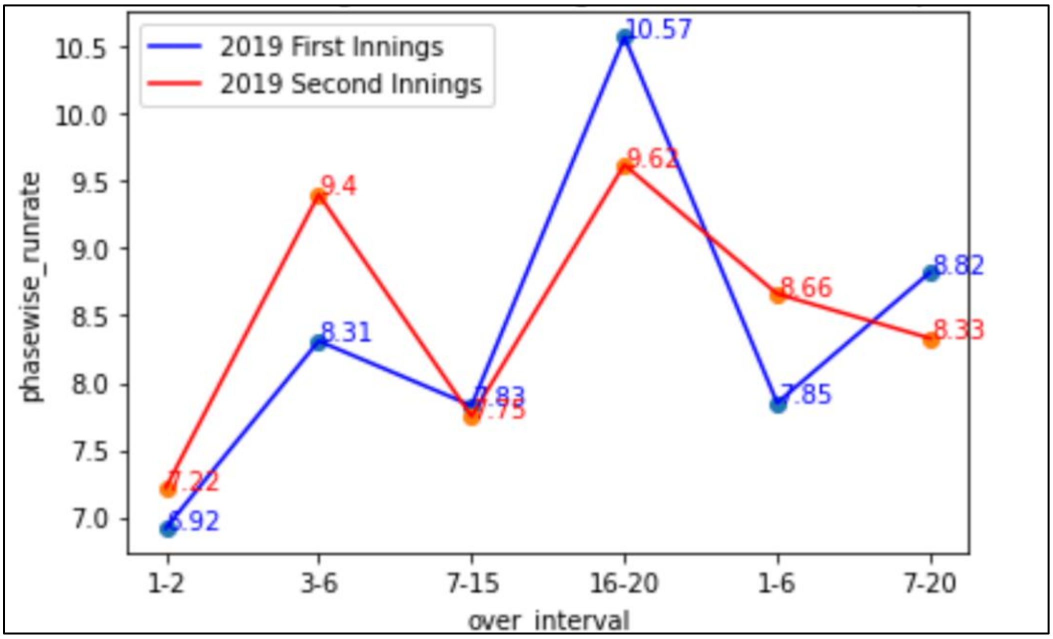


Fig 5:2019 Batting First vs Batting Second Runrate Comparison

Run rate also plays an important role during the match, as it helps the teams to set a target or chase a target. If a team is batting first, they need to score runs at a higher rate to post a challenging total on the board. Similarly we analyzed and compared all the run-rates over the years and making the franchise understand their weakness and strengths in the powerplay,middle-overs and final-overs.

C. Average Run-Rate of Each Franchise Over the Years:

In IPL understanding the teams capacity is the most difficult aspect in order to make it simple we introduced some of the most important factors such as run-rate, strike-rate, etc. we already analyzed year wise runrate over the years but in order to understand deeper, we did analysis on the run-rate for each franchise and help them understand the statistics and finding the weakest part of the team.

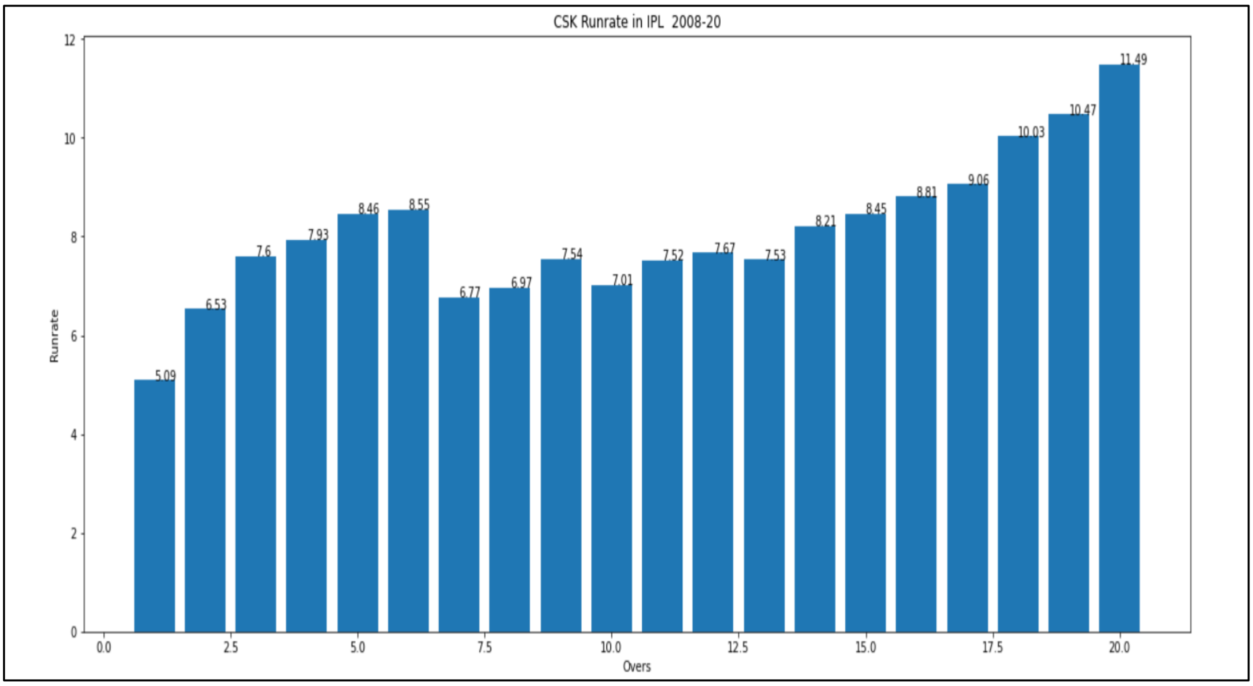


Fig 6: Average Run-Rate of Chennai Super Kings Over the Years

By the above figure 6 in which it describes the average run-rate of Chennai super kings over the years, as we can see run-rate is quite good, in-fact great but there is slight decrement in the middle overs where it is the main region where CSK is struggling, so that they need settled batsmen who can defend the as well as who can hit at least one boundary in an over in the middle overs. CSK is one of the most successful teams in the IPL and the main strength of CSK is powerplay and death-overs in the batting because they have a powerful opening pair and a destructive death batting. The middle overs is the most important phase in the IPL because most of the wickets will fall in this phase if the Batting order collapses then the chances of winning comes close to zero, so having a middle order batsmen is very much important and this can help CSK franchise to produce better performance in the tournament.

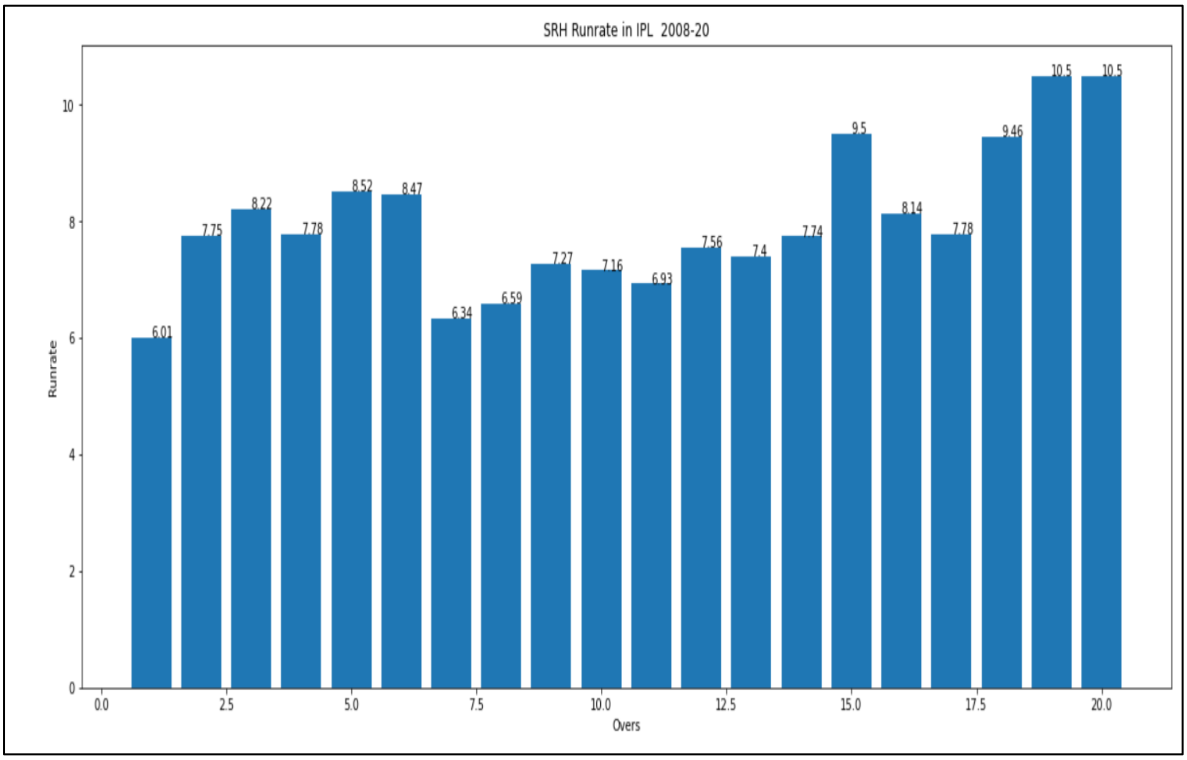
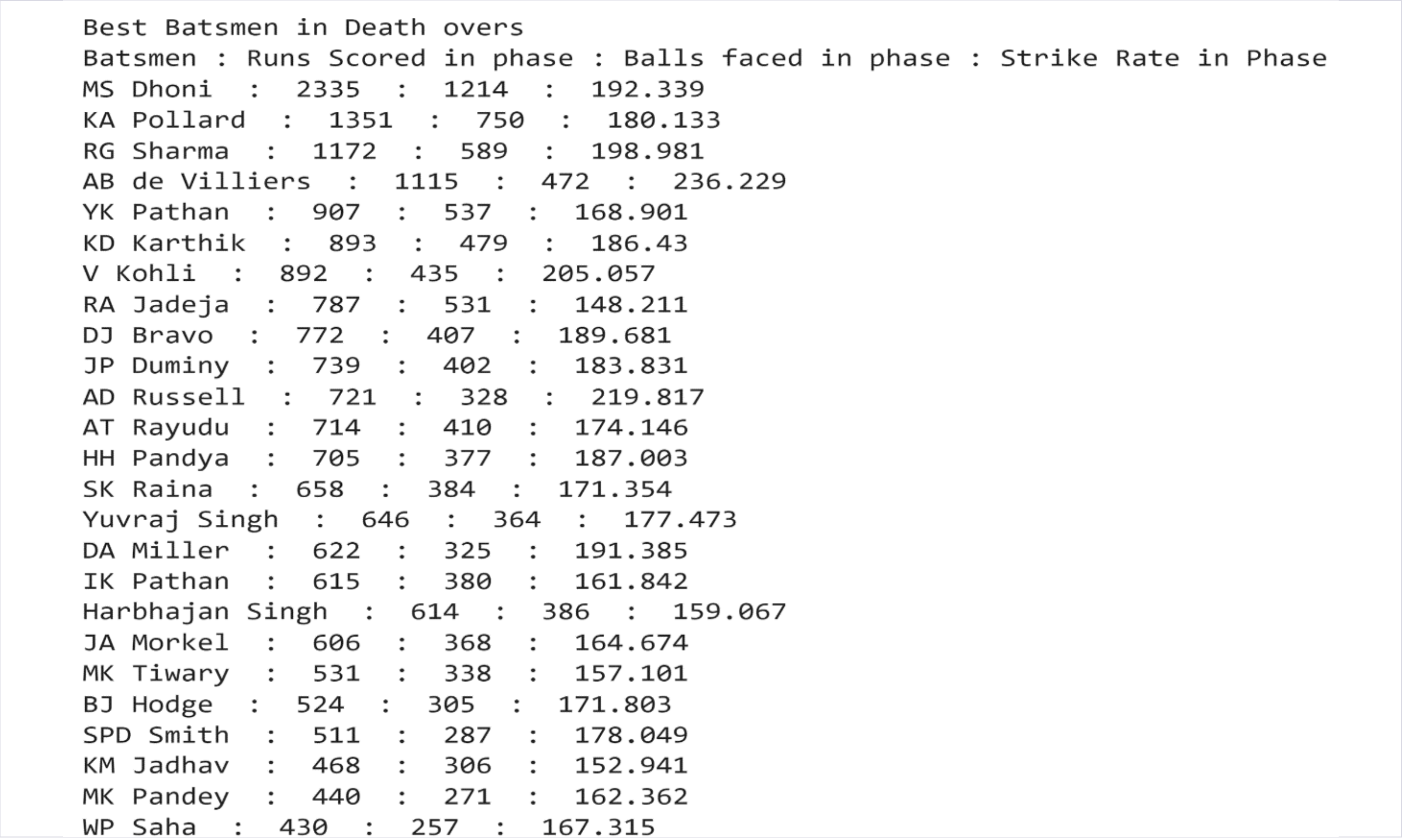


Fig 7: Average Run-Rate of Sunrisers Hyderabad Over the Years

As we can see in the above figure 7 SRH is not only failing in the middle overs but also in the early death overs , SRH is one of the franchises where it depends mostly on the top order batsmen failing to produce a better middle order and death over batsmen in order overcome this problem rather than depending upon only on the top order they need a middle order batsmen who can produce a descent innings and make a strong partnership with the death-over batsmen so that SRH will be able to compete with opponent while chasing a total. When we try compare CSK and SRH both the teams performed almost the same in the powerplay but the major change occurs in the middle and death overs ,where CSK is producing more runs in the death-overs than SRH and turning most of the matches towards them , So this way we can compare one team with each every team and find out the weak links amongst the franchises.

D. Best Batsmen in Death-Overs:

Analysing the best death overs batsmen in IPL is important because it is a crucial phase of the game where the batting team tries to maximize their score, while the bowling team tries to restrict the runs and take wickets. The death overs are typically considered to be the last four to five overs of an innings, during which the batting team tries to score as many runs as possible, while the fielding team tries to prevent that by bowling well and taking wickets. This phase of the game can often determine the outcome of the match, and the ability of a team's batsmen to perform well during the death overs can be a key factor in their success. Analyzing the best death overs batsmen in IPL can help teams make informed decisions about their batting lineups and strategies. It can also help teams identify players who are particularly skilled at scoring runs in the death overs, which can be useful in determining who to retain or recruit for future seasons. Additionally, understanding the strengths and weaknesses of opposing batsmen in the death overs can help bowlers and captains develop effective game plans and field placements to limit their scoring opportunities and take wickets. Overall, analyzing the best death overs batsmen in IPL can provide valuable insights that can help teams improve their performance.

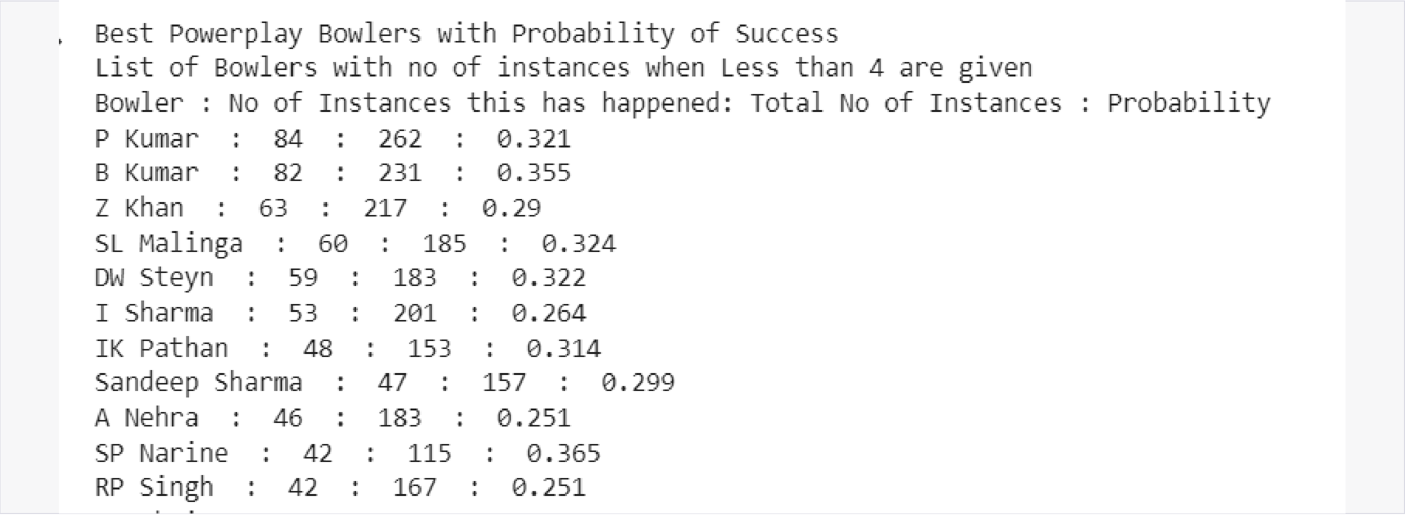


```
Best Batsmen in Death overs
Batsmen : Runs Scored in phase : Balls faced in phase : Strike Rate in Phase
MS Dhoni  :  2335  :  1214  :  192.339
KA Pollard  :  1351  :  750  :  180.133
RG Sharma  :  1172  :  589  :  198.981
AB de Villiers  :  1115  :  472  :  236.229
YK Pathan  :  907  :  537  :  168.901
KD Karthik  :  893  :  479  :  186.43
V Kohli  :  892  :  435  :  205.057
RA Jadeja  :  787  :  531  :  148.211
DJ Bravo  :  772  :  407  :  189.681
JP Duminy  :  739  :  402  :  183.831
AD Russell  :  721  :  328  :  219.817
AT Rayudu  :  714  :  410  :  174.146
HH Pandya  :  705  :  377  :  187.003
SK Raina  :  658  :  384  :  171.354
Yuvraj Singh  :  646  :  364  :  177.473
DA Miller  :  622  :  325  :  191.385
IK Pathan  :  615  :  380  :  161.842
Harbhajan Singh  :  614  :  386  :  159.067
JA Morkel  :  606  :  368  :  164.674
MK Tiwary  :  531  :  338  :  157.101
BJ Hodge  :  524  :  305  :  171.803
SPD Smith  :  511  :  287  :  178.049
KM Jadhav  :  468  :  306  :  152.941
MK Pandey  :  440  :  271  :  162.362
WP Saha  :  430  :  257  :  167.315
```

Fig 8: Best Batsmen in Death Overs

E. Best Powerplay Bowlers:

Analyzing the best powerplay bowler in IPL is important because the powerplay overs, which are the first six overs of a T20 match, are crucial in setting the tone for the rest of the innings. A good powerplay performance can help a team restrict the opposition's scoring rate, take early wickets, and gain momentum. On the other hand, a poor powerplay performance can put a team on the backfoot and make it difficult to recover. By analyzing the best powerplay bowler in IPL, a team can identify the bowler who is most effective in the powerplay overs and use their skills to their advantage. This can help them plan their strategy and tactics better, both while bowling and batting. For instance, if a team knows that a particular bowler is very effective in the powerplay, they can plan their batting order accordingly and try to target other bowlers in the later overs.



```
Best Powerplay Bowlers with Probability of Success
List of Bowlers with no of instances when Less than 4 are given
Bowler : No of Instances this has happened: Total No of Instances : Probability
P Kumar  :  84  :  262  :  0.321
B Kumar  :  82  :  231  :  0.355
Z Khan  :  63  :  217  :  0.29
SL Malinga  :  60  :  185  :  0.324
DW Steyn  :  59  :  183  :  0.322
I Sharma  :  53  :  201  :  0.264
IK Pathan  :  48  :  153  :  0.314
Sandeep Sharma  :  47  :  157  :  0.299
A Nehra  :  46  :  183  :  0.251
SP Narine  :  42  :  115  :  0.365
RP Singh  :  42  :  167  :  0.251
```

Fig 9: Best Bowlers in Powerplay Overs

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14576672

VI. CONCLUSION AND FUTURE SCOPE

Studying franchise cricket in the IPL has given us some fascinating insights into how teams perform, how strategies evolve, and how players shine on the big stage. By digging into large datasets with advanced data mining tools, we uncovered patterns and trends that aren’t immediately obvious. These insights help teams make smarter choices about player selection, match tactics, and overall strategies. They also offer fans and analysts a deeper look into the nuances of the game, making cricket even more exciting to follow.

Using data-driven techniques shows how much of a difference analytics can make in improving team performance. By identifying what works and what doesn’t, teams can fine-tune their approach and boost their chances of success. That said, this kind of analysis isn’t perfect. The results depend a lot on the quality and completeness of the data, and what works for one team might not work for another. Moving forward, improving data quality and expanding the scope of research will make these insights even more powerful and reliable.

REFERENCES

[1]. Haghighat, M., Rastegari, H. and Nourafza, N., 2013. A review of data mining techniques for result prediction in sports. Advances in Computer Science: an International Journal, 2(5), pp.7-12.

[2]. Vistro, D.M., Rasheed, F. and David, L.G., The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics.

[3]. Dey, P.K., Chakraborty, G., Ruj, P. and Sarkar, S., 2012. A Data Mining Approach on Cluster Analysis of IPL. International Journal of Machine Learning and Computing, 2(4), p.351.

[4]. Rastogi, S.K. and Deodhar, S.Y., 2009. Player pricing and valuation of cricketing attributes: exploring the IPL Twenty20 vision. Vikalpa, 34(2), pp.15-24.

[5]. Dey, P.K., Banerjee, A., Ghosh, D.N. and Mondal, A.C., 2014. AHP-neural network based player price estimation in IPL. International Journal of Hybrid Information Technology, 7(3), pp.15-24.

[6]. Dey, P.K., Ghosh, D.N. and Mondal, A.C., 2011. A MCDM approach for evaluating bowlers performance in the IPL. Journal of emerging trends in Computing and Information Sciences, 2(11), pp.563-573.

[7]. D. Thenmozhi, P. Mirunalini, S. M. Jaisakthi, S. Vasudevan, V. Veeramani Kannan and S. Sagubar Sadiq, "MoneyBall - Data Mining on Cricket Dataset," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5.

[8]. K. Jain, M. N. Murty and P. J. Flynn, “Data Clustering: A Review,” ACM Computing Surveys, Vol. 31, No. 3, September 1999, pp. 264-323.

[9]. Kimber, Alan C., and Alan R. Hansford. “A Statistical Analysis of Batting in Cricket.” Journal of the Royal Statistical Society. Series A (Statistics in Society), vol. 156, no. 3, 1993, pp. 443–455. JSTOR, www.jstor.org/stable/2983068. Accessed 28 Feb. 2020.

[10]. Manage, Ananda & Butar Butar, Ferry. (2007). Statistical analysis in one-day cricket. Proc. Amer. Statist. Assoc.. 2600-2605**.**