

Analysis and Prediction of Winning Team and Player Performance in Franchise Cricket using Machine Learning

Tumma Susmitha¹, M Madhu Latha², J Madhumathi³, Dr Rupesh Mishra⁴

¹Computer Science & Engineering

Vasavi College of Engineering

Hyderabad, India

susmitha@staff.vce.ac.in

²Computer Science & Engineering

Chaitanya Bharathi Institute of Technology

Hyderabad, India

madhucse555@gmail.com

³Computer Science & Engineering

Vasavi College of Engineering

Hyderabad, India

madhumathi.jessu@vce.ac.in

⁴Computer Science & Engineering

Gitam Deemed to be University

Hyderabad, India

rmishra3@gitam.edu

Abstract— Sports, especially cricket, produces an enormous amount of statistical data that has led to the evolution of the game. Unlike other sports, cricket involves numerous variables ranging from pitch conditions, weather, boundary length, and several other factors that make each game unique. These variables make every match have its own significance, making cricket an ever-evolving sport that continues to captivate its audience. Over the years, cricket has witnessed a massive transformation with the integration of analytics. The use of data analytics has changed the way players and coaches approach the game. It has brought a new dimension to the game and has opened several new avenues for analyzing and interpreting the game. The Indian Premier League (IPL) has acted as a catalyst in showcasing the potential of cricket to the world and has bridged the gap between different cultures and audiences. In the present-day IPL, every statistic available is being used due to the high level of competitiveness of the tournament. Teams use data analytics to understand their strengths and weaknesses, analyze their opponents' gameplay, and devise strategies to win matches. These analytics provide valuable insights into various aspects of the game, such as player performance, team strategies, and game trends, among others. This paper is a sincere effort to uncover hidden insights in IPL by utilizing data from previous seasons. The vast amount of data available offers a treasure of formation that can help us to understand the nuances of the game and make informed decisions. By analyzing this data, we can identify patterns and trends, assess player performance, and make informed decisions that can impact the outcome of a match.

Keywords:- IPL, Powerplay, Death Overs.

I. INTRODUCTION

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India that was established in 2008. Since its inception, the league has become one of the most popular sporting events in the world, attracting millions of viewers and generating billions of dollars in revenue. The IPL has also become a platform for young and talented cricketers to showcase their skills and earn lucrative contracts. In recent years, the IPL has evolved into a global phenomenon, attracting some of the best cricketers from around the world. The league has also become a hotbed for innovation, with teams experimenting with new strategies,

tactics, and player combinations. As a result, the IPL has become a rich source of data for researchers and analysts who are interested in understanding the dynamics of franchise cricket. The Indian Premier League (IPL) is a highly competitive and popular Twenty20 cricket league that attracts some of the best cricketers from around the world. With so many talented players and teams competing in the IPL, it can be challenging to identify the factors that contribute to success in the league. Moreover, with the evolution of the game and the introduction of new rules and regulations, it can be difficult to keep up with the latest trends and strategies. Therefore, the problem that this

project aims to address is how to use data analysis to identify the factors that contribute to success in the IPL. The project will focus on analyzing the data from the years 2008 to 2019 and identifying patterns and trends in team performance, player performance, and venue performance. The objective is to develop a better understanding of the factors that contribute to success in the IPL and provide recommendations for teams and players to improve their performance in future seasons. The objectives are

- To analyze the IPL data from the years 2008 to 2019 and provide insights into the performance of teams, players, and venues.
- To identify patterns and trends in team performance, player performance, and venue performance.
- To develop best practices for teams and players to improve their performance in the IPL.
- To provide recommendations for teams and players to improve their performance in future seasons.
- To inform strategic decisions for teams and players based on the analysis of the IPL data.
- To contribute to the broader understanding of cricket as a sport and inform future research on cricket analytics.

II. RELATED WORK

In [1], the paper provides a comprehensive survey of various data mining techniques used for predicting sports results. The paper begins by discussing the increasing importance of predicting sports results, which has become an essential part of sports analytics. The authors then provide a brief overview of data mining techniques, including classification, regression, clustering, and association rule mining, and their relevance to sports prediction. The paper then reviews several data mining techniques used in sports prediction, including neural networks, decision trees, support vector machines, and Bayesian networks. The authors also review several case studies that have used data mining techniques for sports prediction. These case studies include predicting the outcome of soccer matches, basketball games, and horse races. The authors provide a detailed analysis of each case study, including the data mining technique used, the performance of the model, and the factors that contributed to the model's success or failure. Finally, the authors summarize the key findings of the paper, including the importance of data quality, feature selection, and model selection in sports prediction. They also highlight the potential future directions for research in this field, including the use of advanced techniques such as deep learning and ensemble methods.

In [2], the paper explores the application of machine

learning and data analytics in predicting the winner of cricket matches. The paper begins by discussing the growing importance of data analytics and machine learning in sports and the potential benefits of using these techniques to predict cricket match winners. The authors then provide a literature review of the existing research in this area, highlighting the various data mining techniques that have been used for cricket prediction. The authors then present their approach for predicting cricket match winners using a combination of machine learning algorithms and data analytics techniques. They use various features such as team rankings, player statistics, and match conditions to train their model and predict the winner of a cricket match. The paper also provides a detailed analysis of the performance of the proposed model, including its accuracy and efficiency. The authors compare their model with other existing models and show that their approach performs better in terms of accuracy and speed. Finally, the authors discuss the potential applications of their model in the field of cricket and sports analytics. They highlight the importance of using data analytics and machine learning techniques in sports prediction and the potential benefits of using these techniques to improve the performance of teams and players.

In [3], the paper explores the use of data mining techniques to perform cluster analysis on the Indian Premier League (IPL). The authors have made use of MATLAB to produce a clustering algorithm based on fuzzy logic to classify the batting statistics of IPL into several clusters. Here they divide the data into four clusters and the goal is to determine the grouping in a set of unlabeled data. The criterion for the classification is run/ball as a parameter. When no. of clusters is 4 the accuracy of classification is 73.48%. They obtained the results for the test set and the accuracy was measured for the 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 analyzing 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.

In [4], The paper discusses the application of data mining techniques on a cricket dataset to predict the outcome of cricket matches. The paper begins by introducing the concept of Moneyball, a strategy in baseball where data analytics is used to select players and build teams. The authors discuss the potential applications of this strategy in cricket and highlight the various data mining techniques that

can be used for this purpose. The authors then present their approach for using data mining techniques to predict the outcome of cricket matches. They use various features such as team performance, player statistics, and match conditions to train their model and predict the winner of a cricket match. The paper provides a detailed analysis of the performance of the proposed model, including its accuracy and efficiency. The authors compare their model with other existing models and show that their approach performs better in terms of accuracy and speed. Finally, the authors discuss the potential applications of their model in the field of cricket and sports analytics. They highlight the importance of using data mining techniques in sports prediction and the potential benefits of using these techniques to improve the performance of players.

III. METHODOLOGY

Data Collection

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 regression analysis, decision trees, neural networks, and association rule mining. The final output of the analysis is an analysis report that summarizes the findings and insights from the analysis. The report may include visualizations such as charts, graphs, and tables to present the data in a user-friendly manner.

Visualization

We can observe how the data appears and what sort of correlation the properties of data hold with the help of data visualization. It's the quickest approach to check if the features match the output.

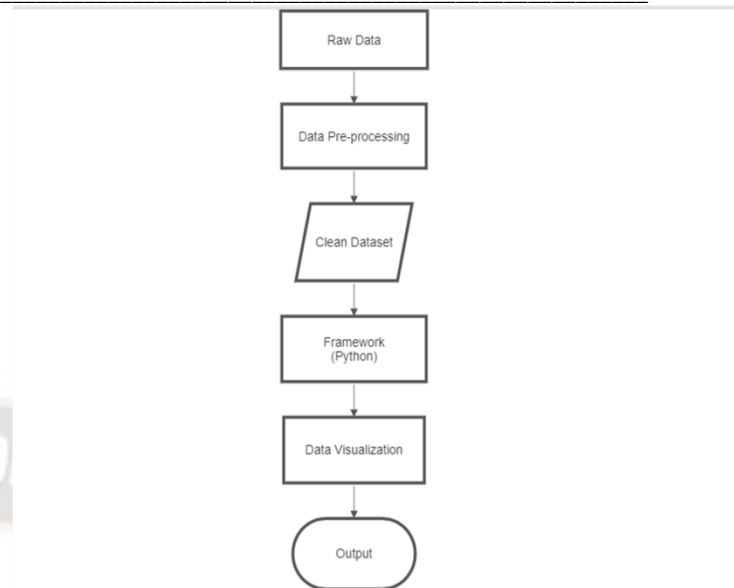


Fig. 1. Block diagram of proposed system

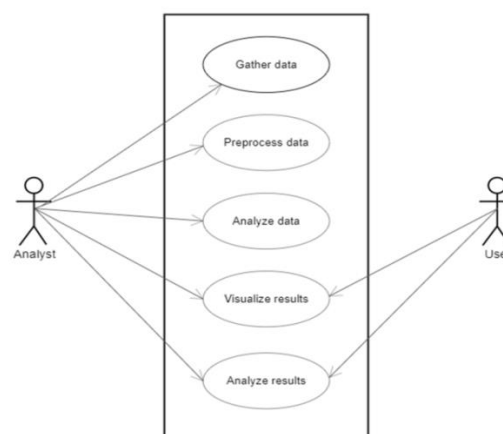


Fig. 2. Use Case Diagram of the Proposed System

Toss Decision vs Result

In Cricket, toss always has its own Importance and has a significant impact on the result of the game. Though from the outside it's just a random event with a half a chance of success probability but when it comes to cricket it had lot to do with the variables involved in cricket especially when it comes to IPL factors like dew come into the picture which makes tough to grip the ball and it effect increase as it becomes darker since most of the matches in IPL are played in dark teams prefer to escape due effect by opting to bowl first which gives a definite advantage, pitch which is another important aspect of the game the behavior of pitch changes from country to country, place to place, stadium to stadium and even day to day, for that matter even its behavior will change over the course of the game too, which makes it a very very important parameter when it comes to taking a decision at toss. Along with these factors, the decision of

toss at times can also depend on the composition of the team and the strengths of the team. And there are so many other factors that make a significant contribution to decide at the toss. Which makes Toss a key element when it comes to cricket, especially IPL. Every team in the IPL does significant work when it comes to Toss due to the above factors. So, Toss might be a random event when you look at it from the outside but when you get into the details you know the significance of it. That's the reason behind this analysis.

In this analysis, we try to find a relation between the toss and the decision taken by the team who won the toss and how much effect it had on the outcome of the game and venue of the match.

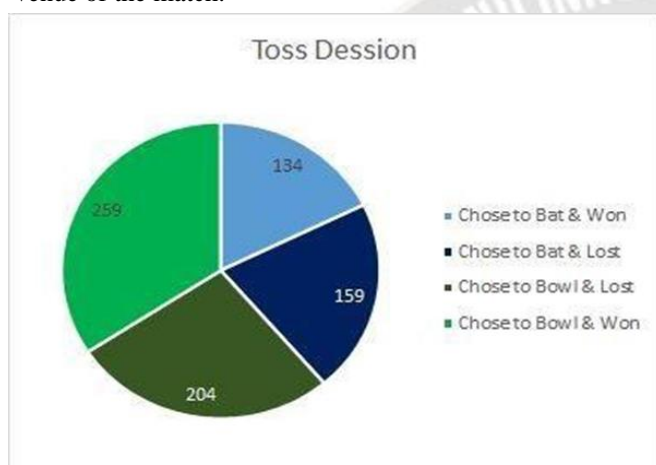


Fig. 3. Toss decision vs Result

First 2 overs vs Next Four overs vs Middle overs vs Death overs(Team Batting First):

In this part we will do an analysis of a team Batting first with its score and wickets in First 2 overs to that of Next Four to that of Middle overs to that of Death overs. The first two over's are mainly for assessing the conditions and adapting to it followed by the next four where teams try to accelerate the score by using power play restrictions. In the middle over generally, teams concentrate on stitching partnerships and rotating strike with odd boundaries so that they can maintain a healthy run rate and get prepared with settled batsmen in order to score big runs in death overs. This helps in finding an approach to the teams while they are batting.

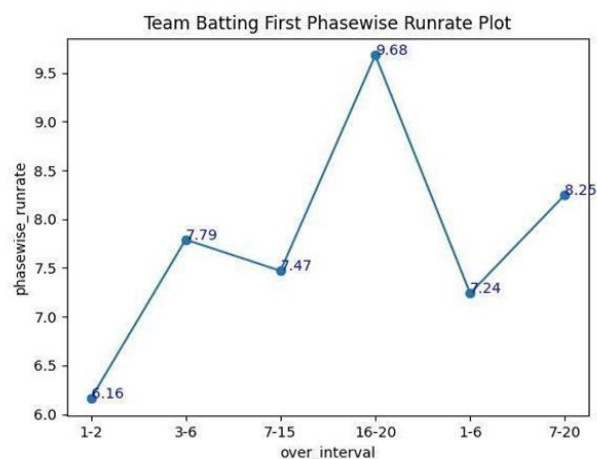
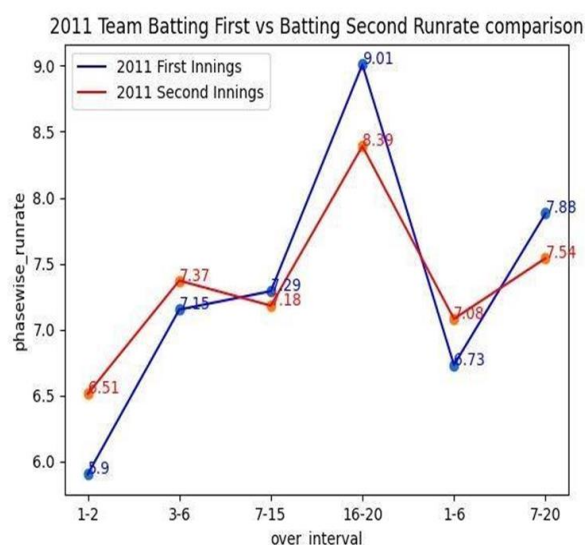


Fig. 4. Over Interval vs Runrate

First 2 overs vs Next Four overs vs Middle overs vs Death overs (Team Batting First and Year wise):

We have discussed this analysis in general but now we try to segregate in a year-wise way so that we can analyze the approach of each team. In this part we will do an analysis of a team Batting first with its score and wickets in First 2 overs to that of Next Four to that of Middle overs to that of Death overs. The first two overs are mainly for assessing the conditions and adapting to it followed by the next four where teams try to accelerate the score by using power play restrictions. In the middle over generally, teams concentrate on stitching partnerships and rotating strike with odd boundaries so that they can maintain a healthy run rate and get prepared with settled batsmen to score big runs in death overs. This helps in finding the approach of a team while they are batting and how it changed over the years.



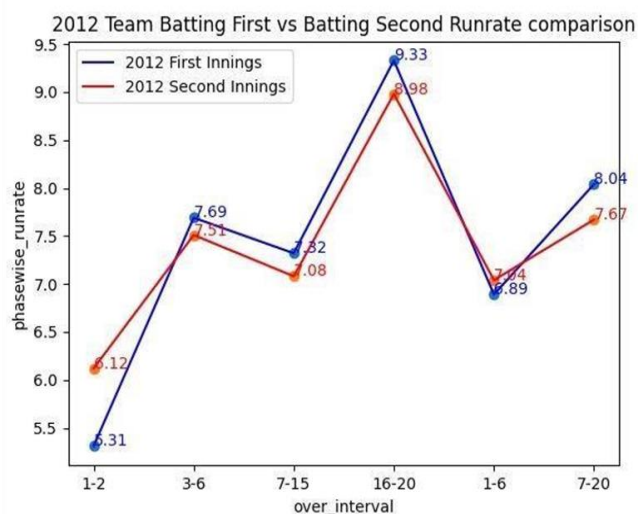


Fig. 5. Over Interval vs Run rate(2011 & 2012)

No of Runs per overs (Powerplay, First Two, Next Four overs vs Middle overs vs Death overs):

In this, we try to find the runs that batsmen are trying to score each ball of the overs in the Powerplay, First Two, Next Four overs vs Middle overs vs Death overs. The first two overs are mainly for assessing the conditions and adapting to it followed by the next four where teams try to accelerate the score by using power play restrictions. In the middle over generally teams concentrate on stitching partnerships and rotating strike with odd boundaries so that they can maintain a healthy run rate and get prepared with settled batsmen in order to score big runs in death overs. This gives an idea of the approach of batsmen in which delivery they want a score.



Fig. 6. Over Interval vs Run rate

Death Overs (16-20 overs) Runs per over analysis (Teamwise and Year wise):

Death Overs (16-20 overs) Runs per over analysis (Team Wise and year wise) helps in finding how teams are playing

their death overs. We have been seeing a gradual increase in the number of runs scored in death overs through the years. As we modern day batsmen are getting better at playing Twenty over cricket and they understand the game well and play accordingly. Modern day batsmen are becoming more powerful physically with the kind of emphasis that is laid on fitness in modern day games which helps in clearing the ropes with ease which is making it difficult for the bowler to bowl. And the players are going with a fearless approach compared to the tedious and risk-free approach of veteran cricketers and they are going all guns blazing at bowler which results in unbelievable bowling assaults and incredible scores. Even the bats are getting better as a result even if the bowler bowls a good delivery and takes the edge of bat it is flowing to the boundary. Even the bowlers are evolving and depending on mix-ups but because of the above factors they are getting hit every now and then. Death Overs are a very important phase of the match where there is a huge chance of the match slipping away from one hand to other hand. So, it is important to look at the way they are progressing.

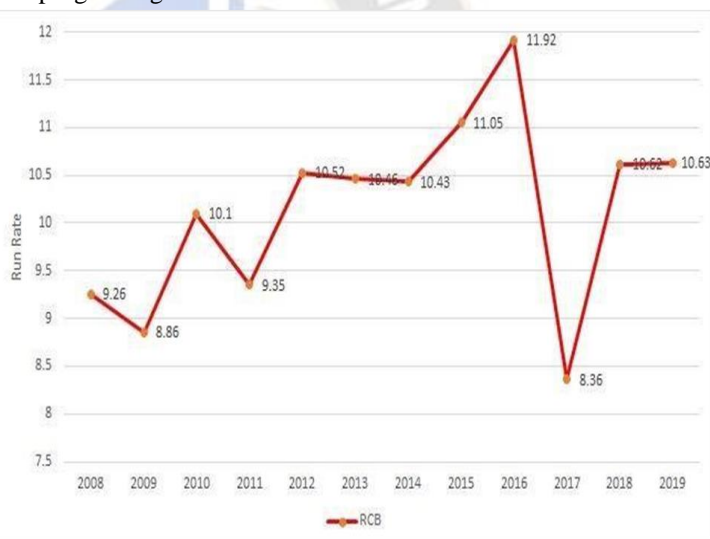


Fig. 7. Team RCB Runrate in Death overs(2008-19)

IV. DATASET DESCRIPTION

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.

V. CONCLUSION AND FUTURE WORK

The analysis of franchise cricket - IPL has revealed valuable insights into the game's performance, team strategies, and player abilities. Using data mining techniques and tools, we were able to analyze vast amounts of data collected from various sources and identify patterns, trends, and relationships that were previously unknown. Our analysis

has shed light on several crucial aspects of the game, including team performance, player selection, match outcomes, and player performance. Our findings highlight the importance of data analysis in understanding the trends and patterns in cricket performance. By identifying the top performers in the IPL, we can help teams and players make informed decisions on strategy and tactics. Additionally, our analysis can be used by cricket fans and enthusiasts to gain a deeper understanding of the game and its players.

The analysis has highlighted the importance of using data-driven insights to make informed decisions in team management, player selection, and match strategy. It has also demonstrated the potential of data mining tools and techniques in generating valuable insights that can help cricket teams, coaches, and administrators to optimize their performance and improve their chances of winning matches. However, it is important to note that the analysis is not without limitations, and the results may not be applicable to all teams or situations. Moreover, the accuracy of the analysis depends on the quality and completeness of the data used.

Overall, the analysis of franchise cricket IPL has demonstrated the potential of data mining techniques and tools in uncovering valuable insights into the game of cricket. It is our hope that this report will serve as a valuable resource for cricket enthusiasts, coaches, and administrators, and contribute to the further development of the game.

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] A. 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