Mini Project Report on

**Cricket Player Analysis**

Submitted in partial fulfillment of the requirements of the

## BACHELOR OF ENGINEERING IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

**Chinmay Deshmukh Roll.no 06**

**Abhijit Patil Roll.no 24**

**Omkar Sarvankar Roll.no 29**

### Name of the Mentor

**Prof. Shilpa Kapure**



**Department of Artificial Intelligence and Machine Learning**

Shivajirao S. Jondhale College of Engineering.

Dombivli (E)

### (Affiliated to University of Mumbai)

(AY 2023-24)

# Contents

##### Abstract i

##### Acknowledgments ii

##### List of Figures iii

##### Introduction 1

* 1. Introduction
  2. Motivation
  3. Problem Statement & Objectives

##### Literature Survey 2

##### Proposed System 3

* 1. Introduction
  2. Architecture/ Framework
  3. Algorithm and Process Design
  4. Details of Hardware & Software
  5. Expected Results
  6. Conclusion

##### 

##### Recommendations 8

##### 

**Appendix 9**

**Glossary 10**

**References 11**

**Abstract**

Cricket, a sport of strategy and precision, thrives on the performances of individual players who can significantly influence the outcome of matches and tournaments. This project presents a comprehensive analysis of cricket player performance using data science methodologies, with a focus on key metrics such as batting averages, strike rates, bowling averages, and economy rates. The objectives encompass player profiling, comparative analysis, and predictive modelling, seeking to uncover insights and trends that can inform team strategies, player selection, and predictive performance assessments. The project begins with data collection and preparation, ensuring the availability of high-quality, comprehensive cricket data. Subsequent steps include exploratory data analysis to unearth patterns, feature engineering to create meaningful performance metrics, and the development of machine learning models for predictive analysis. These models are evaluated and fine-tuned to enhance their predictive accuracy. Data visualization plays a pivotal role in representing performance insights effectively, making it accessible to a wide audience of cricket enthusiasts, analysts, and team decision-makers. The analysis culminates in actionable recommendations for team strategy and player selection, with a focus on enhancing performance in future matches.

**Acknowledgments**

The completion of our mini project could not have been possible without the kind support and help of many individuals. We would like to extend our sincere thanks, and gratitude to all of them. The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mentioning of the people whose constant guidance and encouragement made it possible. We take pleasure in presenting before you, our project, which is a result of studied blend of both research and knowledge. We are highly indebted to our mentor Prof. Anita Shirture our principal sir Dr. P.R. Rodge and project coordinator Prof. Sneha Ingale for their guidance and constant supervision. We would like to express our gratitude towards our parents as without them, the support system: that we needed would never would have been possible. We would also extend our thanks to our batch mates and faculties for their kind cooperation and encouragement which motivated us and helped us in thinking of this project.

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Figures Names** | **Page no** |
| **3.3.1** | **Flow Chart** | **3** |
| **3.5.1** | **Batsmen Dashboard** | **5** |
| **3.5.2** | **Bowler Dashboards** | **6** |
| **3.5.3** | **All-Rounder Dashboard** | **6** |

**1.Introduction**

* 1. **Introduction**

The primary objective of this project is to systematically analyze the performances of cricketers, paying particular attention to metrics like batting average, strike rate, bowling average, and economy rate. Through the lens of data science, we aim to profile players, conduct comparative analyses to identify top performers, and construct predictive models to foresee future outcomes. This analysis represents the intersection of tradition and technology, where cricket's rich heritage is complemented by data-driven decision-making. As we explore the depths of player performance, we discover a new era in cricket analysis—a space where data science leads the way towards understanding the sport's essence, benefiting players, teams, and fans alike.

* 1. **Motivation**

Cricket is not merely a sport; it's a passion that transcends borders and unites millions of fans worldwide. The allure of cricket lies not just in its rich history but in the moments of brilliance, the heroics of individual players, and the tactical intricacies that define each match. It's a sport that combines tradition with innovation, and the time is ripe to leverage the power of data science to unlock its hidden treasures.

* 1. **Problem Statement & Objectives**
* To find Batsmen (Innings Bat, Batting avg, strike rate, boundary%)
* To find All-Rounder (Innings Bat, Batting avg, strike rate, boundary%) Innings Bowl, Economy, Wicket%, Dot Ball%)
* To find Fast Bowlers (Innings Bowl, Economy, Wicket%, Dot Ball%)
* To find Spinners (Innings Bowl, Economy, Wicket%, Dot Ball% )

**2.Literature Survey**

[1]Player data analysis is used in most sports. Sports analysis is full of statistics. This is the present and future of any professional in the field of sports. Both the opening of the stadium aids in player and team analysis and predicts relevant results.

The task at is facing the challenge of predicting the outcome of an IPL cricket match. A total of 644 game statistics were used in this study. Factors such as player strength and luck are used as important factors in prediction. The problem with this study is the dynamic and use of the relevant non-relationship database, the HBase application firmness.

[2]The authors at analyzed the performance of IPL players in terms of runs, the most successful team with wickets, Team performance in general, Man of the match with runs and wickets, throwing winners with runs, Toss winner with wickets, Analysis of Duckworth law winners. The full analysis is based on the presentation using a tableau. The results are predicted from different IPL teams and predicted in extreme analysis so that the winner of the match is predicted in almost any game situation. The accuracy of the selected number of adjectives for each group using feature selection was rated at

[3]In forecasting analytics, Put and Data are used in. The batting and bowling datasets are modelled according to the players' statistics and features. Four multiclass phase algorithms were used and compared. The most accurate classification of both data sets was Random Forest and at least the most accurate was SVM. highlights the performance of athletes especially batters and oversees analysis of Man of the Matches, Maximum Centuries Strikes by Batsmen, Top Batsmen, Batsmen with Top Strike Rate, Top 10 Players with Maximum Runs. Refining and refining of data is done by modification, consolidation. The authors in discussed how to analyse things to study the performance of cricket players and the findings of his study say that the force of battering dominates more than bowling. Studies show that the performance of throwers is one of the most important factors in changing the status quo.

[4] Described the player rating model at the IPL auction. Their model considered factors such as previous player bid price, player information, strike rate etc. Prakash, Patvardhan and Lakshmi described the batting and bowling index to measure the performance of players in their models to predict the results of IPL matches. The mathematical method of proposing correct strike orders for ODI games is shown in.

[5 ]In paper the authors proposed a two-way model using the Naïve Bayes and the Linear Regression Classifier. The first way is to predict the points of the first innings based on the current running rate, etc. The second method predicts the outcome of a given goal by a batting team. The authors in predict the performance of the fourth-season IPL batsmen using the first three seasons. A Multi-Layer perceptron (MLP) neural network is used to predict previous activity.

[6]The outcome of a match by comparing the strength of two teams is predicted by the performance of each player measured. They used algorithms to predict the performance of batsmen and bowlers from past and recent activity data. The so-called Combined Bowling Rate is a combination of three traditional bowling algorithms: bowling rate, strike rate and economy used to analyse bowlers in.

**3.Proposed System**

**3.1 Introduction:**

The primary objective of this project is to systematically analyze the performances of cricketers, paying particular attention to metrics like batting average, strike rate, bowling average, and economy rate. Through the lens of data science, we aim to profile players, conduct comparative analyses to identify top performers, and construct predictive models to foresee future outcomes. This analysis represents the intersection of tradition and technology, where cricket's rich heritage is complemented by data-driven decision-making. As we explore the depths of player performance, we discover a new era in cricket analysis—a space where data science leads the way towards understanding the sport's essence, benefiting players, teams, and fans alike.

**3.3 FrameWork**

Data collection

Using web scrapping

Data cleaning and data transformation

Data transformation in power query

Data modelling and building parameters

Build Dashboard in Power Bi

Analysis and Insights of Player Selecting

**Fig 3.3.1 Flow Chart**

**3.4 Algorithm and Process Design**

Step 1: Define Objectives and Metrics:

Clearly define the objectives of the analysis, such as identifying top performers, understanding batting, or bowling trends, or predicting future performances. Determine the specific metrics to focus on, such as batting average, strike rate, wickets taken, etc.

Step 2: Collect Comprehensive Data:

Gather cricket data from reliable sources, which may include batting and bowling statistics, match outcomes, player profiles, and team information. Ensure the data is comprehensive and covers the scope of your analysis.

Step 3: Data Cleaning and Preprocessing:

Clean the data by handling missing values, removing duplicates, and correcting any inaccuracies. Preprocess the data by organizing it into a structured format, performing feature engineering, and converting it into a suitable data structure for analysis.

Step 4: Data Transformation:

Data transformation is a critical step in the data science process that involves converting raw data of batting strike rate or bowling economy rate based on existing data into a structured and usable format for analysis.

Step 5: Data Modelling and building parameters

This parameter is used as a condition of a decision table or decision tree. So parameter is used as an event source of an event rule. The same parameter is used as a condition and result of a decision table or decision tree of batting average, strike rate, wickets taken, etc

Step 6: Predictive Analysis and Insights:

Use the trained models to make predictions on new data or validate against a separate test dataset. Derive insights from the predictions, such as identifying players likely to perform well in upcoming matches based on historical data.

Step 7: Data Visualization:

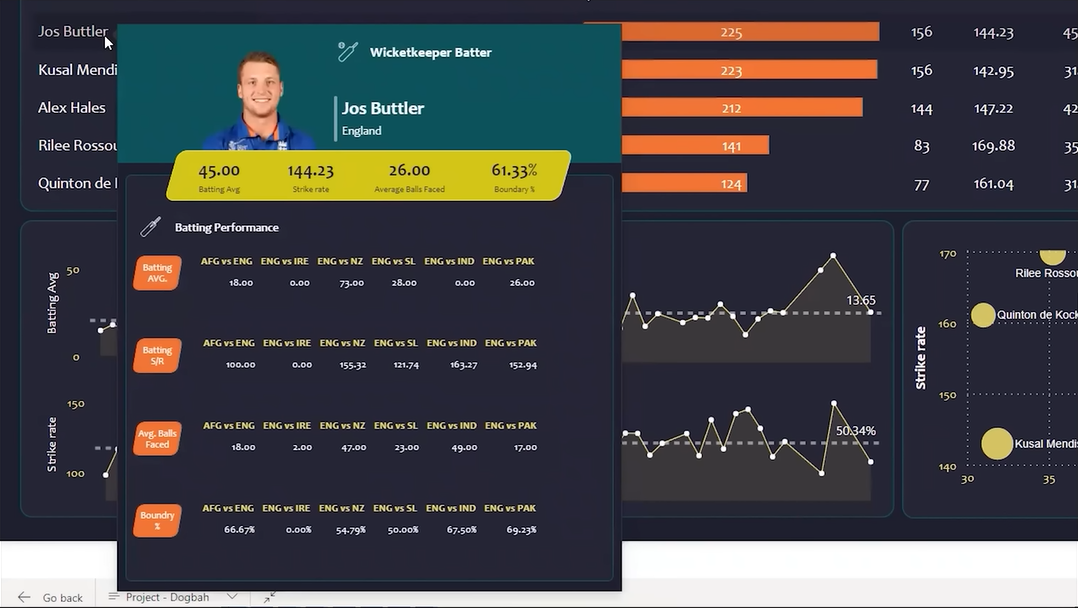
Create visualizations to represent the insights and findings from the analysis. Use appropriate charts, graphs, and dashboards to communicate the results effectively to stakeholders and decision-makers.

Step 8: End of the algorithm

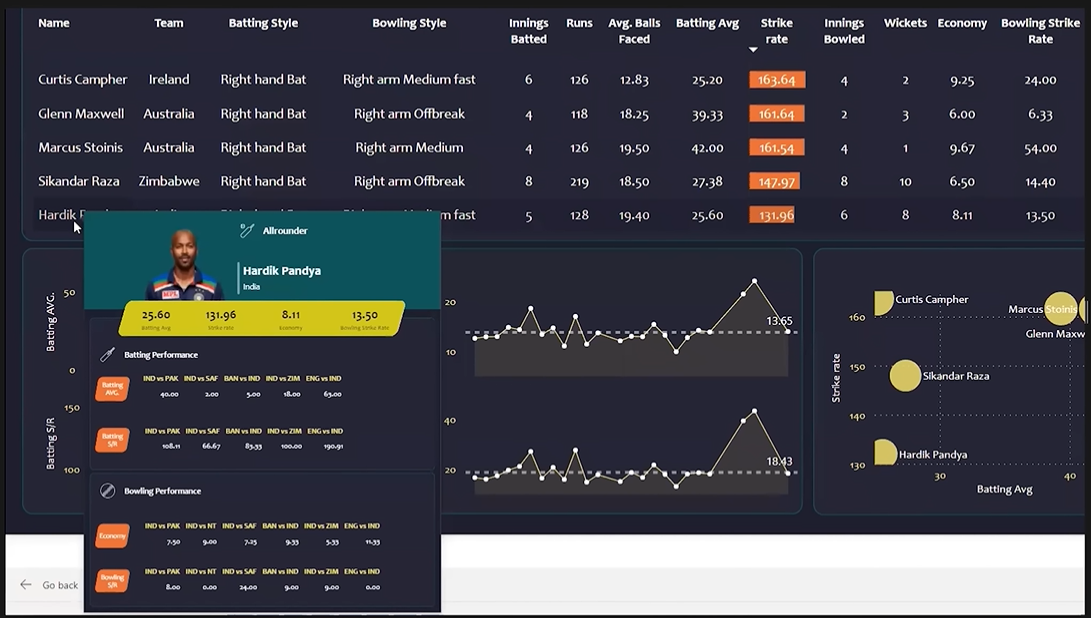
**3.4 Details of Hardware and Software**

* Hardware Requirement: Laptop/PC
* Software Requirement:
* Web Scraping (Web scraping is an automatic method to obtain large amounts of data from websites.)
* Python (Programming Language)
* Pandas (pandas is a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.)

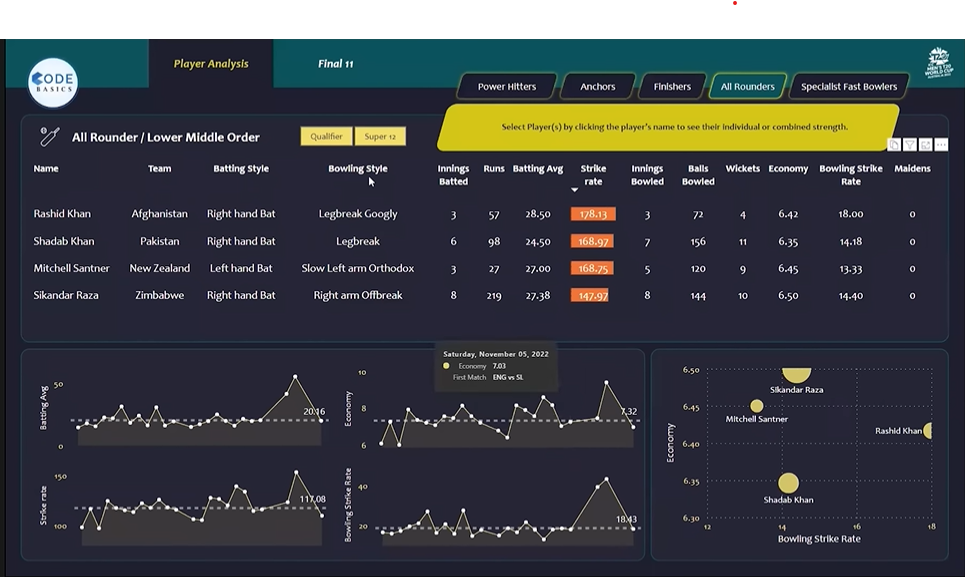
**3.5 Expected Results**

****

**Fig 3.5.1 Batting Dashboard**

****

**Fig 3.5.2 Bowling Dashboard**

****

**Fig 3.5.3 All-Rounder Dashboard**

**3.6 Conclusion**

Our exploration has unveiled patterns and trends in cricket data that have expanded our understanding of the sport. We have shed light on the significance of performance metrics such as batting average, strike rate, bowling average, and economy rate in assessing players' contributions. This deeper understanding can profoundly impact the strategies devised by coaches and analysts, ultimately influencing match outcomes.

The predictive models developed in this project represent a glimpse into the future of cricket analysis. By forecasting player performances, we open the door to a new dimension of preparation and decision-making. Teams can make tactical choices based on anticipated player contributions, offering an edge in an ever-competitive cricketing landscape.

In the end, the stumps remain set, the fielders are poised, and the players are ready, with data as the new ally. This project has paved the way for a new era in cricket analysis, where the game's essence is enriched by data-backed insights, and where cricket continues to captivate the hearts of millions worldwide.

be translated in different language.

# Recommendations

1.A Study on Performance of Cricket Players using Factor Analysis Approach

Authors - Sricharan Shah, Partha Jyoti Hazarika and Jiten Hazarika Department of Statistics Dibrugarh University Dibrugarh, Assam,

We recommend the papers our attempt has been made to study the performance of Cricket Players using Factor Analysis technique. The dataset of 85 batsmen, 85 bowlers; and 95 batsmen, 95 bowlers have been considered from IPL9, 2016 (20 overs) and ICC World Cup, 2015 (50 overs) respectively, and the findings of this study reveals that batting capability dominates over bowling capability which is in conformity with an earlier study on same kind of game.

2.Cricket Player Analytics using DAX

April 2023International Journal of Advanced Research in Science Communication and Technology

Authors:

S Pavan Kumar, P Mannohar Sai, R Sai Murali Krishnam Naidu, P Uma Maheswara Rao, K Ravi Raju

We recommend the above papers as Cricket is a hugely popular sport, the popularity of the shorter forms of cricket, and particularly T20 cricket, is undoubtedly increasing apparently complicated the process of player selection. Visual Insights of players performance help in find out the best players. Data Analysis Expressions and Data Visualization has the potential to revolutionize the pruning process by creating the insights from huge datasets. The goal of the project is to create dashboards using Data Analysis Expressions and Microsoft power bi to determine the player analytics on website that can be easily available for everyone. The project is divided in to five dashboards. The first module focuses on selecting a team from total players. The second dashboards comprise of entire matches summary that exist in the dataset. The third dashboard provides the players who could have the potential to hold the winning possibilities over 90 percent. The fourth dashboard provides the analytics of every player. The final dashboard generates analytics based on the user requirements.

# Appendix

1. Data Quality and Collection: Ensure that data collection is comprehensive and accurate, covering a wide range of player statistics and match details.

2. Exploratory Data Analysis (EDA): Conduct in-depth EDA to uncover hidden patterns, outliers, and relationships in the data.

3. Feature Engineering:

- Create meaningful performance metrics that reflect a player's overall contribution to the team.

- Develop new statistics or indices that can provide additional insights into player performance.

4. Machine Learning Models:

- Develop predictive models that can forecast player performance based on historical data.

- Evaluate and fine-tune these models to enhance their predictive accuracy.

5. Data Visualization:

- Use effective data visualization techniques to represent performance insights.

- Create interactive dashboards for quick and easy access to information by different stakeholders.

6. Player Profiling:

- Develop profiles for individual players, highlighting their strengths, weaknesses, and historical performance trends.

- Identify key performance indicators for each player.

7. Comparative Analysis:

- Compare players' performances within the team and against competitors.

- Identify standout performers and areas for improvement.

8. Team Strategy and Player Selection:

- Provide recommendations for team strategies based on player performance trends.

- Suggest optimal player combinations for specific match situations.

9. Predictive Performance Assessment:

- Offer insights into how specific players are likely to perform in upcoming matches.

- Use predictive models to aid in team selection.

10. Continuous Improvement:

- Emphasize the need for ongoing analysis and adjustment of strategies based on evolving player and team dynamics.

- Recommend periodic updates to the analysis to reflect changing performance patterns.

11. Accessibility:

- Ensure that the analysis is easily accessible to a wide audience, including cricket enthusiasts, analysts, and team decision-makers.

- Consider using user-friendly platforms for sharing insights.

13. Future Work:

- Highlight potential areas for future research or expansion of the project.

- Consider integrating more advanced statistical techniques or data sources for further insights.

# Glossary

* Batting Average: The average number of runs a batsman scores per dismissal.
* Strike Rate: The rate at which a batsman scores runs, often calculated as runs scored per 100 balls faced.
* Bowling Average: A measure of a bowler's effectiveness, calculated as the average number of runs conceded per wicket taken.
* Economy Rate: The average number of runs conceded by a bowler per over bowled.
* Player Profiling: The process of creating comprehensive profiles of cricket players, including their strengths, weaknesses, and historical performance.
* Comparative Analysis: The practice of comparing performances of different players, teams, or matches to identify patterns and trends.
* Predictive Modeling: Using statistical and machine learning models to make predictions about future cricket performances based on historical data.
* Data Visualization: The use of graphs, charts, and other visual aids to represent data and insights effectively.
* Data Collection and Preparation: Steps taken to gather and clean cricket data for analysis.
* Feature Engineering: Creating new features or metrics to enhance the analysis.
* Machine Learning Models: Algorithms used to make predictions and generate insights from the data.
* Actionable Recommendations: Suggestions for improving team strategy and player selection based on the analysis.

**References**

[1] Vidit Kanungo, Tulasi B., “Data visualization and toss related analysis of IPL teams and batsmen performances”, International Journal of Electrical and Computer Engineering, Vol. 9, No. 5, October 2019.

[2] Shubhra Singh, Parmeet Kaur, “IPL Visualization and Prediction Using Hbase”, Information Technology and Quantitative Management, 2017.

[3] S. Sharuka, R.Vani,” Insights on IPL Team Performance using Visual Analytics”, International Journal of Engineering Sciences & Research Technology, November, 2019.

[4] D. Parker, P. Burns and H. Natarajan, \"Player valuations in the Indian Premier League\", Frontier Economics, vol. 116, October 2008.

[5] Tijender Singh, et al., “Score and Winning Prediction in Cricket through Data Mining,” International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Oct 2015.

[6] Hemanta Saikia and Dibyojyoti Bhattacharjee, “An application of multilayer perceptron neural network to predict the performance of batsmen in Indian premier league,” International Journal of Research in Science and Technology, vol. 1, no. 1, 2014.