



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Applying Machine Learning in Cricket for Data Visualization and Prediction

Logeshwari K, Swetha Raju S, Varshini M, Mr. Ravikumar B

UG Student, Department of CSE, Velammal Engineering College, Surapet, Chennai, India

UG Student, Department of CSE, Velammal Engineering College, Surapet, Chennai, India

UG Student, Department of CSE, Velammal Engineering College, Surapet, Chennai, India

Assistant Professor, Department of CSE, Velammal Engineering College, Surapet, Chennai, India

ABSTRACT: Predicting the future sounds like magic whether it be detecting in advance the intent of a potential customer to purchase your product or figuring out where the price of a stock is headed. If we can reliably predict the future of something, then we own a massive advantage. Machine learning has only served to amplify this magic and mystery. Player selection is one the most important tasks for any sport and cricket is no exception. The performance of the players depends on various factors such as the opposition team, the venue, his current form etc. The team management, the coach and the captain select 11 players for each match from a squad of 15 to 20 players. They analyse different characteristics and the statistics of the players to select the best playing 11 for each match. Each batsman contributes by scoring maximum runs possible and each bowler contributes by taking maximum wickets and conceding minimum runs. This paper attempts to predict the performance of players as how many runs will each batsman runrate and how many wickets will each bowler take for both the teams. Both the problems are targeted as classification problems where number of runs and number

Keywords: Datasets, Statistics, Machine Learning, Strike rate, Player Performance.

I.INTRODUCTION

A) Area of Specialization

In Computer Science, Machine learning model predictions allow sports to make highly accurate guesses as to the likely outcomes of a question based on historical data, which can be about all kinds of things— customer churn likelihood, possible fraudulent activity, and more. These provide the business with insights that result in tangible business value. Our model predicts run rate, strike rate, bowler's economy etc. The main aim is to predict the match outcome, performance of each player based on the historical data.

Challenges Cricket can be considered as a very unpredictable sport. The whole outlook of a Cricket match can be changed within a few minutes. Due to the complex nature of the game, decisions on team selection, player performance prediction, match outcome prediction can be tough. As in other sports, every Cricket match leaves behind a huge set of data that can be analyzed and modeled to extract data driven insights of the game. These insights can be very helpful to anyone who's involved in any decision-making process related to the game. Cricket is a dynamic game. A team might seem to be way ahead at the halfway stage or at any stage of the game but an extraordinary performance from one player on the other team can change the outcome of the match within a few minutes. Also, various factors such as natural elements, complex rules regulating the game and the performance of players on a given day etc. play a pivotal role in the outcome of a match. Given the array of factors affecting the game and also its dynamic nature, predicting the outcome of a Cricket match is a challenging task.

II. RELATED WORKS

In paper [1] the author proposed a method Using recursive feature elimination, they considered top 5 parameters for validating their result. Algorithms used were linear regression and support vector machine. Cricket is one of the most popular sports in the contemporary world. The ebullience of securing victory in a particular match has motivated the rudimentary part of this research aspect. In paper [2] the author developed a software tool called CricAI. This tool outputs the probability of victory in an ODI cricket match using input factors such as home game advantage available at the beginning of the match. The CricAI tool can be used in real-world applications by teams playing cricket. It can accordingly be helpful in adjusting certain factors in order to maximize the chances of winning the real game. In paper [3] about the prediction techniques used in all the sports, algorithms used and their efficiency. Since the project we are doing is related to cricket, we concentrated more on cricket in the paper. According to this paper, cricket is the 2nd last of all sports for prediction whereas soccer and basketball tops the list. There were several algorithms described in this paper like Artificial neural network, Decision tree, KNN, Support vector machine, boosting, linear regression, random forest and bayes ensemble.

III. PROPOSED WORK

Hence, we have decided to implement a webpage using HTML and CSS for front end and Python for Backend using four different Algorithms. Dividing the data frame into training and test dataset in the ratio 80:20 respectively using train_test split from sklearn.model_selection. The dependent variable here is strike rate, run rate, economy of the bowler and wickets. Here Ground name and Playing against are string values. But for a machine its hard or impossible to learn the string values, so we assigned an integer value for each string value present in the dataset. We trained the model to calculate run rate, strike rate, wickets, economy of the bowler and to select best playing 11 using four different algorithms namely Decision trees, Naives Bayes algorithm, KNN Algorithm, Gradient Boosting Algorithm, Random Forest Algorithm. The algorithm predicts and the output will be displayed on the webpage based on the user input.

IV. ALGORITHMS USED

A) KNN Algorithm

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using this algorithm. KNN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.

B) Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning

technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

C) Linear Regression

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped straight line representing the relationship between the variables.

D) Decision Tree

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

E) Gradient Boosting

Gradient descent is an optimization algorithm that's used when training a machine learning model. It's based on a convex function and tweaks its parameters iteratively to minimize a given function to its local minimum. Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

V. DESIGN AND ARCHITECTURE

A) DATA COLLECTION

The data set collected for cricket prediction is split into Training set and Test set. The data from 132 matches were taken and put together in the dataset for prediction. The Data Model created using Random Forest, KNN regression, Decision tree, Linear regression and Gradient booster algorithms are applied on the Training set and based on the test result accuracy, Data test set prediction is done.

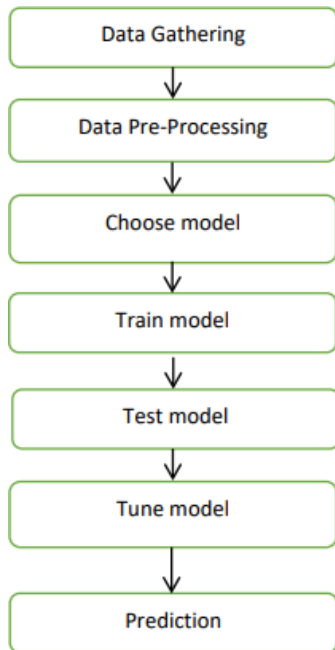


Fig.1. Data Collection

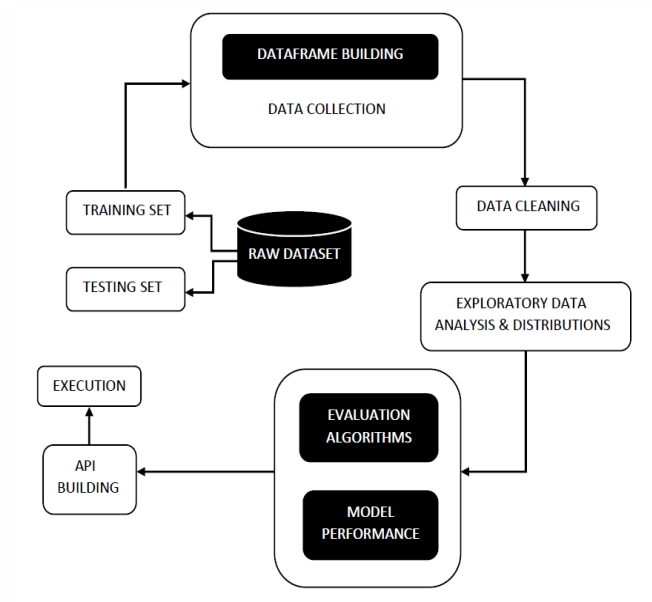


Fig.2. Architecture

For calculating strike rate, run rate of the team, economy of the bowler, wickets and Best 11 players of the team five Algorithms are used. They are.

1. KNN
2. RandomForest
3. LinearRegression
4. DecisionTree
5. GradientBoosting

B) MEASUREMENTS RUNRATE

In cricket, the run rate, or runs per over, is the average number of runs a batting side scores per over. It includes all runs made by the batting side in the innings to that point of the game, both the runs scored by the batsmen and extras conceded by the bowling team.

STRIKE RATE

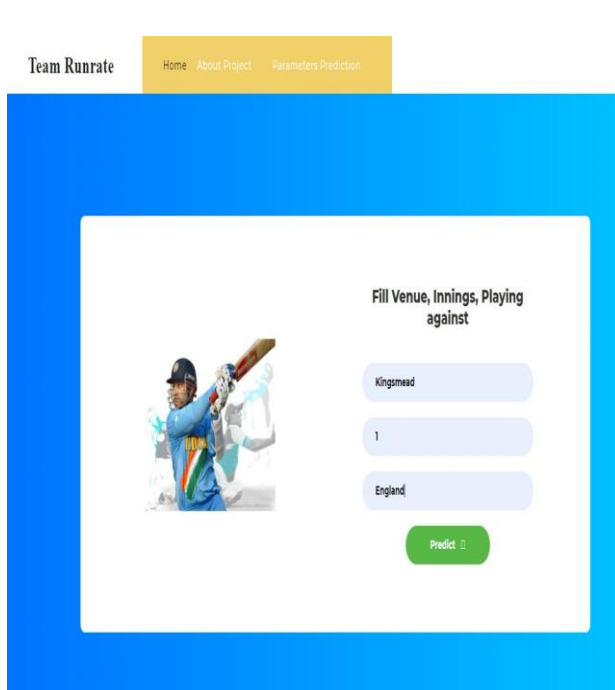
Strike rate refers to two different statistics in the sport of cricket. Batting strike rate is a measure of how quickly a batsman achieves the primary goal of batting, namely scoring runs. Bowling strike rate is a measure of how quickly a bowler achieves the primary goal of bowling, namely taking wickets.

BOWLERS ECONOMY

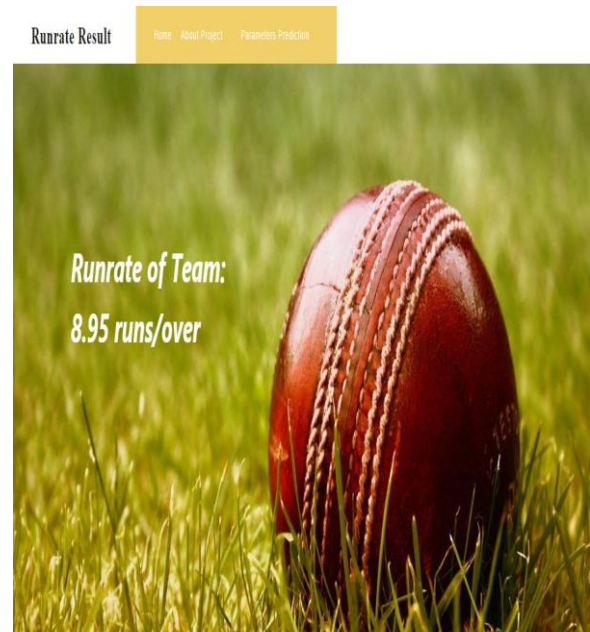
In cricket, a player's economy rate is the average number of runs they have conceded per over bowled. It is one of several statistics used to compare bowlers, commonly used alongside bowling average and strike rate to judge the overall performance of a bowler.

VI IMPLEMENTATION

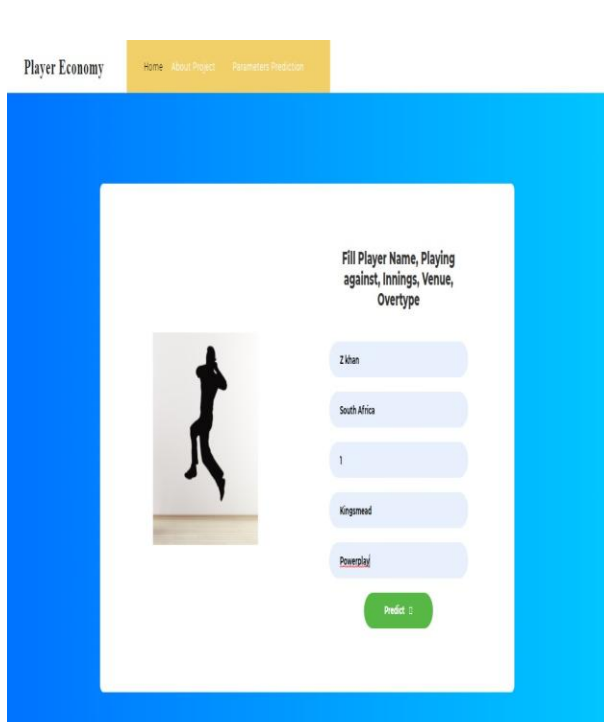
A) Team Runrate



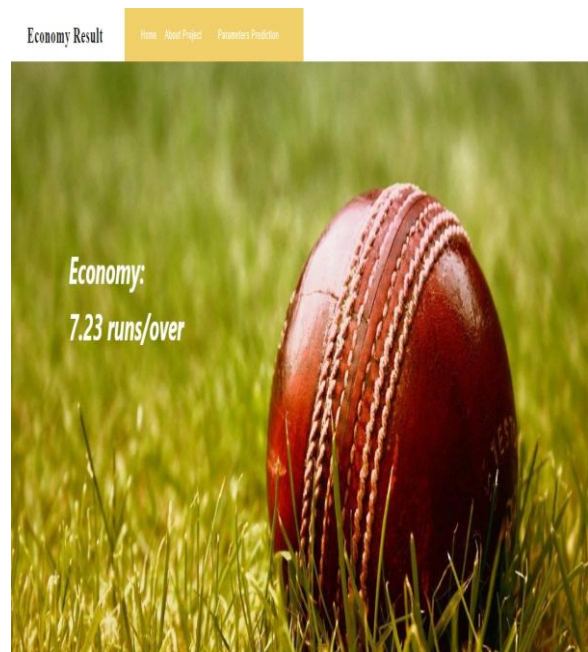
The screenshot shows the 'Team Runrate' prediction interface. It features a navigation bar with 'Home', 'About Project', and 'Parameters Prediction'. The main content area has a blue background and contains a white form. On the left of the form is an image of a batsman. The form prompts the user to 'Fill Venue, Innings, Playing against' and includes three input fields: 'Kingsmead', '1', and 'England'. A green 'Predict' button is located at the bottom of the form.



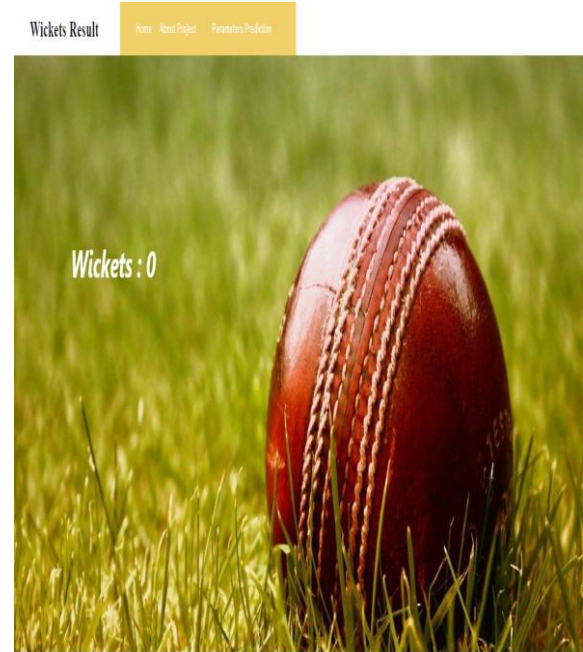
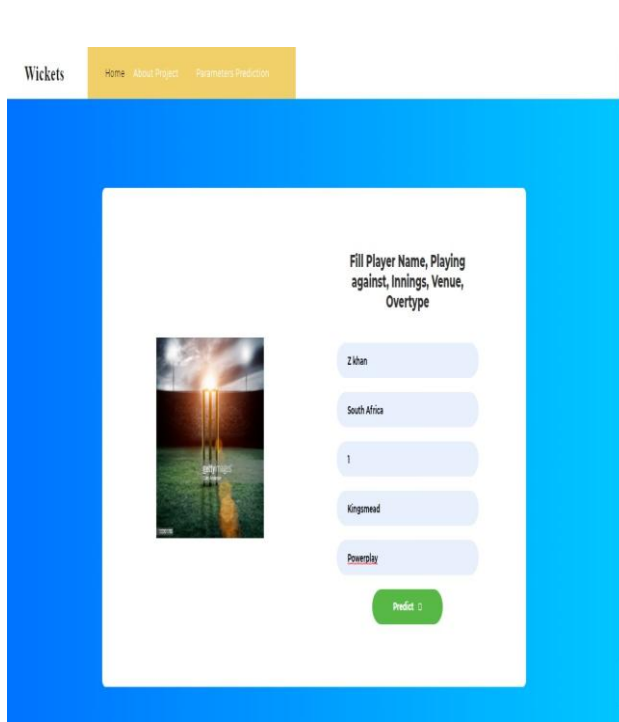
B) Player Economy



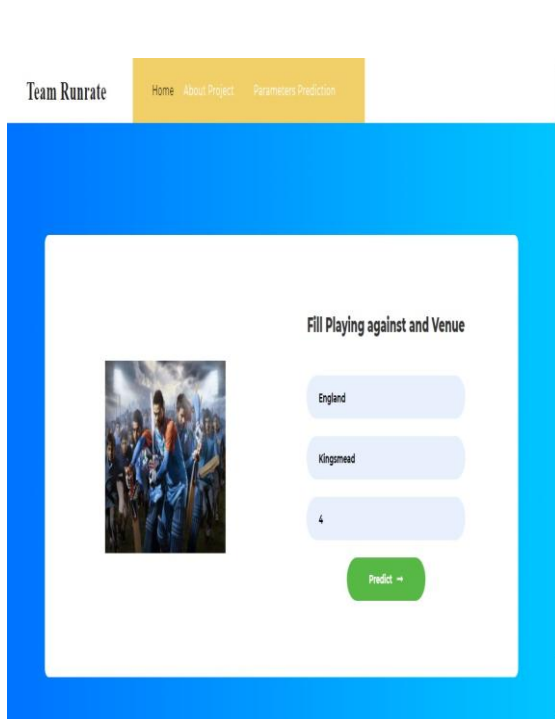
The screenshot shows the 'Player Economy' prediction interface. It features a navigation bar with 'Home', 'About Project', and 'Parameters Prediction'. The main content area has a blue background and contains a white form. On the left of the form is an image of a bowler. The form prompts the user to 'Fill Player Name, Playing against, Innings, Venue, Overtime' and includes five input fields: 'Z Khan', 'South Africa', '1', 'Kingsmead', and 'Powerplay'. A green 'Predict' button is located at the bottom of the form.



C) Wicket



D) Best 11



VII CONCLUSION & FUTURE WORKS

A) CONCLUSION

Player's selection possesses a vital role in the team's triumph. The selection committee board member, coach and captain of the team is responsible for selection of the best players for team for each match. The player's performances depend on various factors such as the location where the match being played, past records, his current form, average rate, strike rate, runs scored at a particular venue, number of innings played against the opposition teams etc. Taking into consideration this information they employ an accurate prediction model which predicts the accuracy of the batsmen and bowlers. In this project we modeled datasets based on player's earlier record. Decision Tree, Naïve Bayes, Random Forest and support Vector Machine supervised machine learning algorithms were evaluated and used. Random forest algorithm found to be produced more accurate and useful outcome among the other classifier algorithms. Whereas the SVM produces unexpected and less useful results. This model works well with further format of cricket i.e., "T20 matches" and "Test series matches" and equivalent procedure can be applied to these 2 formats of game.

B) FUTURE WORKS

1. To improve efficiency, we can possibly use the players' data to assess the quality of each team player.
2. Trying more complex Machine Learning algorithms like XGBoost and fine-tuning the hyperparameters
3. A confusion matrix would be great to analyse which games the model got wrong.
4. We could ensemble that is we could try stacking more models together to improve the accuracy.
5. Going even further and making a model based on player statistics.

REFERENCES

- [1] Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," Published year: 2018
- [2] A. Kaluarachchi and S. V. Aparna, "CricAI: A classification-based tool to predict the outcome in ODI cricket" Published year: 2010
- [3] Brooks, R. D., Faff, R. W., & Sokulsky, D. An ordered response model of test cricket performance. Published Year: 2002
- [4] The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review Published Year: 2019
- [5] Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms Aminul Islam Anik, Sakif Yeaser, +1 author A. Chakrabarty Published Year: 2018
- [6] Kampakis, S. and Thomas, W. (2018). Using Machine Learning to Predict the Outcome of English County Twenty Over Cricket Matches.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details