



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Simulating the Final Score: User-Driven IPL Score Prediction Using Machine Learning

Jadage Sangeeta Lagamna, Jahagirdar Aabha Nandkumar, Pagadyakul Ganesh Anil, Awale

Darshan Amogsiddh, Thamake Mrunal Sachin

Guide, Lecturer, Department of Computer Science, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India^{*1}

Student, Department of Computer Science, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India^{*2}

Student, Department of Computer Science, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India^{*3}

Student, Department of Computer Science, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India^{*4}

Student, Department of Computer Science, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India^{*5}

ABSTRACT: Fans all over the world are captivated by the intense cricket competition known as the Indian Premier League (IPL). For fanatics, there's an additional level of thrill when they can predict the result of these contests. In order to anticipate IPL match scores, this project explores the field of cricket analytics by creating a model based on Python. The model makes use of numerous variables that have a big impact on how a match plays out. Among these are:

Team Dynamics: A significant factor in deciding the ultimate score is the past results of the bowling and batting teams.

Pitch Conditions: A pitch's ability to support batting or bowling can have a big impact on how many runs are scored.

Match Progression: The pressure on the batting team and scoring rate are reflected in the current over (beginning innings, middle over, and death over).

Wicket Loss Instinct.

The model seeks to understand the intricate relationships between different game elements and how they affect the overall score by examining five crucial elements. This research offers a data-driven method for predicting IPL match scores, adding to the expanding field of sports analytics.

KEYWORDS: Machine Learning, sports analytics, linear regression, First Innings, IPL score prediction.

I. INTRODUCTION

The excitement of picking the outcome of your most anticipated IPL match can only be imagined! This project uses machine learning to create a Python-based model for IPL score prediction, bringing that thrill one step closer.

Computers can learn through machine learning, and in this case, the data is all about previous Indian Premier League matches. Information on the teams involved, the kind of surface, and even the number of wickets taken so far will be fed into the algorithm

We'll be employing the specific machine learning method known as linear regression. This facilitates the model's ability to discern connections between the final score and these other elements. The algorithm will learn to "predict" the score for a new match based on comparable circumstances by examining trends in previous matches.

II. DATA RESOURCES

Method Gathering pertinent information is the first stage in forecasting cricket matches. Data from previous matches can offer insightful information on player statistics and team performance. Weather-related information, such humidity and temperature, can also affect how a match turns out also the pitch conditions also affect the final result of the match. A key component of creating a successful ML model for cricket match prediction is choosing the right dataset.

Data Quality: The dataset should be of high quality, with minimal errors and missing values.

Data Availability: The dataset should contain a sufficient number of observations to train the AI model effectively. Ideally, the dataset should cover a wide range of matches.

Relevant Features: The dataset should contain relevant features that are known to impact the outcome of a cricket match.

Data Granularity: The dataset should have the appropriate granularity to support the prediction task.

Overall, selecting a suitable dataset for cricket match prediction using ML requires careful consideration of several factors, including data quality, availability, relevance, granularity, and diversity. By selecting a suitable dataset, ML models can be trained effectively to predict the outcome of cricket matches accurately.

III. PREPROCESSING

The next step in predicting cricket matches is to pre-process the data. Data pre-processing involves cleaning the data, handling missing values, and transforming the data into a format suitable for machine learning algorithms. Cleaning the data involves removing irrelevant data and correcting errors. Here's an example of a cricket dataset that can be used to train AI models for predicting cricket match outcomes:

Match details: This includes information about the match, such as the date, venue, and format (e.g., Test, ODI, T20).

Team statistics: This includes information about the performance of the two teams, such as the win-loss record, batting and bowling averages, and run rate.

Player statistics: This includes information about the performance of individual players, such as batting and bowling averages, strike rates, and player rankings.

Pitch conditions: This includes information about the pitch conditions, such as the type of pitch, grass cover, and weather conditions.

Toss details: This includes information about which team won the toss and chose to bat or bowl first.

Inning details: This includes information about each team's innings, such as the total runs scored, wickets taken, and run rate.

Outcome details: This includes information about the outcome of the match, such as which team won, the margin of victory, and the number of over remaining.

IV. DATASET FEATURES

We are employing an ML-based strategy. Hence, a dataset, algorithmic training of that dataset, and model testing are the fundamental needs of a machine learning algorithm. Thus, we have imported the Kaggle dataset. Later on, using Linear Regression for scoring prediction, the accuracy was calculated and improved.

Score Prediction: For this project, we have collected data of all the IPL matches played from 2008 to 2017. The dataset consists of 76015 numbers of rows. Dataset consists 15 columns over which we applied feature selection techniques and selected 8 features in which all 8 are input feature. The attributes selected were bat team, bowl team, type of pitch over, runs, wickets, runs in previous 5 over, wickets in previous 5 over for score prediction.

Table 1. Dataset attributes and their values

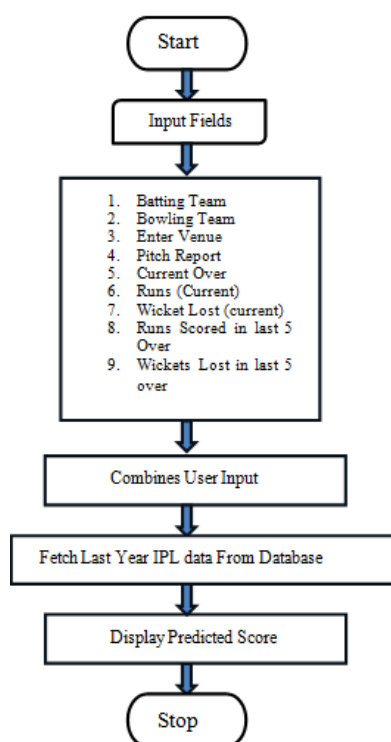
Attributes	Values
Batting team	Batting team names among 8 names in IPL
Bowling team	Bowling team names among 8 names in IPL
Venue	The stadium name
Pitch	The type of pitch among 6 types of pitches
Over	Value > 5 over
Runs	0-770
Wickets	0-10
Runs scored in previous 5 over	0-770
Wickets lost in previous 5 over	1-10
Total runs	0-770

V. FEATURE SELECTION

The features taken in this project are the team name that is the batting and bowling team name, the name of the stadium, the type of pitch, the current over (the current over should be greater than 5 or equal because the previous 5 over are important for prediction), runs scored in current over, wickets lost in current over, runs scored in previous 5 over, wickets lost in previous 5 over. These inputs would help predict the score range of the inputted batting team.

VI. BLOCK DIAGRAM AND METHODOLOGY

Using the pandas library, import the dataset in the first place. Next, pre process the dataset by identifying any null values and replacing them with the mean or median values for each corresponding column. Number values are assigned to the categorical data in the columns. Subsequently, the dataset is subjected to feature selection procedures, which identify only the best collection of characteristics. The dataset defines the set of input features (X) and the output (Y). The output feature is dependent upon the input features, but the input features are independent of one another. Following the import of the library and the creation of the ML Model, the data is split into training and testing subsets using the train-split-test approach. Training the ML Model with training data and making predictions with test data is the next step. Subsequently, the model's correctness can be determined by calculating the ratio between the anticipated and actual testing results. Every ML algorithm does this process again, and the accuracy of each algorithm is determined. Ultimately, an accuracy comparison is conducted between the algorithms to identify the optimal approach for this dataset.



VII. ALGORITHM

We have used the regression algorithm that is the Linear Regression.

Regression : Regression analysis predicts a continuous value by utilizing many algorithms for calculation. The goal variable is the continuous range value, and a specific set of variables are used as the input. Regression methods are utilized in different ways depending on the application. Different regression approaches exist, of which the score is predicted using the ridge, lasso, and linear methods.

Linear regression : Linear regression is used to predict continuous values. The machine learning algorithms are given some known parameters, and their output is a prediction of continuous values. It is not applicable to the classification issues. The suggested model uses linear regression to predict the score.

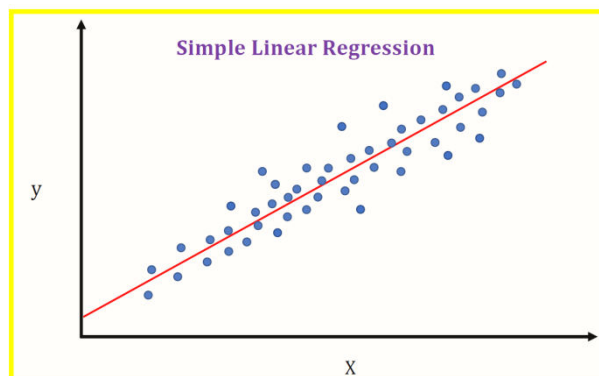


Figure 2 : Linear Regression

Ridge regression: The continuous values can also be predicted using ridge regression. Ridge regression is utilized when multicollinearity (correlations between predictor variables) is present in the data and the variables used for the prediction are more than the observations of multicollinearity.

Lasso regression : This kind of linear regression is employed to forecast continuous values. Within the lasso regression, shrinkage is employed. Shrinkage happens when data values are focused on the focal point. Data values that shrink toward a mean or other center point are referred to as shrinkage. Simple, sparse models are encouraged by the lasso approach.

Score Prediction algorithms : It is found that for the score prediction the linear regression is giving the more accuracy as compared to Ridge regression and Lasso regression.

Table II
Accuracy of score predicting models

Algorithm	Accuracy
Linear Regression	80.92
Lasso Regression	80.45
Ridge Regression	80.84

As we can see, the linear regression produces the best accuracy result for the score prediction. Thus, the following is the linear regression formula used to obtain the theoretical result:

$$y = B_0 + B_1 * x \dots \text{(linear regression equation)}$$

So here, y is the dependent variable

x is independent variable

B₀ is bias coefficient &

B₁ is coefficient of x

VIII. IMPLEMENTATION OF GUI



Figure 3 : Score Prediction Model UI



Figure 4 : Input to Score Prediction Model

The model did not produce the exact output that was predicted. Thus, in order to improve the model's accuracy, we add and subtract 3 to get the score range that is depicted in Fig. 5. Therefore, our model is generally effective.

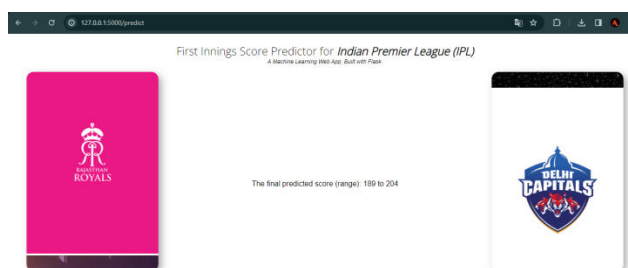


Figure 5 : Score Prediction Result

IX. LIMITATION

There are certain restrictions when it comes to using AI methods to forecast cricket matches. The availability of data is one of the primary constraints. Data can be scarce, particularly for players and teams with less notoriety. In our project the data set which we have used is from 2008 to 2017. The reason for selecting this dataset is the alteration in the team players, in the current years the team configuration changed as there was alteration in the team players. The opaqueness of machine learning models is another drawback. Finding the elements that go into the model's prediction can be tough with machine learning models since they can be intricate and tricky to understand. In real time, as years pass by, the age of players should be taken in consideration. As the players get old, their performance in the matches get affected a bit. Also this prediction carried out on the current features would vary if the team players change dynamically.

X. FUTURE RESEARCH DIRECTION

There is potential for further research in the domain of artificial intelligence approaches applied to cricket match prediction. Subsequent investigations may concentrate on creating novel machine learning models that can manage incomplete data and enhance forecast precision. Research may also concentrate on finding novel characteristics, including athlete fitness levels and injury histories that can raise forecast accuracy. Cricket is a complicated sport with a lot of non-linear feature connections. More precise forecasts and insights can be obtained by investigating non-linear relationships.

XI. CONCLUSION

In conclusion, the use of AI in predicting cricket matches has shown promising results. Various AI techniques such as Logistic Regression have been used to predict the outcome of cricket matches. The data used in cricket match prediction includes team names that is batting team and balling team names, pitch conditions, the current over, runs scored in the current going over, wickets lost in the current going over, runs scored in previous five over and the wickets ost in the previous five over. A case study was presented on using Decision Trees to predict the outcome of T20 cricket matches. While there has been significant progress in using AI for cricket match prediction, there are still many research areas that require further investigation. The use of advanced AI techniques, integration of unstructured data, incorporation of real-time data, and exploring non-linear relationships are some of the potential future research directions. The future of AI in cricket match prediction is bright, and we can expect further advancements in the field in the coming years.

XII. ACKNOWLEDGEMENT

First of all we would like to acknowledge our parents for supporting us in building of this project. They have provided all the materials required for our project and have helped us financially for this project. We would like to thank our guide Prof. S.L. Jadage who guided as throughout the project.

REFERENCES

1. T. Singh, V. Singla and P. Bhatia, "Score and winning prediction in cricket through data mining," 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI), Faridabad, India, 2015, pp. 60-66, doi: 10.1109/ICSCTI.2015.7489605.
2. J. Kumar, R. Kumar and P. Kumar, "Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 343-347, doi: 10.1109/ICSCCC.2018.8703301.
3. Kaluarachchi and S. V. Aparna, "CricAI: A classification based tool to predict the outcome in ODI cricket," 2010 Fifth International Conference on Information and Automation for Sustainability, Colombo, Sri Lanka, 2010, pp. 250-255, doi: 10.1109/ICIAFS.2010.5715668.
4. I. Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT), Dhaka, Bangladesh, 2018, pp. 500-505, doi: 10.1109/CEEICT.2018.8628118.
5. N. Rodrigues, N. Sequeira, S. Rodrigues and V. Shrivastava, "Cricket Squad Analysis Using Multiple Random Forest Regression," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 104-108, doi: 10.1109/ICAIT47043.2019.8987367.
6. M. Jhavar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach", European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Riva del Garda, 2016
7. I. Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT), Dhaka, Bangladesh, 2018, pp. 500-505, doi: 10.1109/CEEICT.2018.8628118.



8. Rameshwari Lokhande, P. M. Chawan "Live Cricket Score and Winning Prediction" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, Volume-5 | Issue1 , February 2018, URL: <http://www.ijtrd.com/papers/IJTRD12180.pdf>
9. H. Barot, A. Kothari, P. Bide, B. Ahir and R. Kankaria, "Analysis and Prediction for the Indian Premier League," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9153972.
10. Basit, M. B. Alvi, F. H. Jaskani, M. Alvi, K. H. Memon and R. A. Shah, "ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques," 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318077.
11. Bandulasiri, "Predicting the Winner in One Day International Cricket", Journal of Mathematical Sciences & Mathematics Education, Vol. 3, No. 1.
12. Analysis and Prediction of Cricket Statistics using Data Mining Techniques Anurag Gangal VESIT, Mumbai Abhishek Talnikar VESIT, Mumbai Aneesh Dalvi VESIT, Mumbai Vidya Zope VESIT, Mumbai Aadesh Kulkarni VESIT, Mumbai.
13. S. Agrawal, S. P. Singh and J. K. Sharma, "Predicting Results of Indian Premier League T-20 Matches using Machine Learning," 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2018, pp. 67-71, doi: 10.1109/CSNT.2018.8820235



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details