Automated Cricket Score Prediction

Polepaka Sanjeeva1*, Jampana Ajith Varma1, Valaparla Sathvik1, Attemla Abhinav Sai Ratan1, Sanjay Mishra2

1Department of CSE (AI & ML), GRIET, Hyderabad, Telangana State, India
2Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, 248007, India

Abstract. Cricket is a popular sport that involves a high degree of variability in terms of game conditions and player performance. The ability to accurately predict cricket scores could provide valuable insights for coaches, analysts, and fans, as well as offer opportunities for sports betting and fantasy games. This paper explores the use of machine learning techniques to predict cricket scores based on a variety of contextual and historical factors. The publicly available cricket dataset is used to build and evaluate several regression models that predict the total runs scored by a team in a limited-overs cricket match. This analysis includes feature engineering to extract and transform relevant input variables, model selection to compare and choose among different regression algorithms, and performance evaluation to assess the accuracy and robustness of the models. This paper also conducts sensitivity analysis to identify the most influential predictors and explore the potential biases and limitations of the models. The results indicate that machine learning techniques can effectively predict cricket scores and provide valuable insights into the factors that contribute to team performance. Automated cricket prediction is useful for cricket teams, coaches, and analysts who seek to improve their game strategies and player selection, as well as for sports betting and fantasy game platforms that seek to provide more accurate experiences for users.

1 Introduction

Cricket is a popular sport that involves a high degree of variability in terms of game conditions and player performance. The ability to accurately predict cricket scores could provide valuable insights for coaches, analysts, and fans, as well as offer opportunities for sports betting and fantasy games. This paper also conducts sensitivity analysis to identify the most influential predictors and explore the potential biases and limitations of the models. The results indicate that machine learning techniques can effectively predict cricket scores and provide valuable insights into the factors that contribute to team performance. The findings have implications for cricket teams, coaches, and analysts who seek to improve their game strategies and player selection, as well as for sports betting and fantasy game platforms that seek to provide more accurate and engaging experiences for users. Automated cricket score prediction has a fascinating history, evolving alongside the

* Corresponding author: sanjeeva1690@grietcollege.com
sport itself. Early on, cricket enthusiasts relied on their intuition and expertise to predict match outcomes. These models continuously learn and refine their predictions, enhancing their accuracy over time. Automated cricket score prediction has become an integral part of the sport, generating excitement and anticipation among fans worldwide.

The working principle of cricket score prediction involves analysing a multitude of variables to estimate the likely outcome of a match. Statistical models and machine learning algorithms process historical data, including team and player performance, venue conditions, weather patterns, and more. These models identify patterns, trends, and correlations within the data to generate predictions. Advanced algorithms continuously learn and adapt based on new information and real-time updates during matches. The advantages of the proposed work are as follows:

1. Enhanced fan experience- Exciting predictions for cricket scores.
2. Informed betting decisions- Reliable predictions for better bets.
3. Fantasy sports strategy- Optimize your fantasy team with accurate predictions.
4. Strategic team decisions- Make informed choices based on predictions.
5. Statistical analysis- Valuable insights for player performance evaluation.

2 Existing methods

Existing approaches for cricket score prediction encompass a range of methodologies. Statistical analysis involves analysing historical data to identify patterns and trends. Machine learning techniques, such as regression models, decision trees, and neural networks, are employed to learn from past data and generate predictions. Some approaches incorporate complex algorithms that consider multiple factors like player performance, team dynamics, venue conditions, and weather. Data-driven models are continuously refined and updated with real-time information during matches.

Authors [6] highlighted the significance of ML in prediction, pattern recognition and error reduction across diverse fields, emphasizing the impact of AI in broad domain. The approach [7] utilized Advanced Deep Learning with global threshold to improve E-commerce product classification, achieving high accuracy and challenging existing technology. Author [8] presented text classification algorithms for various applications and explores the use of machine learning in detecting phishing attacks. Sentiment analysis evaluates feelings of mankind based on reviews or comments, followed by categorizing them as negative, neutral, or positive. The present work explores sentiment analysis architecture and tools for user-friendly opinion mining [9]. Authors [10] discussed the use of machine learning and neural networks, especially CNN, for recognizing handwriting patterns, with a focus on Telugu film industry names, achieving high accuracy (98.3%). Authors [11] presented data-driven prediction techniques, namely, ARIMA and LSTM to forecast COVID-19 cases and deaths. Further, it uses statistical measures to assess accuracy and aims to assist several countries in managing the pandemic.

<table>
<thead>
<tr>
<th>Ref. No.</th>
<th>Description</th>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>Score and Winning Prediction in Cricket through Data Mining Oct 8-10, 2015</td>
<td>Linear Regression algorithm, Naive Bayes classifier</td>
<td>Accuracy of LR is 20% greater than CRR method</td>
</tr>
<tr>
<td>[2]</td>
<td>Money Ball - Data Mining on Cricket Dataset 2019</td>
<td>Naive Bayes classifier, Random Forest method</td>
<td>Linear regression 80.76, Ridge regression 80.69, Lasso regression 81.00</td>
</tr>
</tbody>
</table>

Table 1. Existing approaches.
3 Problem statement and objectives

3.1 Problem statement

In the profession of cricket score prediction, a variety of strategies are used to forecast the innings score of a cricket match. Numerous systems and prediction computations are used to forecast the outcomes of ODI and T20 cricket matches. When predicting the results of cricket matches, the CRR technique is frequently used. The total number of overs in an inning is multiplied by the number of runs scored in an over in the CRR technique. This method excludes all other factors and only considers runs scored in an over.

3.2 Objective

The objective of automated cricket score prediction is to provide accurate and reliable estimates of the final score or match outcome. By leveraging statistical analysis, historical data, and advanced algorithms, the goal is to enhance fan engagement and excitement during matches. Predictions aid bettors in making informed decisions, fantasy sports players in strategizing their teams, and teams in analysing opponents and devising strategies. Additionally, automated score prediction aims to offer valuable statistical insights for player performance evaluation and assist broadcasters in providing informative and engaging commentary.

4 Proposed method

The paper aims to develop an automated cricket score prediction system that utilizes statistical analysis, machine learning, and historical data to generate accurate predictions for match outcomes. The system will consider various factors such as team performance, player statistics, pitch conditions, weather, and other relevant variables. Through the implementation of advanced algorithms and continuous learning, the system will refine its predictions over time. The paper will focus on creating an intuitive user interface for easy access to predictions and integrate real-time updates during matches. The ultimate goal is to provide cricket fans, bettors, and fantasy sports players with valuable insights and enhance their overall experience of the game.
Data collection we will obtain the data-set from the Kaggle data-sets. We'll collect the data in CSV format. The following step will involve cleaning the data that was obtained from the website.

Data cleaning we want to remove all unnecessary columns as part of the data cleaning process, including the match id, location, bowler and batsman names, as well as the striker and non-striker scores. These columns will be omitted as estimation won't need them. Some teams no longer engage in the league, in accordance with IPL data. The IPL does not feature the Deccan Chargers, Gujarat Lions, Pune Warriors India, or Rising Pune Supergiant. As a result, we need to exclude those teams from the data collection and only take into consideration those who are reliable teams. After five overs, we will evaluate the data. The date column, which is available in the data collection in string format, is the subject of a variety of manipulations that we would like to carry out.

Data preprocessing, after data has been cleaned, we will need to prepare our data. The process of data preparation will include the one-shot encoding. One hot encoding covers a lot of the actualization portion. We'll need to redesign the columns in our data collection during the preprocessing phase of the data. Adjusting the columns is essential since we need to appropriately arrange our columns in particular series. We will split the data after it has been collected such that IPL games from before 2016 will be used for the model's training and games from after 2016 will be used as test data.

Algorithms For the forecast, we'll use the Random Forest Regression, the Linear Regression model, and the Lasso Regression model. The model with the greatest precision will be used to make the prognosis. The model that we are going to use for the prediction will be explained during action chapter.

Prediction, the model will process the data before gathering user inputs. After gathering user inputs and comparing them to historical data, we will be able to estimate a score range, or from lower bound to upper bound. The model architecture of the CFP system is shown in the diagram above.
The data set underwent one phase of encoding. One-hot encoding requires logical data organization. Because many machine learning algorithms cannot work with categorical input, we encode data into integers. In our data collection, we have columns for the batting and bowling teams. However, when we supply user input, our model must be able to comprehend it. There are multiple squads in both the batting and bowling sections. Because we don't wish to send the batting team and bowling team's input in string format, we rely on one-hot encoding. The encoded data frame is provided to the model as input. After the user submits the form, the model uses the data to estimate something.

5 Experimental results

The data set contains data from the previous five years to forecast scores. The data used for victory predictions spans a period of seventeen years. We split the data into training and testing portions at a ratio of 70 to 30. 70% of the data is used for training, while 30% is tested. Training - Training can be carried out using training data obtained from the data set. The system will learn from the data about the pattern and different relationships with the aid of this training. The testing data is used to test the data after the training to see whether the machine learning algorithm's prediction or computation was correct or incorrect. To train the system, supervised learning provides data with examples and results.

1. Naive Bayes: They are very ascendable and call for a set of parameters that are linearly spaced out from the number of variables in a learning problem. Most of the time, maximum-likelihood coaching involves linearly evaluating a closed-form phrase.

2. To produce a better forecast, linear regression repeats the same task repeatedly. Future values are predicted using this model.

3. Score model: This model displays a numeric number that has been calculated and predicted using a variety of algorithms.

4. Evaluate model - This model is used to evaluate whether the prediction is right or wrong.

![Table](image)

Fig. 2. Result of a match between Royal Challengers Bangalore and Kolkata Knight Riders.
5.1 Significance of the proposed method

The proposed methods are statistical models that include regression analysis, time series analysis and machine learning algorithms. The second one is scikit learn which is one of the libraries used in python. The next one is tensor flow which is a machine learning library. The next one is XG boost which is a boosting library. And the last one is Simulation Models which use computer programs to simulate the outcomes of a game. Below are some advantages of cricket score prediction.
1. We can predict the outcomes of a match even before a match starts.
2. It helps us in changing the game plans and strategies for the team coaches and supporting staff.
3. Model that can be updated along with the instant modifications and changes.
4. Takes all important features into account the players playing in each match.
5. Good accuracy arrived through both prediction models.
6. It may prove helpful for numerous stakeholders to use machine learning to study cricket matches while taking, player performance, archaeological game data, ecological criteria, preceding game conditions, and other features into account.

6 Conclusion and future enhancements

We can use a model to predict the possibility of chasing in the future. In other words, the algorithm may be able to predict whether a team would be effective in achieving the goal. The model utilized in this paper can be made more accurate. The prediction can take into account variables like the venue, the playing surface, and the opposition team. Additional factors such as batsmen partnerships and pitch conditions, can be introduced to further improve the model's accuracy. The future scope for automated cricket score prediction is promising, with several areas offering opportunities for further development, such as, Advanced Machine Learning Techniques, Real-time Data Integration, Integration with IoT and Wearable Technology, Sentiment Analysis, Utilizing automated score predictions for team management, and Personalized Fan Experience.

References


