Data_Analytics_in_Cricket_- _Research_Paper_- _Milan_Sharma_1.pdf

by cse cu

Artificial Intelligence and Data Analytics in Cricket

By Milan Sharma

ME - CSE, University Institute of Engineering, Chandigarh University
Mohali, Punjab, India

ABSTRACT

As we all know cricket is one of the most popular sports in the world. Selecting the right players, team strategy, events, and plays all play an important role in the outcome of a baseball game. All these factors can be improved through the use of new technology. Artificial intelligence (AI) and data science are helping gaming take a step forward. Technologies like Bat Sensor, Ball Sensor, Hotspot, Ultra Edge and many more other technologies help keep the game fair and competitive. The latest machine learning and deep learning technologies also enable teams to be smarter at selecting players, making connections, and deciding on strategies. This technology helps predict the performance of batsmen, bowlers and even fielding teams based on their past performance and various factors. Teams also use strategies to decide when and where to use certain players. With the help of data, the roles of specific actors became clear. England Cricket Team, Nottinghamshire County Cricket Club, Multan Sultans and many more teams have used this technology and achieved success very quickly with limited resources. This article examines the different types of machine learning and deep learning used in the cricket industry. The main methods used are Linear Regression, K Nearest Neighbors (KNN), Decision Trees and similar methods. From the successes of different teams, it can be seen that the knowledge team is one step ahead of its competitors and has become an important part of their success.

KEYWORDS: Research Paper, Technical Writing, Science, Engineering and Technology, AI, Data Analytics

I. INTRODUCTION

1. Overview

MLB's Oakland Athletics' offering of baseball-related information opens new doors for strategy and planning in sports. After that, baseball changed forever. It took a long time, but now luck has come to cricket and. From selecting 11 players to predicting goals and individual performances, data has revolutionized cricket. The new machine learning team applies the same bowling action and speed using the bowling machine. The model provides unbiased statistics to help select the best players. The information can be used to create ideas package by package. Players have now started watching videos specially prepared for them about their opponents and they are now making plans according to their opponents' weaknesses. War has become one of the most important forms of strategy. Left-arm batsmen are increasingly being used against leg-spinners. There are many recent success stories of teams winning great awards with the help of knowledge. All teams have data analysts on staff and are an integral part of their cabinets. The English football team is a prime example of this theory. After their poor run at the 2015 World Cup, they changed their entire squad and came into the game at the 2019 Cup, aided by the record that helped them win their first world title. Many teams in IPL, Islamabad United and Multan Sultans in PSL, many local teams also follow the same philosophy and have achieved great success in recent times.

2. Problem Statement

Like many other fields, data science and artificial intelligence (AI) are now helping improve the game of

cricket. Information is very important, especially in sports. Information plays an important role in creating matches between players and even creating game strategies. The information makes the task of management and training very easy by selecting the best players and then helping them implement the perfect plan. This study aims to examine the different applications and methods of data science and data analysis used in sports today.

3. Objective

Sports information has many uses. The most important thing in cricket is to select the best player among the available players. All major competitive bidding and design processes are carried out using this information. The group wants to put all personal prejudices aside and choose what is best for them. The documents give them Plans A, B, C and the list goes on. In this way, they are ready to compete with the best teams. The information then helps develop and implement plans to win games in the tournament. Just click here for important information about the opponents' gameplay, their best areas, their weaknesses, free and independent players and many more. The team also runs predictive models to simulate similar situations that will occur during game scenarios. The group can understand the problems and tasks they face based on what has happened in the past. Which player is more valuable to the player at which stage of the game? All this information and more can be extracted and used to support your victory.

4. Limitation and Scope

While data is useful, it always has limitations. It can help you be successful most of the time, But in sports the situation changes with the ball. Opponents may change their game . Many other factors may also be at play. So information is always useful, but you need to have a backup plan to deal with your competitors. It can be very useful, but only for those who use it carefully. Otherwise it will haunt you again.

II. LITERATURE REVIEW

A study in (Singh et al., 2015) discusses the prediction of the total number of runs scored by the team in the first and second innings. Early predictions are based solely on a group's run at a particular time. This study also takes into account location, number of wickets and matches lost. The second innings score prediction must take into account more, which is the target. This technique is used by Naive Bayes and Linear Regression classifiers. All ODIs played between 2002 and 2014 used 5 overs in the 50 Overs of each match. The results show that the linear regression model is less prone to error than the current model. Additionally, from the 5th to the 45th time, the accuracy of Naive Bayes increased from 68% to 91%.

This study from (Wickramasinghe et al., 2014) shows how batters' predictions perform across multiple tests. Data is obtained from competitive tests played over a 5-year period. Take into account the different characteristics of the player and the team he plays for. The 3-layer hierarchical model aims to meet the hierarchical structure of the data. This study concluded that batsman and opponent alignment is the main factor affecting player performance.

Another study in (Passi et al., 2018) discussed the difficulty of batting to select the best 11 players for a particular match to help Cov and try to predict the player's performance. The coach and manager decide. The player's past record, current form, opponent and position play an important role in this prediction. The research attempts to estimate the number of runs scored by batsmen and the number of wickets taken and conceded by bowlers. These problems are divided into different methods and various models such as SVM, Naive Bayes, Decision Tree and Random Forest are used to predict the result. This study concluded that Random Forest has the highest accuracy among all these models.

A study conducted in 2011 (Amin et al., 2014) proposed a team approach to selection using Data Envelopment Analysis (DEA). After the evaluation, players can calculate their DEA score. The study was conducted using a database containing information and details of IPL Season 4 (2011). Players are evaluated using different attributes and their scores are then summed using the linear RIA model.

The study (Pathak et al., 2016) aims to explore data mining and machine learning in sports. This study attempts to predict the outcome of ODI games. Results

often depend on many factors such as course, field, strategy, weather and even time of the game. Use modern techniques such as Naive Bayes, Space Vector Machines (SVM) and Random Forests to predict outcomes. Cricket Outcome Prediction (COP) tools have been developed that show the probability of winning or losing a match using prediction.

The researchers' main goal in this study (UmaMaheswari et al., 2019) is to model automation in order to establish a relationship between specific actions, ideas, and activity patterns. This will ultimately help the teacher make some decisions and give ideas. Data is now huge. Therefore, in order to obtain the complex structure, the relational structure must be used. Use the Research Manager to view data in a smaller but still complete view. Acts as a comparison mechanism by examining active patterns to obtain better results.

Researchers (Elliot et al., 2017) discussed a topic that has developed over the last 8-10 years, the 15-degree bending of the bowling arm. In this article, only the error in the system and the modelling of the reconstruction process with elbow extension tolerance are examined. These rules are determined by the International Cricket Council (ICC). Researchers discuss differences between laboratory and experimental tests of shooter movement. The conclusion is that optical reflection has better accuracy than video systems and is even better when tested in the laboratory.

A study in (Doljin et al., 2015) discussed the kinematics and dynamics of cricket balls. He tried to create smart cricket that would help in better collection of data and understanding of bowlers' movements. In the past, limitations such as electrical design and sensors have hindered the development of these projects. Now is very useful and small items can be used to improve skills. This information will be used to help athletes improve. Therefore, the weight and size of the ball are the same; does not affect the shooter's performance in any case.

Another study in (Foysal et al., 2018) discusses how AI can become a new force in data analysis. Like many other fields, sports have come to rely more on data. The use of deep neural networks for sports data and performance measurement is still evolving. In this paper, they proposed a 13-layer convolutional neural network.

Separating the shots into 6 groups is called "short net". This model records high accuracy with low entropy value.

A study (Sankaranarayanan et al., 2021) presents a data mining approach for cricket simulation and prediction. Unlike other sports such as cricket, baseball and basketball, although it is not very popular in the field of data analysis and research, it has started to develop in this field. Cricket history, matches and other important information are used in this article. It will predict future match events that will result in victory or defeat. A set of comparisons is not used and is modelled using linear regression and nearest clustering algorithms. The article likes to estimate the number of points won in the competition to demonstrate the effectiveness of the model.

III. METHODOLOGY

1. Overview

When it comes to using data science and artificial intelligence in cricket, many different techniques are used. In general, the questions are based on predicting all scores, players' performance and results. Sometimes we want to match different players, comment on different situations, sometimes we want to understand that our best 11 players are based on data. Fantasy cricket and simulations also rely heavily on the concept of intelligence.

2. Techniques Used

Common applications include classification techniques, statistical significance, and data analysis.

2.1 Classification Techniques

Classification usually determines the category to which a new sample belongs. In a basketball match you have to make many decisions. Is the cricketer a batsman or a bowler? Did a team win or lose the game? Is the target greater than or less than 100? And uses categories to answer similar questions. Let's talk about a few commonly used classifiers.

Below image all the fielding positions in the cricket ground.

Fig. 1
(Showing field positions)

Linear Regression

Linear regression is used when the data and attributes are numerical. Instructions for the group are obtained using a linear combination of weight and priority. In cricket, two specific classes are defined, such as whether the wicket falls in the 10th over, and then a linear regression classifier is used to identify the instance of the class it belongs to. Many problems like this are solved using regression methods.

Naive Bayes

This distribution is based on Bayes' probability theorem. Calculate the probability of one event occurring relative to another event. It is used to train the monitoring system so that results can be obtained effectively. An example of using this method can be seen by calculating the probability of batsman No. 3 scoring a hundred . Here we stipulate that the person must be a striker. This is conditional probability. The difference between scoring a hundred and not scoring a hundred can be calculated and batsmen can then be classified as centurions or non-centuries.

2.2 Principal Component Analysis (PCA)

Cricket statistics are varied. PCA is a non-parametric method for extracting meaningful information from data.

This is a common design. First, the information is standardized and predefined according to cricket rules and regulations. After this, the dimensionality of the data is reduced because it is very difficult to analyse large data. Active pattern analysis is then used to look for frequently occurring patterns, which are then summarized to create an overall algorithm. To reveal useful information, organizational analysis is applied to the resulting patterns and different techniques are used to represent the information.

2.3 Data Envelopment Analysis

Data Envelopment Analysis - DEA is a linear programming technique. Efficient and ineffective operators can be determined using DEA scores. DEA points are calculated for different players using a formula specifically designed for this purpose. The RIA score shows the player's value, everything he has, and all the benefits he can provide.
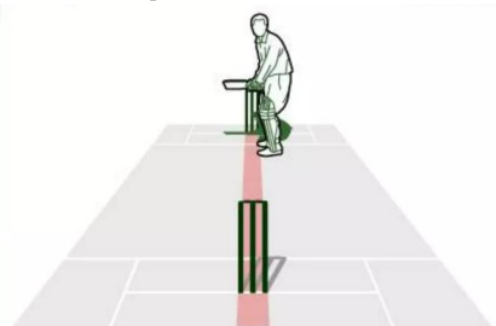

Fig. 2
(Showing Hawkeye stumps line using AI)

IV. EXPERIMENTAL RESULTS

Results from different papers are discussed below. Different models are used to predict the scores of teams, basketball, cooperatives. Some newspapers also estimated the number of wickets conceded and the number of wickets conceded by the bowlers. Some books select the best players from a group of players, sometimes even the best 11 players. Inter-player data was used to create matches and create strategies to use the correct player during the match. Other technologies such as LBW, Hawkeye, Ultraedge also use artificial intelligence to predict the flow of the ball. Bounce, speed and other similar factors are taken into account before making predictions.

Fig.3
(Ultra-edge technology in modern day Cricket)

V. IMPLEMENTATION

Linear regression was used to predict a team's score at the end of each match based on their score in the first 5 rounds of the match. Use the 5-round score range to predict the score of the next 5 rounds. Variables such as current score and wickets lost are reliable to help predict the score accurately. Currently, running speed is used to predict the score, but the results show that the regression function can predict the score better than running. Figure below shows the error in predicting the final score for the two scenarios.
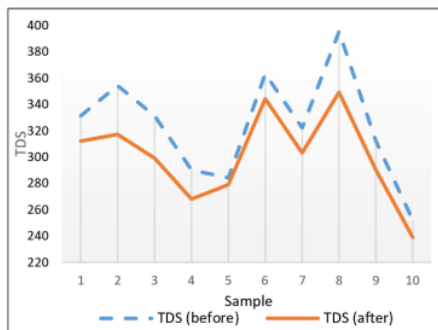
Fig. 4
(Error predicting the final score)

Naive Bayes was used to predict the top 11 players from the player pool. Games, opponents, team composition, etc. to make the final decision. Consider the factors. Players' statistics are compared with each other. Batsmen, bowlers, all-rounders and wicketkeepers are compared with each other and then the best combination is selected.

Classifier	Accuracy (%)			
	60% train 40% test	70% train 30% test	80% train 20% test	90% train 10% test
Naïve Bayes	57.05	57.18	57.48	58.12
Decision Trees	84.40	85.12	85.99	86.50
Random Forest	90.68	91.26	91.80	92.25
SVM	67.45	67.53	68.35	68.78

Fig.5
(Accuracy of different models while predicting runs)

Another paper used different models to compare the accuracy of predicting group scores. Methods such as decision tree classifier, naive Bayes, random forest and support vector machine have been used. As a result, Random Forest is the best choice for predicting batting and bowling. In all cases except the Naive Bayes classifier, the model is augmented as the training data grows.

One study found that a batter's dominant hand plays an important role in predicting scores. Considering the options and match conditions in situations where the player uses his right arm, the left-handed player will generally not perform better than the right-handed player in those situations or against a particular player. The study by (Sankaranarayanan et al., 2014) attempted to improve the prediction of home performance. Compare the Spearman distance measure for the difference between the measures they use. They use Jaccard, Cosine and Hamming measures. It is reported that the Spearman metric performs best, while the object packing method outperforms the nearest neighbor classifier.
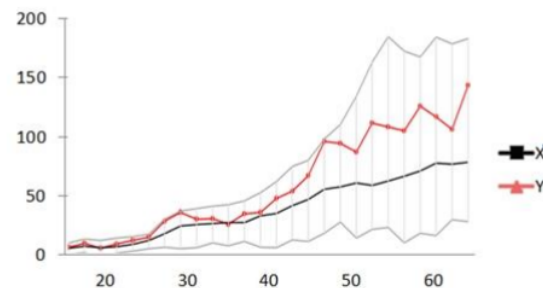

Fig. 6
(Graph showing runs made by a team in the 10 overs interval in a test cricket)

In a study, (Doljin et al., 2015) developed a smart ball to observe the ball's movement in 4 dimensions. Use spinless bowlers especially to collect records. When drawn in 4d it shows that it is better than2d. Above Figure shows a checkbox with color-coded time data.

The image shown below shows the trajectory to check LBW, the ball will be tracked using computer vision application and trajectory will be projected based on angle, and other parameters.
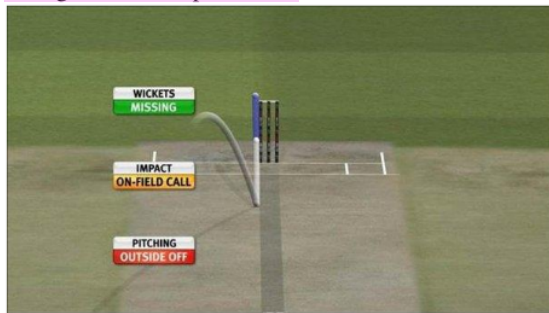
Fig, 7
(Smart ball plotted in 4d with color coded information)

VI. CONCLUSION

Finally, it can be concluded that even the early days of cricket were very difficult. It is becoming increasingly difficult to keep up with all the innovations in the game, but the use of knowledge and intelligence in cricket is better teamwork than ever before. Team insights prioritizes evaluating players' performance and then using these results to select the best players without personal bias.

The data also helps the team develop better strategies and challenges to solve various problems in the game. Before the game started, the team simulated all the situations that would occur during the game. When used correctly, data can put you a few steps ahead of your competitors and help you win games easily. In conclusion, it can be said that the incorporation of knowledge into the sport of cricket has undoubtedly achieved many positive results in the game and has clearly affected the negative ones, if any.

VII. REFERENCES

Amin, G. R., & Sharma, S. K. (2014). Cricket team selection using data envelopment analysis. European journal of sport science, 14(sup1), S369-S376.N.

Doljin, B. and Fuss, F.K. (2015). Upgrade Smart Cricket to advancedperformance analysis of bowling balls. Procedia Technology, 20, 133-137.M. F.

Elliott, B. and Alderson, J. (2007). Laboratory and field testing of cricket bowling: A review of current and past technology models. Sports Biomechanics, 6(1), 99-108.

Foysal, M., Ahmed, F., Islam, M.S., Karim, A. and Neehal, N. (December 2018). Shot-Net: Convolutional neural network for segmentation of multivariate lines. In International Conference on Latest Trends in Imaging and Vision (pp. 111-120). Springer Singapore

Passi, K. and Pandey, N. (2018). Improving cricket match prediction accuracy using machine learning. arXiv preprint arXiv:1804.04226.Pathak, N., &Wadhwa, H. (2016). Use the normal distribution method to predict the results of ODI cricket matches. Proceedings of Computer Science, 87, 55-60.

Sankaranarayanan, V.V., Sattar, J. and Lakshmanan, L.V. (2014, April). Autoplay: Data mining method for ODI cricket simulation and prediction. Proceedings of the 2014 SIAM International Data Research Conference (pp.1064-1072). Industrial Thiab Society for Applied Mathematics.

Singh, T., Singla, V. and Bhatia, P. (2015, October). Score and win predictions for basketball matches from data mining. In 2015 International Conference on Soft Computing Technology and Application (ICSCTI) (pp. 60-66). IEEE no.

UmaMaheswari, P. and Rajaram, M. (2009, March). A new approach to collaborative mining rules in sports data using critical context analysis: a cricketgame perspective. 2009 IEEE International Conference on Advanced Computing(pp. 1074-1080). IB. IEEE

Wickramasinghe, I.P. (2014). Predicting the performance of batsmen in Test cricket. Journal of Human Movement and Exercise, 9(4)

Data_Analytics_in_Cricket_-_Research_Paper_-_Milan_Sharma_1.pdf