**AI AND CRICKET**

**Abstract**

Cricket, as we all know, is one of the most popular sports in the world. Selection of the

right players, team strategy, conditions, and match-ups all play a vital role in the

outcome of a cricket match. Each of these factors can be improved with the use of latest

technologies. Artificial Intelligence(AI) and Data Science have helped take the game a

step forward. Technologies such as Bat Sense, Ball Sensors, Hotspot, Ultra edge and

many others have helped bring fairness as well as competitiveness in the game. Latest

Machine Learning and Deep Learning techniques have also made the teams smarter

while selecting their players, making combinations, deciding strategies and many more.

These techniques help predict the performance of batters, bowlers, or even teams as a

unit based on their previous performances and different factors. Teams also use

strategies on when and where to use specific players. Roles of specific players have been

made much clearer with the help of data. Teams such as England Cricket Team,

Nottinghamshire County Cricket Club, Multan Sultans and many others have used these

techniques and have achieved very quick success with limited resources. This Paper

reviews different types of Machine Learning and Deep Learning techniques being used

in the Cricketing Industry. Mostly the techniques being used are Linear Regression, KWebology (ISSN: 1735-188X)

Volume 19, Number 2, 2022

10120 http://www.webology.org

Nearest Neighbors(KNN), Decision Tree and other similar techniques. From the success

of different teams, it is clear that data is helping teams go a step ahead of their opponents

and is becoming a vital part of their success.

**Keywords**: Data Analytics, Data Science, Cricket, Artificial Intelligence

**Introduction**

**Overview**

The introduction of data into baseball by the Oakland Athletics in Major League baseball

opened a completely new door for strategy and preparation in sports. Baseball changed

completely after that. It took long enough but thankfully it has now entered into cricket as

well. From selection of the playing 11 to predicting the target and individual performances

of players, data has revolutionized cricket completely.

Latest Machine learning techniques have helped teams practice using the same bowling

actions and speeds using bowling machines. Models give unbiased statistics to help select

the best available players. Even ball by ball strategies can be made using data. The players

have now started watching specific videos made for them about the opposition and now

prepare according to the weakness of the opponents. Match-ups have become one of the

most important parts of strategy making. Left arm batsmen are now being used more and

more to cope with leg spinners.

There are a lot of recent success stories of teams who have won major trophies with the

help of data. Every team carries data analysts with them and are a very important part of

their dressing room. The England Cricket team is one of the main examples of this ideology.

After the shameful exit from the 2015 world cup they completely changed their team and

playing method with the help of data which helped them get the first ever world cup in

2019. Many teams in IPL, Islamabad United and Multan Sultans in PSL, and numerous

county teams all follow the same ideology and have seen a lot of success recently

**Objectives**

Data in sports has a lot of uses. In cricket the main thing is selecting the best possible

players from the set of available players. Auction and drafting systems of all the major

leagues are being run using the data. Keeping all the personal bias aside, teams want to

select the best for themselves. Data gives them plan A, B, C and the list goes on. In that

case they are completely ready to go into the tournament with the best available team.

It then occurs that data then helps make and execute the plans to win games in the

tournament. Opponents' playing style, their best areas, their weaknesses, their dependent

and independent players, and much more such useful information is only clicks away.

Teams also run predictive models to simulate what could happen in the game in similar

scenarios. Based on what has happened in the past, teams can get an idea of what challenges

and tasks are ahead of them. Which player is more useful against which player and in what

phase of the game? All this and much more information can be extracted and then applied

to help increase the chances of victory.

**Limitations and Scope**

Data as useful as it can be, will always have limitations. It can help you to succeed most of

the time but the scenario in sports changes in the matter of balls. Opponents can change

their style of play. Many other factors can come into effect. So data can always help but

you must always have backup plans to cope with the opponents.

**Literature Review**

A study in (Singh et al., 2015) talks about the prediction of the total number of runs scored

by a team in the first as well as the second innings. The earlier predictions are based only

on the run rate of the team at a particular moment. This study also takes into consideration

the venue, number of wickets fallen and the opposition. 2nd innings score prediction takes

extra consideration, i.e. the target. These methods have been implemented using Naive

Bayes and Linear regression classifiers. All the ODIs played between 2002 and 2014 were

used at 5 over intervals for the 50 overs of each match. The results showed that the Linear

regression model is less prone to error than the one being used currently. Also, the accuracy

of Naive Bayes increased from 68 percent to 91 percent from 5th to the 45th over.

This study in (Wickramasinghe et al., 2014) shows the prediction of a batsman’s

performance in a test series. Data was taken from the test matches played between a 5-year

gap. Different characteristics of the player and the team he plays were taken into

consideration. A 3 layered hierarchical model is proposed to cater to the hierarchical nature

of the data. The study concludes that the batting hand, and the rank of the opposition are

the major factors that affect a player’s performance.

Another study in (Passi et al., 2018) talks about how difficult it is to select the best 11 for

any specific game and tries to predict the individual player’s performances to help the

coach and the captain in making the right decisions. Player’s previous record, current form,

opposition, and venue, all play an important role in this prediction. The study tries to

predict the runs scored by a batsman and the wickets taken and runs conceded by a bowler.

These problems are classified using different ranges and multiple models such as SVM,

Naive Bayes, Decision Tree, and Random Forest are used to predict the outcomes. It is

concluded in the study that Random Forest has the highest accuracy amongst all these

models.

A study conducted back in 2011 (Amin et al., 2014) proposed a method of a team’s

selection using Data Envelopment Analysis(DEA). After the evaluation the players can be

ranked on the basis of their DEA scores. A dataset consisting the records and details of

IPL season 4(2011) was used to conduct this research. Players were evaluated using

different attributes and their score was then aggregated using a linear DEA model.

The study in (Pathak et al., 2016) looks to explore the field of data mining and machine

learning in sports. This study attempts to predict the outcome of an ODI game. The

outcome usually depends on different factors such as, venue, toss, strategies, weather, and

even the time of the game. Modern techniques such as Naive Bayes, Space Vector

Machine(SVM), and Random Forest are used to predict the outcome. Using the

predictions, a tool Cricket Outcome Predictor(COP) has been developed which tells the

probability of winning or losing the game.

The main aim of the researchers in this study (UmaMaheswari et al. 2009) is the modeling

of an automated framework so that specific movements, strategies, and correlationWebology

between playing patterns can be identified. This will eventually help the coaches in making

certain decisions and strategies. The real time data is humongous. So to get a sophisticated

structure, an Object-relational model is used. Principal component analysis is applied to

take a look at the data in lesser dimensions but still get a decent amount of accuracy. It

works as a comparison mechanism and frequent patterns are analyzed to get interesting

outcomes.

The researchers in (Elliot et al., 2007) talk about a topic that has developed a lot over the

last 8-10 years, i.e. the 15-degree bend in the bowling arm. The paper mainly reviews the

errors in the system and the modeling of the reconstruction process with respect to elbow

extension tolerance. The rules have been set by the International Cricket Council(ICC).

The researchers talk about the differences of Laboratory based testing and on field testing

of the bowler’s action. It is concluded that the opt reflective has a better accuracy than the

video based systems and it is better if the tests are conducted in the laboratory.

A study in (Doljin et al., 2015) talks about the kinematics and dynamics of cricket. It

attempts to develop a smart cricket ball which will help in collecting better data and

understanding of the motions of the bowlers. Previously, technical limitations such as

electric design and sensors have hindered the growth of such types of projects. Now very

useful and tiny sized components are available which can be used to create a smart ball.

The data will be used to help bowlers improve. The ball has the same weight and size thus;

it will not affect the bowler’s performances in any case.

Another study in (Foysal et al., 2018) talks about how AI has become the new powerhouse

of data analytics. Sports, like many other fields, has started to depend more on data.

Applications of deep neural networks in sports data and performance analysis are still

developing. In this paper they have proposed a 13-layer Convolutional Neural Network. It

is called “Short Net” in order to classify the shots into six categories. The model has

recorded a very high accuracy with a relatively low cross-entropy rate.

A study (Sankaranarayanan et al., 2014) provides a data mining approach to cricket

simulation and prediction. Cricket, unlike other games, such as basketball and baseball, is

not very much popular in data analytics and science but it has started to grow in this area.

This paper takes in historical cricket data, state of the match, and other useful information.

It then predicts the future match events which will lead to a win or a loss. A lot of match

parameters are used and are modeled using linear regression and nearest neighbor

clustering algorithms.

**Methodology**

**Overview**

There are a lot of different techniques used when it comes to the use of Data Science and

AI in cricket. Mostly, the problems are based on predicting the total runs, player

performances, and results. Sometimes, we want to make match-ups for different players,

strategies for different scenarios, and sometimes we want to know what our best 11 players

are according to the data. Fantasy cricket and simulation also work mainly on the idea of

AI.

**Techniques Used**

The methods commonly used are classification techniques, principal component analysis

and Data envelopment analysis.

**Classification**

Classification is mainly deciding the class where the new instance belongs. In cricket there

are different types of decisions you have to make. Either a cricketer is a batsman or a

bowler? Did a team win the game or lose it? Is the target more than or less than 100? and

many more similar questions are answered using classification. Let's discuss a couple of

these classifiers which are most commonly used.

**a. Linear Regression**

Linear Regression is used when the data and attributes are numeric. Expression

of a class is obtained using the linear combination of weights and values which

have been predetermined. In cricket mainly two classes are defined such as

whether the wicket falls in 10 overs or not, then using the linear regression

classifier the instance is defined to the class it belongs to. Many other such

problems are solved using the linear regression method.

**b. Naive Bayes**

This classifier is based on Bayes’s probability theorem. Conditional probability

of an event based on another event is calculated. It used to train a supervised

learning setting so that results can be extracted efficiently. An example of how

this method is used can be represented by calculating the probability of number

3 batsman scoring a hundred. Here we have put a condition that the person must

be a one down batsman. This is conditional probability. Probability of scoring a

hundred and not scoring a hundred can then be calculated and the batsman can

then be put into centurion or non-centurion class.

**A. Principal Component Analysis(PCA)**

The data of cricket is very diverse. PCA is a non-parametric method of extracting useful

information from the data. It is a frequent pattern generation technique. Firstly, data is

modeled and preprocessed according to cricketing rules and regulations. Then,

dimensionality of the data is reduced as it is very hard to perform analysis on a high

dimensional data.

occurring patterns and then it is summarized to create a generalized algorithm. Association

analysis is then applied on the received patterns to extract useful knowledge and the

knowledge is represented using different techniques.

**B. Data Envelopment Analysis**

Data Envelopment Analysis - DEA is a linear programming technique. Efficient and Non

Efficient players can be identified using the DEA scores. DEA scores are calculated for

different players using a formula specifically made for this purpose. The DEA score tells

the utility of a player considering everything he offers and the overall contribution they can

offer.

**Experimental Results**

The results from different papers have been discussed below. Different Models have been

applied to predict the runs scored by the teams, batters, partnerships. Runs conceded by

bowlers and wickets taken have also been predicted in some of the papers. Some of the

papers select the best player or sometimes even the best possible 11 from the given players.

Player vs player data has been used to create match-ups and strategies to use the right

player at the right moment in the game. Other technologies such as LBW, Hawkeye,

Ultraedge also use AI to predict the flow of the ball. Bounce, Speed and other similar

factors are taken into context before the prediction is made.

**Implementations**

Linear Regression has been used to predict the runs by the team at the end of the innings

based on how much they have scored in the first 5 overs of the game. Score is predicted for

the next 5 overs using the 5 over interval scores. Variables such as current score and wickets

fallen were considered to help predict the score accurately. Currently, run rate is used to

predict the score, but the results have shown that linear regression can predict the score

better than run rate. Figure 3 below shows the error in predicting the final score in both

cases.

Naive Bayes has been used to predict the best possible playing 11 from the pool of available

players. Factors such as playing conditions, opponents, and team combination were all

considered to make the final decision. Stats of players are compared with each other.

Batters, bowlers, all rounders, and wicket keepers are all compared separately and then the

best combination is selected.

Another paper has used different models to compare the accuracy in predicting the runs

scored by the team. Models such as Decision Tree Classifier, Naive Bayes, Random Forest,

and Support Vector Machines were used. It was concluded that Random Forest is the best

predictor while predicting both batting and bowling. Except for the Naive Bayes Classifier,

Model accuracy for all cases also increased when the size of training data was increased.

The figure 4 below shows the accuracy of different models from different sets of training

and testing data.

One research has found that the handedness of the batter does play a major role in predicting

the score. While considering the options if the opponent is playing a right arm off spin

bowler and the conditions are favorable for spinners, the left handed batsmen are more

likely to fail in these conditions or against that specific player than the right handers. The

research in (Sankaranarayanan et al., 2014) has tried to create a Home run prediction

performance. The Spearman distance metric is compared with the different distance metrics

used by them. They have used Jaccard, Cosine and Hamming metrics. It was reported that

the spearmen metric was the best performer and the attribute bagging performed better than

the nearest neighbor classifiers.Webology (ISSN: 1735-188X)

Volume 19, Number 2, 2022

10127 http://www.webology.org

A smart ball was created by (Doljin et al., 2015) to analyze the movement of the ball in 4d.

Specifically, off spin bowlers were used to collect the data. When plotted in 4d it showed

much better results than in 2d. Figure 5 shows the ball plotted with color coded time

information.

Figure 3: Error in predicting the final score. Taken from (Singh et al., 2015)

Figure 4: Accuracy of different models while predicting the runs. Taken from (Passi et al.,

2018)Webology (ISSN: 1735-188X)

Volume 19, Number 2, 2022

10128 http://www.webology.org

Figure 5: Smart ball plotted in 4d with color coded information. Taken from (Doljin et al.,

2015)

Figure 6: Digital Head of ICC about importance of data in Cricket

Figure 7: Use of AI to predict the route of the ballWebology (ISSN: 1735-188X)

Volume 19, Number 2, 2022

10130 http://www.webology.org

**Conclusion**

Finally, it can be concluded that although the game of cricket was very complex earlier. It

was getting difficult to keep up with all the innovations coming into the game but the use

of data and AI in cricket has helped teams perform much better than they had been doing.

Data has helped teams give value to the player performances and then select the best

possible players using those values without any personal bias.

Data has also helped the team in making better strategies and match-ups to cope with all

sorts of problems in the games. Whatever is going to happen in the game is already being

simulated by the teams before the game even starts. Data, if used intelligently, takes you

many steps ahead of your opponent and helps you win games much easily.

In conclusion, it can be said that the introduction of data into cricket has surely made a

much positive impact in the game which clearly overshadows any negative impact, if it is

there.

**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., & Fuss, F. K. (2015). Development of a smart cricket ball for advanced

performance analysis of bowling. Procedia Technology, 20, 133-137.M. F.

Elliott, B., & Alderson, J. (2007). Laboratory versus field testing in cricket bowling:

A review of current and past practice in modelling techniques. Sports

Biomechanics, 6(1), 99-108.

Foysal, M., Ahmed, F., Islam, M. S., Karim, A., & Neehal, N. (2018, December).

Shot-Net: A convolutional neural network for classifying different cricket shots.

In International Conference on Recent Trends in Image Processing and Pattern

Recognition (pp. 111- 120). Springer, Singapore.

Passi, K., & Pandey, N. (2018). Increased prediction accuracy in the game of cricket

using machine learning. arXiv preprint arXiv:1804.04226.Pathak, N., &

Wadhwa, H. (2016). Applications of modern classification techniques to

predict the outcome of ODI cricket. Procedia Computer Science, 87, 55-60.

Sankaranarayanan, V. V., Sattar, J., & Lakshmanan, L. V. (2014, April). Auto-play:

A data mining approach to ODI cricket simulation and prediction.

In Proceedings of the 2014 SIAM international conference on data mining (pp.

1064-1072). Society for Industrial and Applied Mathematics.

Singh, T., Singla, V., & Bhatia, P. (2015, October). Score and winning prediction in

cricket through data mining. In 2015 international conference on soft

computing techniques and implementations (ICSCTI) (pp. 60-66). IEEE.

UmaMaheswari, P., & Rajaram, M. (2009, March). A novel approach for mining

association rules on sports data using principal component analysis: for cricket

match perspective. In 2009 IEEE International Advance Computing

Conference (pp. 1074-1080). IEEE.B