Role of Data Analytics and AI in Cricket

1st Milan Sharma
University Institute Of Engineering
Chandigarh University
Punjab, India
milansharma449@gmail.com
0009-0007-0644-2802

2nd Manjit Singh
University Institute of Engineering
Chandigarh University
Punjab, India
manjit.e11549@cumail.in

***Abstract*—As everyone is aware, one of the most popular sports in the world is cricket. The appropriate player choices, team tactics, plays, and events all have a significant impact on how a baseball game turns out. New technologies can be used to improve each of these factors. Gaming is progressing thanks to data analytics and artificial intelligence (AI). A number of technologies, including Hotspot, Ultra Edge, Ball Sensor, Bat Sensor, and many more, contribute to the fair and competitive nature of the game. Teams can also choose players, form connections, and choose strategies more intelligently thanks to the most recent advancements in machine learning and deep learning technologies. Based on their historical performance and other variables, this system can estimate how well bowlers, batters, and even fielding teams will perform. Teams also employ techniques to determine which players to utilize and when. Data helped to clarify the responsibilities of individual actors. With little funding, clubs like the Multan Sultans, Nottinghamshire County Cricket Club, and England Cricket Team have exploited this technology to their advantage and succeeded swiftly. The many forms of deep learning and machine learning applied in the cricket industry are examined in this article. The primary techniques employed include Decision Trees, K Nearest Neighbors (KNN), Linear Regression, and related techniques. It is shown by the accomplishments of various teams that understanding team outperforms its rivals and has grown to be a crucial component of their success.**

I. INTRODUCTION

A. 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. Leg-spinners are increasingly being faced with left-arm batsman. 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.

B. 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. By identifying the top players and assisting them in putting the ideal plan into action, the information makes management and training extremely simple. The goal of this study is to investigate the various applications and approaches of data science and data analysis in the modern sports world.

C. 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.

D. 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.

Fig. 1
(Showing field positions)

II. Literature Review

In a study conducted by Singh et al (2015), prediction of total runs in the first and second innings of a cricket match was discussed. The initial predictions are based on a team's current run rate, taking into account factors such as position, wickets and losses. To predict the score of the second half, people will take into account the target score. This prediction method is used using Naive Bayes classifier and linear regression. Rephrase

The study analyzed One Day Internationals (ODIs) that took place between 2002 and 2014, showing that the linear regression model is less error-prone than existing models. Additionally, Naive Bayes accuracy increased from 68Wickramasinghe et al. (2014) conducted a study on predicting batting performance in cricket across multiple tests. Data over a five-year period from competitive tests were analyzed, taking into account the specific characteristics of players and teams. A 3-layer hierarchical model is introduced to accommodate the hierarchical structure of the data. The study concluded that the batsman's lineup and opponents are the main factors influencing a player's performance.

In the study by Passi et al. (2018), the challenge of selecting the best 11 players for a cricket match is discussed. The decision-making process involves taking into account the player's background, current form, quality of opponents and the player's position. The study attempts to estimate the number of runs scored by the batsman and the number of wickets taken and conceded by the bowler. Various models, including support vector machines (SVM), naive Bayes, decision trees, and random forests, are used for prediction.

Research shows that Random Forests have the highest accuracy among these models. A study conducted by a certain researcher (2014) in 2011 proposed a group selection method using data envelopment analysis (DEA). Players are evaluated based on various attributes and their DEA scores are calculated using a linear RIA model. The study used a database containing information from IPL Season 4 (2011).

Pathak et al. (2016) explored data mining and machine learning in sports, focusing on predicting the outcomes of One Day International (ODI) cricket games. Various factors, including pitch conditions, field settings, strategies, weather, and time of the game, influence the results. Modern techniques, like Naive Bayes, Support Vector Machines (SVM), and Random Forests, were used to predict match outcomes. Cricket Outcome Prediction (COP) tools were developed to estimate the probability of winning or losing a match.

In a study by UmaMaheswari et al. (2019), the research goal was to model automation to establish relationships between actions, ideas, and activity patterns in cricket. This modeling aids teachers in making informed decisions and offering suggestions. To manage complex data structures, a relational structure is employed, and a Research Manager is used to view data comprehensively. It acts as a comparative mechanism, analyzing active patterns to yield superior results.

Elliot et al. (2017) examined the topic of 15-degree bowling arm flexion, a topic that has developed over the past 8 to 10 years. The study specifically focused on systematic errors and modeled the reconstruction process with elbow extension, in accordance with the rules set by the International Cricket Council (ICC). The study compared the differences between laboratory tests and tests on shooter movements and concluded that optical reflex systems provide greater accuracy than video systems, especially in laboratory tests. Doljin et al. (2015) examined the kinematics and dynamics of cricket balls and aimed to develop smart cricket technology for better data collection and understanding of bowlers' movements. The study addressed previous limitations, such as electrical design and sensors, that hindered the development of such projects.

Foysal et al. (2018) talk over the a growing belief that role of artificial intelligence (AI) in data analytics, focusing on sports. They proposed a 13-layer convolutional neural network model to measure performance and sports data, achieving high accuracy and low entropy value.

One of the famous researchers Sankaranarayanan et al. (2021) presented a data mining method for simulating and predicting cricket. Cricket, although less popular for data analysis and research than other sports, is gaining ground. The study accustomed cricket history, match data and other historic information to predict future match events. It utilizes linear regression and nearest clustering algorithms to estimate the robustness results and demonstrate the effectiveness of the model.

III. METHODOLOGY

A. Overview

When it comes to using data science and artificial intelligence in cricket, there are many different techniques used. In general, the questions are based on predicting the scores, form

and results of all players. Sometimes we want to play with different players, comment on different situations, sometimes we want to understand that our top 11 players are based on data. Cricket games and fantasy simulations also rely heavily on the concept of intelligence.

B. Techniques Used

Commonly employed methodologies encompass classification techniques, statistical significance, and data analysis.

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.

2) Linear Regression: Linear regression is used when the data and attributes are numeric. Instructions for the group are obtained using a linear combination of weights and priorities. In cricket, two specific classes are defined, for example, if the wicket falls at number 10, then a linear regression classifier is used to determine the instance of the class to which it belongs. Many problems like this are solved using regression.

3) 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.

C. 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.

D. Data Envelopment Analysis

Data envelopment analysis – DEA is a linear programming technique in which efficient and inefficient operators can be determined using DEA scores. DEA points are calculated for different players using a formula specially designed for this purpose. The RIA score shows a player's value, everything he has and every advantage he can bring.

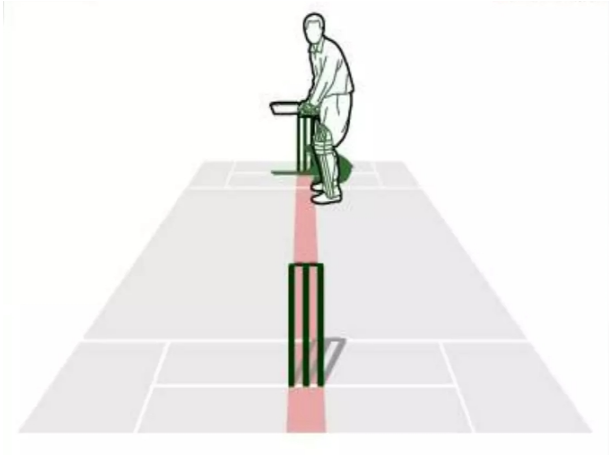


Fig. 2
(Showing Hawkeye stumps line using AI)

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. Data between players is used to create matches and create strategies for using the right players in the match. Other technologies such as LBW, Hawkeye, Ultraedge also use artificial intelligence to predict the direction of the ball's movement. 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 is used to predict a team's score at the end of each match based on their scores in the first 5 innings of the match. Use the 5-round scoring scale to predict the score for the next 5 rounds. Variables such as current points and lost points are reliable to help predict the score accurately. Currently, running speed is used to predict scores, but the results show that the regression function can predict scores better than running. The figure below shows the final score prediction error for both scenarios.

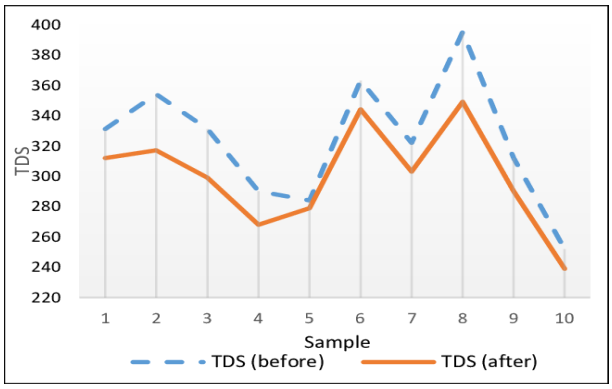


Fig. 4
(Error in predicting the exact score)

Naive Bayes is used to show more accurately, the top 11 players in the player in a team. Games, opponents, team composition, etc. to make the final decision. Let's consider the factors. Players' statistics are compared to each other. The batsman, bowler, all-rounder and wicket-keeper are compared with each other and then the best combination is chosen.

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 group score prediction. Methods such as decision tree classification, naive Bayes, random forests, and support vector machines were used. Therefore, Random Forest is the best choice for predicting strikes and pitches. In all cases except the Naive Bayes classifier, the model gets boosted as the training data increases.

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.

A. MODEL COMPARISON

In another research investigation, an array of models was employed to assess their accuracy in predicting team scores. These models encompassed techniques such as the decision tree classifier, Naive Bayes, random forest, and the support vector machine. The findings underscored that the Random Forest model emerged as the superior choice for predicting both batting and bowling performances. Notably, in all instances, with the exception of the Naive Bayes classifier, the model's performance improved as the training dataset expanded.

Furthermore, an intriguing study revealed the significant role of a batter's dominant hand in predicting scores. When evaluating player performance under specific circumstances where a player employs their right arm, the evidence suggests that left-handed players typically do not outperform right-handed players in those situations or against particular opponents. The study, conducted by Sankaranarayanan et al. in 2014, aimed to enhance predictions of home performance. It entailed a comparative analysis of various metrics, including the Spearman distance measure, Jaccard, Cosine, and Hamming measures. The outcomes indicated that the Spearman metric delivered the most reliable performance, with the object packing method surpassing the nearest neighbor classifier.

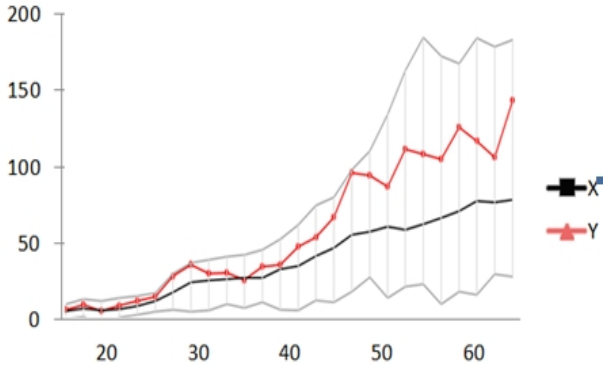
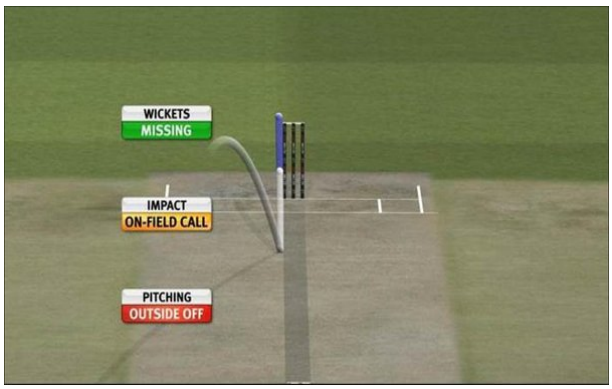


Fig.6
(Graph showing runs made by a team in the 10 overs interval in a test cricket)

In a research endeavor led by Doljin et al. in 2015, an innovative approach was introduced involving the development of a smart cricket ball capable of tracking its movement in four dimensions. This groundbreaking method employed spinless bowlers, specifically selected to gather comprehensive records. The visual representation in four dimensions proved to be a substantial enhancement over traditional two-dimensional representations. The figure presented above exhibits a checkbox adorned with time data color-coded for enhanced comprehension.

Moreover, the image depicted below showcases the trajectory assessment used for LBW (Leg Before Wicket) decisions.

This method entails ball tracking through a computer vision application, with the trajectory projected based on specific angles and other relevant parameters.



Fig, 7
(Smart ball plotted in 3d with color coded information)

VI. FUTURE DIRECTIONS

Looking ahead, the integration of Artificial Intelligence (AI) and Data Analytics into the realm of cricket promises a host of new horizons for advancement and innovation within the sport. Here are some prospective avenues for exploration:

A. Elevated Player Performance Analysis

The amalgamation of AI and data analytics will enable a more comprehensive exploration of player performance, including factors such as mental resilience, fatigue management, and injury prevention. Real-time insights from wearable technology and advanced sensors will inform critical decisions concerning team composition and game strategy.

B. Umpiring Precision and Decision Augmentation

AI holds the potential to significantly enhance umpiring decisions. Beyond the technologies like Hawk-Eye and UltraEdge, AI systems can serve as valuable tools for umpires, leading to more precise adjudication of LBW and edge detection, reducing disputes and augmenting fairness.

C. Fan Engagement and Immersive Experiences

AI offers the capacity to enrich the experiences of cricket enthusiasts. Chatbots, tailored content recommendations, and Virtual Reality (VR) encounters can transform the way fans engage with the sport, providing more immersive and enjoyable interactions.

D. Early Talent Identification

Data analytics can play a pivotal role in identifying promising talents at an early stage. AI models can scrutinize a wide array of performance data, pinpointing young players with exceptional potential, thus assisting cricket academies and teams in making informed investments.

E. Injury Mitigation

Predictive analytics can contribute to injury prevention by evaluating players' workload, fitness levels, and biomechanics. This data-driven approach will enable the creation of personalized training and recovery regimens, reducing the risk of injuries.

F. Match and Tournament Strategy Simulation

AI can simulate various match scenarios, empowering teams to devise strategies tailored to specific conditions and opponents. Coaches can harness AI-derived insights to make real-time decisions during matches.

G. Evolution of the Game

The integration of AI into cricket can pave the way for innovative formats and rule adjustments. Much like the rise of T20 cricket, AI can be instrumental in modeling and forecasting the impact of rule modifications on the game.

H. Advanced Performance Metrics for Bowlers and Fielders

While batsmen have received considerable attention, AI can be employed to formulate more intricate performance metrics for bowlers and fielders. This will assist teams in identifying underrated talents and enhancing fielding strategies.

I. Data Security and Ethical Data Usage

With the growing reliance on data, safeguarding data security and ensuring ethical data utilization is of paramount importance. Future developments should prioritize robust data protection measures and responsible data handling.

J. Global Reach and Accessibility

AI and data analytics can play a pivotal role in advancing the global reach of cricket. By rendering the sport more data-driven and accessible, it has the potential to captivate a broader international audience.

In the forthcoming years, the synergy between AI and data analytics within the realm of cricket is poised to propel the sport into new dimensions. This transformation will render cricket more data-centric, fan-friendly, and player-centric, benefiting not only professional cricket but also fostering grassroots development and community engagement.

VII. CONCLUSION

In summary, it is evident that the early days of cricket presented significant challenges. As the sport continues to evolve, staying abreast of the myriad innovations can be quite demanding. Nonetheless, the integration of knowledge and intelligence into cricket has fostered improved teamwork and player evaluation.

Team insights now take precedence, allowing for the objective assessment of players' performance and, subsequently, the selection of the most suitable candidates, devoid of personal biases. The wealth of data available also empowers teams to craft more effective strategies and address various in-game challenges. In preparation for matches, teams rigorously simulate various scenarios that may unfold during gameplay.

When harnessed effectively, data has the potential to place teams several steps ahead of their competitors, facilitating victories. In conclusion, the infusion of knowledge into the realm of cricket has undeniably yielded numerous positive outcomes for the game while minimizing potential drawbacks, if any.

REFERENCES

[1] 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.

[2] 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.

[3] 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

[4] 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.

[5] 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.

[6] 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.

[7] 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

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