A Survey Paper on-

Data Analytics in Cricket: A Game Changer

Abstract:

Cricket is one of the most popular sport in the world, played by many countries. It is also one of the most played, exciting sport and it has gained huge support from people around the world like us. There are many tournaments conducted by the cricket boards across the world, the most popular tournaments are the One Day International (ODI)-World Cup and T-20 World Cup which is held once in every four years and two years respectively. The winner of the game is usually decided by the highest runs scored by a team in the particular game. To be more precise, it is a game of numbers starting from runs scored by the batsmen to the wickets taken by the bowler. As we know, when dealing with a lot of numeric data, analytics come into light. In the recent years, a lot of Data Analytics, Machine Learning and Artificial Intelligence has become a part of the game. This paper deals with all of the Cricket Analytics and how this has become a game changer in Cricket.

Introduction:

Before digging a little deeper into this topic, what is Data? To answer this, data is generally an information about anything. It can take any forms such as tabular data, structured data, unstructured data, the data that is constantly generated from the social media. In fact, data is living with us, around us, following us everywhere and now arises another question what is Data Science/Analytics? It is a science that involves collecting data, processing the data, observing the data, making predictions and drawing insights from it and finally arriving at a conclusion about the data that we are handling.

Now, all of us from the childhood itself grew by watching or playing some kind of sport at school or at home. Some of sports that are popular are Soccer, Basketball, Cricket, Tennis, Baseball, Hockey etc. As the paper is going to deal about Cricket, the major focus is going to be on Cricket and how Data Science is playing a crucial role in the recent in the game.

Cricket Analytics majorly deals about drawing the insights and making predictions about a game that is to be played, based on the previous games, overall team's performance against a particular team, and also each individual player performance against that team. There are many metrics involved to measure a player's performance. Often there are different metric for batsmen, bowler, wicket keeper, and a fielder on the field.

The use of Analytics in Cricket is not only limited to estimate the outcome of a game or the performance of the players. It is beyond that, there are still various factors to be considered while the team is about to play a game, the pitch on the ground, weather, location of the match etc. Apart, from the Data Analytics, there are various machine learning prediction algorithms that are utterly involved which plays a crucial role as part of the game. In the early times, when the cricket board is hiring for an Analyst, everyone is surprised and doubted about having this position, later on, when there are some amazing predictions about outcome that the analyst is estimating and drawing insights. There are many Cricket Analysts around the world now, almost for every cricket board. There is series streamed on Amazon Prime named "Inside Edge" which is split into two seasons. As a part of the show, they primarily focus on a tournament which is similar to Indian Premier League (IPL) held by BCCI every year, and how the game is played, how the players are selected by a team in the auction, how various betting applications show their predictions about

the game and making money out of it. The major aspect of the series also focuses about an analyst that a top team of that tournament has and how she majorly contributed in changing the team's performance in that season. In this way, Cricket Analyst play a major role and there is so much that role is contributing to the game. The further part of the paper deals with three different sections involving the role of Data Sciences& Big data, Machine Learning and Artificial Intelligence respectively.

 Role of Data Science& Big data in Cricket Analytics:

Firstly, the data analyst needs to gather the information about the players of the team, and then evaluating their performance individually. For a batsman, some of the metrics are- The hard-hitting ability, finishing ability, Consistency of the player, run-time between the wickets.
The numeric equations for the above metrics are as follows:
Hard-hitting Ability: (Fours + Sixes) / Ball played by batsman (The ratio of total no. of fours and sixed hit by a batsman to the total no. of balls played by the batsmen)
Finishing Ability: Not out innings / Total innings played (Not out innings divided by the total innings played)
Consistency of Player: Total Run / No. of times out (the ratio of total runs scored by the batsman to the no. of times he/she was out)
Running b/w the wickets: (Total run – (Fours + Sixes)) / (total ball played – boundary balls)
For instance, if a batsman has got a good score on his/her finishing ability, and a low score on the consistency metric, we can say that he/she is a good finisher, but not consistent.
 Also, not only these metrics there are a lot more metrics for a batsman such as batting average, also the performance a particular batsman is going to bat is affected by the position he is going to bat, which would be as a opener, top-order, middle order, bottom players.Now, when these metrics are being accounted for there is a lot that the analyst has to analyse and make the predictions.
Suppose that we consider the player "Mahendhra Singh Dhoni" one of the greatest captains, player, and a wicket keeper. He has performed differently in various formats under various conditions.
Now, consider two situations where Dhoni bats in different positions i.e. (1-4) and (5-11). Here are some of the stats comparing Dhoni with the other players.

Batting Position 1-4				
Name	**Matches**	**Innings**	**Runs**	**Average**
Mahendra Singh Dhoni	49	48	2449	64.44
Virat Kohli	235	230	11679	60.51
Michael Bevan	63	57	2399	59.97
Babar Azam	73	71	3297	53.17
Faf du Plessis	107	105	4761	52.9
Shai Hope	75	70	3204	52.52
Sir Viv Richards	140	132	5791	52.17
Rohit Sharma	175	173	7983	51.83
Ambati Rayudu	44	41	1494	51.51
Jonathan Trott	68	65	2819	51.25

Considering only those players who has played more than 40 ODI matches batting at 1-4

Batting Position 5-11

Name	Matches	Innings	Runs	Average
AB de Villiers	47	44	2075	79.8
Michael Bevan	169	139	4513	50.7
Lance Klusener	132	98	2483	47.75
Mahendra Singh Dhoni	301	249	8324	47.56
Michael Hussey	151	126	4211	47.31
Kedar Jadhav	69	49	1371	44.22
Rahul Dravid	103	86	2802	43.78
Michael Clarke	82	67	2079	43.31
Misbah-ul-Haq	107	96	3173	42.3
Angelo Mathews	205	175	5421	41.7

Considering only those players who has played more than 40 ODI matches batting at 5-11

As we can see from th above stats, MS Dhoni overtakes every other player when he comes to bat in between the positions 1-4 and he has the highest batting average, unlike the batting position 1-4, he has a low average compared to thr previous average. So, a cricket analyzes this and shares the insights.

Finishing the chasing matches

Name	Matches	Innings	Not Outs	Average
Mahendra Singh Dhoni	189	145	50	51.04
M Azharuddin	173	157	37	37.24
Inzamam-ul-Haq	174	151	34	40.99
Jaques Kallis	169	158	34	44.95
Arjun Ranatunga	150	143	33	34.89

Including all the players who have scored at least 3000 overall runs

This table shows how efficient Dhoni is when it comes to finishing the chasing matches. But, just finishing is not the factor, winning the match also comes into account.

Winning the chasing matches

Name	Matches	Innings	Not Outs	Average
Mahendra Singh Dhoni	116	75	47	102.71
Virat Kohli	89	86	30	96.21
Michael Bevan	75	45	25	86.25
AB de Villiers	64	59	28	82.77
Michael Hussey	51	27	17	74.1

Including those matches in which the team has won match

These stats indicate the Dhoni has a great capability to win a chasing match by finishing the match in his own style. One of the examples that is remarkable one is the 2011- ODI World Cup, finished of the match with a "Six" with a score of 91 in 79 balls.

So, as per the metrics mentioned earlier for a batsman, he will have the highest score on his finishing ability. In this way, the performance of a batsman is estimated and calculated by such metrics.

Now, coming to the bowling metrics, the following are some of them used to evaluate the performance of the bowler. They are:

Economy: Run scored / (No. of ball bowled by bowler / 6).

Wicket taking ability: No. of balls bowled / Wicket taken.

Consistency: Run conceded / Wicket taken.

Crucial Wicket Taking Ability: No. of times four or five-wicket taken / No. of innings played.

For most of the bowlers the first metric and the most important metric is his/her wicket taking ability. Similar to that of batsman mentioned earlier, a bowler may have the best economy rate but if the bowler is not consistent, it affects the game of the team's performance too, so o an average a bowler has to be consistent and with a good economy rate.

Apart from all of this, the analyst has to analyse based on the type of bowling that the bowler possesses, such as fast bowling, pace, spin, etc. For a fast bowler, the speed at which the bowler can bowl is bowling in a particular game, and for a spin bowler the length of the ball is calculated, whether he/she is bowling yorkers, full- length balls, short-length balls etc. has to be calculated and then along with the other metric factors, these are also one of the major criteria.

The International Cricket Council (ICC) gives out rankings for a batsman, bowler and all-rounders (who does both batting and bowling) every year. Their metrics have slightly different metrics such as ranking and rating for a player and they differ from format to format. To be more detailed, the rankings are diverse for a player for ODI's, T20's and test match formats.

The players are rated on a scale from 0 to 1000 and this rating changes after every match. If a player performs better for example the batsman scores more runs than the previous match, his/her rating improves and vice-versa. These ratings are subject to change after every series. The Cricket Analyst here has to keep track of the data and handle this huge data base of stats of players across all the teams in the world and use any Database query language (SQL etc.) to fetch the data whenever needed.

Every year, ICC gives out the ranking based on the ratings. The council issues separate rankings for men and women across all game formats for every batsman, bowler and all-rounder. Here are some of the rankings from the last year.

Men's Batting Rankings

TEST	Test Batting Rankings			ODI	ODI Batting Rankings			T20I	T20I Batting Rankings		
1 — (0)	Joe Root ENG 903			1 — (0)	Babar Azam PAK 873			1 — (0)	Babar Azam PAK 809		
POS	**PLAYER**	**TEAM**	**RATING**	**POS**	**PLAYER**	**TEAM**	**RATING**	**POS**	**PLAYER**	**TEAM**	**RATING**
2 — (0)	Steve Smith		891	2 — (0)	Virat Kohli		844	2 — (0)	Dawid Malan		805
3 — (0)	Kane Williamson		888	3 — (0)	Rohit Sharma		813	3 — (0)	Aiden Markram		796
4 — (0)	Marnus Labuschagne		878	4 — (0)	Ross Taylor		801	4 — (0)	Mohammad Rizwan		735
5 — (0)	Rohit Sharma		805	5 — (0)	Aaron Finch		779	5 — (0)	Lokesh Rahul		729

POS	PLAYER	TEAM	RATING
1 — (0)	Lizelle Lee 🇿🇦 SA		761
2 — (0)	Alyssa Healy	🇦🇺	750
3 — (0)	Mithali Raj	🇮🇳	738
4 — (0)	Tammy Beaumont	🏴󠁧󠁢󠁥󠁮󠁧󠁿	728
5 — (0)	Amy Satterthwaite	🇳🇿	717
6 — (0)	Smriti Mandhana	🇮🇳	710
7 — (0)	Meg Lanning	🇦🇺	699
8 — (0)	Beth Mooney	🇦🇺	690

POS	PLAYER	TEAM	RATING
1 — (0)	Jess Jonassen 🇦🇺 AUS		760
2 — (0)	Jhulan Goswami	🇮🇳	727
3 — (0)	Megan Schutt	🇦🇺	717
4 — (0)	Marizanne Kapp	🇿🇦	715
5 — (0)	Sophie Ecclestone	🏴󠁧󠁢󠁥󠁮󠁧󠁿	701
6 — (0)	Shabnim Ismail	🇿🇦	688
7 — (0)	Katherine Brunt	🏴󠁧󠁢󠁥󠁮󠁧󠁿	666
8 — (0)	Ayabonga Khaka	🇿🇦	643

POS	PLAYER	TEAM	RATING
1 — (0)	Marizanne Kapp 🇿🇦 SA		384
2 — (0)	Natalie Sciver	🏴󠁧󠁢󠁥󠁮󠁧󠁿	372
3 — (0)	Ellyse Perry	🇦🇺	365
4 — (0)	Stafanie Taylor		319
5 — (0)	Deepti Sharma	🇮🇳	299
6 — (0)	Ashleigh Gardner	🇦🇺	275
7 — (0)	Dane van Niekerk	🇿🇦	274
8 — (0)	Hayley Matthews		272

The technology is so much developing that there are devices like sensors to track the data of each and every player right from his/her calory intake to the number of calories burnt, and training levels etc. Apart from this, broadcasting the match and entertaining the audience also involves lot of big data analytics like for example: the broadcasting channel Star Sports asks few questions like "Which player is going to perform well today?" and then audience react to such polls and the channels also provide their choice or the prediction by considering the stat history of the player and then displaying on the screen in-order to create the fun and entertain the audience.

Now, apart from these stats and figures in the form of scores, strike rate, economy etc, the other form of data like video footages of how a particular batsman is reacting and all such data is stored and analysed using some of the big data technologies like Hadoop etc. Data Analytics is also used to track the fitness stats of a player and how likely he/she is prone to have an injury. This helps the team to know that a certain if fit or unfit for the match.

Records

Most consecutive five-wickets-in-an-innings - Test matches

Player	5wi	Bowl	Inns	Team	Opposition	Ground	Match Date	Scorecard
CTB Turner	6	5/44	1	Australia	v England	Sydney	10 Feb 1888	Test # 27
		7/43	3	Australia	v England	Sydney	10 Feb 1888	Test # 27
		5/27	2	Australia	v England	Lord's	16 Jul 1888	Test # 28
		5/36	4	Australia	v England	Lord's	16 Jul 1888	Test # 28
		6/112	2	Australia	v England	The Oval	13 Aug 1888	Test # 29
		5/86	1	Australia	v England	Manchester	30 Aug 1888	Test # 30
T Richardson	5	6/104	3	England	v Australia	Melbourne	1 Mar 1895	Test # 46
		6/39	1	England	v Australia	Lord's	22 Jun 1896	Test # 50
		5/134	3	England	v Australia	Lord's	22 Jun 1896	Test # 50
		7/168	1	England	v Australia	Manchester	16 Jul 1896	Test # 51

Data analytics is not limited just to the scenarios mentioned above but also helps the tools like Insights by 'ESPNCrickinfo 'and 'ScorewithData-IBM game changer', these tools try to make the predictions using the data analytics and algorithms and release their predictions and probability of winning chances for a team just before the match starts and attracting the viewers. Some stats by the ESPNcricinfo is shown in the above figure.Also, the broadcasting partners such as ESPN, Star Sports, Hotstar etc.

Role of Machine Learning in Cricket Analytics:

Over the years Machine Learning and algorithms have made a positive impact in cricket. It helps the team to evaluate accuracy and make predictions and in turn useful to better their strategies. To make predictions on how a player is going to perform in the next upcoming game, past data about the player are considered, and then features are selected carefully depending on the type of player whether the batsman or the bowler, wicket keeper, fielder etc.


Many algorithms and approaches proposed by many people on predictions. Some of them include feature encoding and the convolutional neural network approach. The feature encoding approach involves taking the players from both the teams and the team's compositions as input This approach needs a training set, which consists of all players past stats and their performance in each season and building clusters via K-means Clustering and then combine the overall performance of the players in a certain cluster. All this is a part of VLAD feature encoding scheme. Besides this, a convolutional network is built which involves transferring the features represented for each and every player from the original feature space to a more discriminative feature space

and then a classifier is applied on these representations. To better optimize the network, Focal Loss is used instead of Cross Entropy as Focal Loss perform better for classification problems.

As the training data increases day by day, as the matches are continuously played, the number of layers for the Convolutional Neural Network also increases and hence it performs with better accuracy and produce better predictions.

There is a special focus on cricket taken up by one of the prestigious universities, MIT and there is a page where they develop Machine Learning metrics for cricket on http://cricket.mit.edu/ , one of their proposals is based on the DLS (Duckworth Lewis Stern) method to make a decision when there is a situation of target-revision. Though this method is labelled efficient by the makers, there are many situations from past few years there are times where the DLS result seems to be unjustified. The below figure shows the stats with DLS and without DLS.

	Won by Team Batting First	Won by Team Batting Second	Total
No DLS	865 (48.5%)	916 (51.5%)	1781
DLS	70 (40.6%)	102 (59.3%)	172
Total	935	1018	1953

A group of professors and researchers at MIT proposed a new method based on Novel Algorithm, this estimates an average path to win the game and the number of runs targeted in the maximum overs allocated. Similar to resources remaining used in DLS, some non-linear are taken from these resource data, the new revised scores and par- scores can be calculated and on the estimated decomposition forecasts are made. These are one of the results they got after performing the Novel algorithm.
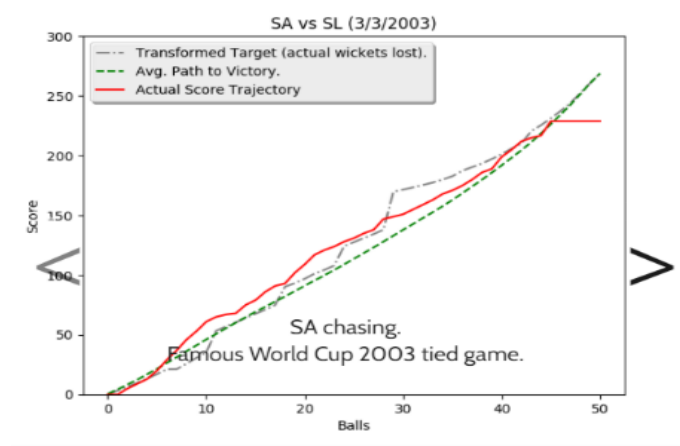
The other approaches include data mining and algorithm approaches to predict the outcomes of the matches. Algorithms like SVM, Naïve Bayes and make some match outcome predictions. The Data is always changing and very complex, so there is a need for highly scalable algorithms. The

training data is very huge, apart from the stats of the players, venue of the match, weather, dew factor, time of the day when the match is played, whether it is a home for the team etc. All these attributes should be taken into account. Some of these play a crucial rule in estimating the team's performance.

Role of Artificial Intelligence in Cricket Analytics:

With the advancement in technologies, the machine can build insights with more depth. This is the era where AI plays a predominant role in almost all sectors. In Cricket, a special bat called "Power Bat" was introduced back in 2017 and most players have been using them since then. This bat has a sensor that is stuck on the handle of the bat which records the speed of the ball, how much the ball is twisted after hitting the bat etc. More recently, a sensor sticker is made by Microsoft IOT team which is much lighter and track records all the values. Based on these value and hitting quality from the batsman, a Power Spek final score is calculated.


The working of the sensor is as follows: the sensor data is connected to a cloud server and there are live analytics shown the screen with the graphs and screen with the help of AI. The 'ESPNcricinfo 'select its super stars with the help of sophisticated AI and machine Learning Algorithms. There is a term called "Luck index" which predicts the toss, faulty empire decisions and no-balls etc.
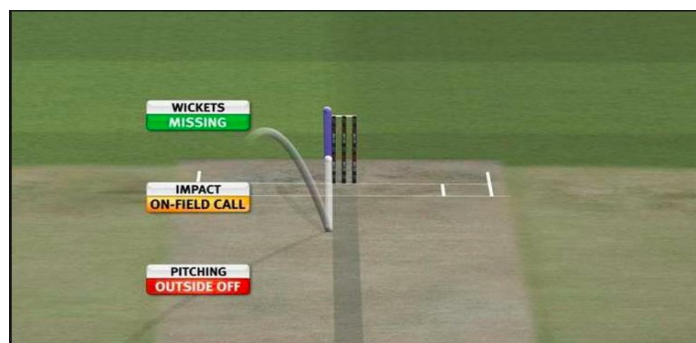
AI is also helping in decision-making for the umpires especially in the scenario when there is an LBW. It is also helping the broadcasters by providing automatic subtitles for the commentary and an optimized camera angles to attract the audience better.

There are lot of devices developed using Artificial Intelligence and Machine Learning. One of the devices is 'AI AirSim' Drones, help us in building a 3-D virtual view of the environment. This helps the third umpire sitting in the dug-out room visualize the things better and make better judgements. The other device is Pose Tracker, with the help of Deep Learning, the tracker calculates the distance of the ball, angle of the batsman, bowler in real time. The below figures shows the AirSim Drone and the created virtual 3-D image.
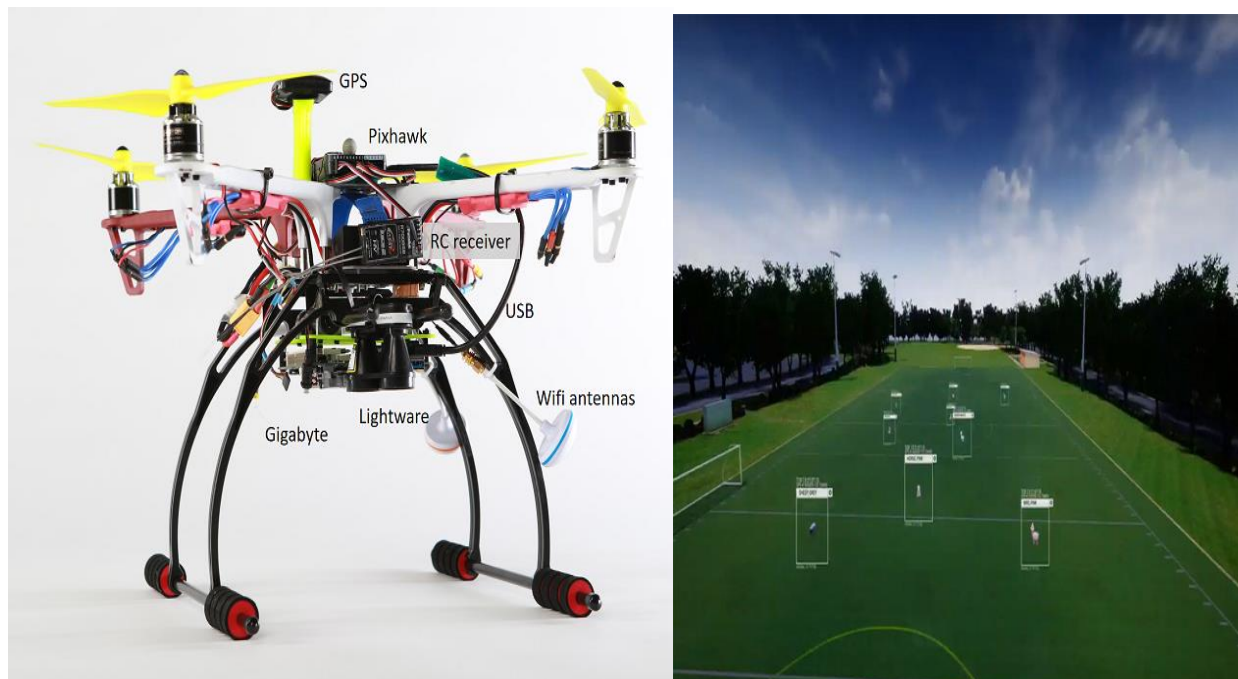
Limitations of Data Analytics, Machine Learning and AI in Cricket:

- The data is very huge and dynamic, so exploratory data analysis becomes hard.
- The predictions may not be accurate all the time, for instance, if the machine predicts that the bowler is going for a slow ball, but the bowlers change their strategy and go for a fast ball depending on the situation, in these cases the algorithms may not predict very much accurately.
- More efficient and highly scalable algorithms are needed, as there is a need for ball-to-ball tracking while playing the game. They have to be precise and quick.

Conclusion and Future Scope:

In conclusion, though there are some limitations with the Analytics in Cricket, the field is still growing and can be expected to have more advancements in the coming future. There can be algorithms built to generate a personalized detailed diet plan for the players to follow based on their physical and mental health conditions, past medical history, the type of food they eat etc. We can expect more sophisticated and advanced AI devices, sensors, fitness trackers which helps players to better themselves.
The only tool that that takes most of the areas or the sectors to a better place is technology and hope that it makes the world we live, into a better place.

References:

- https://www.analyticsvidhya.com/blog/2021/05/the-data-science-behind-ipl/
- https://www.microsoft.com/en-us/ai/ai-lab-airsim-drones
- http://cricket.mit.edu/
- https://www.researchgate.net/publication/316958473_A_Machine_Learning_Approach_to_Analyze_ODI_Cricket_Predictors
- https://medium.com/@springboard_ind/use-of-data-science-in-the-game-of-cricket-ace3fa41c14b
- https://www.projectpro.io/article/big-data-analytics-the-new-player-in-icc-world-cup-cricket-2015/89
- https://madaboutsports.in/ms-dhoni-worth-the-hype/