Analytic Implementation of Big Data in Cricket using Hadoop Framework

Bhawna¹, Seema Baghla²,

¹Student M. Tech. (Comp. Engg.), ²Assistant Professor (Comp. Engg.) Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo, Bathinda, Punjab, India Email: bhawnagarg232@gmail.com¹, seemabaghla.ycoe1@gmail.com²

Abstract-Big data creates a new way in the cricket to predict the results of a game that which team is going to win the game and also help in to improve the performance of the team. In this paper, data mining is performed over the self-created database using the Apache hadoop framework. With the use of mined data, a team can analyse the strategy followed by the opponent team and predict how they are going to play in upcoming matches. The popularity of a particular player can also be analysed by the number of followers on twitter. Players dependent parameters are considered based upon their batting and bowling data to extract the best insights and to classify the batsman and bowler into different categories. In experimental work, analysis of parameters is done using hadoop and hive for all major cricket teams.

Key-words- Cricket team prediction, Apache hadoop, Mapreduce algorithm, HCatalog tool, Beeswax Hive tool.

1 INTRODUCTION

Big data is a term used to refer the data sets that are too big to be handled using the existing database management tools [11]. Only those technologies that can present vast quantities of structured and unstructured data quickly in the right context are the ones that will provide genuine insights. Not only will these insights enhance our knowledge but also lead to proactive decision making which can redefine the future of all our ventures. Internationally, sports like baseball and basketball have already adopted statistical analysis to enhance their performance. Today cricket has been added to the list of technologically enhanced sports with the adoption of the advanced big data analytics. Big data creates a new way in the field of cricket to predict the outcome of a game by analysing the numerous amounts of data generated in every match of cricket. This data can be generated from the various wearable devices like fitness trackers that present the real-time stat of each player including the speed of batsman hitting the ball and the speed of bowlers throwing the ball, their heart rate and the acceleration. According to the data generated coaches can analyse the performance of each batsman and the bowler.

2 LITERATURE REVIEW

Gandomi and Haider (2015) [1] explained the concept of analytics to gain valuable and valid insights from big data. They reviewed various

analytics techniques for video, text, social media data, audio and predictive analytics. Che et al. (2013) [2] discussed the platforms for managing and processing big data as well as efforts anticipated on the big data mining. They pointed out the privacy issue which is collapsed by big data. They proposed and expressed the garbage mining which is a serious issue that is not realized and addressed by other in the era of big data. Bhosale and Gadekar (2014) [3] described the challenges faced during the processing of big data like heterogeneity, lack of structure, and privacy and these challenges must be addressed for the efficient and fast processing of big data. They use the hadoop as the solution for the processing of big data. Dean and Ghemawat (2004) [4] explained the implementation of MapReduce that processes many terabytes of data on a large cluster of commodity machines. Khan et al. (2014) [5] created the data-life cycle using the terminology and technologies of big data. This cycle enhances the efficiency of data management by converting raw data to published data. This paper also discussed the issue of data integrity. Rein and Memmert (2016) [6] introduced the big data in the field of elite soccer and also assist in other sports science field as enormous data become an extensive phenomenon. Shweta and Garg (2013) [7] used Weka to find out the best association rule. Here author consider three association rule algorithms and compare them: Apriori Association

Rule, Predictive Apriori Association Rule and Tertius Association Rule. Author compares the result of these three algorithms which shows that Apriori Association algorithm performs better than any other algorithm. Anjali et al. (2015) [8] explained the use of big data analysis in the field of sports. They use the real time data of cricket tweets to get the accurate results. For the analysis of data, map reduce algorithm is used. They suggest to use the big data analysis in other sports also like football, baseball etc. Agarwal et al. (2017) [9] explained the use of statistical modeling approach in the field of cricket to predict the best appropriate team to be lined up for a specific match. Manikandan and Ravi (2014) [10] explained that the use of big data tools like mapreduce algorithm over hadoop and hdfs helps the organization in taking best business Mukherjee and Shaw (2016) [11] discussed the practical and theoretical challenges that are hindering in the development of big data analytics. Phaneendra and Reddy (2013) [12] explained the hadoop architecture consisting HDFS to handle big data systems. The authors also focused on various challenges faced by enterprises while handling big data. Ahlawat et al. (2016) [13] defined the ecosystem, developed various models and categorized the elements on the basis of big data. They are able to identify various techniques and technologies for big data analytics. They analysed the areas for the usage of big data analytics. The various data forms are also discussed in this paper.

3 METHODOLOGY

The role played by big data for obtaining the objectives are elaborated with the working of MapReduce algorithm and Hortonworks Sandbox 2.2.0. Initially the pre-processed and the compiled (csv) data files are uploaded in hadoop based Hortonworks platform using file browser tool. These files are then accessed using HCatalog tool to form the tables. To obtain the desired results, scripts and queries are written on Beewax Hive tool. Twenty- four parameters are considered based upon the batting and bowling data of a player. Table 1 shows the description of the batting and bowling attributes of cricket.

Table 1 Cricket attributes with descript	ion
--	-----

Attribute	Description
Player	Name of the players
Matchtype	Specify the type of match like "ODI" or "T20"
Country	Name of the countries

Startingyear	Players started to play the match
Last year	Specify the year till the player play the match
Matches	How many matches are played by the player
Innings	divisions of a match during
6	which one team takes its turn
	to bat
Notout	Player will be not out if he
Deres	comes out to bat in an innings
Runs Striltorate	kuns scored by the player
i Batting	a batsman achieves
strikerate	the primary goal
ii. Bowling	of batting, namely
strikerate	scoring runs
	ii. how frequently
	a bowler achieves the
	primary goal
	of bowling, namely
TT' 1	taking wickets
Hignestscore	High number of scores made
	by the player in specific matchtype
Centuries	Score of 100 attain by the
Conturios	batsman
Halfcenturies	Score of 50 attain by the
	batsman
Ducks	When the player is dismissed
	without facing a ball
Overs	an over consists of six
	consecutive balls bowled by
Maidanaan	a <i>single</i> bowler
Waldenovers	no runs are scored
Wickets	Number of wickets taken by
	the bowler
Economy	It is the average number of runs
	agreed for each over bowled
Fourwickethaul	Refers to a bowler taking
	four wickets in a single innings
Fivewickethaul	Refers to
	a bowler taking wickets in a
Catches	Number of times howler
Cateries	dismissed the batsman by
	catching a ball
Stumpings	Method to dismiss the player
Ballsfaced	Number of balls faced by the
	player
Average	i. Average is the total
i. Batting	number of runs that
average	they have scored
II. DOWIING	number of times they
average	have been out
	ii. It is the ratio
	of runs conceded
	per wickets taken.

4 RESULTS AND DISCUSSIONS

The process is initiated by collecting the data on the basis of some attributes as shown in table 1 related to the field of cricket, to enhance the winning chances of a team in the matches of "ODI" and "T20". Figure 1 represents the database of cricketers based upon their batting performances.

	-					,											
	A	В	C	D	E	F	G	н	1	J	K	L	M	N	0	P	
1	Player	Matchtyp	Country	Startingy	Last year	Matches	Innings	Notout	Runs	Highestse	Average	Ballsface	Strikerati	centuries	halfcenturi	Ducks	
2	Abdullah	ODI	Afghanistan	2010	2010	2	- 1	0	3	3	3	16	18.75	0	0		٤
3	Afser Zeze	ODI	Afghanistan	2014	2017	17	16	1	264	60	17.6	488	\$4.09	0	2	1	ł.
4	Afsar Zazz	T20	Afghanistan	2013	2013	1	3	0	9	9	9	21	42.85	0	0		٥
5	Aftab Alar	ODI	Afghanistan	2010	2015	12	8	4	25	14	6.25	38	65.78	0	0	- 1	1
6	Aftab Alas	T20	Afghanistan	2012	2018	9	3	2	1	1*	1	5	20	0	0	1	4
7	Ahmed Sh	ODI	Afghanistan	2009	2009	1	3	0	2	2	2	10	20	0	0		٤
8	Amir Ham	ODI	Afghanistan	2012	2017	31	16	9	23	7	3.28	94	24.46	0	0	1	4
3	Amir Ham	T20	Afghanistan	2013	2017	31	9	5	40	21	10	34	105.26	0	0	1	4
10	Asghar St	ODI	Afghanistan	2009	2018	86	81	6	1608	101	21.44	2580	62.32	1	7	5	ŝ.
11	Asghar St	T20	Afghanistan	2010	2018	51	46	4	907	62	21.59	853	106.33	0	3		2
12	Dewlet Al	ODI	Afghanistan	2009	2010	3	0	0	0	0	0	0	0	0	0		5
13	Dawlat A	T20	Afghanistan	2010	2010	2	3	1	2	2*	0	3	66.66	0	0		٥
14	Dawlat Za	ODI	Afghanistan	2011	2018	72	49	25	500	47*	20.83	624	80.12	0	0		6
15	Dewlet Ze	T20	Afghanistan	2012	2017	33	16	7	68	13	7.55	42	161.9	0	0	1	i.
16	Fareed At	ODI	Afghanistan	2014	2017	5	2	2	1	1*	0	6	16.66	0	0		٥
17	Fareed At	T20	Afghanistan	2016	2017	7	0	0	0	0	0	0	0	0	0		5
18	Gulbedin	ODI	Afghanistan	2011	2018	41	34	5	615	82*	21.2	851	72.26	0	4		ł.
19	Gulbadin	T20	Afghanistan	2012	2018	38	31	9	432	56*	19.63	357	121	0	1	1	2
20	Hamid Ha	ODI	Afghanistan	2009	2016	32	17	4	92	17	7.07	171	53.8	0	0	1	4
21	Hamid Ha	T20	Afghanistan	2010	2016	22	9	6	50	22	16.66	46	108.69	0	0		ð.
22	Hashmat	ODI	Afghanistan	2013	2017	17	17	3	322	72	23	556	57.91	0	1	1	ė.
23	Hashmat	T20	Afghanistan	2013	2013	1	3	1	1	1*	0	1	100	0	0		5
24	Hasti Gul	ODI	Afghanistan	2009	2009	2	1	1	23	23*	0	11	209.09	0	0		ð.
25	Hazratulk	T20	Afghanistan	2016	2016	1	1	0	18	18	18	24	75	0	0	0	5
26	Ihsanulla	ODI	Afghanistan	2017	2018	10	10	1	170	54	18.88	246	69.1	0	2		ė.
27	Izatullah	ODI	Afghanistan	2010	2015	5	3	1	7	6*	3.5	27	25.92	0	0	1	i.
28	Izatullahi	T20	Afghanistan	2012	2012	4	1	1	0	0*	0	3	0	0	0	0	5
29	Javed Ahr	ODI	Afghanistan	2010	2018	39	36	0	920	81	25.55	1161	79.24	0	7	1	i.
30	Javed Ahr	T20	Afghanistan	2012	2017	3	2	0	6	6	3	7	85.71	0	0	1	i.
31	Karim lan	001	Afrhanistan	2017	2017	1	1	0	9	9	9	33	27.27	0	0		1

Fig. 1. Batting database of cricket

С	1	Ŧ	: ×	$\sqrt{-f_x}$	Cou	untry										
4	A	в	С	D	Е	F	G	н	1	J	К	L	м	N	0	Ρ
1	Player	Matchty	Country	Startingy	Last year	Matches	Innings	Notout	Runs	Highests	c Average	Ballsface	Strikerate	centuries	halfcenturi	Ducks
2	Abdullah	ODI	Afghanistan	2010	2010	2	1	0	3		3	16	18.75	0	0	
3	Afsar Zaza	ODI	Afghanistan	2014	2017	17	16	1	264	60	17.6	488	54.09	0	2	
4	Afsar Zaza	T20	Afghanistan	2013	2013	1	1	0	9	9	9 9	21	42.85	0	0	
5	Aftab Ala	ODI	Afghanistan	2010	2015	12	8	4	25	14	6.25	38	65.78	0	0	
6	Aftab Ala	T20	Afghanistan	2012	2018	9	3	2	1	1*	1	5	20	0	0	
7	Ahmed Sh	ODI	Afghanistan	2009	2009	1	1	0	2	2	2 2	10	20	0	0	(
8	Amir Harr	ODI	Afghanistan	2012	2017	31	16	9	23	1	3.28	94	24.45	0	0	
9	Amir Harr	T20	Afghanistan	2013	2017	31	9	5	40	21	1 10	38	105.26	0	0	
0	Asghar St	ODI	Afghanistan	2009	2018	86	81	6	1608	101	21.44	2580	62.32	1	7	1
1	Asghar St	T20	Afghanistan	2010	2018	51	46	4	907	63	21.59	853	106.33	0	3	
2	Dawlat Al	ODI	Afghanistan	2009	2010	3	0	0	0	() 0	0	0	0	0	(
3	Dawlat Al	T20	Afghanistan	2010	2010	2	1	1	2	2*	0	3	66.66	0	0	(
4	Dawlat Za	ODI	Afghanistan	2011	2018	72	49	25	500	47*	20.83	624	80.12	0	0	
5	Dawlat Za	T20	Afghanistan	2012	2017	33	16	7	68	13	7.55	42	161.9	0	0	
6	Fareed Al	ODI	Afghanistan	2014	2017	5	2	2	1	1*	0	6	16.66	0	0	(
7	Fareed Al	T20	Afghanistan	2016	2017	7	0	0	0	0	0 0	0	0	0	0	(
8	Gulbadin	ODI	Afghanistan	2011	2018	41	34	5	615	82*	21.2	851	72.26	0	4	
9	Gulbadin	T20	Afghanistan	2012	2018	38	31	9	432	56*	19.63	357	121	0	1	
20	Hamid Ha	ODI	Afghanistan	2009	2016	32	17	4	92	17	7.07	171	53.8	0	0	
1	Hamid Ha	T20	Afghanistan	2010	2016	22	9	6	50	23	16.66	46	108.69	0	0	(
22	Hashmat	ODI	Afghanistan	2013	2017	17	17	3	322	73	23	556	57.91	0	1	
23	Hashmat	T20	Afghanistan	2013	2013	1	1	1	1	1*	0	1	100	0	0	(
4	Hasti Gul	ODI	Afghanistan	2009	2009	2	1	1	23	23*	0	11	209.09	0	0	(
25	Hazratull	T20	Afghanistan	2016	2016	1	1	0	18	18	18	24	75	0	0	(
26	Ihsanulla	ODI	Afghanistan	2017	2018	10	10	1	170	54	18.88	245	69.1	0	2	
27	Izatullah	ODI	Afghanistan	2010	2015	5	3	1	7	6*	3.5	27	25.92	0	0	
28	Izatullah	T20	Afghanistan	2012	2012	4	1	1	0	0*	0	3	0	0	0	(
29	Javed Ahr	ODI	Afghanistan	2010	2018	39	36	0	920	81	25.55	1161	79.24	0	7	
80	Javed Ahr	T20	Afghanistan	2012	2017	3	2	0	6	(; 3	7	85.71	0	0	
31	Karim lar		Afehanistan	2017	2017	1	1	0	9		9	33	27.27	0	0	

Fig. 2. Bowling database of cricket

Figure 2 represents the dataset of cricketers based upon their bowling performances.



Fig. 3. Script 1 written in hive tool

Figure 3 shows how the script is written to mine the data. The resultant output of script 1 is shown in figure 4 which is in the tabular form.

🔄 💐 🗿 🤏 💈 📣 💆 Query Editor - Mr Queries - Saved Queries	0 Hist	I (3 Settings									å he•
,												
DOWNLOADS	F	lesuts Que	ery Log	Columns	Visualizations							
Download as XLS	÷	+ player	matchtyp	e ‡count	try 🕴 centurie	s 🕴 runs	+ averag	+ startingyear	+ lastyea	r 🕴 strikerate	† inning	s 🕴 notout
Save	0	AM Rahane	ODI	India	3	2962	35.26	2011	2018	78.63	87	3
	1	AT Rayudu	ODI	India	2	1055	50.23	2013	2016	76.28	30	9
Did you know? If the result contains a X	2	5 Kumar	UDI	India	0	379	14.57	2012	2018	16.25	40	14
large number of columns, click a row to	3	IK Pathan	001	Infa	0	1044	23.39	2004	2012	13.54	6í 67	17
select a column to jump to. As you type into the field, a drop-down list displays	*	MS Dhoni	00	Inda	9	9793	510	2004	2010	87.94	269	77
column names that match the string.	6	P Kumar	ODI	Inda	0	292	13.9	2007	2012	88.21	33	12
	7	R Ashwin	ODI	India	0	675	16.07	2010	2017	86.98	61	19
	8	RA Jadeja	ODI	India	0	1914	31.37	2009	2017	85.29	93	32
	9	RG Sharma	ODI	India	17	6594	44.55	2007	2018	86.96	174	26
					_		_				_	
_	F	ig.	4. R	lesi	ults (of	scri	pt 1	in t	abu	lar	
	F	ig.	4. R	lesi	ults (of	scri	pt 1	in t	abu	lar	
	F	ig.	4. R	lesi	ults o	of :	scri	pt 1	in t	abu	lar	-
	F	ig.	4. R	lesi	ults o	of :	scri	pt 1	in t	abu	lar	
	F	ig.	4. R	lesi	ults o	of	scri	pt 1	in (abu	lar	
	F	ig.	4. R	Resi	alts o	of	scri	pt 1	in t	abu	lar	
	F	ig.	4. R	Rest	ults o	of :	scri	pt 1	int	abu	lar	
	F	ig.	4. R	Resi	alts o	of :	scri	pt 1	in (abu	lar	
	F	ig.	4. R	Resu	alts o	of	scri	pt 1	in t	abu	lar	
	F	ig.	4. R	Resu		of	scri	pt 1	in t		lar	
Centuries n	F	rig.	4. R	Rest	alts of	of :	scri	pt 1	int	rabu	lar	
	F	rig.	4. R	Rest	ults of states	of : ating	scri	pt 1	int	rabu	lar	i
Centuries r. Regend	F	ig.	4. R	Rest	ults of starts	of : ating	year	pt 1	int	abu	lar	ŕ
Centuries r. K legend	F	rig (Pathan	4. R	rage	alts of	of : ating	year	pt 1	int	abu J	lar	ŕ
Centuries II. Regend D Kataka II.S. Tkayuku IB Kunar	F	Cig. (Patan Astwin	4. R	rage	alts of	of :	year	pt 1	in t	abu J	lar	i
Centuries II RAItatae AT Rayudu B Kumar D Katha Nis Dicoi P Kumar RA Ladeje Ra Stama RY Uthago	F JINS	Cig. (Patan Ashvin Dhavan	4. R	rage	ults of states o	of :	year	pt 1	in t	abu Strike	lar	ir
Centuries II. Vegend Mi Ratare AT Rayudu Bi Kunar Di Katha Mis Ditori Pi Kunar RJ, Loteje ARG Stama RV Uttago SK Bata V Vehi W Shaka	F JINS	Fig (Patan Ashvin Dhavan	4. R	Rest	ults of states of the states o	of :	year	pt 1	in t	abu strike	lar	ŕ



Figure 5 represents the result of script 1 in pie chart form.

Figure 6 represents the script 2 written in the beewax hive tool to mine the data from bowling database of cricket.

	Quary Editor : main1	
DATABASE	Query Eultor . Indiff	
default .	0	
SETTINGS	1 dren table if exists ball1:	
	2 create table ball1 as select	
ACO	3 player,	
	5 country,	
FILE RESOURCES	5 startinglean, 7 lastvean.	
Add	8 overs,	
	9 maidenovers, 18 wirkets.	
USER-DEFINED FUNCTIONS	11 average,	
101	12 strikerate from boxiling 13 where matchtwoe="T20" and	
	14 country="India" and	
DEDAMETED/74TVW	15 Star Lingren Patrex and 16 Lastyear<-2018 and	
G Cashin Damastrutation	17 overs)=48 and	
	19 wickets>+5 and	
ENAL NOTECATION	38 average)=5 and 11 strikerata5:	
Email me on completion	The second second	

Fig. 6. Script 2 written in hive tool

Figure 7 shows the result of script 2 in the tabular form.

Cuery Editor My Queries	Saved Queries	🖸 🔳 🌡 History Tables	selins	åhue∙
Query Resu	lts: ball	1		
DOWNLOADS Downicad as CSV	¢	Results Q	erj Log Columns Visualizations	

388 0 A Helta 720 India 208 207 980 2 34 223 172 1 1 Bumer 720 India 208 204 94 2 28 2247 186 2 HardapanSingh 720 India 206 2016 920 5 1 25 252 244 3 HF Pangia 720 India 206 2016 920 5 1 35 2434 165 4 KPbani 720 India 206 000	Download as XLS		i bulle.		(waited	amonglear	- margan	Vitia	Theresis are a	Therese	anenage	
1 Funce 72 Hela 2012 2018 914 2 28 2017 195 2 HelangerStep 720 Hela 2016 2015 1 25 25 252 244 3 HEPangerStep 720 Hela 2016 2015 1 28 244 185 4 KPather 720 Hela 2016 2015 211 218 2016 13 218 2017 15 4 KPather 720 Hela 2016 2015 2012 77.0 1 28 2017 15 6 100-max 100 100 2016 2016 100	Save	0	A Nehra	T20	India	2009	2017	98.0	2	34	22.29	17.2
2 https://righting.instruction 1016 2016 102 5 15 252 244 3 HFParting 121 1016 2016 2016 1 15 244 165 4 KFParting 121 1016 2016 2017 1 128 2017 165 4 KFParting 120 1016 2016 101 218 2107 165 4 KFParting 120 1016 2016 120 1 218 2017 165 5 120 1016 2016 1017 1 218 120 101		1	8 Kumar	T20	Itóa	2012	2018	91.4	2	28	22.07	19.6
3 #PiPada 120 tota 205 2010 055 1 35 444 165 4 KAdata 120 tota 205 2012 170 1 33 220 65 5 Concernent 120 tota 205 2012 170 1 33 200 65 5 Concernent 120 tota 205 2010 103 1 35 185 13 1 Tota 1 20 tota 205 2010 113 1 35 185 13 1 Tota 1 20 tota 205 2010 113 1 35 185 13 1 Tota 1 20 tota 205 2010 113 1 35 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 13 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 113 1 105 185 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 10 10 10 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 2010 114 10 100 10 1 Tota 1 20 tota 205 200 10 10 10 10 10 10 10 10 1 Tota 1 20 tota 205 200 10 10 10 10 10 10 10 10 10 10 10 10 1		2	Harbhajan Singh	T20	India	2006	2016	102.0	5	25	25.32	24.4
$4 \\ 4 \\ 7 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1$		3	HH Pandya	T20	India	2016	2018	80.5	1	26	24.84	18.6
		4	IK Pathan	T20	India	2006	2012	77.0	1	28	22.07	16.5
Fig. 10. Result of query in tabular form		¢	11 Dummh	750	ladia	4542	9548	400.4	1	н	50.45	47.0
Image: res	Fig		10. F	Resu	lt o	f que	erv	in	tabul	ar f	orn	1
8 YSCharla T21 hda 2016 2018 813 1 35 19.5 19.3		,	телоходи	169	11919	L		1996-1	IV.	vi	VI.I	44.1
V O O O O O O O O O O O O O O O O O O O		8	YS Chahal	T20	Itóa	2016	2018	81.3	1	35	18.45	13.9
VyEttr MyOurie Streto Dense Helloy Tuble Settings		_										
Ny Cetar My Cares Seet Cares Hittiny Tales Settings	🕺 🖲 🖷 💈 🔌 🐌			3								
(third e and e and and e and e and e and	ery Editor My Queries Saved Queries	H	itory Tables	Settings								
												tikezta

Fig. 8. Results of script 2 in area covered

Figure 9 shows the query written in hive tool to analyis the popularity of players on social media site like twitter based upon follower> 3000000.

(4) 😵 🖲 🚾 📓 🔕 🚺	O M 9 3
Query Editor My Queries Saved Queries	History Tables Settings
DATABASE default SETTINGS Add FILE RESOURCES Add USER DEFINED FUNCTIONS Add PARAMETERIZATION Email motification Email wotification	Query Editor 2 2 2 3 4 4 4 4 4 4 4 4 4 4 4 4 4
	Execute Save as Explain or create a New query

Fig. 9. Query written in hive tool

8 😵	U 🖌	4 💈	0	\$	8	å hue 🔹
Query Editor	My Queries	Saved Queries	History	Tables	Setlings	

Query Results: Unsaved Query

10000000 L				
Download as CSV Download as XLS	+	¢ player	0 matchtype	followers
Save	0	AM Rahare	ODI	4800000
MP IOR (I)	1	G Gambhir	ODI	8300000
job 201807060554 0054	2	MS Dhoni	ODI	7100000
	3	R Ashuin	ODI	0000088
	4	R Shama	ODI	12100000
id you know? If the result contains a 🕺	5	S Dhawan	ODI	300000
elect a column to jump to. As you type	6	SK Raina	ODI	1470000
to the field, a drop-down list displays	1	SR Tendukar	CDI	2550000
olumn names that match the stilling.	8	V Kahli	ODI	25700000
	9	V Sehwag	CCI	1750000



F Fig. 11. Result of query in line form n and figure 11 shows the same resultant output in line graph form.

5 Conclusions and future scope

Even though the use of Big Data analytics has been a game-changer in the world of sports, it should be for informative purpose only. And for professional cricketers these results can take on a whole new meaning. If technologies behind the data are customized for cricket teams they can pave the way towards a higher level of performances. Similar systems could easily condense years of data and help teams make better decisions when it comes to team selection and game strategy. By combining data driven insights with years of experience and practice, even the most experienced professionals can make better decisions. If a team stops depending on its skills and gut instincts, the whole excitement of the game would vanish. So, there should be a distinctive gap between sports and Big Data analytics.

So, it can be concluded that

- Teams should effectively consume data to derive insights.
- Predicting the entire outcome of the game is difficult, but technology can assist in doing so.
- Human decision making still dominates technology as most teams have similar data but performance varies.

In future, with the help of Machine Learning algorithms can be used to identify complex yet meaningful patterns in the data, which then allows us to predict or classify future instances or events. We can use data from the first innings, such as the number of deliveries bowled, wickets left, runs scored per deliveries faced and partnership for the last wicket, and compare that against total runs scored. Machine learning techniques like SVM, Neural Network, Random Forest can be used to create a model from the historical first innings data, considering the teams playing the match. The same model can be used to predict the second innings which is interrupted by rain. This will give a more accurate prediction than the D/L (Duckworth-Lewis system) method which is the most popular use of mathematics in cricket, as we are using a lot of historical data and all relevant variables.

REFERENCES

- [1] A. Gandomi, M. Haider (2015), "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, 35(2), pp. 137-144.
- [2] D. Che, M. Safran, Z. Peng (2013), "From big data to big data mining: Challenges, issues,

and opportunities", International Conference on database systems for Advanced Applications, 7827, pp. 1-15.

- [3] H. S. Bhosale, D. P. Gadekar (2014), "A review paper on big data and Hadoop" International Journal of Scientific and Research Publications, 4(10), pp. 1-7.
- [4] J. Dean, S. Ghemawat (2004), "MapReduce: Simplified data processing on large clusters", Communications of the ACM, 53(1), pp. 1-13.
- [5] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, A. Gani (2014), "Big data: Survey, technologies, opportunities, and challenges", The Scientific World Journal, pp. 1-18.
- [6] R. Rein, D. Memmert (2016), "Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science", SpringerPlus, 5, pp. 1-13.
- [7] Shweta, K. Garg (2013), "Mining efficient association rules through Apriori algorithm using attributes and comparative analysis of various association rule algorithms", International Journal of Advance Research in Computer Science and Technology, 3(6), pp. 306-312.
- [8] S. Anjali, V. Aswini, M. Abirami (2015), "Predictive analysis with cricket tweets using big data", International Journal of Scientific & Engineering Research, 6(10), pp. 78-83.
- [9] S. Agarwal, L. Yadav, S. Mehta (2017), "Cricket team prediction with Hadoop: Statistical Modeling Approach", Procedia Computer Science, 122, pp. 525-532.
- [10] S.G Manikandan, S. Ravi (2014), "Big data analysis using apache Hadoop", IEEE International Conference on IT Convergence and Security, pp. 1-4.
- [11] S. Mukherjee, R. Shaw (2016), "Big data-Concepts, applications, challenges and future scope", International Journal of Advance Research in Computer and Communication Engineering, 5(2), pp. 66-73.
- [12] S.V. Phaneendra, E.M. Reddy (2013), "Big data- Solutions for RDBMS problems- A Survey", International Journal of Advance Research in Computer and Communication Engineering, 2(9), pp. 3686-3691.
- [13] T. Ahlawat, Dr. R. K. Rambola (2016), "Literature review on big data", International Journal of Advancement in Engineering Technology, Management & Applied Science, 3(5), pp. 21-30.