



IPL

**BIG DATA
(UE15CS314)**

**CLASS PROJECT - BIG DATA
PROJECT REPORT**

IPL SCORE PREDICTION

Project Title: IPL SCORE PREDICTION

Project Group Members :

SRN	NAME	EMAIL
01FB15ECS069	Bhoomika R Rao	bhoomikarao19@gmail.com
01FB15ECS071	Harshita Biridepalli	harshitabiridepalli@gmail.com
01FB15ECS192	Niriksha Kunder	nirikshak97@gmail.com
01FB15ECS214	Preetika	preetika101@gmail.com

Problem Statement

Given two cricket teams, batting order and bowling order and who bats first predict the score , predict ball by ball and handle wickets seperately.

APPROACH

All the data is extracted from cricsheets.org. From the zip file, all the 2016 ball-to-ball match csv are then filtered from the rest. From those csvs, batsmen and bowlers statistics are created and player vs player details are also made. Using those player details, clusters of batsmen and bowlers are created. K-means clustering algorithm is used for clustering. The clusters are used to calculate each ball's outcome probability. This probability is used to predict the overall match summary and thereby to determine which team is most likely to win.

ALGORITHM

Step 1

Extraction of Data - Extracted player vs player data for every match from The data describes performance of each batsman against every respective bowler. The data fields consists of Batsman name, Bowler name, number of 0s, 1s, 2s, 3s, 4s, 5s, 6s and 7s scored, followed by total number of runs scored, number of balls and strike rate. Batsmen and bowler statistics were also downloaded from Cricinfo

Step 2

Clustering of Batsmen/Bowler - Using the Spark MLlib's K-Means Clustering, and selecting Runs scored, Number of innings played, Batting Average and Strike Rate as parameters for clustering of batsmen. And Selecting Runs conceded, Wickets taken, Economy, Strike Rate and Four Wickets as parameters for clustering of bowlers. It clustered the datasets with similar characteristics into a single cluster. Ten such clusters were obtained for both Bowler and Batsmen.

Step 3

Computation of probabilities - Done using MapReduce. Player vs player data was taken. Probabilities for every batsman scoring a 0 or 1 or.... Or 7 against a given bowler was computed. Map phase involved reading each record and dividing the number of 0s scored by total number of balls to find the probability of scoring a 0. Reduce phase outputted the probabilities into a CSV file. A map reduce script was written which matches each player in the player vs player data to a cluster. The average of the probabilities of the runs scored obtained from the above mentioned CSV file, for that all the players of that cluster is found. These averages are then used to find the averages for each cluster combination.

Step 4

A python script was written to do the match score prediction. First we find the cumulative of the cluster probabilities. For example if the probability of scoring 0 runs for 0.3 and the probability of 1 run is 0.2 then the cumulative probability of scoring 0 runs would be 0.3 and that for 1 run would be 0.3+0.2 which is 0.5. Next given a batsman and bowler their respective clusters are obtained. We then generate a random number between 0 and 1 and check for which of the cumulative probabilities in the batsman-bowler cluster combination is it less than or equal to. That will be the runs scored for that ball. Similarly this is repeated for all the balls of that innings and this process is again repeated for the second innings. The winner of the match would be the team that scored the most runs.

EXPERIMENTAL RESULTS

Step1: Obtaining Batsman, Bowler and Player Vs Player Data

Snippet of Batsmen

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Pos	Player	Mat	Inns	NO	Runs	HS	Avg	BF	SR	Hundreds	Fifties	Fours	Sixes
2	1	Virat Kohli	139	131	23	4110	113	38.05	3151	130.43	4	26	359	149
3	2	Suresh Raina	147	143	21	4098	100	33.59	2957	138.58	1	28	360	160
4	3	Rohit Sharma	142	138	23	3874	109	33.68	2941	131.72	1	29	323	163
5	4	Gautam Gambhir	132	131	13	3634	93	30.79	2927	124.15	0	31	422	51
6	5	Chris Gayle	92	91	12	3426	175	43.36	2235	153.28	5	20	279	251
7	6	Robin Uthappa	135	130	15	3390	83	29.47	2635	128.65	0	17	341	104
8	7	David Warner	100	100	12	3373	109	38.32	2372	142.2	2	32	338	134
9	8	MS Dhoni	143	128	45	3271	70	39.4	2354	138.95	0	16	236	140
10	9	AB de Villiers	120	109	26	3257	133	39.24	2181	149.33	3	21	275	140
11	10	Shikhar Dhawan	113	112	16	3082	95	32.1	2546	121.05	0	25	348	62
12	11	Yusuf Pathan	134	122	32	2761	100	30.67	1881	146.78	1	12	230	140
13	12	Virender Sehwag	104	104	5	2728	122	27.55	1755	155.44	2	16	334	106
14	13	Ajinkya Rahane	95	89	13	2675	103	35.19	2212	120.93	1	23	285	51
15	14	Shane Watson	94	91	13	2551	104	32.7	1813	140.7	2	14	251	121
16	15	Dinesh Karthik	138	121	14	2542	86	23.75	2046	124.24	0	12	256	60

Snippet of Bowlers

A	B	C	D	E	F	G	H	I	J	K	L	M
Pos	Player	Mat	Inns	Overs	Runs	Wkts	BBI	Avg	Econ	SR	FourW	FiveW
1	Lasith Malinga	98	98	381.3	2546	143	5/13	17.8	6.67	16	4	1
2	Amit Mishra	112	112	401	2918	124	5/17	23.53	7.27	19.4	3	1
3	Dwayne Bravo	106	103	336.2	2754	122	4/22	22.57	8.18	16.54	2	0
4	Piyush Chawla	123	122	407.1	3082	120	4/17	25.68	7.56	20.35	2	0
5	Harbhajan Singh	125	123	445.3	3119	119	5/18	26.21	7	22.46	1	1
6	Vinay Kumar	102	101	346.4	2870	101	4/40	28.41	8.27	20.59	1	0
7	Ravichandran Ashwin	111	108	381.4	2500	100	4/34	25	6.55	22.9	1	0
8	Ashish Nehra	82	82	296.5	2298	98	4/10	23.44	7.74	18.17	1	0
9	Dale Steyn	90	90	343	2306	92	3/8	25.06	6.72	22.36	0	0
10	Zaheer Khan	89	88	326.3	2469	92	4/17	26.83	7.56	21.29	1	0
11	RP Singh	82	82	295.5	2338	90	4/22	25.97	7.9	19.72	2	0
12	Pragyan Ojha	92	90	316.3	2332	89	3/11	26.2	7.36	21.33	0	0
13	Sunil Narine	66	66	262	1619	85	5/19	19.04	6.17	18.49	6	1
14	Bhuvneshwar Kumar	76	76	277.5	1970	85	4/14	23.17	7.09	19.61	2	0
15	Albie Morkel	91	87	287.1	2359	85	4/32	27.75	8.21	20.27	1	0

Snippet of Player Vs Player data

Format: Batsman, Bowler, 0s, 1s, 2s, 3s, 4s, 5s, 6s, 7+, Dismissal, Runs, Balls, SR

A	B	C	D	E	F	G	H	I	J	K	L	M	N
A Ashish Reddy	TG Southee	1	1	1	0	0	0	0	1	0	9	4	225
A Ashish Reddy	CH Morris	0	1	0	0	0	0	0	0	0	1	1	100
A Ashish Reddy	PV Tambe	1	3	0	0	0	0	0	0	0	3	4	75
AM Rahane	B Kumar	5	4	1	0	3	0	0	0	0	18	13	138.46
AM Rahane	TA Boult	7	2	1	0	3	0	0	0	0 bowled	16	13	123.07
AM Rahane	P Kumar	1	3	0	0	0	0	0	0	0	3	4	75
AM Rahane	KV Sharma	4	4	2	0	1	0	0	0	0	12	11	109.09
AM Rahane	RS Bopara	7	4	0	0	2	0	0	0	0	12	13	92.3
AM Rahane	A Ashish Reddy	1	1	0	0	0	0	0	0	0	1	2	50
SV Samson	B Kumar	5	3	1	0	0	0	0	0	0	5	9	55.55
SV Samson	TA Boult	2	1	0	0	0	0	0	0	0	1	3	33.33
SV Samson	P Kumar	6	2	4	0	2	0	0	0	0	18	14	128.57
SV Samson	KV Sharma	1	2	0	0	0	0	0	0	0	2	3	66.66
SV Samson	RS Bopara	1	0	0	0	0	0	0	0	0 caught	0	1	0
SPD Smith	TA Boult	1	2	0	0	1	0	0	0	0	6	4	150
SPD Smith	KV Sharma	1	0	0	0	1	0	0	0	0 caught	4	2	200

Step2: Clustering Batsman and Bowler

Snippet of a batsman cluster

Format: Player, Position, Mat, Inns, Not Outs, Runs, HS, Avg, BF, SR, Hundreds, Fifties,

Fours, Sixes, Cluster

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Akshath Reddy	151	12	10	0	164	42	16.4	160	102.5	0	0	15	2	0
Shivli Kaushik	426	7	2	1	0	0	0	1	0	0	0	0	0	0
Ajantha Mendis	390	10	3	1	3	2	1.5	6	50	0	0	0	0	0
Azhar Bilakhia	220	7	7	2	69	22	13.8	85	81.17	0	0	5	0	0
Krishmar Santokie	446	2	0	0	0	0	0	0	0	0	0	0	0	0
R Bishnoi	309	3	3	0	19	18	6.33	17	111.76	0	0	1	2	0
Sachithra Senanayake	347	8	4	3	10	7	10	17	58.82	0	0	0	0	0
Travis Head	231	3	3	1	54	37	27	40	135	0	0	3	2	0
Michael Nesar	423	1	0	0	0	0	0	0	0	0	0	0	0	0
Raju Bhatkal	434	1	1	0	0	0	0	2	0	0	0	0	0	0
Vijay Shankar	456	1	0	0	0	0	0	0	0	0	0	0	0	0
Asad Pathan	259	8	7	2	39	14	7.8	23	169.56	0	0	5	2	0
Ben Hilfenhaus	433	17	1	1	0	0	0	0	0	0	0	0	0	0

Step3: Obtaining probabilities

Snippet of probabilities

Format: Batsman, Bowler, Prob:0s, Prob:1s, Prob:2s, Prob:3s, Prob:4s, Prob:5s, 6s, 7+

A	B	C	D	E	F	G	H	I	J
A Ashish Reddy	A Nehra	0.5	0.5	0	0	0	0	0	0
A Ashish Reddy	AD Mathews	0.25	0	0.25	0	0.25	0	0.25	0
A Ashish Reddy	Anureet Singh	0	1	0	0	0	0	0	0
A Ashish Reddy	CH Morris	0	1	0	0	0	0	0	0
A Ashish Reddy	DJ Bravo	0.5	0	0	0	0.5	0	0	0
A Ashish Reddy	Harbhajan Singh	0.375	0.375	0.125	0	0	0	0.125	0
A Ashish Reddy	Imran Tahir	0	1	0	0	0	0	0	0
A Ashish Reddy	JP Duminy	0	1	0	0	0	0	0	0
A Ashish Reddy	KA Pollard	0.3333333333	0.3333333333	0.3333333333	0	0	0	0	0
A Ashish Reddy	MG Johnson	0	0.3333333333	0.3333333333	0	0.3333333333	0	0	0
A Ashish Reddy	NM Coulter-Nile	1	0	0	0	0	0	0	0
A Ashish Reddy	PV Tambe	0.25	0.75	0	0	0	0	0	0
A Ashish Reddy	SL Malinga	0.5	0.25	0.25	0	0	0	0	0
A Ashish Reddy	Sandeep Sharma	0	0.3333333333	0	0	0	0	0.6666666667	0
A Ashish Reddy	TG Southee	0.25	0.25	0.25	0	0	0	0.25	0
A Mishra	A Nehra	0	0.5	0.5	0	0	0	0	0
A Mishra	D Wiese	0	1	0	0	0	0	0	0

Step4: The final result:

ANALYSIS AND ACCURACY

In order to precisely check the accuracy of the simulated matches , we compared the simulated results with that of actual results.We calculated accuracy by taking in results of 10 IPL 2016 matches.

1. 3rd Match

KXIP vs GL

Simulated match : KXIP : 110

GL : 113

GL won

Actual match : KXIP : 161

GL : 162

GL won

RESULT : Successful Prediction

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/lpl/run
nf:P Sahu
** Ball - 86
** Run - 4
PREDICTED :
Target : 110
Score : 113
Winner : Team 2
Team 1
('M Vijay', '-', 41)
('M Vohra', '-', 31)
('WP Saha', '-', 28)
Team 2
('AJ Finch', '-', 53)
('SK Raina', '-', 21)
('KD Karthik', '-', 21)
ACTUAL :
('Team 1', 161)
('Team 2', 162)
('Innings 1', 'M Vijay', '-', 41)
('Innings 1', 'AJ Finch', '-', 74)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/lpl/run$
```


2. 8th Match

SH vs KKR

Simulated match : SH : 126
 KKR : 131
 KKR won

Actual match : SH : 142
KKR : 146
KKR won

RESULT : Successful Prediction

```

bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 102      name.py      1.yani      2.yani      3.yani
** Run - 6
name: yani
score: random

PREDICTED :
Target : 126
Score : 131
Winner: Team 2 = dict()
Winner = {}

Team 1
('EJG Morgan', '-', 42)
('NV Ojha', '-', 37)
('DA Warner', '-', 16)
.strip() = d[s]

for i in bowl.readlines():
Team 2 d = dict()
('G Gambhir', '-', 59)
('RV Uthappa', '-', 55)
('MK Pandey', '-', 12)
.strip() = d[s]

Winner

ACTUAL :
Team 1 = random.random()
('Team 1', 142)
('Team 2', 146)
run = 0
while r <= 11[1]:
('Innings 1', 'EJG Morgan', '-', 51)
('Innings 1', 'G Gambhir', '-', 90)
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run$

```


3. 18th Match

SH vs KXIP

Simulated match : SH : 115
KXIP : 108
SH won

Actual match : SH : 143
KXIP : 146
SH won

RESULT : Successful Prediction

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 112
** Run - 6
data_version: 0.9
created: 2018-09-29
revision: 1
PREDICTED :
Target : 115
Score : 108
Winner : Team 1
Team 1 :
('SE Marsh', '-', 37)
('H Vohra', '-', 31)
('AR Patel', '-', 19)
Team 2 :
('DA Warner', '-', 36)
('S Dhawan', '-', 34)
('EJG Morgan', '-', 14)
ACTUAL :
('Team 1', 143)
('Team 2', 146)
```

4. 19th Match

GL vs RCB

Simulated match : GL : 210/4
RCB : 208/3
GL won

Actual match : GL : 180
RCB : 182
RCB won

RESULT : Failed Prediction

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 116
** Run - 2
data version: 0.0
created: 2018-08-24
revision: 1
PREDICTED :
Target : 154
Score : 156
Winner: Team 2
Team 1 tickets: 1
('V Kohli', '-', 82)
('KL Rahul', '-', 44)
('AB de Villiers', '-', 21)
Team 2 tickets: 1
('V Kohli', '-', 82)
('KL Rahul', '-', 44)
('AB de Villiers', '-', 21)
Team 1:
('V Kohli', '-', 82)
('KL Rahul', '-', 44)
('AB de Villiers', '-', 21)
Team 2:
('KD Karthik', '-', 72)
('BB McCullum', '-', 36)
('SK Raina', '-', 22)
ACTUAL :
('Team 1', 180)
('Team 2', 182)
Winner: Team 2
data version: 0.0
created: 2018-08-24
revision: 1
('Innings 1', 'V Kohli', '-', 100)
('Innings 1', 'KD Karthik', '-', 50)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$
```

5. 24th Match

MI vs KKR

Actual match : MI : 143

KKR : 99

MI won

Simulated match : MI : 174

KKR : 178

MI won

RESULT : Successful Simulation

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 113
** Run - 1
bowl = yam
bats = random
PREDICTED : yam1
Target : 143
Score : 99
#cluster input/bats1cluster.csv, r
#cluster input/balls1cluster.csv, r
#cluster input/runs = dict()
#cluster input/runs = dict()
Winner :Team 1
innings = 1
Team 1
('G Gambhir', '-', 76)
('AD Russell', '-', 21)
('SA Yadav', '-', 20)
strip() = d[0]
for i in bowl.readlines():
    i = i.split(',')
Team 2
('RG Sharma', '-', 46)
('KA Pollard', '-', 32)
('AT Rayudu', '-', 21)
strip() = d[0]
Winner =
ACTUAL : #Team(player):
random.random()
('Team 1', 174)
('Team 2', 178)
innings = 1
Innings 1
('Innings 1', 'G Gambhir', '-', 59)
('Innings 1', 'RG Sharma', '-', 68)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$
```

6. 32nd Match

KKR vs KXIP

Simulated match : KKR : 152

KXIP : 153

KXIP won

Actual match : KKR : 164

KXIP : 157

KKR won

RESULT : Failed Simulation

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 109
** Run - 1
Data version: 0.1
created: 2018-05-01
revision: 1
PREDICTED :
Target : 152
Score : 153
Winner : Team 2
Team 1:
('RV Uthappa', '-', 55)
('G Gambhir', '-', 53)
('YK Pathan', '-', 33)
Team 2:
('GJ Maxwell', '-', 67)
('DA Miller', '-', 32)
('WP Saha', '-', 17)
ACTUAL :
('Team 1', 164)
('Team 2', 157)
Innings 1:
('Innings 1', 'RV Uthappa', '-', 70)
('Innings 1', 'GJ Maxwell', '-', 68)
```

7. 34th Match

SH vs GL

Simulated match : SH : 140

GL : 128

SH won

Actual match : SH : 126

GL : 129

SH won

RESULT : Successful prediction

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 117
** Run - 0
PREDICTED :
Target : 140
Score : 128
Winner :Team 1
Team 1
('AJ Finch', '-', 42)
('DJ Bravo', '-', 31)
('BB McCullum', '-', 27)
Team 2
('S Dhawan', '-', 57)
('DA Warner', '-', 21)
('Yuvraj Singh', '-', 14)
ACTUAL :
('Team 1', 126)
('Team 2', 129)
('Innings 1', 'AJ Finch', '-', 51)
('Innings 1', 'S Dhawan', '-', 47)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$
```

8. 44th Match

RCB vs GL

Simulated match: RCB : 137

GL : 136

RCB won

Actual match : RCB : 248

GL : 104

RCB won

RESULT : Successful Prediction

```

bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run
nf:S Kaushik
nf:Sachin Baby
** Ball - 120

** Run - 0

match: yasl
match: random

PREDICTED : yasl*

Target : 137
Score : 136 (['Kohli','AB de Villiers','CH Gayle','AJ Finch','PV Tambe','RA Jadeja'])
Winner : Team 1
Inns : 1

Team 1
('V Kohli', '-', 73)
('AB de Villiers', '-', 45)
('CH Gayle', '-', 18)
.strip() = d[s]

for i in bowl.readlines():
Team 2 d = i.split()
('AJ Finch', '-', 59)
('PV Tambe', '-', 27)
('RA Jadeja', '-', 17)
.strip() = d[s]

Inns = 2

ACTUAL : (['Team 1', 248]
['Team 2', 104])
run = 0
while run <= 120:
('Innings 1', 'AB de Villiers', '-', 129)
('Innings 1', 'AJ Finch', '-', 37)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$

```

9. 50th Match

RCB vs KXIP

Simulated match : RCB : 118

KXIP : 93

RCB won

Actual match : RCB : 211

KXIP : 120

RCB won

RESULT : Successful Simulation

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 89
** Run - 0
Team 1: Yod
Team 2: random
PREDICTED : Team 1
Target : 118
Score : 93
Team 1: Yod
Team 2: random
Winner : Team 1
Team 1:
('V Kohli', '-', 56)
('CH Gayle', '-', 45)
('KL Rahul', '-', 11)
Team 2:
('Gurkeerat Singh', '-', 33)
('WP Saha', '-', 15)
('MM Sharma', '-', 14)
Winner : Team 1
ACTUAL : Team 1
Team 1:
('Team 1', 211)
('Team 2', 120)
Winner : Team 1
Innings 1:
('Innings 1', 'V Kohli', '-', 109)
('Innings 1', 'WP Saha', '-', 23)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$
```


10. 58th Match

SH vs KKR

Simulated match : SH : 146
KKR : 119
SH won

Actual match : SH : 162
KKR : 140
SH won

RESULT : Successful Prediction

```
bhoomi@bhoomi-Inspiron-3543: ~/Downloads/CricketPrediction-master/ipl/run
** Ball - 120
** Run - 1
Team 1
Team 2
PREDICTED :
Target : 146
Score : 119
Winner : Team 1
Team 1
('Yuvraj Singh', '-', 38)
('DA Warner', '-', 31)
('DJ Hooda', '-', 27)
Team 2
('G Gambhir', '-', 36)
('MK Pandey', '-', 29)
('SA Yadav', '-', 22)
ACTUAL :
('Team 1', 162)
('Team 2', 140)
('Innings 1', 'Yuvraj Singh', '-', 44)
('Innings 1', 'MK Pandey', '-', 36)
bhoomi@bhoomi-Inspiron-3543:~/Downloads/CricketPrediction-master/ipl/run$
```

Out of 10 simulated matches , 8 matches were predicted successfully .

FUTURE ENHANCEMENTS

- 1) We have used kmeans clustering for creating clusters. A different kind of clustering like agglomerative or hierarchical clustering could be used.
- 2) Experimenting with different criteria for finding clusters.

REFERENCES

- 1) CricInfo

<http://www.espncricinfo.com/indian-premier-league-2015/engine/series/791129.html>

- 2) IPL Website

- 3) Spark MLlib:

<https://blogs.msdn.microsoft.com/bigdatasupport/2015/09/24/a-kmeans-example-for-spark-mllib-on-hdinsight/>