RESEARCH ARTICLE OPEN ACCESS

IPL-2017 Cross Country Cluster Analysis

Mr. Chirag Goyal
Assistant Professor
Department of Computer Science & Engineering,
JCDM College of Engineering
Sirsa

ABSTRACT

In this paper, an attempt has been made to study the performance of Cricket Players playing in IPL from different countries using analysis of Key Performance Indicators and to know the insights regarding which country players plays well in the League. Indian players excluded from the research as it is an Indian League in which seven out of possible eleven should be from India. Some other countries are also not considered i.e. Bangladesh, Afghanistan, Sri Lanka as count of players from these countries are very less and they are outliers for our analysis. The dataset of Players has been considered from IPL10, 2017 (20 over's), over which cluster analysis has been applied and the findings of this study reveals that based on the Key Performance Indicators, players of England performs well from their counterparts. These kinds of research could help the franchisees to decide on their next year's strategy although these analyses could be done at more historical data say all the IPL versions for more accuracy.

Keywords- IPL , KPI, Cricket, Cluster Analysis, Buckets

I. INTRODUCTION

Cricket or the gentleman's game is a very old, widespread and uncomplicated pastime game. Historically, cricket's origins are uncertain and the earliest definite reference is in south-east England in the middle of the 16th century. It spread globally with the expansion of the British Empire, leading to the first international matches in the second half of the 19th century and yet the most popular game of the today's world. It is a game of uncertainty. One cannot predict outcome of the game till the last moment of the game though the possible results are known to all, therefore, an appropriate probability model can be applied to predict the result. Cricket is played in a standard format called a test match for a long period. The test match is a two innings per team contest that is played over five days. The long duration bores the audience as well as viewers in the television then a newer format evolved. The newer format shortened the duration to one where each team plays one innings with limited number of over's. This format was commercially successfully and spectators enjoyed shorter version of the cricket. But, in shorter format game with limited number of over's i.e.,Twenty20, played over a few hours with each team having a single innings of 20 over's (i.e. 120 legal deliveries), the player's performance is one of the major factors for team selectors. So, one of the attempts has been taken to analyze the performance of players in 20 over's matches to draw the conclusion. The auction prices of a player are very much dependent on his current and past form and how quickly one could adapt to the team demands.

The 2017 season of the Indian Premier League, also known as IPL 10, was the tenth edition of the IPL, a professional Twenty20 cricket league established by the BCCI in 2007[1]. The tournament featured the eight teams. The 2017 season started on 5 April 2017 and finished on 21 May 2017.It is the biggest cricketing tournament and one of the world's most viewed sporting events. It is a tournament where renowned international cricketers come together on one stage & budding Indian players are groomed under their guidance. IPL is where talent meets opportunity. The cricket team is a group of 11 (eleven) players consisting of batsmen, bowler, wicketkeeper, and all-rounder. The team should be balanced and diversified to enhance the probability of the success. In addition, the success can also depend on the type of pitch, winning of toss, and sequence of batting or bowling. Besides this, the performance of batsmen and bowlers is the key factor of the results of a particular match. Nowadays, research is going on to study the performance of such factor using different statistical (probabilistic/ stochastic) approach. Kimber and Hansford [2] studied batting average of batsmen with the help of

different statistical technique (mainly, geometric distribution). A graphical method given by Van Staden [3], for comparison of cricket player's batting and bowling performances. Whereas, Sharp et al., [4] used an integer programming to determine the optimal team based on player's performance in twenty20 cricket. Lemmer [5] shows how integer optimization, scientific method, can be used to aid in selecting a cricket team. Keeping these points in mind, in this paper a study has been carried out.

II. METHODOLOGY AND DATA

Cluster analysis is a technique which discovers the substructure of a data set by dividing it into several groups. Clustering plays an important role in data analysis and interpretation. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way"[6].

A cluster is therefore a collection of objects which are"similar" between them and are "dissimilar" to the objects belonging to other clusters. In this case, we easily identify the five clusters into which the performance of teams can be divided; the similarity criterion is Key Performance Indicator: two or more teams belong to the same cluster if they are "close" according to their measured value.

Clustering could also be done based on the equal sized categories but in this analysis, we prepared the clustering based on the benchmarks values of individual KPI [7]. It might be the case, one or more team have the similar measure value, hence they would be lie in the same group.

We could also use the ranking approach but the pitfall of that approach is a minor difference would costs more. For summarized results, classification clustering technique is more appropriate. In this paper, clusters have been described by the following parameter:

Bucket= (Max value- Min value)/5.

To study the batting and bowling performance of players in IPL-10, 2017 the seven important measures of batting and bowling statistics such as Batsman Strike Rate, Batsman Average, Batsman Boundary Hit Ratio, Batsman Sixer Hit Ratio and three bowling measures such as Bowler Economy Rate, Bowler Average, and Bowler Strike Rate has been considered. The Key Performance Indicator have been tabulated below.

Batting Statistics	**Description**
Batting Average	the ratio *R*/N, where *R* denotes the number of runs scored and *N*the number of times the batsman was out.
Batting Strike Rate	the ratio R/BF, where *R* denotes the number of runs scored and *BF*denotes the number of balls faced by a batsman.
Balls per Six	Average Number of balls taken by batsman to score a six.
Balls per Four	Average Number of balls taken by batsman to score a four.

Table 1: Key Performance Indicators of Batting.

Bowling Statistics	**Description**
Bowling Average	TR/W, where *TR* is the total runs conceded by a bowler and *W*is the total number of wickets.
Bowling Strike Rate	TB/W, where *TB* is the total number of balls bowled by a bowler and *W*is the total number of wickets.
Bowler's Economy Rate	*T*R/O, where *TR* is the total number of runs conceded by a bowler and *O*is the total number of overs bowled by a bowler.

Table 2: Key Performance Indicators of Bowling.

To fit the analysis, data has been collected from IPL season 10, 2017 which are freely available in the website: www.espncricinfo.com [8].

The IPL, a professional league for 20 over's cricket competition in India was initiated by the Board of Control for Cricket in India (BCCI) and is supervised by BCCI Vice-President, who serves as the league's chairman and commissioner. The IPL10 T20 was concluded in the month of April-May, 2017. In the tenth season of IPL, there were a total of 8 teams, namely Delhi Daredevils (DD), Gujarat Lions (GL), Kings XI Punjab (KXIP), Kolkata Knight Riders (KKR), Mumbai Indians (MI), Rising Pune Super giants (RPS), Royal Challengers Bangalore (RCB), Sunrisers Hyderabad (SRH) on the names of famous cities of India. These teams select players (both Indian and foreign) through an auction. The maximum number of foreign players to be played into a team is four. The final was played on 21 May between Mumbai Indians and Pune Supergiant, in which Mumbai Indians (MI) wins the trophy.

The player statistics has been considered from IPL10, 2017 respectively in our study. In our study, there are some outliers which are not considered in analysis. Countries like Afghanistan, Sri Lanka and Bangladesh have not been considered because the count of the players from those countries is very few. Each team has been assigned to any of the five clusters (Outstanding, Very Good, Good, Satisfactory and Poor). Based on their measure in the Performance Indicators, each of the clusters has a specific score like Outstanding Cluster having a score of 5, Very Good having 4, Good having 3, Satisfactory having 2 and Poor having score 1. At the end, the total score of the countries are calculated and based on that conclusions have been drawn.

III. RESULTS AND DISCUSSION

To perform the analysis, the Key Performance Indicators (KPI) of bowling and batting statistics are mined from the dataset. The result dataset of Batting and Bowling

Key Performance Indicators (KPI's) are shown in the Fig.1 and Fig.2.

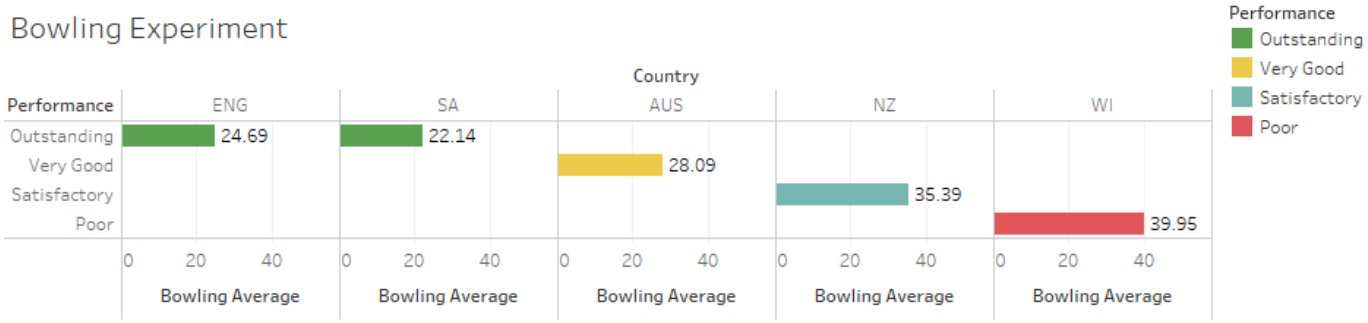
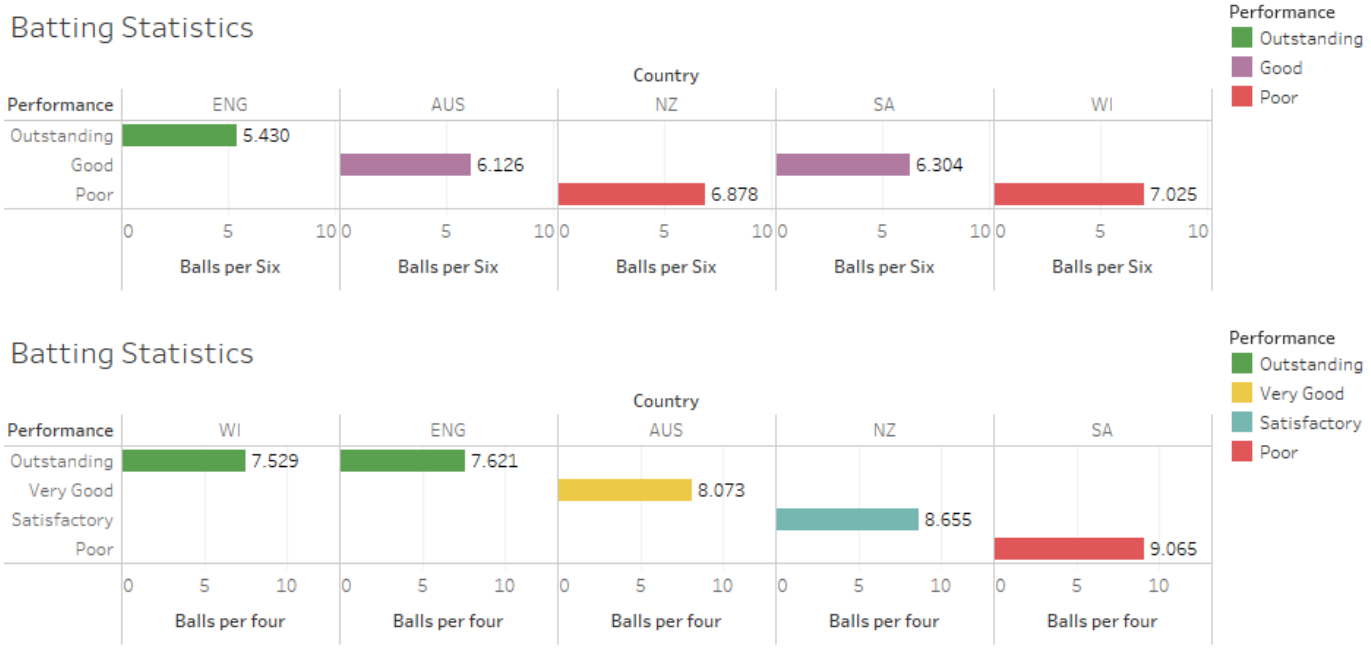


Fig.1 Bowling Clusters of players of different countries.



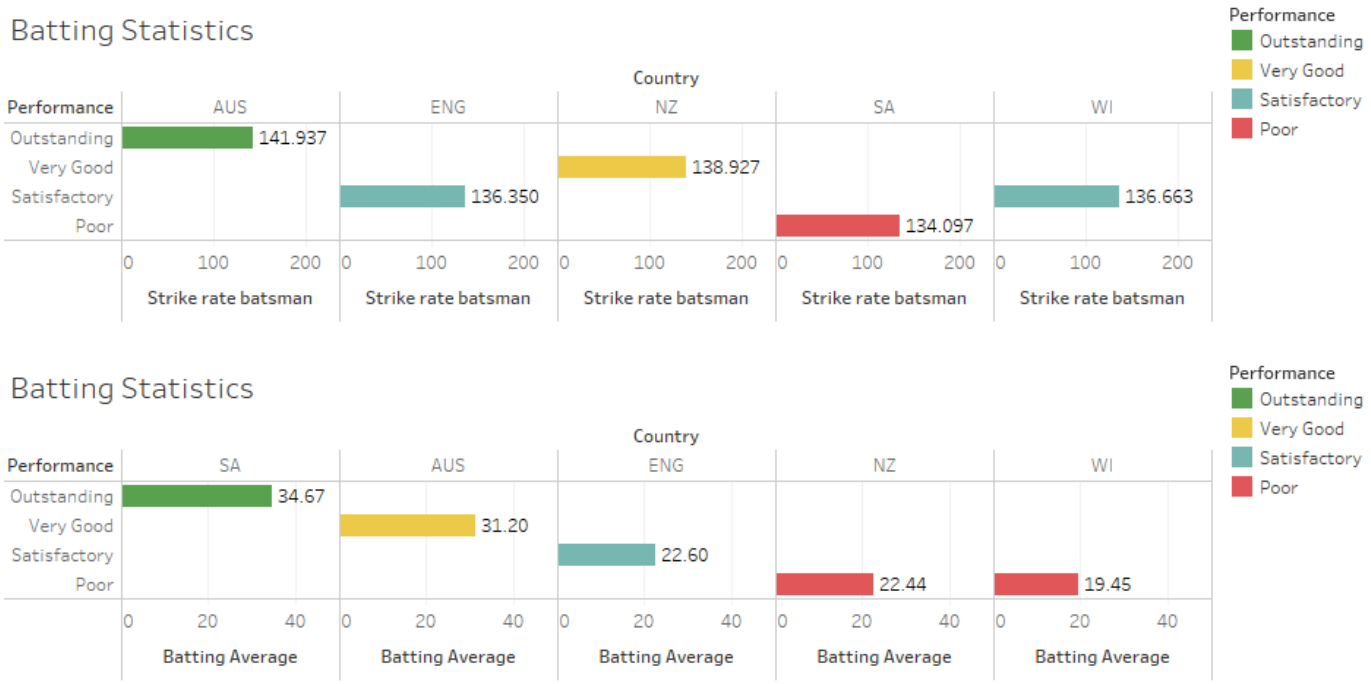


Fig. 2: Batting Clusters of players of different countries.

As shown in the Fig.1 and Fig. 2 the graphs are plotted for bowling and batting statistics where each statistics has been clustered. Different colour bars represent different clusters. The lower the bowling statistics means the better the bowler is performing. Some of the batting statistics indicators are better for higher values i.e. Strike Rate and Batting Average, while some of the indicators are better for lower values i.e. Balls per Six and Balls per four.

The score of the individual countries players have been calculated and after the calculation the score of the different countries based on the score on which cluster they are lying are tabulated below.

Country	Bowling Average	Bowling Strike Rate	Economy Rate	Balls per Six	Balls per Four	Batting Average	Batting Strike Rate	Total Score
Australia	4	4	4	3	4	4	5	28
England	5	5	5	5	5	2	2	29
New Zealand	2	3	1	1	2	1	4	14
South Africa	5	5	5	3	1	5	1	25
West Indies	1	1	5	1	5	1	2	16

Table 3: Score of players of countries based on Key Performance Indicators.

IV. CONCLUDING REMARKS

In this paper the performance of cricket players in IPL season 10, 2017 has been analyzed. The statistical technique has been employed to explore the interrelationship among various Key Performance Indicators (KPI's) of batting and bowling. Based on the above analysis, the England players are performing well as a group and New Zealand Players are the lowest performers. This kind of analyses could help the franchisee teams to invest their money in a more intelligent way and pick the right set of players.

Although it might happen, the performances of 2-3 key players could impact this analysis but that is the reason we excluded those teams, which have short stay in the league or have very less player participation. Also this analysis could also be impacted due to the number of international players in the team respectively, but in terms of leagues we would not be very much concerned about that. The addition of more KPI's and more historical data further strengthen the analysis and could give better heuristic analysis. The reason of picking the Clustering approach over the other methodologies is like based on the T20 KPI benchmarks we could assume the par scores of the each and every KPI and based on that data could be quantized. For example, if the Strike Rate of a batsman is 100, it would be considered as Satisfactory in terms of a league, whereas the same performance is Good in One Day and Very Good / Outstanding in Test Matches. Similarly, if the Economy Rate of a bowler is 7 he would be very good in league but would be treated Poor in other formats of the game.

Based on the time, pitches natures and ground situations these benchmarks could be adjusted for more concrete results.

REFERENCES

[1] Preston, I. and Thomas, J.: Batting strategy in limited overs cricket, Statistician, 49(1), p. 95–106 (2000).
[2] Kimber, A.C. and Hansford, A.R.: A statistical analysis of batting in cricket, Journal of the Royal Statistical Society, Series A 156, p. 443-455 (2013) .
[3] Van Staden, P. J.: Comparison of cricketers' bowling and batting performances using graphical displays, Current Science, 96(6), p. 764–766 (2009).
[4] Sharp, G.D., Brettenny, Gonsalves,J.W., Lourens, M. and Stretch, R.A.: Integer optimization for the selection of a Twenty20 cricket team, Journal of the Operational Research Society, 62, p. 1688-1694 (2011).
[5] Lemmer, H.H.: Team selection after a short cricket series, European Journal of Sport Science, DOI: 10.1080/17461391.2011.587895 (2013).
[6] Pabitra Kumar Dey, Gangotri Chakraborty, and Suvobrata Sarkar, "Cluster Detection Analysis using Fuzzy Logic.
[7] Sricharan Shah, Partha Jyoti Hazarika and Jiten Hazarika " A Study on Performance of Cricket Players using Factor Analysis Approach", International Journal of Advanced Research in Computer Science, Volume 8, No. 3,(2017) ISSN No. 0976-5697.
[8] www.espncricinfo.com/IPL2017