

Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015

Sanmoy Bhattacharjee
Department of Mathematics
IIT Kharagpur
Kharagpur, India
sanmoy22@gmail.com

Jayakrushna Sahoo
Department of Mathematics
IIT Kharagpur
Kharagpur, India
jayakrushnas@gmail.com

Adrijit Goswami
Department of Mathematics
IIT Kharagpur
Kharagpur, India
goswami@maths.iitkgp.ernet.in

Abstract— Performance analysis in every sport is essential to find out the weaknesses and strengths of the players. In a team game like cricket, analysis of career-data is indispensable to get the insight of the players' performance, which helps the selectors to do their job flawlessly and also helps the players' themselves to identify their weaknesses and their strengths. And, when the time comes for world cup cricket, every team looks for their best team-combination to be available on the ground to achieve the desired result in their favour. Association rule mining techniques reveal the unknown information from a huge set of data, and this technique can be used in extracting the information from the performance-data of the players. In this paper, we have applied association rule mining technique on individual Indian players' career-record to obtain the underlying unknown relations of several factors impacting the players' performances. This analysis could help in selecting the best-suited team-combination in a given match condition. The relation among various intrinsic factors such as venue of the match, batting first or second, batting position (of batsmen), strike-rate and runs (of batsmen), economy-rate and wickets taken (of bowlers) is analyzed. The result of this study could be useful for Indian team-captain and team-manager in decision making and strategy planning and could also boost the chances of success for team India for the world cup 2015.

Keywords- Association rule mining; Apriori algorithm; Cricket; Performance analysis; Strategy planning; ICC Cricket World Cup 2015

INTRODUCTION

Cricket is a game between two teams with eleven players on each side (no substitutions) and two on-field umpires and a third off-field umpire refereeing the game. The game is basically played in three different formats namely T-Twenty, One Days Internationals (ODI) and Test match Cricket. The rules of cricket are governed by the International Cricket Council (ICC) and the Marylebone Cricket Club (MCC). The game is most popular in Oceania, England, the Indian subcontinent, the West Indies and Southern Africa.

A World Cup Tournament (WCT) is held every four year for the ODI version (T-Twenty WCTs are also held, no World Cup is held for Test matches). However the focus of this paper will be the WCTs for the ODI version. The popularity and sponsorship of the World Cup Cricket has risen considerably with time and viewership of cricket now stands in the third

position only after FIFA and The Summer Olympics. According to the timeline of this paper the next World Cup is going to be played in the months of February-March of 2015.

In our study, we have done some analysis on the career performances of the Indian cricket players, who are most probably going to be the part of the Indian cricket team for the 2015 World Cup, using a data mining technique called as *Association Rule Mining* (ARM).

Association Rule Mining (ARM) is a data mining model which was first proposed by Agarwal et al [1]. The first usage of this model was in the Market Basket Analysis to find how items purchased by customers are related. ARM figures out all the rules correlating the presence of one set of items with that of another set of items from a database of transactions. A transaction is a list of items purchased by a customer in one visit. If M and N are the sets of items in a transaction database, then an expression of the form $M \Rightarrow N$ denotes an association rule. The simple interpretation of this expression is that for a transaction containing M tends to contain N.

For the formation of association rules we look for frequently occurring patterns while analysing data, and through the usage of *support* and *confidence* criteria we figure out the most important relationships. *Support* tells us how frequently the items appear in the database and *confidence* is the indication of the number of times the associated items, forming the rule, have been found in the transaction database. For an association rule $M \Rightarrow N$, the confidence is the support of the set of all items that appear in the rule ($\text{support } S_1 = \text{MUN}$) divided by the support of the antecedent of the rule ($\text{support } S_2 = \text{M}$).

Another important factor taking care of the statistical dependency of the antecedent and consequent is called as the *lift* [2]. For an association rule, lift measures how many times more often its antecedent (M) and consequent (N) are together, than expected, if they were statistically independent. From a mathematical point of view, lift is the ratio of observed support to the expected support. A lift=1, indicates that the probability of occurrence of the consequent is independent of the antecedent. No rule can be drawn involving two independent events. A lift < 1 indicates a negative correlation between the antecedent and consequent. A lift > 1 tells us as to what extent the occurrences

are dependent on each other and these are the rules which are important for predicting the consequent in future data sets.

In this paper, we used the career records of the probable Indian cricket players for the world cup 2015, as the transaction dataset. We were able to obtain interesting relations among various factors impacting the players' performance. The venue of a match, innings order of the match (batting first or bowling first), batting position of a batsman in the batting line-up, strike-rate (S/R) of a batsman directly affect a batsman's score in a match, a detailed analysis of which is presented in this paper. Bowlers' performance with respect to number of wickets and economy-rate has also been analyzed in our study.

METHODOLOGY

Documented information on different types of mining algorithms and usage of various types strategies and data structures for implementing these algorithms exist [3] [4]. Although the algorithms may differ in terms of their efficiency and memory usage but primarily they all lead to the same set of rules. Among all data mining algorithms apriori algorithm was the first one to be introduced. We have been a little conventional in our approach; we used this apriori algorithm for all our data analysis.

The way apriori algorithm works involves two steps. In the first step, frequent itemsets in the transactional database are obtained. The cardinality of these frequent item sets varies from 1 to k (k-itemset) and in the second step these obtained frequent item sets are used for determining the association rules. These association rules pin points the general trends in the database.

For our work, the career-data of every player has been visualized similar to the transaction data in retail industry. The factors such as venues and innings order (setting target or chasing target), batting position, strike rate and runs scored by a batsman, economy-rate and wickets taken by a bowler; has been considered similar to examining a transaction database. Rules giving the associations of affecting factors (venues, innings order, batting position, strike-rate) in antecedent and performance measures (runs scored by a batsman, economy-rate and wickets taken by a bowler) of players in consequent, gives the desired result by which we can analyze the players' performance at different conditions.

DATASETS STUDIED

Data of the players required for our analysis were obtained from *espncricinfo* website. Players who had played more than 30 One Day Internationals (ODIs) for India were only considered. Two data sets were taken into consideration. Dataset-1 comprises of the career-records in all ODI matches played by the batsmen till 20th October, 2014 and Dataset-2 comprises of the career-records in all ODI matches played by the bowlers till 20th October, 2014. Abandoned matches and the matches in

which a player did not get a chance to perform have been excluded. To measure a player's performance we studied the following parameters.

1) *Venue*: Whether a match is being played in the home country of a team or in an away venue is a very important factor that affects a player's performance. Players of every team find it difficult to play outside their home country as they have to adjust to different weather conditions as well as different pitch conditions. For our purpose we will refer to the matches played in India as the home venue, matches played in Australia-New Zealand as the Oceania venue and rest of the world as the away venue.

2) *Innings Order*: Innings order (i.e. whether a team bats first or bowls first) plays a crucial role in determining a player's performance. Some of players are psychologically stronger when it comes to chasing while others are more comfortable while setting a target. It is a general belief that setting a target is much easier than chasing. However with more day/night matches coming into picture with the dew factor coming into play somewhat belies this hypothesis.

3) *Batting position (pos) of a batsman*: A batsman may be assigned a position in the batting line-up according to his batting style and according to the match situations of a team. Some players are seen to perform well for a specific batting position while others can adjust to any number. Analysing a batsman's performance in terms of his batting position is always helpful for a team and it may often change outcomes of a match.

4) *Strike-rate (S/R) and runs scored by a batsman*: Runs scored and strike rate (runs scored per 100 balls) of a batsman are important parameters that needs to be studied for a batsman. Strike rate of a batsman may decide the team total (runs scored by the team) in an ODI match. Scoring more runs with a healthy strike rate is most desirable from the batsman's as well as team's point of view. For our analysis we considered three slabs for runs scored (< 40, 40-70 and >70) and S/R (< 70, 70-100 and >100). An S/R < 70 and a score < 40 is considered low while an S/R > 100 and a score > 70 is considered high.

5) *Economy-rate and wickets taken by a bowler*: Although in cricket batsmen are the main focus of attention, however one cannot deny the role played by the bowlers. A lower economy rate (number of runs conceded per over) and taking more number of wickets is desirable from the point of view of a bowler. Different match conditions affect a bowler's economy and wickets taken, which in turn measures his performance. For our analysis we considered three slabs for wickets taken (0-1, 2-3 and > 3) and economy rate (< 5, 5-6 and > 6).

EXPERIMENT AND ANALYSIS

We used R (version 3.1.2) statistical software package for data analysis. We varied the support of the antecedent of the rule from 0.1 to 1.0. Performance analysis of batsmen and bowlers are depicted in Table 1 and Table 2 respectively.

TABLE I
ANALYSIS OF THE BATSMEN PERFORMANCE (DATASET 1)

Shikhar Dhawan Performance	Confidence	Lift
{Opening pos bat, S/R below 70} \Rightarrow Scores below 40	93%	1.41
{Batting second, Opening pos bat, S/R above 100} \Rightarrow Scores above 70	60%	3.26
{Oceania venue, Opening pos bat} \Rightarrow Scores below 40	100%	1.52
{Oceania venue, Opening pos bat, Scores below 40} \Rightarrow S/R below 70	75%	2.04
{Away venue, S/R below 70} \Rightarrow Scores below 40	88%	1.33
{Away venue, Scores 40-70} \Rightarrow S/R 70-100	50%	1.46
{Home venue, Scores above 70} \Rightarrow S/R above 100	50%	1.73
{Home venue, Scores above 70} \Rightarrow S/R 70-100	50%	1.46
Rohit Sharma Performance	Confidence	Lift
{Opening pos bat, Scores below 40} \Rightarrow S/R below 70	89%	1.57
{Batting first, S/R below 70} \Rightarrow Scores below 40	86%	1.28
{Batting first, S/R 70-100} \Rightarrow Scores 40-70	55%	2.95
{Batting second, S/R below 70} \Rightarrow Scores below 40	97%	1.43
{Batting second, 5th pos bat, S/R 70-100} \Rightarrow Scores above 70	50%	3.54
{Batting second, 4th pos bat} \Rightarrow Scores below 40	80%	1.19
{Batting second, 4th pos bat, Scores below 40} \Rightarrow S/R below 70	83%	1.47
{Oceania venue, 4th pos bat} \Rightarrow Scores below 40	90%	1.34
{Oceania venue, Batting second} \Rightarrow Scores below 40	92%	1.37
{Oceania venue, Batting second, Scores below 40} \Rightarrow S/R below 70	75%	1.33
{Away venue, Batting first} \Rightarrow Scores 40-70	45%	2.46
{Away venue, Batting second, Opening pos bat} \Rightarrow Scores below 40	80%	1.19
{Away venue, Batting second, Opening pos bat} \Rightarrow S/R below 70	80%	1.42
Virat Kohli Performance	Confidence	Lift
{3rd pos batting, S/R below 70} \Rightarrow Scores below 40	100%	1.68
{Away venue, Batting second, 3rd pos batting} \Rightarrow S/R below 70	50%	1.12
{Home venue, Batting first, 3rd pos batting} \Rightarrow S/R below 70	50%	1.12
{Batting second, 3rd pos batting, Scores above 70} \Rightarrow S/R above 100	53%	2.79
{Batting second, 3rd pos batting, Scores above 70} \Rightarrow S/R 70-100	47%	1.30
{3rd pos batting, Scores above 70} \Rightarrow S/R above 100	50%	2.64

{3rd pos batting, Scores above 70} \Rightarrow S/R 70-100	50%	1.38
{3rd pos batting, S/R above 100} \Rightarrow Scores above 70	80%	2.90
{Away venue, Batting second} \Rightarrow Scores above 70	50%	1.81
Ajinkya Rahane Performance	Confidence	Lift
{Opening pos bat, S/R 70-100} \Rightarrow Scores 40-70	57%	2.53
{Opening pos bat, S/R below 70} \Rightarrow Scores below 40	92%	1.29
{Batting first, S/R below 70} \Rightarrow Scores below 40	91%	1.28
{Batting second, S/R below 70} \Rightarrow Scores below 40	100%	1.41
{Batting second, 4th pos bat} \Rightarrow Scores below 40	100%	1.41
{Batting second, 4th pos bat, Scores below 40} \Rightarrow S/R below 70	80%	1.38
{Oceania venue, 4th pos bat} \Rightarrow Scores below 40	100%	1.41
{Oceania venue, 4th pos bat} \Rightarrow S/R below 70	75%	1.29
{Away venue, S/R 70-100} \Rightarrow Scores 40-70	75%	3.32
{Away venue, S/R below 70} \Rightarrow Scores below 40	100%	1.41
{Away venue, Batting first} \Rightarrow Scores 40-70	50%	2.21
{Home venue, Batting first} \Rightarrow Scores below 40	80%	1.13
{Home venue, Batting first, Scores below 40} \Rightarrow S/R below 70	87%	1.51
Suresh Raina Performance	Confidence	Lift
{6th pos bat, S/R below 70} \Rightarrow Scores below 40	94%	1.30
{5th pos bat, S/R below 70} \Rightarrow Scores below 40	95%	1.32
{Batting first, S/R below 70} \Rightarrow Scores below 40	89%	1.23
{Batting second, S/R below 70} \Rightarrow Scores below 40	100%	1.38
{Batting first, Scores 40-70} \Rightarrow S/R above 100	50%	1.72
{Batting second, 5th pos bat} \Rightarrow Scores below 40	81%	1.13
{Batting second, 5th pos bat, Scores below 40} \Rightarrow S/R below 70	50%	1.29
{Away venue, Batting second} \Rightarrow Scores below 40	82%	1.14
{Away venue, 6th pos bat} \Rightarrow Scores below 40	86%	1.18
Mahendra Singh Dhoni Performance	Confidence	Lift
{Batting first, Scores above 70} \Rightarrow S/R above 100	52%	1.75
{Batting first, Scores above 70} \Rightarrow S/R 70-100	48%	1.45
{Batting first, 6th pos bat, Scores below 40} \Rightarrow S/R below 70	63%	1.68
{Batting second, 6th pos bat, Scores below 40} \Rightarrow S/R below 70	55%	1.48
{6th pos bat, S/R below 70} \Rightarrow Scores below 40	86%	1.51
{Batting second, 6th pos bat, S/R 70-100} \Rightarrow Scores 40-70	45%	1.69
{Batting first, S/R below 70} \Rightarrow Scores below 40	93%	1.64
{Batting second, 6th pos bat} \Rightarrow Scores below 40	63%	1.10
{Away venue, Batting second} \Rightarrow Scores below 40	82%	1.43
{Away venue, S/R below 70} \Rightarrow Scores below 40	89%	1.57
Ravindra Jadeja Performance	Confidence	Lift
{8th pos bat, Scores below 40} \Rightarrow S/R above 100	50%	1.69
{Batting second, 8th pos bat} \Rightarrow S/R above 100	50%	1.69
{Batting second, 7th pos bat} \Rightarrow Scores below 40	90%	1.10
{Batting first, 7th pos bat, S/R below 70} \Rightarrow Scores below 40	92%	1.13

<i>{Batting second, 7th pos bat, S/R below 70} ⇒ Scores below 40</i>	100%	1.23
<i>{Batting second, 7th pos bat} ⇒ S/R below 70</i>	60%	1.20
<i>{Oceania venue, Batting second, 7th pos bat} ⇒ Scores below 40</i>	100%	1.23
<i>{Oceania venue, Batting second, 7th pos bat} ⇒ S/R below 70</i>	57%	1.14
<i>{Away venue, Scores below 40} ⇒ S/R below 70</i>	56%	1.11
<i>{Home venue, S/R below 70} ⇒ Scores below 40</i>	94%	1.16

Interesting rules are written in italics

From the above data we can infer that when Dhawan plays at a slower rate he scores below 40 in 93% of the matches. Quite interestingly, when it comes to India chasing a total, if Dhawan plays with an S/R above 100, he scores above 70 in 60% of the matches, irrespective of the venue. When we talk about Oceania, he scores below 40 in all the matches played by him.

From Rohit's performance we have seen that he generally likes to play at a slow pace at the start of his innings. This often leads to his downfall as we can easily see that if he plays slowly, he gets out cheaply, irrespective of whether India is batting first or second. The story remains the same when it comes to away venues, which includes Australia and New Zealand. Another interesting fact to be noted when it comes to away venues is that Rohit's performance has been very poor when India chases but quite contrastingly he moderately scores good runs in 45% of the matches when India bats first. From the tabulated data we can also figure out that batting at no. 4 has not been a prolific run scoring position for Rohit.

For Virat Kohli, batting at No. 3, the position where he generally bats, 100% of the times when his strike rate falls below 70 he gets out with less than 40 runs. It is found that batting at No. 3, in second innings of away venues and first innings of home venues, his strike rate has been on the lower side. Whenever he scores above 70, his strike rate has been above 70. When he plays with a S/R greater than 100 he scores above 70, which is found in 80% of the cases. In away venues with India chasing, Virat scores above 70 in 50% of the cases, an interesting bit of fact for an interesting player.

Studying the case of Ajinkya Rahane we found out that his chances of getting out increases quite dramatically when he plays with a low strike rate irrespective of India chasing or batting first. When India chases, Rahane, batting at No. 4, gets out below 40 in every match and 80% of the times his S/R is found to be below 70. In Oceania venues, it's not good news for Rahane. His performance has been very poor, an issue that has to be handled carefully by the team manager as well as Rahane himself during the world cup matches. However there is some respite for Rahane, as his performance gets better in away venues when India bats first, a scenario which is quite just the opposite in home venues.

Turning our attention to Suresh Raina, we see that he gets out quickly most of the times when his S/R falls below 70. This

holds for him independent of his batting position and independent of India batting first or second. With India chasing and Raina batting at No.5, 81% of the times he scores below 40 and half of the times his S/R is below 70. Most important of all, a fact concerning the world cup is that in away venues his performance has been very bad in most matches when India chases and he bats at No. 6.

Looking at Dhoni's stats we find that he scores good runs when he plays at a good pace. More often when he gets out at a low score his S/R is found to be low. Although we know Dhoni to be a "good finisher", his stats tells quite a different story. Playing at no. 6 he is found to throw his wicket away with a score less than 40 in 63% of the matches with India chasing. In away venues when the team chases, he scores below 40 in 82% of the matches. Also for away venues if he plays with a S/R below 70, he scores below 40 in 89% of the cases.

Coming to "Sir" (as his team mates call him) Ravindra Jadeja, in half of his matches he scores below 40 with a high S/R, batting at No. 8. An interesting aspect of Jadeja's batting is that when he bats at No. 7 with his team chasing, in 90% of the matches he gets out with a low score. Maintaining a batting position at No. 7, when he plays with a low S/R, in most cases he scores below 40. Coming to Australia-New Zealand venue, his performance has been poor in all matches when his team chases.

TABLE II
ANALYSIS OF THE BOWLERS' PERFORMANCE (DATASET 2)

Ishant Sharma Performance	Confidence	Lift
<i>{Bowling first, Economy 5-6} ⇒ 2-3 wickets</i>	60%	1.50
<i>{Oceania venue, Bowling first} ⇒ Economy below 5</i>	56%	1.50
<i>{Away venue, Bowling first} ⇒ 0-1 wickets</i>	62%	1.13
<i>{Away venue, Bowling second} ⇒ 2-3 wickets</i>	71%	1.79
<i>{Home venue, Bowling first} ⇒ 2-3 wickets</i>	57%	1.41
<i>{Home venue, Bowling second} ⇒ 0-1 wickets</i>	67%	1.23
Bhuvneshwar Kumar Performance	Confidence	Lift
<i>{Bowling second, Economy above 6} ⇒ 2-3 wickets</i>	50%	1.82
<i>{Oceania venue, Bowling first} ⇒ Economy above 6</i>	50%	2.22
<i>{Oceania venue, Bowling first} ⇒ 0-1 wickets</i>	100%	1.43
<i>{Away venue, Bowling first} ⇒ Economy below 5</i>	63%	1.39
<i>{Away venue, Bowling second} ⇒ Economy below 5</i>	50%	1.11
<i>{Home venue, Bowling first} ⇒ 0-1 wickets</i>	100%	1.43
<i>{Home venue, Bowling first} ⇒ Economy 5-6</i>	50%	1.54
<i>{Home venue, Bowling second} ⇒ 2-3 wickets</i>	55%	1.98
<i>{Home venue, Bowling second} ⇒ Economy below 5</i>	55%	1.21
Mohammad Shami Performance	Confidence	Lift
<i>{Bowling first, Economy below 5} ⇒ 0-1 wicket</i>	80%	1.78
<i>{Bowling second, Economy below 5} ⇒ 0-1 wicket</i>	67%	1.49

<i>{Bowling second, Economy above 6} ⇒ 2-3 wickets</i>	67%	1.49
<i>{Oceania venue, Bowling first} ⇒ Economy above 6</i>	100%	1.71
<i>{Oceania venue, Bowling first, Economy above 6} ⇒ 2-3 wickets</i>	50%	1.12
<i>{Away venue, Bowling first} ⇒ 2-3 wickets</i>	71%	1.59
<i>{Away venue, Bowling first, Economy above 6} ⇒ 2-3 wickets</i>	75%	1.67
<i>{Home venue, Economy below 5} ⇒ 0-1 wicket</i>	83%	1.86
<i>{Home venue, Economy above 6} ⇒ 2-3 wickets</i>	50%	1.12
Umesh Yadav Performance	Confidence	Lift
<i>{Bowling second, Economy above 6} ⇒ 2-3 wickets</i>	60%	1.58
<i>{Away venue, Bowling first} ⇒ Economy above 6</i>	75%	1.21
<i>{Away venue, Bowling first} ⇒ 0-1 wicket</i>	100%	1.61
<i>{Away venue, Bowling second} ⇒ 2-3 wickets</i>	60%	1.58
<i>{Home venue, Bowling first} ⇒ 2-3 wickets</i>	60%	1.58
<i>{Home venue, Bowling second} ⇒ 2-3 wickets</i>	60%	1.58
<i>{Home venue, Economy above 6} ⇒ 2-3 wickets</i>	50%	1.32
Ravindra Jadeja Performance	Confidence	Lift
<i>{Bowling first, 2-3 wickets} ⇒ Economy below 5</i>	92%	1.60
<i>{Bowling second, 2-3 wickets} ⇒ Economy below 5</i>	65%	1.12
<i>{Oceania venue, Bowling first} ⇒ 0-1 wicket</i>	91%	1.47
<i>{Away venue, Bowling first} ⇒ Economy below 5</i>	67%	1.16
<i>{Home venue, Bowling second} ⇒ 2-3 wickets</i>	46%	1.42
<i>{Home venue, Bowling second} ⇒ Economy below 5</i>	65%	1.13
Ravichandran Ashwin Performance	Confidence	Lift
<i>{Oceania venue, Bowling first} ⇒ 0-1 wicket</i>	73%	1.40
<i>{Away venue, Bowling first} ⇒ Economy below 5</i>	67%	1.22
<i>{Away venue, Bowling first} ⇒ 0-1 wicket</i>	67%	1.28
<i>{Away venue, Bowling second} ⇒ Economy below 5</i>	58%	1.07
<i>{Home venue, Bowling first} ⇒ 2-3 wickets</i>	52%	1.08
<i>{Home venue, Bowling second} ⇒ 2-3 wickets</i>	67%	1.39
<i>{Home venue, Bowling second} ⇒ Economy below 5</i>	67%	1.22

Interesting rules are written in italics

Looking at the bowling stats, Ishant Sharma in away conditions performs well in terms of wickets, with India bowling in the second innings. Ishant's performance completely reverses when team India plays in home venues. Considering the Oceania venues, Ishant bowls quite economically in the first innings.

Bhuvneshwar Kumar manages to get wickets with an economy above 6 in 50% of the cases, provided India bowls second. In away grounds he has been economical in more than half of the matches irrespective of the innings. But when it comes to Oceania venues his performance is not quite satisfactory in the first innings. In home conditions he performs well when his team bowls second, but his performance goes down in the first innings.

Mohammed Shami can consider himself a bit unlucky as he is devoid of wickets even after maintaining a good economy. This happens for more than half of the matches that he plays. But quite surprisingly when his economy exceeds 6 he manages to pick 2 or 3 wickets in most cases. This happens irrespective of the venue and innings.

Umesh Yadav manages to pick up wickets in 60% of the matches when he bowls in the second innings with an economy greater than 6. In away venues he performs very poorly in terms of both wickets and economy when India bowls first. Also in away venues he manages to take 2-3 wickets in 60% of the matches when India bowls second. In home venues he performs well in terms of wickets irrespective of the innings.

Considering Ravindra Jadeja's performance as a bowler we figure out that in home venues he performs well in the second innings. In away venues in first innings, he bowls economically with less than 5 runs per over in 67% of the cases.

In away matches Ashwin is very economical in more than half the matches irrespective of India bowling first or second. In Oceania venues, he bowls poorly in the first innings as he takes less than 2 wickets in 73% of the matches. Ashwin is found to be more successful in the second innings in home venues.

RELATED WORK

Researches on sports done so far are mostly statistics based while the ones that are based on data mining are very few in numbers. To discover interesting knowledge from the game of football, Carson et al [13] presented a sports data mining approach. For foretelling the future state of a game of cricket, Bailey and Clarke [11] and Sankarnarayanan et al [12] have implemented machine learning methods. Current status of a game and the background of the two teams playing the game are the two key requirements for future prediction of a game through this method. Association rule mining technique was used by Raj and Padma [6] for analysing India's performance in ODIs. A combination of association rule mining and Bayes classifier was used by Kaluarachchi and Varde [7] for finding out the factors that contribute to a win.

Social Network Analysis (SNA) was employed by Mukherjee [8] to analyse a player's performance. Team and captain rankings have been studied through PageRank Algorithm [10]. Amin and Sharma have used a two stage method, comprising of ordered weight averaging followed by regression for deciding the rank, to measure the performance of a batsman [9] and have also proposed a data envelopment method for finding the rank of the cricket players [5].

We have presented ARM technique in this paper to find out the weaknesses and strengths of players at different playing conditions. Associations among various factors have been pin pointed to identify the suitability of players at various conditions.

CONCLUSIONS

Using the ARM data mining technique we figured out various facts about the Indian players who are most likely to be the part of the 2015 cricket world cup. Some interesting facts we observed are as follows:

- Batsmen who get out in low scores are most likely to have a lower strike rate.
- Performance of the players is not that great in the Oceania and away venues.

We are able to figure out areas of weaknesses and strengths of the players through our analysis. This analysis would prove to be helpful for the players and the team in strategy planning and improving on weakness areas.

However the ARM approach that we used does not tell us the whole story, it has a few limitations. While analysing we have not taken into consideration the quality of the opposite team. Secondly we did not consider the pitch conditions, which govern, to some extent, whether a match will be low scoring or high scoring. This will have a direct impact on the player's performance. So there exists a lot of scope for improvement. However we have to be a little careful as well, as, if we examine too many factors in the transaction database then generated rules may be of low support, which cannot be utilized flawlessly in strategy-planning. In future work, these points would be summoned so as to have a much stronger understanding of the player's performance.

ACKNOWLEDGMENT

We are thankful to espnricinfo.com which provides such useful information on cricket. Without the help of this website our analysis would not have been possible. We appreciate other such websites as well which are the storehouses of such huge data and information.

REFERENCES

- [1] Agrawal, Rakesh, Tomasz Imieliński, and Swami, Arun. "Mining association rules between sets of items in large databases." In *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207-216. ACM, 1993.
- [2] Tsay, Li-Shiang. "Interestingness Measures for Actionable Patterns." In *Rough Sets and Intelligent Systems Paradigms*, pp. 277-284. Springer International Publishing, 2014.
- [3] Hipp, Jochen, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. "Algorithms for association rule mining—a general survey and comparison." *ACM sigkdd explorations newsletter* 2, no. 1 (2000): 58-64.
- [4] Dunham, Margaret H., Yongqiao Xiao, Le Gruenwald, and Zahid Hossain. "A survey of association rules." *Retrieved January 5* (2001): 2008.
- [5] Amin, Gholam R., and Sujeet kumar Sharma. "Cricket team selection using data envelopment analysis." *European journal of sport science* 14, no. sup1 (2014): S369-S376.
- [6] Raj, K. Antony Arokia Durai, and Panchapakesan Padma. "Application of Association Rule Mining: A case study on team India." In *Computer Communication and Informatics (ICCCI), 2013 International Conference on*, pp. 1-6. IEEE, 2013.
- [7] Kaluarachchi, Amal, and Aparna S. Varde. "CricAI: A classification based tool to predict the outcome in ODI cricket." In *Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on*, pp. 250-255. IEEE, 2010.
- [8] Mukherjee, Satyam. "Quantifying individual performance in Cricket—A network analysis of batsmen and bowlers." *Physica A: Statistical Mechanics and its Applications* 393 (2014): 624-637.
- [9] Amin, Gholam R., and Sujeet Kumar Sharma. "Measuring batting parameters in cricket: A two-stage regression-OWA method." *Measurement* 53 (2014): 56-61.
- [10] Mukherjee, Satyam. "Identifying the greatest team and captain—A complex network approach to cricket matches." *Physica A: Statistical Mechanics and its Applications* 391, no. 23 (2012): 6066-6076.
- [11] Bailey, Michael, and Stephen R. Clarke. "Predicting the match outcome in one day international cricket matches, while the game is in progress." *Journal of sports science & medicine* 5, no. 4 (2006): 480.
- [12] Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan. "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction." *Proceedings of SIAM International conference on Data mining*, pp. 1064-1072, SIAM, 2014.
- [13] Carson K. Leung, Kyle W. Joseph, Sports Data Mining: Predicting Results for the College Football Games, *Procedia Computer Science*, 35(2014), 710-719.