

# Analyzing the performance of the Indian Cricket Team using Weighted Association Rule Mining

Deepak Saraswat<sup>1</sup>

Department of Physics and Computer  
Science  
Dayalbagh Educational Institute  
Agra, India  
<sup>1</sup>deepakcsdei@gmail.com,

Vijay Dev<sup>2</sup>

Department of Physics and Computer  
Science  
Dayalbagh Educational Institute  
Agra, India  
<sup>2</sup>vijay.dev.dei@gmail.com,

Preetvanti Singh<sup>3</sup>

Department of Physics and Computer  
Science  
Dayalbagh Educational Institute  
Agra, India  
<sup>3</sup>preetvantisingh@gmail.com

**Abstract**—This paper presents a Weighted Association Rule Mining algorithm for analyzing the Indian cricket team in one day international cricket matches against Sri Lanka and South Africa. This analysis is used by the team for framing game winning plans. The factors such as result of toss, taking part in home condition or away condition, first bat or second bat, and the game result are taken into consideration.

**Keywords**— Association rules, confidence, weighted support, weighted confidence.

## I. INTRODUCTION

Data mining in today's world is very important as it helps in extracting important pattern from a given data set. Association, classification, clustering, and outlier deduction are various techniques for effective data mining.

Association rule mining is applied to find associations in a volumetric data. Earlier this technique was applied to discover consumer purchasing behaviors. However, today it is used in several other fields such as marketing (Xu, Frankwick, & Ramirez [1]), finance (Olson, & Wu [2]), telecommunication (Le [3]), and management (Choi, Chan, & Yue, 2017). Association rule mining is also used for sports management (Zhao & Chen [4]).

Ofoghi, Zeleznikow, MacMahon&Raab [5] developed data mining tools to architecturally link the game and data mining areas. Asanka [6] identified the factors that influence the result of the extra deliveries. Bhattacharjee, Sahoo&Goswami [7] employed association rule mining method in cricket to find the fundamental unidentified relations of the factors impacting the performance of a player. Singh, Singla & Bhatia [8] proposed a model using Linear Regression Classifier and Naïve Bayes Classifier. First innings score was predicted using linear regression classifier on the basis of current run rate, number of wickets fallen and location of the game. The result of the game in the second innings was predicted using naïve bayes classifier. Verma&Izadi [9] proposed a Cricket Prognostic System, which is an analytical research framework for cricket on the basis of ball-by-ball data and historic data using machine learning and statistical methods. Zhao & Chen [10] developed a sport model by combining multiple small models with different data mining techniques for improving the accuracy of sport model result. Nair [11] developed a model for predicting target score. Satao, Tripathi, Vankar, Vaje, & Varekar [12] developed a prediction system using the K-means clustering algorithm. The system used historical match data, performance of the player and spectator to score prediction and match result.

The general framework of Association rule mining assumes that all items have the same significance which is not always the case. Some products may be more profitable, as compared to others. Hence rules regarding them are of greater value. In order to handle such situations weights are assigned to the variable based on their significance.

The paper presents an algorithm for weighted association rule mining problem and demonstrates its use in cricket management.

## II. MATERIAL AND METHODS

An association rule is denoted in the form of  $X \rightarrow Y$ , ( $X$  implies  $Y$ ) where  $X$  and  $Y$  are the two disjoint itemsets.  $X$  is an antecedent and  $Y$  is consequent. Support and Confidence are the two important parameters of Association rule. The Support measures how a rule is applicable to the given dataset and is computed by using the following equation:

$$\text{Support}(XY) = \frac{\text{Support count of } XY}{\text{Total number of transaction record in database}} \quad (1)$$

Confidence is the strength of the association rule and is used to find out how frequently items from itemset  $Y$  appear in transaction's record that contains itemset  $X$ . This is computed using the equation:

$$\text{Confidence} \left( \frac{X}{Y} \right) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \quad (2)$$

Or

$$\text{Confidence} \left( \frac{X}{Y} \right) = \frac{P(X \cup Y)}{P(X)} \quad (3)$$

The Overall structures of Association rule mining consider that all items have the same significance which is not always the case. There are cases where some items are more profitable than others. Thus weights are assigned based on their importance in the transaction. This form of rule mining is called Weighted Association Rule Mining.

The process contains of following phases:

1. Compute the Confidence of the itemsets.
2. Assign weights to each pair of (attribute, attribute value).
3. Compute Weighted Support and Confidence.

## III. THE PROPOSED ALGORITHM

In order to perform weighted association rule mining this section discusses the developed algorithm.

Step 1- Identify the attributes ( $a_i$ ) and the value of the attribute ( $v_i$ ) for a given domain.

Step 2- Identify the transaction set  $T = (a_i, v_i)$  from the data of last 10 years.

Step 3- For these transactions generate rules using inspection method.

Step 4- Compute the Confidence using equation (3) for the generated rules.

Step 5- Assign weights ( $w_i$ ) to the pair ( $a_i, v_i$ ) after discussion with experts and observing the Confidence values.

Step 6- Compute itemset transaction weight using the equation (4).

$$ITW(ABC) = W_A * W_B * W_C \quad (4)$$

Step 7- Compute the Weighted Support (WS) and Weighted Confidence (WC) using the equation (5) and equation (6) respectively.

$$WS(ABC) = \frac{ITW(ABC)}{NO. \text{ Of } ITW} \quad (5)$$

$$WC(ABC) = \frac{WS(ABC)}{NO. \text{ Of } ITW} \quad (6)$$

#### IV. CASE STUDY

This section demonstrates the implication of the proposed algorithm in the area of Indian Cricket.

The data set includes all one day international matches played by Indian cricket team against South Africa and Sri Lanka from 2008 to 2017. This data was collected from the <http://www.espncriinfo.com>.

The association between different factors representing playing condition and the match result is examined.

Different attributes considered are in Table 1 as:

##### A. Toss result (Toss lose, Toss Win)

Toss result decides which team will play (bat or field) first. Therefore mostly the toss result impacts the result of a match. Mostly seaming situation help the bowlers for seaming ball in the morning session which becomes more difficult for the batsman to play, but in the evening session or under the artificial light it becomes difficult to grip the wet ball (because of dew) for spinners and fielders.

##### B. Ground (Home Condition, Away Condition)

The ground condition is also analyzed to examine team India performance. Generally, the winning result of team is good in home condition than winning result of team in away condition.

##### C. Innings Order (Batting first, Batting second)

The innings order is also an important factor for a cricket team. Generally chasing the target by any team is more stressful than to establish a target in a game.

##### D. Match Outcome (Win, Lose)

Probable results for a game is win, lose, tie, and no outcome. This study analyzes the winning matches only.

TABLE I. ATTRIBUTE NAME AND THEIR VALU

Attribute	Value
Toss	Win, Lose
Ground	Home, Away
Inning Order	First, Second
Match Out Come	Win, Lose

On the basis of these attributes values following transaction data set are generated:

TABLE II. TRANSACTION DATA SET

Attribue name with its value	Item
India win toss	A
India lose toss	B
Home condition	C
Away condition	D
Bat first	E
Bat second	F
India win	G
India lose	H

Next the rules are generated as shown in Table 3(India vs Sri Lanka) and Table 4 (India vs South Africa)

TABLE III. GENERATED RULE (INDIA VS. SRI LANKA)

Transaction	India vs Sri Lanka
T <sub>1</sub>	AED→G
T <sub>2</sub>	AFD→G
T <sub>3</sub>	AEC→G
T <sub>4</sub>	BED→G
T <sub>5</sub>	BEC→G
T <sub>6</sub>	BFC→G
T <sub>7</sub>	AD→G
T <sub>8</sub>	BC→G
T <sub>9</sub>	BD→G
T <sub>10</sub>	FC→G

TABLE IV. GENERATED RULE (INDIA VS. SOUTH AFRICA)

Transaction	India vs South Africa
T <sub>1</sub>	AED→G
T <sub>2</sub>	AEC→G
T <sub>3</sub>	BFC→G
T <sub>4</sub>	BED→G
T <sub>5</sub>	BFD→G
T <sub>6</sub>	AC→G
T <sub>7</sub>	AD→G
T <sub>8</sub>	BD→G
T <sub>9</sub>	BC→G
T <sub>10</sub>	AD→G
T <sub>11</sub>	AC→G
T <sub>12</sub>	FD→G

The Confidence is now calculated as shown in Table 5 and Table 6. For example the confidence for the Transaction 1 from the Table 3 is(India wins toss, Bat first, Away Condition)(Win=3, lose=2)

Confidence = 60% for (India vs. Sri Lanka) using equation (3).

TABLE V. CONFIDENCE (INDIA VS. SRI LANKA)

Transaction	India vs Sri Lanka	Confidence (%)
T <sub>1</sub>	AED→G	60
T <sub>2</sub>	AFD→G	60
T <sub>3</sub>	AEC→G	100
T <sub>4</sub>	BED→G	66
T <sub>5</sub>	BEC→G	75
T <sub>6</sub>	BFC→G	100
T <sub>7</sub>	AD→G	61
T <sub>8</sub>	BC→G	100
T <sub>9</sub>	BD→G	66
T <sub>10</sub>	FC→G	100

TABLE VI. CONFIDENCE (INDIA VS. SOUTH AFRICA)

Transaction	India vs South Africa	Confidence(%)
T <sub>1</sub>	AED→G	33
T <sub>2</sub>	AEC→G	100
T <sub>3</sub>	BFC→G	33
T <sub>4</sub>	BED→G	33
T <sub>5</sub>	BFD→G	40
T <sub>6</sub>	AC→G	100
T <sub>7</sub>	AD→G	33
T <sub>8</sub>	BD→G	37
T <sub>9</sub>	BC→G	20
T <sub>10</sub>	AD→G	50
T <sub>11</sub>	AC→G	100
T <sub>12</sub>	FD→G	40

Weights ( $w_i$ ) are assigned to the pair ( $a_i, v_i$ ). In Table 7 (India vs Sri Lanka ) and Table 8 (India vs South Africa)

TABLE VII. WEIGHT ASSIGN (INDIA VS. SRI LANKA)

Attribue name with its value	Item	Weight
India win toss	A	0.6
India lose toss	B	0.4
Home condition	C	0.7
Away condition	D	0.3
Bat first	E	0.5
Bat second	F	0.5
India Win	G	0.7

TABLE VIII. WEIGHT ASSIGN (INDIA VS. SOUTH AFRICA)

Attribue name with its value	Item	Weight
India win toss	A	0.6
India lose toss	B	0.4
Home condition	C	0.9
Away condition	D	0.1
Bat first	E	0.7
Bat second	F	0.3
India Win	G	0.4

In next step the Item transaction weight is computed using the equation (4).

Thus ITW (AED) =  $0.6 * 0.5 * 0.3 * 0.7 = 0.063$  for (India vs. Sri Lanka)

Computed ITW is displayed in Table 9 and Table 10.

TABLE IX. ITEMSET TRANSACTION WEIGHT (INDIA VS. SRI LANKA)

Transaction	Itemset Weight
T <sub>1</sub>	0.063
T <sub>2</sub>	0.063
T <sub>3</sub>	0.147
T <sub>4</sub>	0.042
T <sub>5</sub>	0.098
T <sub>6</sub>	0.098
T <sub>7</sub>	0.126
T <sub>8</sub>	0.196
T <sub>9</sub>	0.084
T <sub>10</sub>	0.245

TABLE X. ITEMSET TRANSACTION WEIGHT (INDIA VS. SOUTH AFRICA)

Transaction	Itemset Weight
T <sub>1</sub>	0.0168
T <sub>2</sub>	0.1512
T <sub>3</sub>	0.0432
T <sub>4</sub>	0.0112
T <sub>5</sub>	0.0048

T <sub>6</sub>	0.2160
T <sub>7</sub>	0.0240
T <sub>8</sub>	0.0160
T <sub>9</sub>	0.1440
T <sub>10</sub>	0.0240
T <sub>11</sub>	0.2160
T <sub>12</sub>	0.0120

Next the Weight Support (WS) is computed using the equation (5) and is displayed in Table 11 and Table 12.

TABLE X. WEIGHT SUPPORT (INDIA VS. SRI LANKA)

Transaction	Weight Support
T <sub>1</sub>	0.0378
T <sub>2</sub>	0.0378
T <sub>3</sub>	0.1470
T <sub>4</sub>	0.0277
T <sub>5</sub>	0.0735
T <sub>6</sub>	0.0980
T <sub>7</sub>	0.0768
T <sub>8</sub>	0.1960
T <sub>9</sub>	0.0554
T <sub>10</sub>	0.2450

TABLE XI. WEIGHT SUPPORT (INDIA VS. SOUTH AFRICA)

Transaction	Weight Support
T <sub>1</sub>	0.0055
T <sub>2</sub>	0.1512
T <sub>3</sub>	0.0356
T <sub>4</sub>	0.0142
T <sub>5</sub>	0.0019
T <sub>6</sub>	0.2160
T <sub>7</sub>	0.0079
T <sub>8</sub>	0.0059
T <sub>9</sub>	0.0288
T <sub>10</sub>	0.1200
T <sub>11</sub>	0.2160
T <sub>12</sub>	0.0048

Finally Weight Confidence is computed using the equation (6) and is displayed in Table 13 and Table 14.

TABLE XII. WEIGHT CONFIDENCE (INDIA VS. SRI LANKA)

Transaction	Weight Confidence
T <sub>1</sub>	0.0226
T <sub>2</sub>	0.0250
T <sub>3</sub>	0.1470
T <sub>4</sub>	0.0282
T <sub>5</sub>	0.0551
T <sub>6</sub>	0.0980
T <sub>7</sub>	0.0228
T <sub>8</sub>	0.1960
T <sub>9</sub>	0.0365
T <sub>10</sub>	0.2450

TABLE XIII. WEIGHT CONFIDENCE (INDIA VS. SOUTH AFRICA)

Transaction	Weight Confidence
T <sub>1</sub>	0.0018
T <sub>2</sub>	0.1512
T <sub>3</sub>	0.0118
T <sub>4</sub>	0.0047
T <sub>5</sub>	0.0058
T <sub>6</sub>	0.2160
T <sub>7</sub>	0.0026
T <sub>8</sub>	0.0028
T <sub>9</sub>	0.0058
T <sub>10</sub>	0.0620
T <sub>11</sub>	0.2160
T <sub>12</sub>	0.0019

## V. RESULTS AND ANALYSIS OF THIS RESEARCH

This section discusses the results produced from the above Tables. The Table XII shows that if opponent is Sri Lanka then performance of Indian Cricket team is excellent in Home condition but in away condition Performance of India cricket team is not so bad. The Table XII also shows that mostly matches are against Sri Lanka, there is no impact of toss outcome, ground condition and inning order because Indian cricket team is superior to Sri Lanka.

In case of South Africa, the table (XIII) shows that India win more matches when India win the toss, Bat first and home ground and India loses more matches when play in away condition.

## VI. CONCLUSIONS

This paper presents the development of Weighted Association Rule mining process and demonstrates its application for Indian cricket. The effect of the toss, venue, and batting first or second are also discussed in this research paper. This research can be helpful for Indian team selectors, coaches, Skipper for making plans during or before the matches and for selecting the best player. There are some limitations of this research such as effect of team arrangement has not been measured. Total matches played by India, South Africa, Sri Lanka are not taken against other teams.

## REFERENCES

- [1] Xu, Zhenning, Gary L. Frankwick, and Edward Ramirez. "Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective." *Journal of Business Research* 69, no. 5 (2016): 1562-1566.
- [2] Olson, David L., and Desheng Dash Wu. "Data Mining Models and Enterprise Risk Management." In *Enterprise Risk Management Models*, pp. 119-132. Springer, Berlin, Heidelberg, 2017.
- [3] Le Jouan, H. (2017). U.S. Patent No. 9,699,133. Washington, DC: U.S. Patent and Trademark Office.
- [4] Choi, Tsan-Ming, Hing Kai Chan, and Xiaohang Yue. "Recent development in big data analytics for business operations and risk management." *IEEE transactions on cybernetics* 47, no. 1 (2017): 81-92.
- [5] Ofoghi, Bahadorreza, John Zeleznikow, Clare MacMahon, and Markus Raab. "Data mining in elite sports: a review and a framework." *Measurement in Physical Education and Exercise Science* 17, no. 3 (2013): 171-186.
- [6] Asanka P,D. (2014), Outcome of the Extra Delivery in Cricket Data warehousing and Data mining approach, In *International Journal of Engineering Research & Technology (IJERT)*, Vol. 3 Issue 10,(990-999)
- [7] Bhattacharjee, Sanmoy, JayakrushnaSahoo, and AdrijitGoswami. "Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015." In *Advances in Computing and Communication Engineering (ICACCE)*, 2015 Second International Conference on, pp. 616-621. IEEE, 2015.
- [8] Singh, Tejinder, Vishal Singla, and Parteek Bhatia. "Score and winning prediction in cricket through data mining." In *Soft Computing Techniques and Implementations (ICSCTI)*, 2015 International Conference on, pp. 60-66. IEEE, 2015.
- [9] Verma, Aman, and Masoumehlzadi. "Cricket Prognostic System: A framework for real-time analysis in ODI cricket." In *KDD Workshop on Large-Scale Sports Analytics*. 2016.
- [10] Zhao, Baojin, and Lei Chen. "Prediction Model of Sports Results Base on Knowledge Discovery in Data-Base." In *Smart Grid and Electrical Automation (ICSGEA)*, International Conference on, pp. 288-291. IEEE, 2016.
- [11] Nair, Prasant. "Editing the worm graph alias ball by ball data and predicting the winner for cricket." (2016).
- [12] Satao, Preeti, AshutoshTripathi, JayeshVankar, BhaveshVaje, and VinayVarekar. "CRICKET SCORE PREDICTION SYSTEM (CSPS) USING CLUSTERING ALGORITHM."