

Selection for Balanced Cricket Team Fourth coming ICC Championship 2017

Utsav Jagdishbhai Solanki

PG Scholar: Department of IT

G.H.Patel Collage of Engineering and Technology,
V.V.Nagar, Anand, Gujarat, India.
er.utsavsolanki27@gmail.com

Prof. Jay Vala

Assistant Professor: Department of IT

G.H.Patel Collage of Engineering and Technology,
V.V.Nagar, Anand, India.
jayvala@gcet.ac.in

Abstract—Sport grown from their levels of country to international, essential part of any sport is measure performance analysis of players based on past statistics. Selection of team like cricket, based on player performance as strength and weakness is useful for team selector to make decision. Arrival of world cup, every team finds their best team combination to present on ground to get maximum result in their favour. Large dataset contain much useful information to identify that information uses association rule mining technique. This paper contains association rule mining technique based on individual specific Indian players and make set of players for cricket team. Player's statistics as past records, unknown relation of factors impact player performance, analyzing is helpful to identify best team as per terms. The factors as grounds, venue, strike rate, average, match played by particular player, runs scored, ball faced, high score, not out, match played, numbers of fours and sixes, number of hundreds and fifty's, position of batsman, dismissal, for bowlers factors as wicket taken, maiden's, over bowled, runs given(per match), economy rate, inning played. The result of this study is useful for Indian team selector and manager of team, for strategic planning and decision making and increase winning for team India for the ICC world cup 2017.

Keywords—Data mining; hash apriori algorithm; Association rules (AR); support; confidence; ICC world cup.

I. INTRODUCTION

Cricket is popular sport in many countries as India, Australia, New Zealand, etc... Cricket is play between two teams of eleven players. Players are selected based on performance of past records and role to play in team such as batsman, bowler and wicket keeper, player who have grip in batting and bowling too, is known as all-rounder. Cricket organize in different way as Test match, T-Twenty match which player for twenty overs that over contain six balls to play, One Day Internationals matches contain fifty overs to play in match.

International cricket council is committee for arrange tournament in India. This paper is focused for One day International matches, also societal and grown with extent viewers. The paper is for ICC championship 2017.

Analysis for particular Indian player based on ODI matches for championship 2017 technique using Association rule mining.

Data mining is used for information for large amount of database, there are many techniques for identify meaningful information from data, Association rule is one of techniques that identify relation in itemset based on frequent itemset, association rule [9] is generated based in rules uses to identify pattern.

In data mining, association rule is defined by Agarwal [2][10]. AR is uses in market basket for identify customer purchased items with related items, with in database of number of transactions. Database transaction contains number of items taken by visitors in market. In transaction database P and Q are sets of item sets that symbol $P \rightarrow Q$ shown as association rule, represents that transaction in database have P with Q in it. Association rule is measure using support [6][7] (frequent items gets form dataset) and confidence (particular item found from database) a criterion, another measure is lift, ratio of antecedent and consequent, and comes to gather. Lift result is measure equal to one, less than one and greater than one [2], dramatically shows relation as independent, negative and use for future prediction of items with another item in next year's database.

This paper contains data of cricket players ICC championship 2017 of India held in England. Paper containing analysis of batmen players performance as batting opposition, runs scored, strike rate for batsman, for bowlers maidens, runs given, wickets, etc.

II. METHODOLOGY

Data mining contain different types of techniques and algorithms, based on that help to identify memory consumption, efficiency and execution time. Association rule, apriori is base algorithm to identify meaningful patterns [5], because of large database it take more time and overhead. We

formally change to use hash apriori algorithm [3] to perform on our definitive data.

Hash apriori algorithm applies in steps, using partition algorithm it reduces record occupation time [4][8], dataset is scan only once after frequent itemset would inflect 1 to numbers of k itemset, after that using hash value in hash function it generate hash key that store in hash table for further use. Using these itemset induce association rule.

This paper batsmen and bowler records is considered same as transaction record [1]. Batsman affecting factor are strike rate, runs scored (ability to play on pitch), ball faced, average, venue, opposition team, dismissal, minute played and for bowler affecting factors are runs given, wickets, maidens, economy rate, rest of factors are same in both players records such as ground, match result and position, based on these factors identifying performance, effort that ability carry out from players.

III. DATASET UNDERSTANDING

Collections of dataset repository are espninfo, sports.ndtv and cricbuzz website. Players selected based on conditionally satisfied for at least 10 ODI matches played for India. A player dataset 1 is contains batsmen match played till 29 October 2016 and dataset 2 is for bowler match played till 29 October 2016. Players performance parameters details given following.

- 1) Runs scored and Strike rate: Batsmen runs score count based on 100 balls called S/R, also it flows overall team effort in runs. Runs criteria for cover all boundary for dataset is less than seventy, seventy to hundred and greater than hundred
- 2) Venue: player performance is change with different weather, pitch and environment. Venue contain for home as played in India, for Asian Venus played in Sri Lanka, for ocean country are New Zealand and Australia, and for rest are away country.
- 3) Opposition: this attribute is stats to player and team played with which opponent team.
- 4) Ground: this attribute shows which ground player played in different matches also home, away, ocean, or Asian venues.
- 5) Minute played: how many minutes of player played on this ground record is contain in it, criteria for that is <70, 70-140, >140.
- 6) Ball faced: this attribute is ball played by batman in per match, criteria for that is <60, 60-120, >100.
- 7) Average: this attribute is numbers of runs division by number of back to pavilion.
- 8) Position: Batsman batting position matter as player is match opener, one down, second down or middle order to play match. Position role in match is coordination with players and plan to play.
- 9) Match result: cricket match is winning, loss, tied and no result matches are covered in records.

- 10) Runs given: bowler given runs in match contain records.
- 11) Wickets: bowler taken wickets in matches are given in this record.
- 12) Maidens: bowlers bowled maidens in number of matches record (no runs given in six balls) called maiden.
- 13) Economy rate: Bowler's effort is known by wicket taken in matches, wicket criteria are 0-1 wickets, 2-3 wickets and above 3 wickets.
- 14) Over bowled: numbers of overs bowled by bowlers in numbers of matches.
- 15) Inning played: match inning played important role for player as first inning or second inning.

IV. ANALYSIS AND RESULTS

Dataset analysis environment is based on Java NetBeans IDE, taking antecedent, as support from minimum 0.1 to maximum 1.0 for batsmen and bowlers Tables are given in it. Result mentioned confidence 0.68(multiplying with 100) is means to 68 percentage of that association rule.

TABLE I. BOWLER DATASET ASSOCIATION RULE TO IDENTIFY PERFORMANCE

Dhawal Kulkarni Performance	
AssociationRule	[[Position 8, inning 1, v Zimbabwe] -> [Match_result won]: confidence=0.6875: lift=1.0]
AssociationRule	[[Position 8, Match_result won] -> [Maidens 0-1]: confidence=1.0: lift=1.0]
AssociationRule	[[Match_result won, v Zimbabwe] -> [Position 8]: confidence=1.0: lift=1.0]
AssociationRule	[[Position 8, Match_result won, Maidens 0-1] -> [inning 1]: confidence=1.0: lift=1.0]
AssociationRule	[[Position 8, Maidens 0-1, v Zimbabwe] -> [Over_bowled >8]: confidence=1.0: lift=1.0]
AssociationRule	[[Position 8, Maidens 0-1, v Zimbabwe] -> [Wickets 1]: confidence=1.0: lift=1.0]
Hardik Pandya Performance	
AssociationRule	[[inning 1, Over_bowled 4-8, home , v Zimbabwe] -> [Position 8]: confidence=0.75: lift=1.0]
AssociationRule	[[Position 8, Over_bowled 4-8, Maidens 0-1] -> [Match_result won]: confidence=0.75: lift=1.0]
AssociationRule	[[Position 8, Wickets 1, home , Maidens 0-1] -> [Over_bowled 4-8]: confidence=0.6666666666666666: lift=0.8888888888888888]
AssociationRule	[[Wickets 1, home , Maidens 0-1] -> [Economy_rate >6]: confidence=1.0: lift=1.3333333333333333]
Ishant Sharma Performance	
AssociationRule	[[Position 8, Match_result won, v Zimbabwe] -> [Wickets 1]: confidence=1.0: lift=1.0]
AssociationRule	[[Position 8, Over_bowled >8, Maidens 0-1, v Zimbabwe] -> [Match_result won]: confidence=0.6407766990291263: lift=1.0]
AssociationRule	[[Position 8, Match_result won, Maidens 0-1] -> [Over_bowled >8]: confidence=0.8918918918918919: lift=0.9669985775248933]
AssociationRule	[[Over_bowled >8, Match_result won, v Zimbabwe] -> [Position 8]: confidence=0.7184466019417476: lift=1.0]

AssociationRule [[Position 8, Maidens 0-1, v Zimbabwe] -> [Match_result won]: confidence=0.9223300970873787: lift=1.0]
Yuvraj Singh Performance
AssociationRule [[Over_bowled <4, Match_result won, v Zimbabwe] -> [home]: confidence=0.8125: lift=1.099264705882353]
AssociationRule [[Position 8, Match_result won, Wickets 1, Economy_rate <5, v Zimbabwe] -> [Over_bowled <4]: confidence=0.7058823529411765: lift=1.2488687782805432]
AssociationRule [[inning 1, Position 8, Match_result won, v Zimbabwe] -> [Wickets 1]: confidence=0.72727272727273: lift=0.983957219251337]
AssociationRule [[Position 8, Over_bowled <4, Wickets 1] -> [home]: confidence=0.8125: lift=1.099264705882353]

TABLE II. BATSMEN DATASET ASSOCIATION RULE TO IDENTIFY PERFORMANCE

Cheteshwar Pujara Performance
AssociationRule [[Position 10, Runs_scored <40, Average <70, Wellington] -> [Match_result won]: confidence=1.0: lift=1.0]
AssociationRule [[Position 10, Minute_played <70, Match_result won] -> [v Zimbabwe]: confidence=0.8: lift=1.0]
AssociationRule [[Minute_played <70, Strike_rate 70, Match_result won, Average <70, v Zimbabwe] -> [Wellington]: confidence=0.8: lift=1.0]
AssociationRule [[Minute_played <70, Runs_scored <40, Ball_faced <60, Average <70, v Zimbabwe] -> [Strike_rate 70]: confidence=0.8: lift=1.0]
Wridhiman Saha Performance
AssociationRule [[Position 10, Ball_faced <60, Match_result won, v Zimbabwe] -> [Wellington]: confidence=1.0: lift=1.0]
AssociationRule [[Position 10, Runs_scored <40, Match_result won] -> [home]: confidence=1.0: lift=1.0]
AssociationRule [[Strike_rate 70, Match_result won, Wellington] -> [Runs_scored <40]: confidence=0.77777777777778: lift=1.0]
AssociationRule [[Position 10, Runs_scored <40, Match_result won, Ball_faced <60, Wellington, v Zimbabwe] -> [Average <70]: confidence=1.0: lift=1.0]
MS Dhoni Performance
AssociationRule [[Minute_played <70, Strike_rate 70, Average <70, v Zimbabwe] -> [Match_result won]: confidence=0.62162162162162: lift=1.0]
AssociationRule [[Minute_played <70, Runs_scored <40, Match_result won, Average <70] -> [Ball_faced <60]: confidence=0.7368421052631579: lift=1.2982456140350878]
AssociationRule [[Minute_played <70, Strike_rate 70, Wellington] -> [v Zimbabwe]: confidence=0.62162162162162: lift=1.0]
AssociationRule [[Position 10, Runs_scored <40, Average <70, Wellington] -> [Ball_faced <60]: confidence=0.8070175438596491: lift=1.105912930474334]
AssociationRule [[Position 10, Runs_scored <40, Average <70, Wellington] -> [Match_result won]: confidence=0.72972972972973: lift=1.0]
AssociationRule [[Runs_scored <40, Wellington] -> [Minute_played <70]: confidence=0.84: lift=1.0535593220338983]
AssociationRule [[Position 10, Strike_rate 70, Average <70, v Zimbabwe] -> [Runs_scored <40]: confidence=0.864406779661017: lift=1.0153349475383373]
Virat Kohli Performance
AssociationRule [[Position 10, Runs_scored <40, Average <70, Wellington] -> [Ball_faced <60]: confidence=0.9107142857142857: lift=1.5061813186813187]
AssociationRule [[Position 10, Runs_scored <40, Average <70, Wellington] -> [Match_result won]: confidence=0.6046511627906976: lift=1.0]

AssociationRule [[Runs_scored <40, Wellington] -> [Minute_played <70]: confidence=0.9074074074074074: lift=1.3454661558109835]
AssociationRule [[Minute_played <70, Strike_rate 70, Match_result won, Average <70, v Zimbabwe] -> [Wellington]: confidence=0.6162790697674418: lift=1.0]
AssociationRule [[Position 10, Ball_faced <60, Match_result won, v Zimbabwe] -> [Wellington]: confidence=0.6511627906976745: lift=1.0]

Partially rules for selection of players

Figures shows execution time for taking support values from 0.1 to 1.0 and based on analysis confidence is taken as 0.5 for identifying rules, Figure (a) and Figure (b) shown below it for bowler and batsman dataset execution time in it.

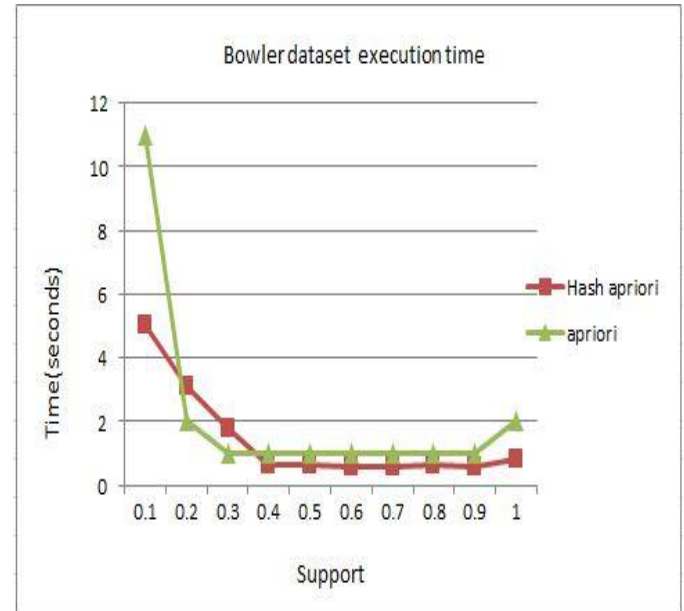


Fig. 1. Support for bowler dataset with time.

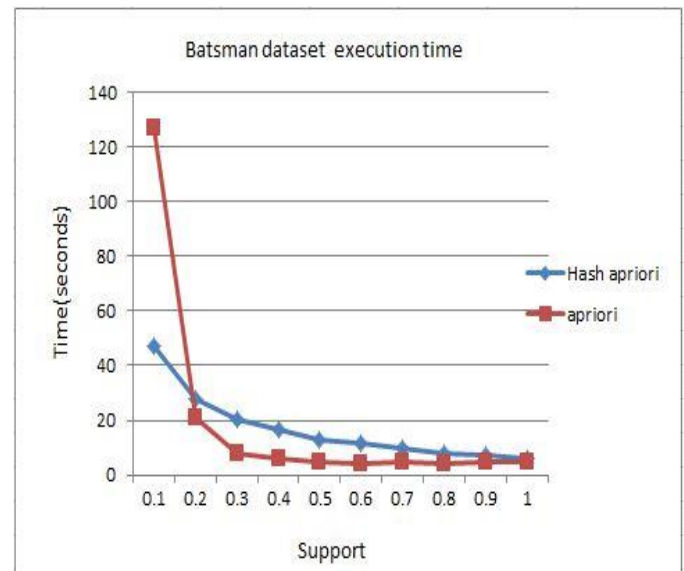


Fig. 2. Support for batsmen dataset with time.

For balanced cricket team all player role and effort must know by selection committee, here are balanced team suggestions for play in ICC 2017 based on dataset and association rule generated on past record statistics.

Cricket team is combination of four batsmen, five bowlers and two all-rounders with wicket keeper, in past records player selected based on only their main job as batsman or bowler not as all-rounder players but now a days particular players have batting and bowling in their job as playing in team. This paper suggested team containing all-rounder with multiple job players in cricket team.

TABLE III. FIRST SUGGESTED TEAM FOR ICC 2017

Players
hardik pandya(pace)- all rounder
amit mishra(spin)- all rounder
ms dhoni (wc)-batsman
rohit sharma- batsman
virat kohli- batsman
shikhar dhavan- batsman
yuzvendra chahal(spin)- bowler
yuvraj singh(spin)- bowler
dhawal kulkarni(pace)- bowler
ishant sharma(pace)- bowler
mohammed shami(pace)- bowler

TABLE IV. SECOND SUGGESTED TEAM FOR ICC 2017

Players
hardik pandya(pace)- all rounder
amit mishra(spin)- all rounder
yuvraj singh(spin)- all rounder
ms dhoni (wc)-batsman
rohit sharma- batsman
virat kohli- batsman
shikhar dhavan- batsman
gautam gambhir- batsman
yuzvendra chahal(spin)- bowler
dhawal kulkarni(pace)- bowler
ishant sharma(pace)- bowler

V. CONCLUSION

Selection for balance cricket team for is hard to choose, but performance of player can show way for environment condition, pitch condition scored by players, easy to choose in team to play.

For away venue performance of player are averages and also away ground player scoring middle not to extreme runs or high score in away Venue also batsman are scored low compare to home venue

Bowler job is to player for last bowling order that can perform well as maidens, low runs given also wicket taken as one. Batsman played well in away opponent teams as Zimbabwe, performance of batsman strike rate, run scored are average to count and ball played by batsman with minute played by then are low.

Association rule mining is telling us patterns based on past records for specific batsman and bowler, player physical defects as injury, medical issue and past fracture that can affect during player performance. Weather environment that player match in day or night matches in different grounds are affecting factors.

ACKNOWLEDGMENT

I wish to acknowledg my parents, Jagdishbhai Solanki and Sarlaben Solanki and my family, also Prof. Jay Vala (GCET) and all the staff of Department of IT, G.H.Patel Collage of Engineering and Technology, for supporting and appreciate me, for cricket data, special thanks to espninfo.com, sports.ndtv.com and cricbuzz.in, for collected accurate records stored with in large time span that is appreciable and useful.

REFERENCES

- [1] Sawant, Vinaya, and Ketan Shah. "Performance Evaluation of Distributed Association Rule Mining Algorithms." *Procedia Computer Science* 79 (2016): 127-134.
- [2] Bhattacharjee, Sanmoy, Jayakrushna Sahoo, and Adrijit Goswami. "Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015." *Advances in Computing and Communication Engineering (ICACCE)*, 2015 Second International Conference on. IEEE, 2015.
- [3] S. Aguru, "A Hash Based Frequent Itemset Mining using Rehashing," no. December, pp. 4198-4204, *IJRITCC*, 2014.
- [4] Prasanna K., M. Seetha, and AP Siva Kumar. "CApriori: Conviction based Apriori algorithm for discovering frequent determinant patterns from high dimensional datasets." *Science Engineering and Management Research (ICSEMR)*, 2014.
- [5] Mazid, Mohammed M., ABM Shawkat Ali, and Kevin S. Tickle. "Finding a unique association rule mining algorithm based on data characteristics." *Electrical and Computer Engineering*, 2008. *ICECE* 2008. International Conference on. IEEE, 2008.
- [6] Heaton, Jeff. "Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms." *SoutheastCon*, 2016. IEEE, 2016.
- [7] Dongre, Jugendra, Gend Lal Prajapati, and S. V. Tokekar. "The role of Apriori algorithm for finding the association rules in Data mining." *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014 International Conference on. IEEE, 2014.
- [8] Sumangali, K., et al. "Mining interesting itemsets from transactional database." *Computational Intelligence and Computing Research (ICCIC)*, 2014 IEEE International Conference on. IEEE, 2014.
- [9] Abbas, Wan Faezah, Nor Diana Ahmad, and Nurlina Binti Zaini. "Discovering Purchasing Pattern of Sport Items Using Market Basket Analysis." *Advanced Computer Science Applications and Technologies (ACSAT)*, 2013 International Conference on. IEEE, 2013.
- [10] Jamsheela, O., and G. Raju. "Frequent itemset mining algorithms: A literature survey." *Advance Computing Conference (IACC)*, 2015 IEEE International. IEEE, 2015.