iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** IX **Month of publication:** September 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Novel Algorithm to Reduced Computational Data Sets for Fuzzy Association Rule

Rohit Miri[1], Priyanka Tripathi[2], Keshri Verma[3], S.R.Tandan[4]
[1,4]Department of Computer Science & Engineering
Dr. C.V. Raman University, Kota, Bilaspur, India (C.G.)
[2]Department of Computer Technology, NITTTR, Bhopal (M.P.)
[3]Department of Computer Science and Application, NIT, Raipur (C.G)

Abstract*— The Data Mining Association rule is suitable for small or medium scale data sets. For large data sets it becomes more complex because while applying association rule finding candidate and large item set again and again amongst variables. So we reduce the computational time or data complexity by applying the proposed algorithm. The large data set requires more computational time on data operations. If we apply fuzzy association rule then it becomes a complex task. So it is necessary to reduce the large data set into smaller data set by applying the proposed Novel Algorithm. After that we can apply a fuzzy association rule. We work on reduction of unwanted data sets from large data sets that are not important for making the decision. This research paper tries to eliminate those item sets which are not important for finding any association rule. We apply our pseudo code or algorithm to achieve our task.

Keywords*—Fuzzy Logic, Data Mining, Novel algorithm, Row counts

I. INTRODUCTION

Now a day's fuzzy logic and data mining are the two most important techniques on the field of research area. Fuzzy logic uses linguistic variable to provide the answer that human being are unable to give. Whereas data mining is used to find out the essential things from large data sets, then apply some association data mining algorithm to find the relationships among the item sets. This research paper finds out the range of linguistic variable by using standard deviation and mean in place of assuming its values. Secondly we apply the concept of row values and column values. The row value plays an important role in the field of fuzzy logic. Column values are also important but do not play any role in this algorithm. In future it will use on clustering technique. It makes our task easier to calculate data mining association rule. Row values count helps to eliminate unwanted data sets that are not important for making association rule. This proposed algorithm helps the fuzzy mining association rule to reduce their computational time or data sets. Before applying Mining association we reduce the given data sets into smaller data sets.
Later we can apply the clustering technique to calculate the same thing. This algorithm is also useful for performing prediction or forecast of past data sets.

II. LITERATURE REVIEW

Mohammed Al-Maoleg1[1] et al chooses those item sets that are frequently present in particular transaction id. Its algorithm works on low support 5.5 and reduces the time complexity of the program very easily. Zhiyong ma [4] et al converts all the item sets into Boolean matrix by using CP tree method and reduces the time for the task.Arpna Shrivastava [5] et al , in this paper the authors have used the codes for all the items and remove the duplication by using data cleansing technique. This is also most efficient as compared to simple Apriori algorithm. K. Sathesh Kumar and M. Hemalatha [3] , this paper reduces the operational time carried out by Apriori algorithm by using artificial Bee colony optimization method (FABCO).Mehmet Kaya et al [24] , in this paper the author find the efficient algorithm by carried out mining fuzzy clustering algorithm (CURE). They found out the centroid by CURE for triangular membership function.so that they can range the fuzzy membership method correctly and also reduces the computational time.
As we know we need some kind of association rule to perform data mining algorithm. Getting this Agrawal and his co-worker carried out some mining algorithm based on the large data sets, which is difficult to find association mining rule [9-18]. These break the mining steps into two phases. In the first phase candidate of item sets are obtained and counted by scanning the transactions. The number of item set must support the minimum pre-defined threshold value called minimum support. Then later we make the pair of

item sets and apply the association rule for getting the required output. Srikant and Agrawal also proposed partitioned based mining association algorithm. Cai at al proposed weighted mining rule of data sets [19]. Yue et al, extended the fuzzy concept based on vectors [22].Most of them are find out the range of triangular fuzzy membership function directly , means they assumed the range of linguistic variable. But on my paper we have find out the range of linguistic variable by using mean and standard deviation.

Mining association rule was also performed by Hang [21]. The aim of his research is digging out the essential or useful item from very large data set. He also wants to improve the data mining algorithm in terms of time complexity. We used [9],[15] for comparison with my algorithm. They have done on fuzzy mining association rule to reduce the computational time. They all used the simple mining association rule for doing the task, the TRApriori mining association technique is used from the the paper [16] .

III. PROPOSED METHODOLOGY

The first step is to take an exemplary data set of bank industry to evaluate the data sets. We have taken the data sets from KEEL data repository in the field of marketing. We have selected 8 important different attributes with having 20 row values of bank data to check its association amongst the attributes. The data a set is shown below on table no 1.1.

S.NO.	MARITAL STATUS	AGE	EDUCTION	TYPES OF HOME	OCCUPTION	HOUSEHOLD MEMBER	UNDER 18	INCOME
1.	1	5	5	1	5	5	3	9
2.	1	3	5	3	1	3	2	9
3.	5	1	2	1	6	4	3	1
4.	5	1	2	1	6	4	3	1
5.	1	5	4	1	8	2	1	8
6.	5	2	3	3	9	3	2	1
7.	3	3	4	3	3	1	1	6
8.	1	5	3	3	8	3	1	2
9.	1	5	4	3	8	2	1	4
10.	5	2	4	3	9	1	1	1
11.	2	2	3	1	2	2	1	4
12.	1	5	3	1	5	4	1	7
13.	1	5	4	1	1	2	3	7
14.	3	3	3	3	2	2	2	1
15.	1	5	3	3	5	2	1	8
16.	3	4	2	3	3	2	1	2
17.	1	4	4	1	2	5	4	9
18.	1	4	4	1	1	5	4	8
19.	3	5	3	5	2	3	1	4
20.	5	1	2	1	6	4	2	1

Table No 1.1. Marketing Data Set from KEEL Repository

STEP:1 Take data set for performing fuzzy mining association rule and find out the standard deviation and mean. It is helpful in finding the domain or range values of fuzzy Linguistic variable. i.e low , middle and high

STEP:2 Create or build triangular membership function and define or assume the range of linguistic variable with the help of first step (standard deviation and mean) of given data sets i.e low, middle and high . There are eight different attributes in the data set. Attributes are also categories in to two types. The first type of attributes contains five distinct values and second type of attribute contains nine distinct values. So we must define two membership functions for two distinct types of attributes or variables. After that we combined their fuzzy values in one table. The two membership function diagram is shown below named as figure no 1.1 and

Volume 3 Issue IX, September 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

figure no 1.2 respectively. The fuzzy values of these two membership function is shown below in table 1.2

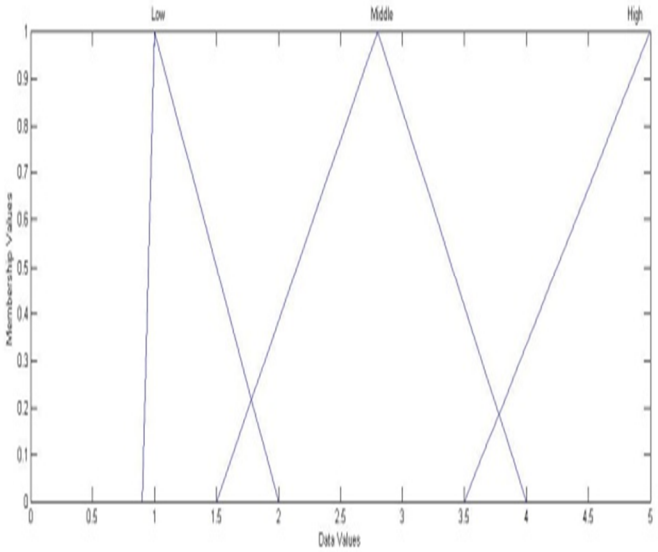


Figure No 1.1 (Membership function of five Distinct Values)

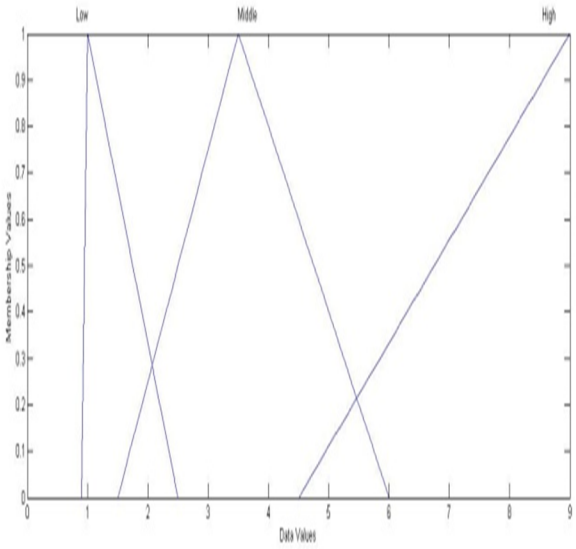


Figure No 1.2 (Membership function of Nine Distinct Values)

STEP:3 Find out the fuzzy values of given data sets

In this, we have used triangular membership function because of its easiness and computational efficiency. We can also use Gaussian membership function. We can categories it's as Low, Middle and high. Thus we have used three fuzzy membership values produced for each result obtained is as follow:-

According to the predefined membership functions for this transaction data in table No 1.1

Transform the actual values of each attribute into fuzzy sets. Take MarritalStatus value in case 1 as an example. The value 1 is replaced by a fuzzy set (1/low + 0.0/middle + 0/ high) Using the given membership functions. This step is repeated for all the data sets.

S.NO.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MARITAL STATUS	L	1	1	0	0	1	0	0	1	1	0	0	1	1	0	1	0	1	1	0	0
	M	0	0	0	0	0	0	.9	0	0	0	.38	0	0	.9	0	.9	0	0	.9	0
	H	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
AGE	L	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	M	0	.9	0	0	0	.38	.9	0	0	.38	.38	0	0	.9	0	0	0	0	0	0
	H	1	0	0	0	1	0	0	1	1	0	0	1	1	0	1	.34	.34	.34	1	0
EDUCATION	L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	.38	.38	0	.9	0	.9	0	0	.9	.9	0	.9	.9	.38	0	0	.9	.38
	H	1	1	0	0	.34	0	.34	0	.34	.34	0	0	.34	0	0	0	.34	.34	0	0
TYPES OF HOME	L	1	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1
	M	0	.9	0	0	0	.9	.9	.9	.9	.9	0	0	0	.9	.9	.9	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
OCCUPTION	L	0	1	0	0	0	0	0	0	0	0	.34	0	1	.34	0	0	.34	1	.34	0
	M	.42	0	0	0	0	0	.73	0	0	0	.24	.42	0	.24	.42	.73	.24	0	.24	0
	H	.12	0	.34	.34	.77	1	0	.77	.77	1	0	.12	0	0	.12	0	0	0	0	.34
HOUSEHOLD MEMBER	L	0	0	0	0	.34	0	1	0	.34	1	.34	0	.34	.34	.34	.34	0	0	0	0
	M	.42	.73	.83	.83	.24	.73	0	.73	.24	0	.24	.83	.24	.24	.24	.24	.42	.42	.73	.83
	H	.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.12	.12	0	0
UNDER 18	L	0	.34	0	0	1	.34	1	1	1	1	1	1	0	.34	1	1	0	0	1	.34
	M	.73	.24	.73	.73	0	.24	0	0	0	0	0	0	.73	.24	0	0	.83	.83	0	0.24
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INCOM	L	0	0	1	1	0	1	0	.34	0	1	0	0	0	1	0	.34	0	0	0	1
	M	0	0	0	0	0	0	0	.24	.83	0	.83	0	0	0	0	.24	0	0	.83	0
	H	1	1	0	0	.77	0	.34	0	0	0	0	.56	.56	0	.77	0	1	.77	0	0

Table no 1.2. The fuzzy values Transformation table

STEP:4 Count its row values and column values of fuzzy values data sets. i.e

Row Values: it is a valid value present in the row except zero or it is a number of non-zero values in a particular row.

Column Values: it is a valid value present in the column except zero or it is a number of non-zero values in a particular column.

STEP:5 Select the row which has maximum number of row values or contains more than 50 %

Count Value of Row: Row values/ Total no of items * 100.

I.e The count value of MarritalStatus(low)= (10 *100)/20 =50 %

The following table no 1.3 is useful data sets

S.NO.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Sum
Age	H	1	0	0	0	1	0	0	1	1	0	0	1	1	0	1	.34	.34	.34	1	0	11
Education Occupation Household	M	0	0	.38	.38	0	.9	0	.9	0	0	.9	.9	0	.9	.9	.38	0	0	.9	.38	11
	H	.12	0	.34	.34	.77	1	0	.77	.77	1	0	.12	0	0	.12	0	0	0	0	.34	11
	M	.42	.73	.83	.83	.24	.73	0	.73	.24	0	.24	.83	.24	.24	.24	.24	.42	.42	.73	.83	18
UNDER 18	L	0	.34	0	0	1	.34	1	1	1	1	1	1	0	.34	1	1	0	0	1	.34	14

Table No 1.3 Useful Data Sets

STEP:6 If the sum of row values of remaining rows has > maximum sum values of outputted row. Add this row as in important data sets.

STEP:7 Select only those values that satisfy the step 5 and Step 6. This data set is important or essential data sets used for building fuzzy mining association rules. Discard the remaining data sets because they are not important for mining association rule.

STEP:8 Select the column which has more than 50% of count value. This column values plays an important role in making data mining association rules in future work.(optional step)

STEP:9 Apply data mining fuzzy association rule on the above important or useful data sets.

STEP: 9.1 Define the minimum support value and confidence threshold i.e. support=5.5 and confidence threshold=.7

STEP: 9.2: Generate the candidate item set C1

Taking the linguistic value AGE.High as an example .The scalar cardinality will be (1 + 1+1+1+1+1+1+.34+.34+.34+1)= 9.03 . The remaining candidate C1 item sets for are shown below.

{ (AGE.High, 9.03),(Education.Middle,7.8), Occupation.High,5.69), (HouseholdMember.Middle,9.1), (Undereighten.Low,11.36) }

STEP:9.3 Generate L1 Large data sets.

The L1 large item sets stands on the greatest value for individual attribute and is equal to or greater than minimum support value. So l1 large item sets are as follows:-

{ (AGE.High, 9.03),(Education.Middle,7.8), (HouseholdMember.Middle,9.1), (Undereighten.Low,11.36) }

STEP 9.4Generate the candidate itemset Ci+1 from Li.

The candidate item set is generated from large itemset Li. Here comparison operations take the minimum membership value/degree from newly form candidate item set. In below example, we compare the linguistic variable values Age.High with Education.middle

Volume 3 Issue IX, September 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

that is equivalent to (Age.High, Education.middle). the result are as shown below.

Table 1.4 the linguistic value (Age.High, Education.middle)

S.No	(Age.High,	Education.middle).	(Age.High, Education.middle).
1	1	0	.0
2	0	0	.0
3	0	.38	.0
4	0	.38	.0
5	1	0	.0
6	0	.9	.0
7	0	0	.0
8	1	.9	.9
9	1	0	.0
10	.0	.0	.0
11	0	.9	.0
12	1	.9	.9
13	1	0	.0
14	0	.9	.0
15	1	.9	.9
16	.34	.38	.34
17	.34	.0	.0
18	.34	.0	.0
19	1	.9	.9
20	0	.38	.0
			3.94

Table No 1.4 C2 Candidate Item Sets

The C2 candidate item sets are as follow

S.N	**Itemsets**	**Count**
1	Age.High, Education.middle).	3.94
2	Age.High, HouseholdMember.Middle).	4.59
3	Age.High, Undereighteen.Low).	6.34
4	Education.middle, HouseholdMember.Middle	5.12
5	Education.middle, Undereighteen.Low	5.90
6	HouseholdMember.Middle, Undereighteen.Low	4.75

Table No 1.5 Complete C2 Candidate Item Sets

STEP 9.5: Generate L2 Large data sets.

The L2 large item sets stands on the greatest value for individual attribute and is equal to or greater than minimum support value. So l1 large item sets are as follows:-

{ (Age.High, Undereighteen.Low)., 6.34(Education.middle, Undereighteen.Low,5.90), }

STEP 9.6: Repeat the steps 9.4 and 9.5 until L1+1 large item set is null

STEP 9.7Construct the fuzzy association rule.

(a) Construct the association rules for all thelarge item set .there are three possible association rules.

AGE(H) ,UnderEighteen(L)/ AGE(H)= 6.34 /9.02=0.702

Education.(M), Undereighteen(L)/ Education.(M), = 5.90 /7.8=0.7544

UnderEighteen(L), AGE(H) / UnderEighteen(L)/ = 6.34 /11.36=0.5580

Undereighteen(L),Education.(M)/ Undereighteen(L)/ = 5.90 /11.36=0.519

The fuzzy association rules are listed on table 1.6

S.N	Fuzzy Association Rule	Confidence Value
1	AGE(H)→ ,UnderEighteen(L)	70.2%
2	Education.(M)→Undereighteen(L)	75.44%

Table No: 1.6 Summary of Association rule

(b)We can also find the confidence of the entire rule. Suppose our minimum threshold is .70 for confidence. Its confidence value is calculated as:

The confidence values of the other are shown below.

“If Age = High, then Undereighteen = Low” has a confidence value of 70.2%;

In plain English it means that the old man with having age more than 60 years does not have the under 18 year child.

“If Education = Middle, then Undereighteen = Low” has a confidence value of 76.64%;

In plain English it means that the old man with having qualification 12 standard or 1 to 3 years of college does not have the less than 18 year child.

IV. CONCLUSION

It a simple Novel algorithm to find the fuzzy association rule for large data sets. Complexity of data is reduced before applying mining algorithm. So we find the candidate itemsets and large item sets again and again on less number of variables. Prediction results are more accurate if we apply on large data sets.

Makes mining algorithm simple due to smaller data sets .i.e. The proposed algorithm reduces the large data sets into smaller data sets. We apply mining association rule after this task.

The Proposed algorithm is very useful when Apriori algorithm fails to find the mining association rule due to large data sets.

It reduces the complexity of data so it also reduces the time complexity of data or mining algorithmi.eFinding candidate and large item sets again and again are impossible due to large data sets.

We must predict the past data set quickly and more accurate on this proposed algorithm..

V. FUTURE SCOPE

For future work, we can predict the result of association rule without applying the Data Mining association algorithm based on the fuzzy classification or clustering technique.

We can also apply the concept of operational research technics or method for finding useful data sets.

VI. APPENDIX

Appendixes, if needed, appear before the acknowledgment.

VII. ACKNOWLEDGMENT

We would like to express our sincere thankfulness to our guide Dr PriyankaTripathi and also to Dr KeshriVerma, for their helpful support and valuable advice during our research work. We also thank to KEEL data repository for their valuable data sets.

REFERENCES

[1] N S Nithya and K Duraiswami “Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface” vol 39 part 1 ,febuary 2014 , pp 39-52 .

[2] K Sathesh Kumar and M Hemlatha “An innovative potential on rule optimization using fuzzy artificial bee colony” Research journal of applied science , engineering and Technology.. volume 2627-2633 , ISSN:2040-7459, E-issn:2040-7467, April 05 2014.

[3] zhiyong MA , QingyingQiu and PeieenFeng “CPM algorithm for mining associaotion rules from databases of engineering design+n instances.” Journal of information science and engineering vol30 , 463-481 (2014)

[4] Arpnashrivastava , R . C Jain and A K shrivatava “Generating 3rd level association rule using fast aprioriimplementation.”British journal of mathematics and computer science. Volume 4(2):241-251, 2014

[5] Neelukhare ,neeruadlakha, K R pardasani “An algorithm for mining multidentional fuzzy association rule” international journal of computer science and information security vol 5, no-1 ,2009

[6] Thomas sudkomp “Refinement of temporal constraints in fuzzy association ” volume 48(2008) 14 June 2007

[7] Tiberiu Ban , Fuzzy computing for complexity level of evaluation tests. StudiaUniv, Babes-Bolai, Informatics, Volume LVIII number 1, 2013

[8] Tzung-Pei Honga,Chan-Sheng Kuob,Shyue-Liang Wangc “A fuzzy AprioriTid mining algorithm with reduced computational time” Volume 5, Issue 1, December 2004, Pages 1–1

[9] Tzung-pei-Hang, Guo-Cheng lan, Yi-Hsinlin and Shing-tai pan. An effective gradual data reduction strategy for fuzzy item set mining international journal of fuzzy systems ,vol 15 no2 June 2013.

[10] Usha Rani, R VijayaPrakkash, Dr A. Gavardhan,Mining Multi-level association Rules using fuzzy, International journal of emerging technology and advanced engineering. Volume 3, issue 8, August 2013.

[11] RuchiBhargava ,Shrikant lade , Effective positive negative association rule mining using improved frequent pattern tree. International journal of advanced research in computer science and software engineering.Volume3 , issue 4 April 2013 .

[12] DharmendraKumar , Deepak Bhardwaj rise of data mining: Current and Future application areas. IJCSI international journal of computer science issue vol 8 issue 5 No1 September 2011.

[13] KanuPatel ,Vatsalshal , jitendrapatel , jayna donga comparision of various association rule mining algorithm on frequent item sets. International journal of advanced research in computer Science and software engineering Vol-3 issue II Nov 2013 .

[14] JrShianchen hung liehchou, ChingHsue Cheng, jen-ya Wang CPDA Based Fuzzy association rules for learning achievement mining, 2009 international conference on machine learning and computing IPCSIT vol3 (2011) IACSIT PRESS SINGAPUR.

[15] E Ramaraj, K Rameshkumar, N Venkatesan”A better performed transaction Reduction algorithm for mining frequent item set from large voluminous database”Volume 5, Issue 1, December 2004, Pages 1–10

[16] R. Agrawal, T. Imielinksi and A. Swami, “Mining association rules between sets Submitted manuscript 24 of items in large database,“ The 1993 ACM SIGMOD Conference, Washington DC,USA, 1993.

[17] R. Agrawal, T. Imielinksi and A. Swami, “Database mining: a performance perspective,” IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No.6, 1993, pp. 914-925.

[18] C. H. Cai, W. C. Fu, C. H. Cheng and W. W. Kwong, “Mining association rules with weighted items,” The International Database Engineering and Applications Symposium, 1998, pp. 68-77.

[19] T. P. Hong and C. Y. Lee, "Induction of fuzzy rules and membership functions from training examples," Fuzzy Sets and Systems, Vol. 84, 1996, pp. 33-47.

Volume 3 Issue IX, September 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

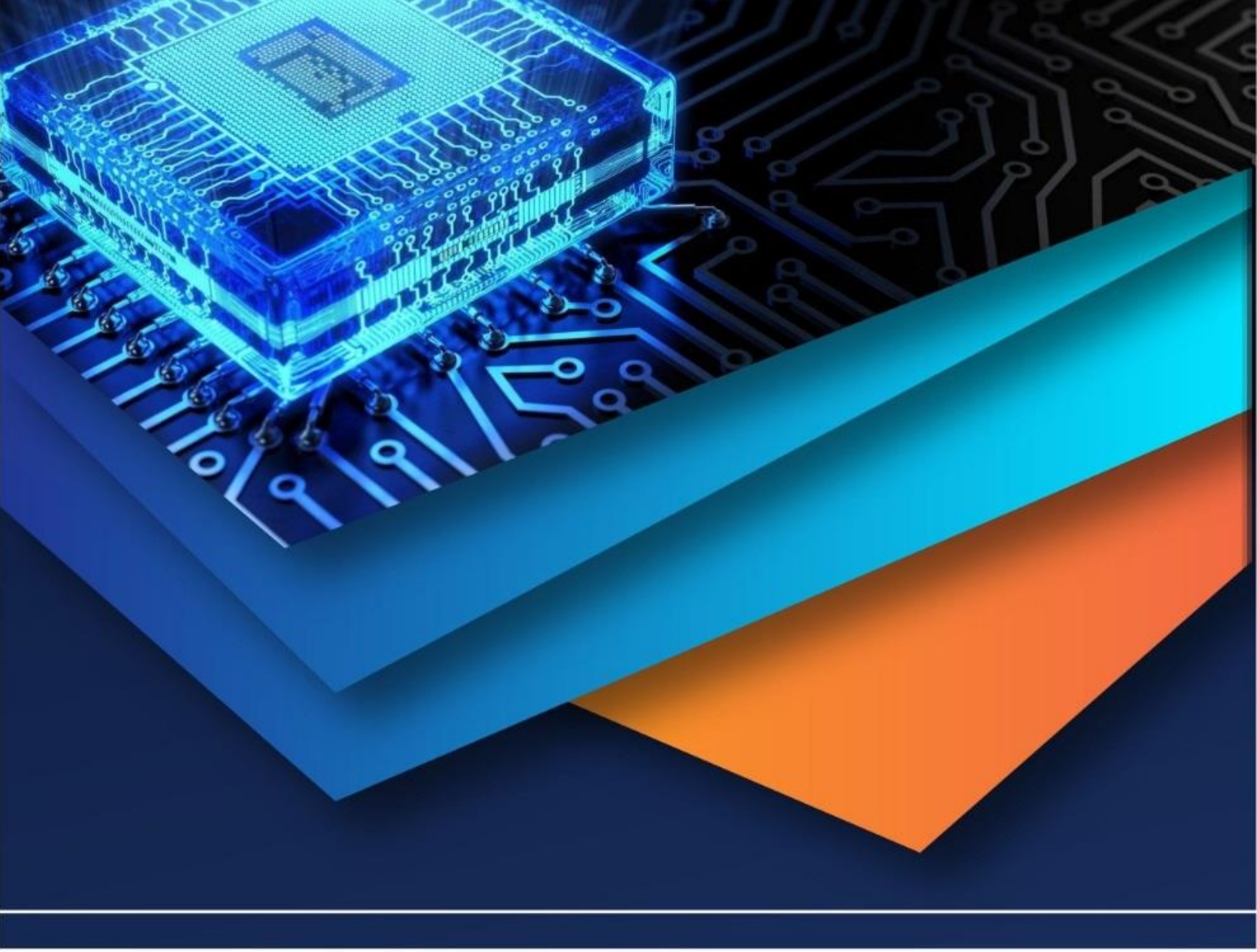
[20] T. P. Hong, C. S. Kuo, and S. C. Chi, "Mining association rules from quantitative data," Intelligent Data Analysis, Vol. 3, No. 5, 1999, pp. 363-376. A. Kandel, Fuzzy Expert Systems, CRC Press, Boca Raton, 1992, pp.8-19. Submitted manuscript 26
[21] S. Yue, E. Tsang, D. Yeung and D. Shi, "Mining fuzzy association rules with weighted items," The IEEE International Conference on Systems, Man and Cybernetics, 2000, pp. 1906-1911.
[22] L. A. Zadeh, "Fuzzy sets," Information and Control, Vol. 8, No. 3, 1965, pp. 338-353.
[23] Mehmet Kaya1,*, Reda Alhajj1,**, Faruk Polat2, and Ahmet Arslan3 "Efficient Automated Mining**ofFuzzyAssociation Rules** "DEXA 2002, LNCS 2453, pp. 133-142, 2002.

RohitMiri is currently Ph.D.scholar and H.O.D of Computer Science and Engineering and Information Technology Department, Dr. C.V. Raman University, Bilaspur, India. He received his B.E. degree in Computer Science and Engineering from the NIT, Raipur (formally known as Government Engineering College, Raipur) in 2004, and M. Tech degree in Computer Science from College of Engineering, Pune Maharashtra, India in 2008 .His research interests include application of Web Technologies, Data mining & Warehousing ,Fuzzy Logic and Cyber Security.

Dr Priyanka Tripathi is working as an Associate Professor in the Department of Computer Application at National Institute of Technical Teachers Training and Research, Bhopal. She has also worked on various projects in Tata Consultancy Services. She has done MCA in Govt Engineering College Raipur. Ph.D from Maulana Azad National Institute of Technology, Bhopal. Working in the area of web Technology,networking, agile computing. Presented various paper in International Conferences at USA, Thailand etc. Also Chaired sessions in international conferences.

Dr Keshri Verma is an Assistant Professor in the Department of Computer Science in National Institute of Technology, Raipur. Research interests include application of Data Mining and Image Processing.

S.R. Tandanis currently Ph.D.scholar and Assistant Professor in the Department of Computer Science and Engineering, Dr. C.V. Raman University, Bilaspur, India. He received his B.E. degree in Computer Science and Engineering from the NIT, Raipur (formally known as Government Engineering College, Raipur) in 2006, and M. Tech degree in Computer Science from Birla Institute of Technology, Mesra Ranchi, India (JH) In 2009. His research interests include application of Artificial intelligence in robotics, Soft Computing, Dynamic path planning and Mobile robot navigation in



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY