iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** V **Month of publication:** May 2015

DOI:

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

Prediction of Data Sets by using Fuzzy Mining Association Rule

Sushmita Acharjee[1],Rohit Miri[2], Asha Miri[3]
[2,3]*Department of Computer Science & Engineering,* [1]*Dr. C.V. Raman University, Kota, Bilaspur, India (C.G.)*
[2]*Department of Computer Science, MCKV , Howrah (W.B.)*

Abstract— This paper is proposed to reinforce fuzzy association rules, Association rule mining is one of the fundamental tasks of data mining. The conventional association rule mining algorithms, using crisp set, are meant for handling Boolean data. However, in real life quantitative data are voluminous and need careful attention for discovering knowledge. Therefore, to extract association rules from quantitative data, the dataset at hand must be partitioned into intervals, and then converted into Boolean type. In the sequel, it may suffer with the problem of sharp boundary. This problem will be solved by standard deviation and mean of the given data sets. The boundary values of the entire linguistic variable are calculated by the first phase of our algorithm.
Keywords— fuzzy logic, Data Mining, Support , Confidence

I. INTRODUCTION

Previous studies often partitioned the length of interval in equal length and ignored the distribution characteristics of datasets. Therefore, this paper focus on improving the persuasiveness in determines the universe of discourse and membership functions of fuzzy association rules. In empirical case study, we use an exemplary dataset to be simulation data, which contains the learning achievement data of 7 different schools. The remaining content of this paper is organized as follows. Related work is described in below section. Section 3 presents our proposed approach. Finally conclusion of this paper stated in section 4.

A. Reducing large unwanted data set without affecting the outcomes of fuzzy mining association algorithm.
B. Reducing the operational time of item set up to 30%. As compared to previous result from different research paper.
C. More fruitful for any business prediction process.

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 .3 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 I have find out the range of linguistic variable by using mean and standard deviation.

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

A. Fuzzy Association Rule

Agrawal et al. Introduced Apriori approach to explore behaviors in market basket data, those transactions are all binary values. However, any real-world database may also contain quantitative attributes. To overcome their shortcomings, fuzzy association rule mining approaches were proposed to handle both the quantitative and categorical attributes. Fuzzy association rules-based methods transform quantitative values into a fuzzy set for each attribute; and use the membership degree operations to find rules. These rules can be expressed like the one: If age is young, then salary is low. Detailed overviews of fuzzy association rules mining methods can be found in.

III. THE PROPOSED APPROACH

In this section, the fusion model is proposed to improve the persuasiveness in determining the universe of discourse and membership functions of fuzzy association rules. Firstly, we build the membership function based on standard Deviation and mean for each conditional attribute. Secondly, we use the AprioriTid mining algorithm to extract fuzzy association rules. An exemplary dataset used throughout this paper is shown in Table 1. The dataset contains 10 instances, which is characterized by the following attributes (STATE board & CBSE board school):

A. R.S. Nursery English School (RSNS)
B. Holy Nursery English School(HNES)
C. Bhibhuti English Medium School(BES)
D. D.A.V Public School (DAV)
E. Kendriya Vidyalaya (K.V)
F. Loyala English School (LES)
G. Bharat Mata English Medium School (BMES)

All these attributes are numerical values. The proposed table is introduced in detail as follows:

STUDENT NO.	RSNS	HNES	BES	DAV	K.V	LES	BMES
1	67	56	86	77	86	71	68
2	61	74	72	79	89	77	80
3	84	63	61	89	86	79	89
4	73	67	84	86	65	84	62
5	70	78	70	89	87	72	79
6	65	83	68	77	86	61	87
7	67	72	70	87	75	71	80
8	86	65	85	63	64	84	86
9	75	63	57	65	79	87	88
10	79	74	79	63	63	85	89

The task or our algorithm is categorized in two three part or three phase.

First phase: To determine the range of linguistic variable by using standard deviation and mean. The mean of all the data sets will be the Centre portion middle value of all the subjects. We indicate that phase in to red colors in the fuzzy triangular membership function to the below diagram of figure no 1.1. the boundary of the middle linguistic variable will be Mean-2*SD(left) and Mean+ *SD (Right). The outer boundary of the triangular membership function will be Minimum (Low) and maximum (High) value of the assume data sets of all the schools. The maximum value of the low linguistic variable will be the Mean-SD(Right) and the minimum value of starting range of the high linguistic variable will be Mean + SD (Left).

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

Second Phase: In the second phase , we calculated the fuzzy values of all the variable by using triangular membership function. In this phase we also assume the minimum support=1.5 and minimum confidence .6 .

Third Phase: in the Third phase we apply the AprioriTid mining association rule. This is our final phase where we got the output of association rules. We can also predict about which school will provide the best result on the basis of their past result. The relationship between the school will be outputted with their confidence value.

	RSNS			**HNES**			**BES**			**DAV**			**K.V**			**LOYALA**			**BMES**		
No.	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**
1	.02	.04	.00	.08	.00	.00	.18	.34	.00	.00	.66	.00	.00	.00	.64	.09	.64	.00	.54	.00	.00
2	.03	.06	.00	.00	.09	.00	.00	.87	.00	00	.47	.13	.00	.00	.93	.00	.63	.00	.00	.70	.00
3	.00	.00	.00	.07	.00	.00	.78	.00	.00	.00	.00	1.0	.00	.00	.64	.00	.39	.18	.00	.00	.85
4	.05	.01	.06	0.6	.07	.00	.00	.00	1.0	0.0	0.0	.76	.39	.31	.00	.00	.00	.64	1.0	.00	.00
5	.04	.01	.00	.00	.03	.00	.00	.87	.00	.00	.00	1.0	.00	.00	.74	.00	.77	.00	.00	.81	.00
6	.01	.02	.00	.00	.00	.09	.67	.00	.00	.00	.66	.00	.00	.00	.64	.97	.00	.00	.00	.00	.65
7	.00	1.0	.00	.02	.00	.00	.00	.87	.00	.00	.00	.84	.00	.92	.00	.09	.64	.00	.00	.70	.00
8	.00	.00	1.0	.07	.07	.00	.00	.09	.00	.70	.01	.00	.80	.00	.00	.00	.00	.64	.00	.05	.54
9	0.0	.02	.09	.05	.05	.00	.02	.00	.00	.52	.20	.00	.00	.66	.00	.00	.00	.91	.00	0.0	.75
10	.00	.90	.00	.00	.94	.00	.02	.12	.36	.70	.01	.00	.88	.00	.00	.00	.00	.73	.00	.00	.85

IV. ALGORITHM AND DATA ANALYSIS

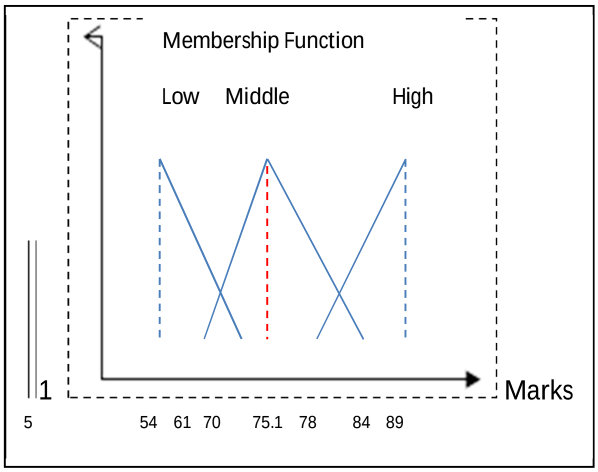
The all the steps of our proposed algorithm are depicted bellows:

STEP 1: take the input data sets and calculate their mean and standard deviation.

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

The membership function is built by employing triangular fuzzy number and the triangular fuzzy number model of Table 2 is shown in Fig. 2. The membership function as shown in Fig. 2 can be established based on Standard Deviation and Mean.

STEP 2: find out the range and boundary values of all the linguistic variable. To determine the range of linguistic variable by using standard deviation and mean. The mean of all the data sets will be the Centre portion middle value of all the subjects.
We indicate that phase in to red colors in the fuzzy triangular membership function to the below diagram of figure no 1.1. the boundary of the middle linguistic variable will be Mean-2*SD(left) and Mean+ *SD (Right).
The outer boundary of the triangular membership function will be Minimum (Low) and maximum (High) value of the assume data sets of all the schools.
The maximum value of the low linguistic variable will be the Mean-SD(Right) and the minimum value of starting range of the high linguistic variable will be



STEP 3: Predefine the minimum support value and confidence threshold.

Users need to predefine the minimum support value and the confidence threshold. The minimum support value is set to 1.5 and the confidence threshold is set at 0.7 in this case.

STEP 4: Generate the candidate item set c1.

Taking the linguistic value RSNS Low as an example, the scalar cardinality will be (0.02 + 0.03 + 0.05 + 0.04 + 0.01) = 0.15. The $C1$ candidate item set for this example is shown as follows:

{(RSNS.low,0.15), (RSNS.middle,2.06), (RSNS.high,1.15), (HNES.low,0.35), (HNES.middle,1.25), (HNES.high,0.09), (BES.low,1.75), (BES.middle,3.16), (BES.high,1.00), (DAV.low,1.92), (DAV.middle,2.01), (DAV.high,3.73), (KV.low,1.68), (KV.middle,1.98), (KV.high,3.59) , (LES.low,1.15), (LES.middle,3.08), (LES.high,3.1), (BMES.low,1.54), (BMES.middle,2.26), (BMES.high,3.64), }
STEP 5: Generate the $L1$ large item set.
The $L1$ large itemset rides on the largest count value for each attribute and is equal to or greater than the minimum support value. In this exemplary dataset, $L1$ can be denoting as follows:
{(RSNS.middle,2.06), (HNES.middle, 1.25), (BES.middle,3.49), (DAV.high, 3.73), (KV.high, 3.59) ,(LES.high, 3.1) ,(BMES.high, 3.64) , }.

STEP 6: Generate the candidate item set

The candidate itemset is generated from Li. Here, comprasion operation is used to return the lesser membership degree of a

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

newly formed candidate itemset. In the example below , we compare the linguistic values RSNS.middle with HNES.middle that is transformed into RSNS.middle,HNES.middle.Tthe result is shpwn below :

NO.	RSNS.middle	HSNS.middle	(RSNS.middle, HNES.middle)
1	0.04	0.00	0.00
2	0.06	0.09	0.06
3	0.00	0.00	0.00
4	0.01	0.07	0.01
5	0.01	0.03	0.01
6	0.02	0.00	0.00
7	1.00	0.00	0.00
8	0.00	0.07	0.00
9	0.02	0.05	0.02
10	0.09	0.00	0.00

Taking the linguistic value (RSNS.middle,HNES.high) as an example , the scalar cardinality will be (0.06 + 0.01 + 0.01 + 0.02) = 0.1. The c2 candidate itemset for this exemplary dataset is shown in table below:

ITEMSET	COUNT
RSNS.middle,HNES.middle	0.1
RSNS.middle,BES.middle	0.33
RSNS.middle,DAVhigh	0.92
RSNS.middle,KV.high	0.94
RSNS.middle,LES.high	0.12
RSNS.middle,BMES.high	0.13
HNES.middle,BES.middle	0.31
HNES.middle,DAV.high	0.19
HNES.middle,KV.high	0.05
HNES.middle,LES.high	0.92
HNES.middle,BMES.high	0.97
BES.middle,DAV.high	1.84
BES.middle,KV.high	0.87
BES.middle,LES.high	0.68
BES.middle,BMES.high	0.54
DAV.high,KV.high	0.84
DAV.high,LES.high	0.82
DAV.high,BMES.high	0.85
KV.high,LES.high	1.3
KV.high,BMES.high	0.66
LES.high,BMES.high	2.2

STEP 7: Generate the L i+1 large itemset from candidate itemset C 1+i.

The L2 large itemset rides on the count value, which is equal to or greater than the minimum support value. The L2 large itemset rides for this exemplary dataset is shown as follows :

{((BES.middle,DAV.high)1.98) , ((LES.high,BMES.high)2.2)}

No.	BES.middle	DAV.high	LES.high	(BES.middle,DAV.high,LES.high)
1	0.34	0.00	0.00	0.00
2	0.87	0.13	0.00	0.00
3	0.00	1.00	0.18	0.00
4	0.00	0.76	0.64	0.00
5	0.87	1.00	0.00	0.00
6	0.00	0.00	0.00	0.00
7	0.87	0.84	0.00	0.00
8	0.09	0.00	0.64	0.00
9	0.00	0.00	0.91	0.00
10	0.12	0.00	0.73	0.00

Table6.The linguistic value (BES.middle,DAV.high,LES.high).

STEP 8: Repeat the steps 6 & 7 until large itemset is null

In the example below, we generate the candidate itemset C3. The count value of linguistic value (BES. middle,DAV.high,LES.high) is described in table 6.the c 3 candidate itemset is listed in table 7. All of the count value in c 3 are smaller than the minimum support value. Therefore , the L3 large itemset is null.

ITEMSET	COUNT
BES.middle,DAV.high,LES.high	0.00
BES.middle,DAV.high,BMES.high	0.85
BES.middle,LES.high,BMES.high	0.84
DAV.high,LES.high,BMES.high	0.18

STEP 9: Construct the fuzzy association rules.

Construct the fuzzy association rule from L2. In the following example,

BES.middle $\Rightarrow$ DAV.high, its confidence value is calculated as follows:

$$\frac{(BES.middle,DAV.high)}{BES.middle} = \frac{1.84}{3.16} = 0.58$$

The fuzzy association rule is listed below

Fuzzy association rules	Confidence value
BES.middle DAV.high	0.60
DAV.high BES.middle	0.50
DAV.high LES.high	0.22
LES.high DAV.high	0.27

After this calculation we can check whether the confidence values of the above association rules (listed in table 8) are larger or equal to the predefined threshold. The final resulting rules are thus obtained: If the score of data structure is high, then the score of management information system is high.

Output is if BES middle then DAV is High

V. CONCLUSION

This paper proposes a fusion model to improve the persuasiveness in determining the universe of discourse and membership functions, and eliminate unwanted data sets from given data sets. Prediction of data sets is also performed from fuzzy mining association rule.

VI. FUTURE SCOPE:

We can find the result of Data Mining Association Rule without using the actual data mining association rule . the result obtained from the proposed techniques will also support up to maximum accuracy by modifying the proposed algorithm..

VII. APPENDIX

Appendixes, if needed, appear before the acknowledgment.

VIII. ACKNOWLEDGMENT

We would like to express our sincere thankfulness to our guide Dr Priyanka Tripathi, for their helpful support and valuable advice during our research work.

REFERENCES

[1] MOHAMMED AL-MAOLEGI ,BASSAMARKOK "AN IMPROVED APRIORIALGORITHM FOR ASSOCIATION RULES" INTERNATIONAL JOURNAL ON NATURAL LANGUAGE COMPUTING VOLUME 3 , NO-1 , FEBUARY 2014.

[2] 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 .

[3] 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.

[4]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)

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

[6] 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

[7]Thomas sudkomp "Refinement of temporal constraints in fuzzy association " volume 48(2008) 14 June 2007

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

[9]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

[10] 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.

[11]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.

[12] 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.Volume 3 , issue 4 April 2013 .

[13] 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.

[14] 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 .

[15] 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.

[16]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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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

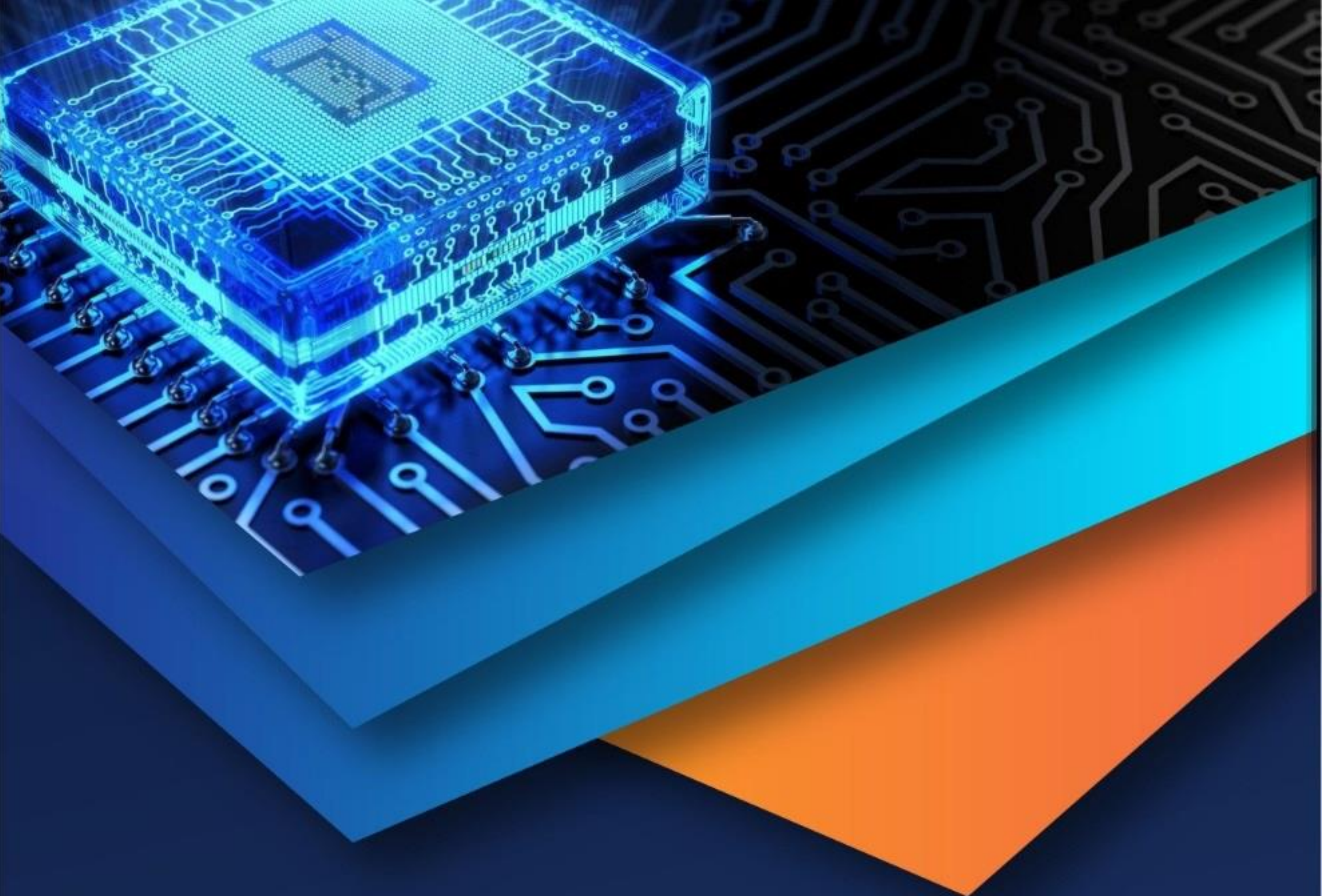
[22] S. Yue, E. Tsang, D. Yeung and D. Shi, "Mining fuzzy association rules with weighted items," The IEEE International Conference on Systems, Man and

Cybernetics, 2000, pp. 1906-1911.
[23] L. A. Zadeh, “Fuzzy sets,” Information and Control, Vol. 8, No. 3, 1965, pp. 338-353.
[24]Mehmet Kaya[1,*], Reda Alhajj[1,**], Faruk Polat[2], and Ahmet Arslan[3] “Efficient Automated Mining **of Fuzzy Association Rules** “DEXA 2002, LNCS 2453, pp. 133-142, 2002.

Sushmita Acharji is currently M tech. scholar of Computer Science and Engineering, MCKV institute of engineering, Howrah. SHe received his B.E. degree in Computer Science and Engineering from Dr. C.V. Raman University, Bilaspur, India. Working in the area of web Technology,Data Mining and Fuzzy logic

Rohit Miri 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

Asha Miri is student of M Tech final year student in Dr C V raman University from computer science and engineering departments. She is completed her bachelor of engineering in Dr C V raman University, kota bilaspur. Working in the area of web Technology,Data Mining and Fuzzy logic



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY