Dr. K.Kavitha, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 2, February- 2016, pg. 01-08 **ISSN: 2321-8363**

RANKING INTERESTING RULES FOR RECOGNIZING BEST TEACHER USING MULTIDIMENSIONAL DATACUBE

Dr. K.Kavitha
Assistant Professor, Department of Computer Science,
Mother Teresa Women's University, Kodaikanal
kavitha.urc@gmail.com

Abstract
Data Mining provides many technologies to detect the hidden patterns and predict an interestingness rules. Decision making and cluster is an important concept to group the related information and makes the better decision accordingly. Quality of education depends largely on teacher's ability. This paper focuses the key problem in the teacher's performance. The proposed method evaluates the Interesting Rules levels for recruiting best teacher. This method allows the user for to finding the rules classification to determine whether teacher can be recruited or not. The proposed idea is tested by extracted data from institution and results proven that it provides better accuracy.

Keyword: *Association Rule, Interesting Measures, Cluster, Multi-Dimensional Data, DataCube*

1. Introduction
In education field, quality education is a tedious process to decide. Every institution is having its own criteria. Based on the criteria, best teacher's has to be selected but it is very difficult to identify the qualified teachers. Information's are collected along with their experiences. The key problem is to extract the information and classification, rule prediction and interesting rule classification to perform the required task. Data mining plays a major role in knowledge discovery using various domains. Decision making and clustering is an important concept to group the related information and make the better decision accordingly. Clustering and multi-dimensional data cube classification provides some additional tools to reduce the outcome uncertainty for enhancing the educational quality. The objective of best teacher's recruitment is to evaluate the ranks so that the process makes better decision making. It helps to improve the speed and consistency of recruitment application.

The rest of the paper is organized as follows. Related work is explained in Section 1.The proposed work is presented in Section 2. Section 3 contains the comparative analysis results obtained. Finally, the conclusion is drawn at section4.

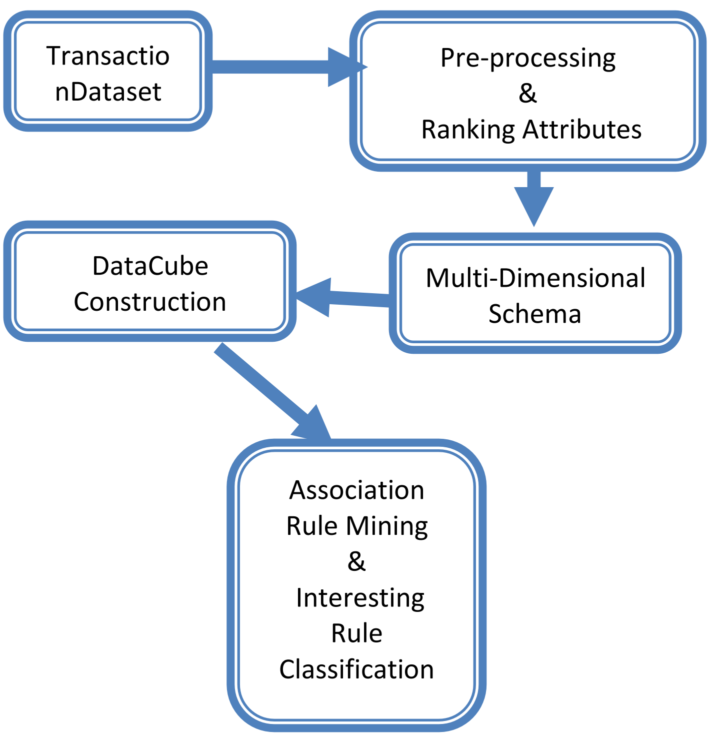
2. Related Work
This section deals with the work related to risk evaluation and association rules. Usman.et.al proposed a method to select a subset of informative dimension and fact identifiers from initial candidate sets. Knowledge Discovered from standard approach for mining original data[1]. Pears.et.al presented a methodology to incorporate semi-automated extraction methods. Binary tree of hierarchical clusters are constructed for each node. This paper presented new method to generates rank for nominal attributes and generate candidate multidimensional schema[2].

Liu.et.al recommended a system to generate association rules for personal financial schemas. The data cubes generated based on financial information and multidimensional rules are generated[3]. Chiang.et.al proposed a model to mine association rules of customer value. Ward's method initiated to partition the online shopping market into three markets. Here Supervised learning is employed to create association rules[4].

Herawan.et.al presented an approach for regular and maximal association rules from transactional datasets based on soft set theory. Here Transactional datas are transformed into Boolean values based on information system[5].

3. Predicting Best Teachers based on Classifying Interesting Rules in Multidimensional Dataset

Multidimensional schema is generated by merging the benefits of hierarchical and multi-dimensional scaling techniques based on the neighbour cluster, nominal variables and numeric variables are ranked. It represents the relationship between higher dimensional and performance measures. Highly ranked dimensions and facts can be used to construct a data cube. It helps to provide intercutting rule classification techniques. The following figure shows the main steps involved in proposed work. The real world teacher's recruitment dataset is used to diversity the association rule. The dataset is initially pre-processed to attain the highest quality of the dataset. Initial step removes the unnecessary portions from the dataset. Existing methods applies clustering techniques to categories the rules. But in the proposed method provides better performance than the clustering technique.



3.1 Attribute Ranking

To reduce the size of the dataset data mining uses some technologies. It can be classified in to two such as Nominal Attributes and Numerical Attributes.

3.2 Ranking Numeric Variables

Numeric variables are ranked by Principal Components Analysis (PCA) in term of the degree of variance. It transforms a given set of attributes into smaller set of attributes. Here dataset can be reduced with and numeric variables. The following table depicts the numerical attributes with the rank measure based on the calculated value.

Attributes	Calculated value	Rank
Experiences	0.156	2
Age	0.090	4

PCA transforms the original dataset into a smaller set which is attained by finding a set of eigen vectors V_1.
Ranking numeric value is evaluated by $Li=\sum^{n}_{j=1} V_{ij}$ $V_i= 1,2,3,.....n$

3.3. Ranking Nominal Attribute

Nominal Attributes are ranked by using Information gain. Entropy and information gain measures the correlation. Entropy measures the uncertainty among random variables in the dataset. Information gain is used to select the best attributes splitting criteria.

$Entropy(N) = \sum^{M}_{i=1} S_i \log_2 S_i$

Where m is the number of classes in the dataset and N is the number of recording in N

Attributes	Information Gain	Rank
Marital status	0.100	3
Academic	0.255	1
Category	0.248	2
Sex	0.078	4

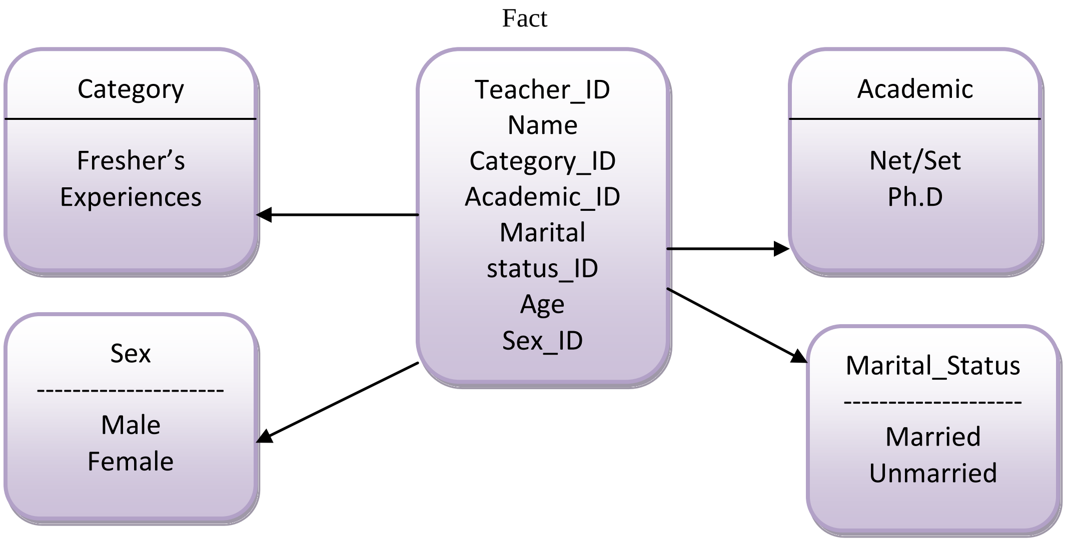
Information gain can be calculated as follows

$IG(NA)=Info(D)-Info_{NA}(D)$ where IG is the information gain of nominal attributes
$Info(D)=\sum^{M}_{i=1} P_i \log_2(p_i)$ Where P_i represents the Probability with m- number of class
$Info_{NA}|D|=\sum^{l}_{k=1} |D_k| / |D| \, Info(D_k)$
where|D| is the total no of observation in dataset D and D_k is the value of all attributes. Nominal attributes are ranked based on the highest Information gain.

3.4 Multi-Dimensional Scaling

To achieve natural grouping for each of the nominal attributes, Multi dimensional scaling is applied to find the semantic relationship between the calculated values for each nominal attribute. Relationship can be easily visualised by parallel co-ordinates based on the similar semantic values, Assigned values are grouped together. Multi-Dimensional schema formation.

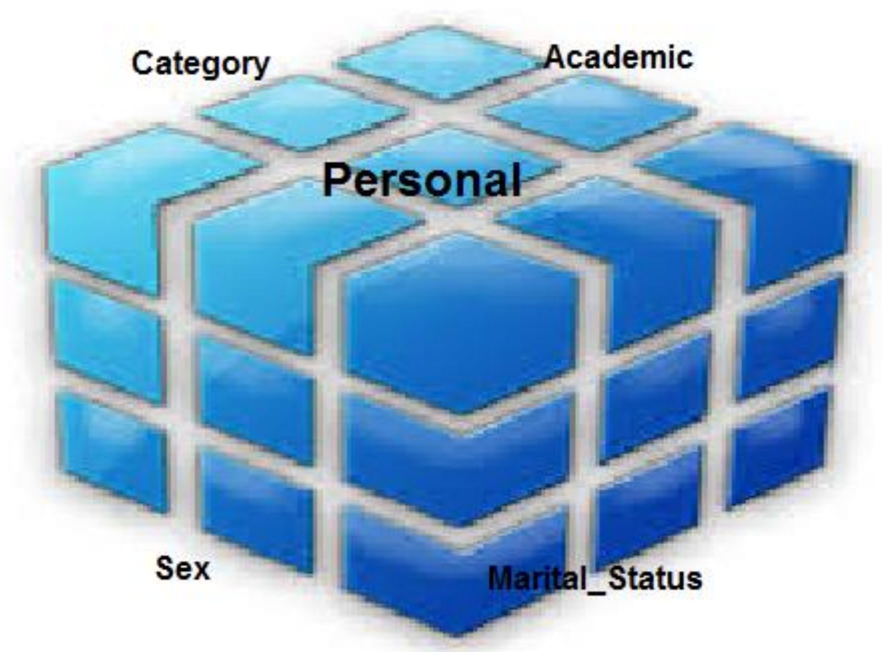
Multi-Dimensional schema is constricted by facts(numeric variables) and dimension values(nominal attributes). Grouping defines the hierarchical dimensions. Each dimension in a cluster has group and value level.This schema is represented in the form of nominal attributes as dimensional and numeric attributes as facts. Multi-dimensional schema for teachers recruitment dataset is shown in figure.



The above Multidimensional schema contains all the dimensional and facts in the teacher's recruitment dataset. This schema is used to construct the Information Data cubes.

3.5 Data cube Construction

Highly ranked facts and dimensional are used to construct data cube. It helps the user to show the rank positions. The key point in a data cube stores combined measure of the dimension values in a multi-dimensional space. OLAPP operations are used to explore the data into meaningful pattern. Data cubes can be constructed by providing any number of dimensions and particular for the Information cubes.

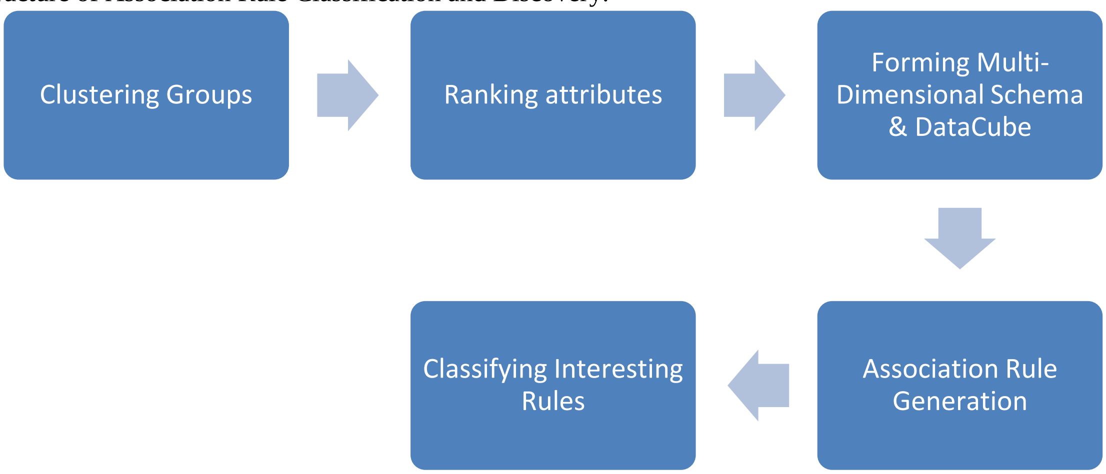


Dr. K.Kavitha, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 2, February- 2016, pg. 01-08 ISSN: 2321-8363

Association Rule mining is a descriptive approach to explore data which helps to identify relationships among values in a database.

3.6 Association Rule Discovery Schema

The association rule is defined as $P \rightarrow Q$ where P and Q are disjoint and frequent items in the dataset respectively. Most frequent rules are extracted based on user specified constraints such as threshold limt. It finds the interesting relationships and association among the attribute sets present in the dataset. The following figure represents the structure of Association Rule Classification and Discovery.



3.7 Interested Rules Classification

Discovered Rules are classified into highly instructed, medium and low interrupted rules. To classify the rules, mean value is fixed based on the highest ranked interested rules. The proposed algorithm classifies and predictions both certain and uncertain date. Accordingly it combines the uncertain data into rule based mining.

Here the calculated Rule(CR) is gathered as input and it is classified based on the importance value. The interested rule are classified as low interested (LI), medium(MI) and highly interested rule(HI). The mean value M is calculated based on the importance among the rules. Weight denotes the maximum importance value of the rule.

Input: Calculate_Rule CR, i=1,2,.......,n

Output: LI, MI, HI

Find maximum importance (weight) for CR
    Calculate M for CR
    M = weight / N
For each $r_i \in$ CR
    Get $weight_i$ for r_i
if ($weight_i \leq \mu$)
Add r_i into LI
if ($weight_i > \mu$ && $r_i \leq 2\mu$)
Add r_i into MI
if ($weight_i > 2\mu$)
Add r_i into HI
End

Dr. K.Kavitha, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 2, February- 2016, pg. 01-08 ISSN: 2321-8363

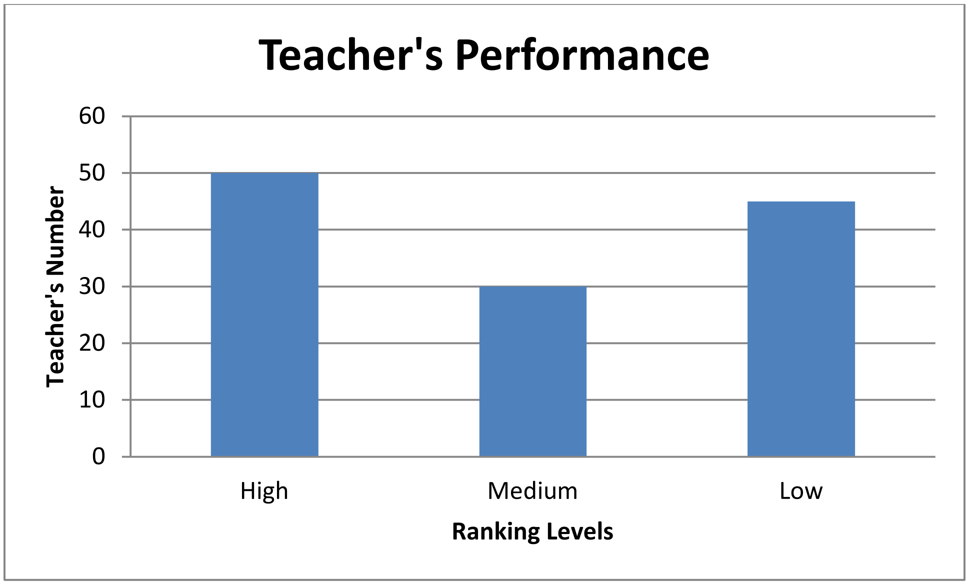
The algorithm categorise the given rule CR into three sets LI, MI and HI. The following table classifiers the interested rules with their importance value as LI, MI and HI.

Rules	Importance	Class
Academic Group= Group2Category_type Group = Group2 → Marital Status=married	0.996	Highly Interested Rules
Academic Group= Group2, Category type Group Group1 → Marital Status=Unmarried	0.978	Highly Interested Rules
Academic Group= Group2, Category type Group Group2 → Marital Status=Unmarried	0.876	Medium Interested Rules
Academic Group= Group2, Category type Group Group1 → Marital Status=Unmarried	0.788	Medium Interested Rules
Academic Group= Group1, Category type Group Group1 → Marital Status=Unmarried	0.531	Low Interested Rules

3.8 Performance Evaluation

To evaluate the effectiveness of the classifying interested rules, performance evaluation is carried out based on highly ranked Interesting rules. Experimental datasets are grated with 100 applicants with their personal academic qualification and experience details. The applicants are segmented based on subject-type and experience. Totally it consists of attributes among attributes, attributes are selected by pre-processing method.

Institution proposes different rules for teachers recruitment based on subject type highly ranked MR & LR Interested rules are classified based on threshold limit. Based on percentage of rank, teachers are classified as high, medium and low as shown in figure.



Finally, teachers are selected based on high and medium ranked attributes. Here low ranked rules are eliminated and highly medium are considered for further proceedings.

4. Conclusion

The Proposed approach generates multidimensional schema and classifies the interesting Rules. This work supports the institution to rank the applied teachers performance and helps to identify the best teachers. This method extracts more information association rules. Accuracy prediction is generated an interesting rules by multi-dimensional schema. It is proven that the proposed work provides better accuracy than the existing system.

References

[1] M.Usman, R.Pears and A.Fong, "Discovering diverse Association Rules from Multidimensional Schema,",2013

[2] R.Pears,M.Usman and A.Fong, "Data guided approach to generate Multi0Dimensional schema for targeted Knowledge discovery", 2012.

[3] G.Liu, H.Jiang, R.Geng and H.Li, "Application of Multidimensional association rules in personal financial services" in Computer design and Applications(ICCDA), 2010 International Conference ,pp V5-500-503

[4] W.Y.Chiang "To mine association rules of Customer values via data mining procedure with improved model:An empirical case study'Expert Systems with Applications, Vol 38, pp 1716-1722,2011

[5] T.Herawan and M.M.Deris, "A soft set approach for association rule mining Knowledge based Systems", vol 24, pp 186-195,2011.

[6] G.Francesca, "A Discrete-Time Hazard Model for Loans: Some evidence from Italian Banking system", American Journal of Applied Sciences, vol. 9, p. 1337, 2012

[7] D.Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation", Information Technology and Control, Vol. 36, pp. 98-102, 2007.

[8] M.L.Bhasin, "Data Mining: A Competitive tool in the Banking and Retail Industries" Banking and Finance, vol 588,2006

[9] B. Bodla and R. Verma, "Credit Risk Management Framework at Banks in India," ICFAI Journal of Bank Management, Feb2009, vol.8, pp. 47-72, 2009.

[10] R. Raghavan, "Risk Management in Banks," CHARTERED ACCOUNTANT-NEW DELHI-, vol. 51, pp.

841-851, 2003.

[11] M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi, and C. S. Pattichis, "Assessment of the risk factors of coronary heart events based on data mining with decision trees," Information Technology in Biomedicine , IEEE Transactions on, vol. 14, pp. 559-566, 2010.

[12] M. Anbarasi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," International Journal of Engineering Science and Technology, vol. 2, pp. 5370-5376, 2010.

[13] M. Du, S. M. Wang, and G. Gong, "Research on decision tree algorithm based on information entropy," Advanced Materials Research,vol. 267, pp. 732-737, 2011.

[14] X. Liu and X. Zhu, "Study on the Evaluation System of Individual Credit Risk in commercial banks based on data mining," in Communication Systems, Networks and Applications (ICCSNA), 2010 Second International Conference on, 2010, pp. 308-311.

[15] B. Azhagusundari and A. S. Thanamani, "Feature selection based on information gain," International Journal of InnovativeTechnology and Exploring Engineering (IJITEE) ISSN, pp. 2278-3075.

[16] M. Lopez, J. Luna, C. Romero, and S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums," Educational Data Mining Proceedings, 2012.