

# Pertaining the Concept of Risk Evaluation and Prediction for Multi-Dimensional Clustering

Dr. K.Kavitha

Assistant Professor, Department of Computer Science,  
Mother Teresa Women's University, Kodaikanal

## Abstract

Data mining technique has some major improvement in the field of knowledge discovery clustering is an important techniques to group the similar items without advance knowledge. Risk assessment is an important task of passport sanction through E\_corner. Risk evaluation process is used to identify the applicant's behaviour. Risk evaluation and prediction is done based on decision making approach. This method allows the user to generate the risk percentage can be sanctioned or not. This paper concentrates mainly the concept of multi-dimensional data clustering for Risk evaluation and prediction.

**Keyword:** Risk Evaluation, Prediction, Association Rule, Interesting Measures, Feature Extraction

## 1. INTRODUCTION

Data mining discovers useful and interested patterns from huge volume of data. Where the data can be stored in database, data warehouse and other information repositioned. It comprises integration techniques from various disciplines. Such as data visualization database technology, Information retrieval high performance computing machine learning and pattern reposition classification of multi-dimensional data is the most challenging area in Data mining. In classification problem, each object is defined by its attribute values in multi-dimensional space. The existing system consider the data analysis might identify the set of candidate data cubes for exploratory analysis based on domain knowledge.

Rest of this paper is structured as follows. Research works related to Risk Assessment are discussed in Section2. Detailed explanations of implementing project grants using Prediction Algorithm presented in section 3.Experimental results are reported in Section4 to prove the accuracy and finally section 5 concludes this research work.

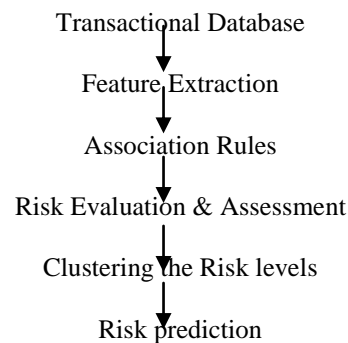
## 2. RELATED WORK

Credit Risk evaluation is an interesting management problem in financial structure. Francesca.et.al proposed hazard model for predicting loan population which involves different probability of risk factors. Probability is modelled into two groups such as good and bad borrowers [1].

Zakrzewska.et.al proposed a technique allows building of different rules for different type of customers; each applicant is assigned to similar group of concerned part [2]. Bhasin.et.al presented a model to extract efficient information from existing data and enables the system to make better decisions [3].

## 3. RISK EVALUATION AND PREDICTION FOR MULTI-DIMENSIONAL CLUSTERING DATA

Risk assessment is one of the existing problems in passport sector. The decision for the passport sanction to an applicant should be evaluated properly so that it may not lead to criminal activities in our country. ERPCA method aids the passport sector to make the evaluation for passport sanction in an enhanced manner. The overall flow of the proposed work is shown in figure.

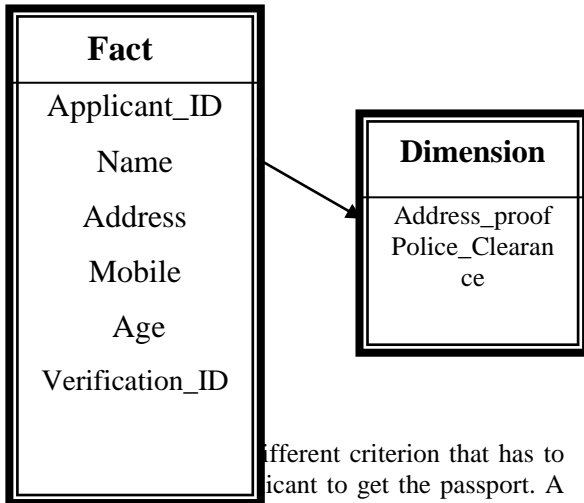


Risk Assessment is done in two levels such as primary and secondary and finally the applicants are clustered based on the prediction as good (or) bad applicant. Here, each applicant who needs a passport has to provide their professional details to the office. These details are stored in the database for further access. The applicant details are to be prepared in such a manner suitable for data mining. The dataset contains attributes like age, employment status, qualification etc., the details of the applicants are collected in the database and the segmented based on applicant then the valuable attributes are selected using feature extraction.

### 3.1 Feature Extraction for Passport Dataset

Data sets for analysis may contain many attributes which contains irrelevant data to the mining. Though it is possible for a domain expert to pick out the essential attribute it may consume time and

make the task more difficult. Keeping of irrelevant attributes leads to confusion in the mining process and also increases the data size. Thus the attributes here to be reduced to decreases the size. The main goal of this process is to find a minimum set of attributes. For reducing computational complexity, here feature extraction is done by calculating information gain.



different criterion that has to be significant to get the passport. A part from the rules created earlier, new rules can also be introduced in the method. To receive a passport, the applicant has to satisfy particular touch stones like correct address proof, police clearance etc., The appreciates of these attributes are altered based on the criteria's and user information enforced by the applicant. Rule list is framed by the sector as shown in the table

Cust - Id	Age	Correct Address proof	Qualification	Police Clearance
1	30	Yes	PG	Yes
2	35	Yes	UG	No
3	25	No	PG	Yes
4	20	Yes	UG	No
5	40	No	SSLC	yes

**3.3 Risk Assessment and Evaluation**

In order to access a passport, officer should be capable of measuring the risk attached to the information. Risk assessment is done by measuring certain attributes. Risks are separated into the categories. Such as Level I and Level II type risk. They can be considered as primary and secondary risk. The primary risk is calculated by considering three attributes such as criminal clearance, Address -Proof-Correction. Secondary risk is calculated by using the attribute age & qualification. These attributes are selected based on the values obtained from Infogain. Based on the predicated rules, values are assigned for the attributes.

**3.4 Clustering Technique**

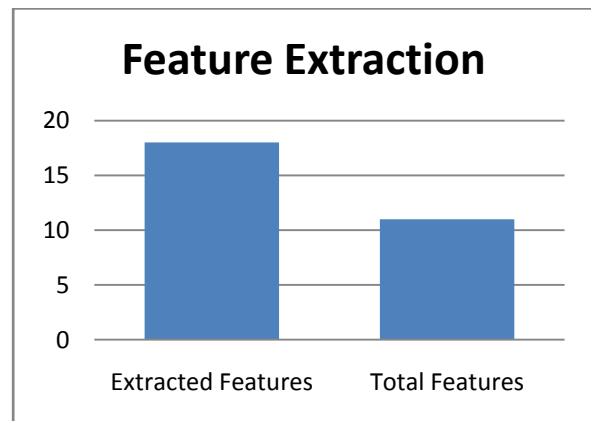
It groups the set of objects and finds whether there is some relationship between the objects. In this clustering group of data elements can be long to more than one cluster which is associated with each element is a set of membership levels. The variables are classified as Low, medium and high. Based on these three vectors, the data's are clustered.

**3.5 Risk prediction**

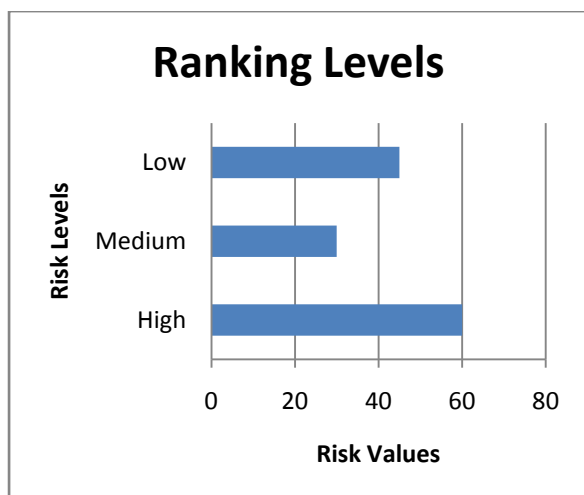
For passport sanction, threshold value of risk is set by the user. Risk is predicted based on this value (ie) officer can decide whether to sanction the passport or not(i.e) If the percentage of risk for an applicant is greater than the threshold value, the application is rejected otherwise loan sanctioned. Approval and Rejection list are classified based on threshold value.

**4. EXPERIMENTAL RESULT**

To evaluate the effectiveness of the Risk evaluation techniques series of experiment is percolated by performance validation. To start off with the method, the experimental dataset are generated with 1000 applications. Personal information a collected from users and verification information are collected from corresponding department. This information is used to estimate and measure how a random variable is able to describe and impact on other variable. Figure depicts the number of features present before and after extraction process.



After feature extraction, rules are predicted based on primary and secondary levels, percentage of risk is calculated. Based on the percentage value for an applicant is classified as Low, Medium and high as shown in figure



### CONCLUSION

Risk assessment is an important task in the passport sector. This paper performs the risk evaluation of multi-dimensional data dependency on an efficient risk prediction. Risk assessment is performed in two levels such as primary & secondary for enhancing the accuracy of risk computation. The method identifies the risk percentage to determine whether the passport can be sanctioned to an applicant or not. The clustering algorithm categorizes the risk levels as low medium and high dependency on the percentage of acquired risk values.

### REFERENCES

- [1] 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
- [2]D.Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation", Information Technology and Control, Vol. 36, pp. 98-102, 2007.
- [3]M.L.Bhasin, "Data Mining: A Competitive tool in the Banking and Retail Industries" Banking and Finance, vol 588,2006
- [4] M.Usman, R.Pears and A.Fong, "Discovering diverse Association Rules from Multidimensional Schema,"2013
- [5] R.Pears,M.Usman and A.Fong, "Data guided approach to generate Multi0Dimensional schema for targeted Knowledge discovery", 2012.
- [6] 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
- [7] 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
- [8] T.Herawan and M.M.Deris, "A soft set approach for association rule mining Knowledge based Systems", vol 24, pp 186-195,2011.
- [9]A. Ghatge and P. Halkarnikar, "Ensemble Neural Network Strategy for Predicting Credit Default Evaluation."
- [10]S. Purohit and A. Kulkarni, "Credit evaluation model of loan proposals for Indian Banks," in Information and Communication Technologies (WICT), 2011 World Congress on, 2011, pp. 868-873.
- [11]D. Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation,"

- Information Technology and Control, vol. 36, pp. 98-102, 2007.
- [12]M. L. Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries," Banking and finance, vol. 588, 2006.
- [13]N. İkizler and H. A. Guvenir, "Mining interesting rules in bank loans data," in Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.
- [14]J. Nassali, "A Loan Assessment System for Centenary Rural Development Bank," 2005.
- [15]T. Jacobson and K. Roszbach, "Bank lending policy, credit scoring and value-at-risk," Journal of banking & finance, vol. 27, pp. 615-633, 2003.
- [16]G. Kabir, I. Jahan, M. H. Chisty, and M. A. A. Hasin, "Credit Risk Assessment and Evaluation System for Industrial Project."
- [17]B. Bodla and R. Verma, "Credit Risk Management Framework at Banks in India," ICAFI Journal of Bank Management, Feb2009, vol. 8, pp. 47-72, 2009.
- [18]R. Raghavan, "Risk Management in Banks," CHARTERED ACCOUNTANT-NEW DELHI-, vol. 51, pp. 841-851, 2003.
- [19]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.