

# Predicting Risk for Evaluating Eligible Students using Multi-Dimensional Dataset

**Dr. K.Kavitha**

*Assistant Professor*

*Department of Computer Science*

*Mother Teresa Women's University, Kodaikanal*

*Kavitha.urc@gmail.com*

## Abstract

Data Mining extracts required information from large dataset. The Key Concept of Data mining is to classify the desired data according to posterior probability. It utilizes data classification and prediction technique generating students rank list for college admission is an important risk in all the institution eligible students for particular discipline is evaluated based on specified constraints here risk evaluation and prediction is performed which allows the user to generate the risk predicting and determine whether seats can be granted or not. This paper concentrates mainly the concept of multidimensional data classification and clustering for predicting the risks.

**Keywords:** Association risk, prediction, clustering, classification, Multi dimensional, Feature Extraction.

## Introduction

Due to high contention in an educational field, the Institution has to determine and select the best students based on the rank to upgrade institutional the education status. Management analyser the bulk data and classify them based on cut off marks and category. Data mining provides many technologies to study and discover the interesting figure. This technique classifier the student's dataset according to their posterior ability. Here Data mining offers classification and prediction method to improve the task.

Seat allotment decisions are crucial task in the institution. Due to high risk, association with improper grant decisions may reflect in poor results and reduce the university rank numbers.

To limit the risk, screening the student's educational background. Before selecting the students, Institution has to take various steps such as performance of the students by determining the educational background along with category. Sometime with bulk data and technology slackness, decision may be wrong and improper selection failure. Here data mining offers clustering technique which applies intelligent information for providing tools with the required data to reduce uncertainty.

Student eligibility is defined by statistical method to predict the probability which helps to select whether desired seat allotted or not. The objective of students rank list generation manages the risk involved in providing selection risks. So that the system makes better decision quickly. Eligibility score helps to increase the speed and consistency of the student's 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 section 4.

## 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].

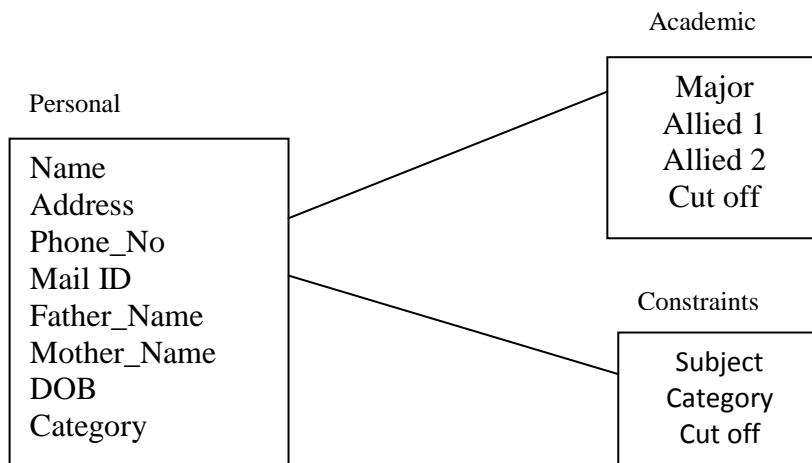
Herawan.et.al presented an approach for regular and maximal association rules from transactional datasets based on soft set theory. Here Transactional datas are transformed into Boolean values based on information system[5].

### Proposed Risk Assessment Algorithm

Risk Assessment is an existing problem in the rank list generation for eligible students. The decision for student's eligibility should be evaluated properly so that it may not lead problems. Here ERPCA method aids to make the evaluation and generate the report in an enhanced manager. The overall flow structure of risk assessment is



Here, each student who needs particular discipline has to provide their personal and educational details in the institution. The corresponding details are entered in the database for further processing. The database contains attributes like age, major, allied1, allied2, cut\_off, category, discipline etc.



The details of the students are collected and then segmented based on academic score and category then the valuable attributes are selected using feature selection concepts.

#### (i) Feature Extraction

The main aim of this concept is to find the minimum attribute set for understanding the data and reduces the complexity. Here irrelevant data can be removed to reduce the dataset size. It measures the uncertainty of collected data when the outcome is not known. Information gain is used to select the best criterion based on decision tree. In Student database primary and secondary data's are evaluated and eliminate the secondary data such as Father \_Name, Mother \_Name, DOB etc.,

### Rule Formation

Every institution has different standards that have to be satisfied by the students to opt the desired field. To receive the discipline, student admission has to satisfy particular constraints like cut off mark and category. The following algorithm is used to evaluate the eligibility of the student based on Risk.

### Algorithm

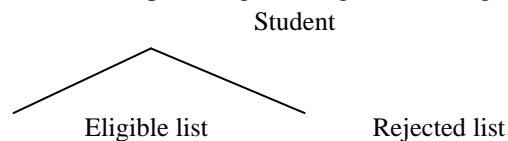
1. Set the criteria for each subject
2. Extract the required attributes
3. Clustering the attributes based on criteria similarity
4. Generate Multi-Dimensional selection formation
5. Evaluate risk lists for each students based on major & Allied Marks.
6. Apply Association Rule mining and identifies interpreting m
7. Classification technology applied to separate the student category such as eligible and reject.

### Risk evaluation

Risk Evaluation is performed by measuring certain attributes based on risk categories. Primary risk consider cut off mark and category and secondary risk focuses constrained relation. Based on the predicated rules, specific values are assigned to the attributes for selection and convert it into binary format.

### Clustering

This method groups the set of objects and finds whether there is a relationship between an object. It shows the intensity of the relationship between information elements. Here risk percentage is categorized as eligible and rejected list.



### Rule Prediction

For eligible students, threshold value is set. Risk is predicted based on the specified threshold limit. If the risk percentage is greater than the threshold limit, then the students are rejected otherwise accepted for admission.

### Conclusion

Risk Assessment is an important task in educational institution. The proposed method performs the risk evaluation of multi-dimensional data clustering. By using this method, eligible students are assessed easily based on the risk percentage and also enhance the accuracy of the computation. This method easily identifies the percentage of risk to determine whether the students are eligible or not for admission.

### 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 (ICCD), 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 Innovative Technology 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.