ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

GENERATING STUDENT'S ELIGIBILITY RANK LIST BASED ON RISK ASSESSMENT USING CLUSTERING TECHNIQUES

Dr. K.Kavitha
Assistant Professor
Department of Computer Science
Mother Teresa Women's University, Kodaikanal

Abstract

Decision making and clustering are the important task to make the better decisions accordingly. Risk Assessment and evaluation is a significant task of institutions those provide college seats to students. The students eligibility list generation depends largely on institution criteria to determine the risk effectively. This paper focuses the key problem in the educational institution such as eligible and rejected list. Rule prediction method is applied to evaluate the risk of granting seats in the desired subjects. The proposed method allows the users for finding the risk percentage to determine whether students eligible to admit or not. The proposed idea is tested by extracted data and results proven that it provides better accuracy.

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

Introduction

Data mining provides many technologies and tools to detect the hidden patterns and predict interesting rules. In educational field, selecting eligible students for admission are tedious process to decide. Every institution has its own criteria for each discipline. Based in the criteria, student's eligibility has to be generated. But it is very difficult to identity the best students. So Personal Information are collected along with academic details from students. The key problem is to extract the information and data classification based on criteria. Data mining offers classification, rule generation and prediction method to perform the required task.

Risk evaluation of seats grant is essential for an institution because inappropriate decision may produce great losses in an institution because if the problem leads in the institution then affects the admission numbers.

To determine this risk type, students have to be sorted based on academic proofing and category, institution has to decide and allot the seats which meet their standards. Clustering algorithm provides some extra tool to thin out the outcome of the product for quality enhancement. The objective of seat allotment is to appraise the risk so that the process makes better decision making. It serves to improve the 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 section4.

ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

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 generate 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 shown in Fig. (1)

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.

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 datas 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.

ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

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.

Experimental Result

To assess the effectiveness of the risk evaluation, performance evaluation is carried out based on risk percentage. The experimental data set is generated with 500 applicants with their personal and academic details. It consists of 15 Attributes. The applicants are segmented based on desired subject type. Next, feature selection is performed to extract the attributes which is used to eliminate the redundancy and identify the irrelevant dates. Among 15 attributes, 3 attributes are selected by feature extraction process. Rules are predicated after feature extraction process. Institution proposes different rules for students based on subject type. Then risk assessment is accessed based on primary and secondary attributes risk percentage is calculated based on their attributes. Students are classified an eligible and rejected as shown in Fig. 2.

Finally, students are separated based on constraints relation such as eligible among particular subject as shown in Fig. 3. Here rejected students are eliminated and eligible students are considered for further proceedings.

Conclusion

Risk assessment is a crucial task in the educational industry. This report offered a theoretical account for risk evaluation where the bulk quantity of data is engendered and risk appraisal is practiced based on data mining techniques. Rule prediction is performed in evaluating the student's eligibility rank list. Each Risk levels are measured based on specific standards and properties are determined eligible and rejected students list are judged based on the risk percentage which is compared with specified threshold value.

ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

Tables and Figures

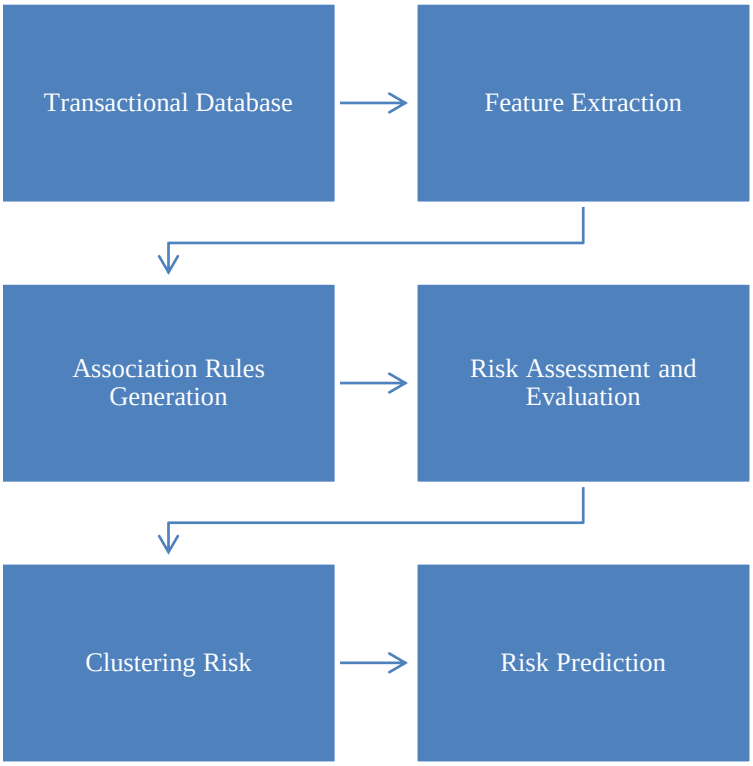

Fig. 1 Flow structure of risk assessment

Students	Eligible	Rejected
500	350	150

Student's category

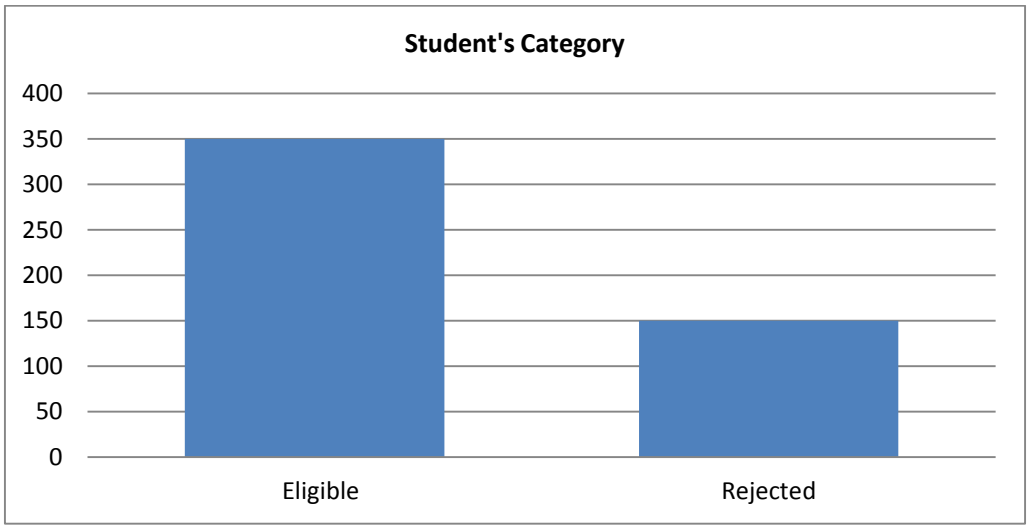

Fig. 2 Classification of eligible and rejected students

ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

Subject	Eligible	Reject
Tamil	50	10
English	75	25
Maths	100	20
Comp.sci	150	50
Physics	30	20

Subject wise risk status

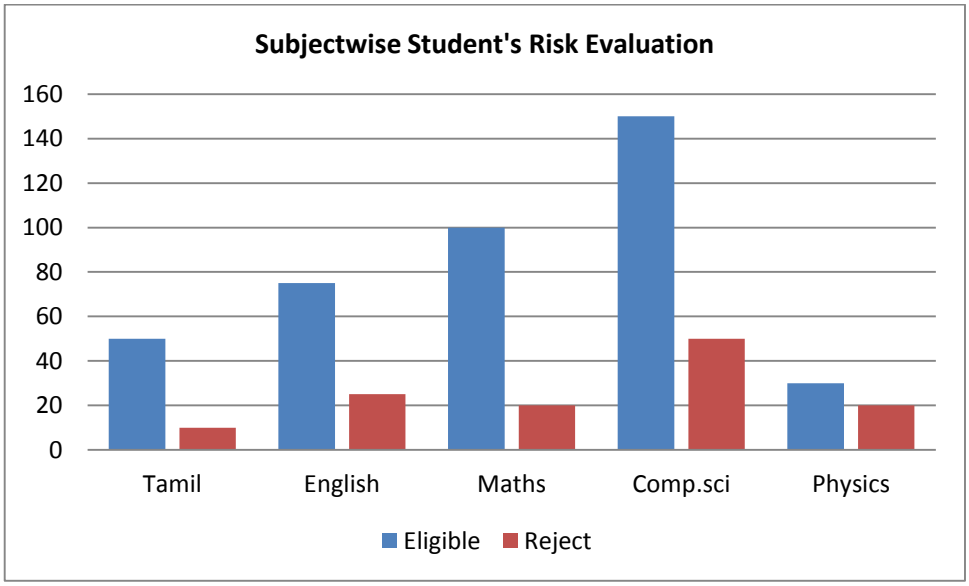

Fig. 3 Separation of students based on constraints

References

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

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

III. 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

IV. 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

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

VI. 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

ISSN (online): 2250-141X
Vol. 5 Issue 4, December 2015

VII. D.Zakrzewska, “On integrating unsupervised and supervised classification for credit risk evaluation”, Information Technology and Control, Vol. 36, pp. 98-102, 2007.

VIII. M.L.Bhasin, “Data Mining: A Competitive tool in the Banking and Retail Industries” Banking and Finance, vol 588,2006

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

X. R. Raghavan, "Risk Management in
a. Banks," *CHARTERED ACCOUNTANT-NEW DELHI-* , vol. 51, pp. 841-851, 2003.

XI. 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*
a. *Biomedicine , IEEE Transactions on,* vol. 14, pp. 559-566, 2010.

XII. M. Anbarasi, E. Anupriya, and N.
a. 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.

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

XIV. X. Liu and X. Zhu, "Study on the
a. 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.

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

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