

Comparison of Different Classification Techniques for Educational Data

Kavita Pabreja, Department of Computer Science, Maharaja Surajmal Institute (an affiliate of GGSIP University), New Delhi, India

ABSTRACT

Data mining has been used extensively in various domains of application for prediction or classification. Data mining improves the productivity of its analysts tremendously by transforming their voluminous, unmanageable and prone to ignorable information into usable pieces of knowledge and has witnessed a great acceptance in scientific, bioinformatics and business domains. However, for education field there is still a lot to be done, especially there is plentiful research to be done as far as Indian Universities are concerned. Educational Data Mining is a promising discipline, concerned with developing techniques for exploring the unique types of educational data and using those techniques to better understand students' strengths and weaknesses. In this paper, the educational database of students undergoing higher education has been mined and various classification techniques have been compared so as to investigate the students' placement in software organizations, using real data from the students of a Delhi state university's affiliates.

KEYWORDS

Bayesian Network, Classification, Decision Tree, Educational Data Mining, Efficiency in Academics, Prediction, Students' Databases, WEKA

INTRODUCTION

In India, Universities are operating in a very complex and highly competitive environment. With numerous courses on offer in the universities, the students have a great choice available to them. The main challenge for these universities is to appropriately understand their strengths, analyze their performance, identify their uniqueness, and build a strategy for further development as per the dynamic expectations and requirements of the students and the industry. Kabakchieva (2013) stated that University management should focus on the profile of admitted students, their pre-university characteristics and the data generated as a result of the association of the students with them.

Data mining offers promising results for classification of students based on different objectives of a university as explained by Romero & Ventura (2007). According to various studies, there are different educational objectives for using classification, such as: to discover potential student groups with similar characteristics and reactions to a particular educational strategy as per Chen, Liu, Ou, & Liu (2000), to detect students' misuse of the learning system as mentioned by Romero, Ventura, Espejo, & Hervás (2008), to group students who are hint-driven or failure-driven and find common misconceptions that students possess (Yudelson, Medvedeva, Legowski, Castine, Jukic, & Rebecca, 2006). Other objectives that may be of interest to a university are identification of learners with low impetus and to find remedial actions to improve results as suggested by Cocea & Weibelzahl (2006), to predict/classify students while using intelligent tutoring systems as explained by Hämmäläinen & Vinni (2006), etc.

The specific objective of the proposed research work is to find out if there are any patterns in the students' data that could be useful for predicting the scope of his/her employability as employment plays the most important role for any student pursuing a professional programme. The predictions for employment are based on the students' personal, pre-university and operational characteristics while pursuing a course. An attempt has been made to classify students into employed or unemployed class followed by a comparison of different data mining techniques for classification. This study is the first attempt of its kind as far as any state university of India is concerned. The paper is arranged as follows: first section focuses on the related work done, followed by section on description of the research methodology (data collection and its pre-processing). Next section explains and compares the results of the various classification techniques applied, finally, the discussions and conclusions have been mentioned.

LITERATURE REVIEW

Educational Data Mining (EDM) is a relatively new stream in the data mining research and there are only a few studies by researchers in this field. Extensive literature reviews of the EDM research field have been done by Romero & Ventura (2007), covering the research efforts in the area, between 1995 and 2005, and by Baker & Yacef (2009), for the period after 2005. Romero & Ventura surveyed the application of data mining to traditional educational systems, particularly web-based courses, well-known learning content management systems, and adaptive and intelligent web-based educational systems. They have applied data mining techniques *viz.* statistics and visualization; clustering, classification and outlier detection; association rule mining; pattern mining and text mining.

The potential applications of data mining in higher education have been explained by Luan in his work in 2002. The author has also discussed how data mining saves resources while maximizing efficiency in academics. Ma, Liu, Wong, Yu, & Lee (2000) focussed on understanding student types and then opting for targeted marketing by applying different data mining models. Tair & El-Halees (2012) have used educational data mining to improve graduate students' performance, and to overcome the problem of low grades of graduate students. The authors have tried to extract useful knowledge from graduate students data collected from the college of Science and Technology for a period of fifteen years [1993-2007].

Yadav, Bharadwaj, & Pal (2012) used decision tree classifiers to find the best classifier for retention data to predict the student's drop-out possibility. Classification technique has been applied on student database to predict the students' division on the basis of previous database by Bharadwaj & Pal (2011). Priya & Senthil (2013) also applied similar techniques to improve the students' performance and help, to achieve the goal by extracting the discovery of knowledge from the end semester marks. The application of K-means clustering algorithm to predict students' result has been demonstrated by Bhise, Thorat, & Supekar (2013).

Kumar & Chadha (2012) used association rule mining in enhancing the quality of students' performances at Post Graduation level. Prediction of grades to classify them in five classes: A, B, C, D and E or F, from test scores has been done using neural networks by Fausett & Elwasif (1994).

There are different types of classification methods and artificial intelligent algorithms that have been applied to predict student academic score. The prediction of academic success (classes that are successful or not) has been done using discriminant function analysis in a study by Martínez (2001); classification of students to predict their final grade has been done using genetic algorithms by Minaei-Bidgoli & Punch (2003); prediction of a student's academic success (to classify as low, medium and high risk classes) has been explained using different data mining methods by Superby,

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/comparison-of-different-classification-techniques-for-educational-data/165419

Related Content

Building Consumer Trust for Internet E-Commerce

George Yee (2007). *Trust in E-Services: Technologies, Practices and Challenges* (pp. 218-234).

www.irma-international.org/chapter/building-consumer-trust-internet-commerce/30459

Toward Cloud Federation: Concepts and Challenges

Antonio Celesti, Francesco Tusa and Massimo Villari (2012). *Achieving Federated and Self-Manageable Cloud Infrastructures: Theory and Practice* (pp. 1-17).

www.irma-international.org/chapter/toward-cloud-federation/66224

Improvement of Measurement Contribution for Asset Characterization in Complex Engineering Systems by an Iterative Methodology

Giulio D'Emilia, Antonella Gaspari and Diego Pascual Galar (2018). *International Journal of Service Science, Management, Engineering, and Technology* (pp. 85-103).

www.irma-international.org/article/improvement-of-measurement-contribution-for-asset-characterization-in-complex-engineering-systems-by-an-iterative-methodology/201474

The Challenge of Service Level Scalability for the Cloud

Luis M. Vaquero, Juan Cáceres and Daniel Morán (2011). *International Journal of Cloud Applications and Computing* (pp. 34-44).

www.irma-international.org/article/challenge-service-level-scalability-cloud/53141

Information Needs of Logistics Service Providers

Ulla Tapaninen, Hennariina Pulli and Antti Posti (2010). *Service Science and Logistics Informatics: Innovative Perspectives* (pp. 113-127).

www.irma-international.org/chapter/information-needs-logistics-service-providers/42638