Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue

Pooja Thakar
Assistant Professor
VIPS, GGSIPU
Delhi, India

Anil Mehta, Ph.D
Professor
University of Rajasthan
Jaipur, India

Manisha, Ph.D
Associate Professor
Banasthali University
Jaipur, India

ABSTRACT

In this era of computerization, education has also revamped itself and is not limited to old lecture method. The regular quest is on to find out new ways to make it more effective and efficient for students. Nowadays, lots of data is collected in educational databases, but it remains unutilized. In order to get required benefits from such a big data, powerful tools are required. Data mining is an emerging powerful tool for analysis and prediction. It is successfully applied in the area of fraud detection, advertising, marketing, loan assessment and prediction. But, it is in nascent stage in the field of education. Considerable amount of work is done in this direction, but still there are many untouched areas. Moreover, there is no unified approach among these researches. This paper presents a comprehensive survey, a travelogue (2002-2014) towards educational data mining and its scope in future.

General Terms

Data Mining, Education

Keywords

Educational Data Mining (EDM)

1. INTRODUCTION

In last decade, the number of higher education universities/institutions have proliferated manifolds. Large numbers of graduates/post graduates are produced by them every year. Universities/Institutes may follow best of the pedagogies; but still they face the problem of dropout students, low achievers and unemployed students.

Understanding and analyzing the factors for poor performance is a complex and incessant process hidden in past and present information congregated from academic performance and students' behavior. Powerful tools are required to analyze and predict the performance of students scientifically.

Although, universities/institutions collect an enormous amount of students' data, but this data remains unutilized and does not help in any decisions or policy making to improve the performance of students.

If, Universities could identify the factors for low performance earlier and is able to predict students' behavior, this knowledge can help them in taking pro-active actions, so as to improve the performance of such students. It will be a win-win situation for all the stakeholders of universities/institutions i.e. management, teachers, students and parents. Students will be able to identify their weaknesses beforehand and can improve themselves. Teachers will be able to plan their lectures as per the need of students and can provide better guidance to such students. Parents will be reassured of their ward performance in such institutes. Management can bring in better policies and strategies to enhance the performance of these students with additional facilities. Eventually, this will help in producing skillful workforce and hence sustainable growth for the country.

Analysis and prediction with the help of data mining techniques have shown noteworthy results in the area of fraud detection, predicting customer behavior, financial market, loan assessment, bankruptcy prediction, real-estate assessment and intrusion detection. It can be very effective in Education System as well. It is a very powerful tool to reveal hidden patterns and precious knowledge, which otherwise may not be identified and difficult to find and comprehend with the help of statistical methods.

Substantial work is done towards the usage of data mining techniques in Education, but still there are many untouched areas and no unified approach is followed. This paper presents a comprehensive literature review of relevant researches done in last decade from year 2002 to 2014.

This paper is divided into following sections. Second section presents the details of researches done in the area of education in tabular form describing methodologies and findings of each research. Third section summarizes these researches. Fourth section concludes and identifies the areas; where more work is required; hence describes the future scope.

2. COMREHENSIVE REVIEW OF LITERATURE

A comprehensive literature review of various significant researches in the area of Educational Data Mining ranging from Year 2002 to 2014 is presented below in a categorized tabular form (Table 1).

These researches can be broadly classified into five areas:

2.1 Survey of papers published in Educational Data Mining.

2.2 Predicting Academic Performance with Pre/Post Enrollment Factors.

2.3 Comparison of Data Mining Techniques in predicting academic performance.

2.4 Correlation among Pre/Post Enrollment Factors and Employability.

2.5 Other areas of Education.

Table 1

Category	Year, Author(s)	Methodology	Key Findings
Survey of papers published in Educational Data Mining	2014, Peña-Ayala, Alejandro	Statistical and Clustering Processes	Identified kinds of educational systems, disciplines, tasks, methods, and algorithms.
	2010, Romero, Cristóbal, and Sebastián Ventura		Listed tasks in educational area resolved through data mining and future lines. Suggested to develop more unified and collaborative studies.
	2009, Baker, Ryan SJD, and Kalina Yacef		Identified key features of researches in EDM as discovery with models, emergence of public data, tools.
	2007, Romero, Cristóbal, and Sebastian Ventura		Presented survey on application of data mining on traditional educational systems. Emphasized on the need of much more specialized work.
Predicting Academic Performance with Pre/Post Enrollment Factors	2014, Saranya, S., R. Ayyappan, and N. Kumar	Naive Bayes Algorithm	Graphically represented Institutional Growth Prognosis and Students' Progress Analysis.
	2014, Archer, Elizabeth, Yuraisha Bianca Chetty, and Paul Prinsloo	Experimental Usage of Employee Profiling Software	Experimented the usage of a commercial product generally used for employee profiling in corporate, for higher education environment.
	2014, Hicheur Cairns, Awatef, et al.	Clustering Technique	Professionals' data was analysed during training of a consulting company.
	2014, Arora, Rakesh and Dharmendra Badal	Association Analysis Algorithm	Found set of weak students based on graduation and post graduation marks.
	2012, Osmanbegović, Edin, and Mirza Suljić	Chi-Square Test, One R-Test, Info Gain and Ratio Test, Naive Bayes, DTree	Found predicting model for academic performance that is user friendly for professors or non-expert users.
	2012, Sukanya, M., S. Biruntha, Dr S. Karthik, and T. Kalaikumaran	Bayesian Classification Method	Analysed and assisted the low academic achievers in higher education.
	2011, Torenbeek, M., E. P. W. A. Jansen, and W. H. A. Hofman	Structural Equations Modeling, Correlation Matrix	Examined two variables, Pedagogical approach and skill development in the first 10 weeks of enrollment
	2011, Yongqiang, He, and Zhang Shunli	Association Rules Analysis	Guidance provided for scientific management and comprehensive evaluation of students.
	2011, Sakurai, Yoshitaka, Tsuruta, and Rainer Knauf	Decision Tree	Estimated success chances of curricula by implementing student profiling with storyboard system.
	2011, Aher, Sunita B., and L. M. R. J. Lobo	Classification and Clustering	Analyzed the performance of final year students.
	2010, Ayesha, Shaeela, Tasleem, Ahsan, Inayat	K-Means Clustering	Analyzed students' learning behavior to check the performance of students and predicted weak students.
	2010, Kovacic, Zlatko	Classification Tree Models	Investigated enrolment attributes to pre-identify success of students.
	2010, Al-shargabi, Asma A., and Ali N. Nusari	Clustering, Association Rules and Decision Trees	Analyzed students' academic achievement, students' drop out, and students' financial behavior.
	2010, Yan, Zhi-min, Qing Shen, and Bin Shao	Rough Set Theory	Students' grades were analyzed.
	2010, Ningning, Gao	Neural Network, Rough Set Theory	Predicted drop outs from course
	2010, Knauf, Rainer, Yoshitaka Sakurai, Setsuo Tsuruta, and Kouhei Takada	Decision Tree	Analyzed successful Storyboard (e-learning system) success paths for students.
	2010, Wu, X., Zhang, H., & Zhang, H.	Decision Tree	Suggested comprehensive evaluation method of that can objectively distinguish the grades of students.
	2010, Youping, Bian Xiangjuan Gong	Decision Tree	Evaluated the high school students and studying effectiveness.
	2010, Liu, Zhiwu, and Xiuzhi Zhang	Decision Tree	Built forecasting model for students' marks to identify negative learning habits or behaviors of students.
	2009, Zhu, Li, Yanli Li, and Xiang Li	Association Rule	Predicted student's achievement systematically and improved teaching management.
	2009, Nayak, Amar, Jitendra , Vinod, Shadab	Proposed Use Of Ontology, RDF, XML	Proposed enterprise framework to identify suitable semantic data related to students, faculties and courses.
	2009, Wang, Pei-ji, Lin Shi, Jin-niu Bai, Yu-lin Zhao	Apriori Algorithm	Improved algorithm used to mine the students' data table.
	2009, Ramasubramanian, Iyakutti, and Thangavelu	Rough Set Theory	Predicted weak students.
	2008, Selmoune, Nazih, and	Association Rules,	Found the success and failure factors of students.

Zaia Alimazighi	Apriori Algorithm	
2008, Shangping, Dai, and Zhang Ping	Genetic Algorithm, Novel Spatial Mining Algorithm	Predicted final grade based on features extracted from log data in web-based system.
2008, Bresfelean, Paul, Mihaela, Nicolae, Comes	J48 and farthest first algorithms	Provided managerial information on understanding, predicting and preventing academic failure.
2008, Zhang, Xiaolong, and Guirong Liu	Statistical Approach, Association Rules,	Found hidden patterns for students to avoid becoming low performer ones.
2008, Villalon, Jorge J., and Rafael A. Calvo	Grammar Trees, Regular Expressions	Concept Map Mining (CMM) done from essays written by students to judge their knowledge.
2007, Hien, Nguyen Thi Ngoc, and Peter Haddawy	Bayesian Network	Concluded that socio-economic environment can play a major role in the performance of students
2006, Radaideh, Qasem, Shawakfa, Mustafa	Decision Tree, Rough Set Theory, Naïve Bayes	Produced system generated rules to predict the final grade in a course under study.
2005, Rasmani, Khairul A., and Qiang Shen	Fuzzy Approaches	Fuzzy approaches used to classify students academic performance
2005, Delavari, Naeimeh, Reza Beikzadeh, and Somnuk Phon-Amnuaisuk	Decision Tree	Described a roadmap for the application of data mining in higher education by pre- identifying weak students.
2004, Salazar, A., J. Gosalbez, I. Bosch, R. Miralles, and L. Vergara	Clustering and Decision Rule	Concluded that more variables are required for realistic analysis of academic performance.
2003, Minaei-Bidgoli, Behrouz, and William F. Punch	Genetic Algorithm	Predicted final grade of student based on features extracted from logged data in an education Web-based system.
2002, Kotsiantis, S., C. Pierrakeas, and P. Pintelas	Decision Tree, Naive Bayes Algorithm, NN	Concluded that learning algorithms could enable tutors to predict student performance long before final examination.
Comparison of Data Mining Techniques in predicting academic performance		
2011, Sharma, Mamta, and Monali Mavani	Decision Tree, Sota, Naïve Bayes	Comparison of three algorithms in terms of prediction of students result.
2009, Siraj, Fadzilah, and Mansour Ali Abdoulha	Cluster Analysis, NN, Logistic Regression and Decision Tree.	Compared three techniques for understanding undergraduate's student Enrolment data.
2009, Wook, Muslihah, Yuhanim, Norshahriah, Rizal, Isa, Nor Fatimah, and Hoo Yann Seong	Artificial Neural Network, Clustering and Decision Tree	Compared two data mining techniques for predicting and classifying students' academic performance.
2008, Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás	Statistical Classifier, Decision Tree, Rule Induction, Fuzzy, NN	Concluded that classifier model is appropriate for educational use in terms of accuracy and comprehensibility for decision making.
2007, Nghe, Nguyen Thai, Paul Janecek, and Peter Haddawy	Decision Tree and Bayesian Network	Decision tree proved to be consistently 3-12% more accurate than the Bayesian network in predicting academic performance.
Correlation among Pre/Post Enrollment Factors and Employability		
2014, Pool, Lorraine Dacre, Pamela Qualter, and Peter J. Sewell	Exploratory and Confirmatory factor analyses, t-test	Emotional Intelligence, Self-Management, Work & Life Experience are found to be important factors for Employability Development Profile.
2014, Vanhercke, Dorien, Nele De Cuyper, Ellen Peeters, and Hans De Witte	Conceptual	Described that perceived employability is tied to competences and dispositions.
2013, Potgieter, Ingrid, and Melinde Coetzee	Co-relational Statistics, Multiple Regression	Significant relationships found between the Participants' personality preferences and their employability attributes.
2013, Finch, David J., Leah K. Hamilton	Two-phase, mixed-methods study	Revealed employers place the highest importance on soft-skills and the lowest importance on academic reputation.
2013, Bakar, Noor Aieda Abu, Aida Mustapha, and Kamariah Md Nasir	Clustering Analysis Using K-Means and Expectation Maximization Algorithms	Found that across industries, graduates have average interpersonal communication, lacks in creative and critical thinking, analytical and team work.
2012, Jackson, Denise, and Elaine Chapman	online survey	Significant differences found between academic and employer skill ratings suggesting prominent skill gap between institutes and corporate.
2011, Gokuladas, V. K	Statistical : Correlation And Multiple Regressions	GPA and proficiency in English language are important predictors of employability and female students are better performers.
2010, Gokuladas, V. K	Statistical Analysis: Correlation And Regression Analysis	Concluded that Graduates need to possess specific skills beyond general academic education to be employable.

Other areas of Education	2013, Jantawan, Bangsuk, and Cheng-Fa Tsai	Bayesian Method and Decision Tree Method	Presented comparison of classification accuracy between two algorithms to evaluate employees' performance.
	2012, Dejaeger, Karel, et al	Decision Trees, NN, SVM, Logistic Regression	Determined student satisfaction for a particular course.
	2012, Srimani, P. K., and Malini M. Patil	Classifiers, Random Tree, Random Forest	Faculty evaluation based on different parameters
	2012, Yadav, Surjeet Kumar, and Saurabh Pal	Decision Tree	Assisted in selecting Students for enrollment in a particular course.
	2011, Pandey, Umesh Kumar, and Saurabh Pal	Association Rule	Concluded that mix medium class is more preferred over Hindi and English medium class.
	2010, Balakrishnan, Julie M. David	Decision Tree and Clustering	Predicted learning disabilities of school-age children.
	2009, Linjie, Qu, and Lou Lanfang	Modified Apriori Algorithm	Proposed to build evaluation index system and teaching index method based on data mining.
	2009, Ahmed, Almahdi Mohammed, Norita Md Norwawi, and Wan Hussain Wan Ishak	Association Rules	Found the patterns in matching organization and student interests, where they meet each other's requirements.
	2008, Dimokas, Nikolaos, Nikolaos Mittas, Alexandros Nanopoulos, and Lefteris Angelis	Data warehouse	Proposed data warehouse to facilitate and provide thorough analysis of department's data.
	2008, Zhao, Hua-long	OLAP, Data warehouse	Analyzed curriculum's establishment from many angles.
	2008, Pumpuang, Pathom, Anongnart Srivihok, and Prasong Praneetpolgrang	Bayesian Network, Decision Forest	Nbtree identified as the best classifiers to predict student sequences for course registration planning
	2008, Ranjan, Jayanthi, and Saani Khalil	Conceptual	Described data mining process in management education and focusing on academic aspects of admission and counseling process.
	2007, Bresfelean, Vasile Paul	Decision Trees	Analyzed students' choice in continuing their education with post University studies (Master degree, PhD)
	2006, Aksenova, Svetlana S., Du Zhang, and Meiliu Lu	Support Vector Machines and Rule-Based Approach	Proposed an approach to predict the enrollment headcount by a predictive model built from new, continued and returned students.
	2004, Talavera, Luis, and Elena Gaudioso	Naive Bayes	Described the usage of Data Mining Techniques to support evaluation of collaborative activities.

Noteworthy research papers and their findings are mentioned below in each category.

2.1 Survey of papers published in Educational Data Mining

Recent paper published in 2014 in Elsevier titled "Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works" presented the survey of published papers from 2010-2013 and divided Educational Data Mining approaches in kinds of educational systems, disciplines, tasks, methods, and algorithms. Author identified that each Educational Data Mining approaches can be organized according to six functionalities student modeling, student behavior modeling, assessment; student performance modeling; student support and feedback versus curriculum-domain knowledge-sequencing, mostly focusing on academic performance [6].

Romero and Ventura, in 2010 published a paper in IEEE, which listed most common tasks in the educational environment resolved through data mining and some of the most promising future lines. Educational Data Mining community remained focused in North America, Western Europe, and Australia/New Zealand. They mentioned that there is a considerable scope for an increase in educational data mining's scientific Influence. They also suggested developing more unified and collaborative studies [45]. In another paper by them in year 2007 titled "Educational data

mining: A survey from 1995 to 2005" surveyed the application of data mining to traditional educational systems. They concluded that much more specialized work is needed in order for educational data mining to become a mature area.[69]

Paper titled "An Empirical Study of the Applications of Data Mining Techniques in Higher Education" published in year 2011, listed potential areas in which data mining can be applied in higher education [31].

2.2 Predicting academic performance with Pre/Post Enrollment Factors

Most of the published research papers belong to this category. Latest work published in International Journal of Computer Science and Mobile Computing, 2014 describes the process of finding the set of weak students based on graduation and post graduation marks [5]. Another paper published in European Journal of Scientific Research in 2010 also analyzed students' learning behavior to predict weak students. [34]. P. Ramasubramanian, K. Iyakutti and P. Thangavelu, in year 2009 also predicted weak students using rough set theory. [54]

To assist in selecting students for enrollment in a particular course Surjeet Kumar Yadav and Saurabh Pal used Decision Trees technique of data mining in 2012 [22]. In another paper

in 2010 Zlatko J. Kovačić also investigated enrolment attributes to pre-identify success of students.[35]

In 2012, M. Sukanya, S. Biruntha, Dr.S. Karthik and T. Kalaikumaran analyzed and assisted the low academic achievers in higher education using Bayesian Classification Method of Data Mining. [21]

A comprehensive evaluation method for undergraduates; that can objectively distinguish the grades of students was developed by Xiewu, Huacheng Zhang, Huimin Zhang in year 2010 [40]. Another study by Dai Shangping, Zhang Ping, in year 2008 predicted final grades of students based on features extracted from log data in web-based system and published their work in IEEE [56]

Paper published in IEEE in 2011 titled "Success Chances Estimation of University Curricula Based on Educational History, Self-Estimated Intellectual Traits and Vocational Ambitions" integrated student profiling with storyboard system (Learning Management System) and concluded that success chance heavily depends on individual properties.[26]

Jorge J. Villalon and Rafael A. Calvo did a novel study named as Concept Map Mining (CMM) from essays written by students to judge their knowledge and published their work in IEEE, 2008 [61]

2.3 Comparison of Data Mining Techniques in predicting academic performance of students

In 2011, Springer published a paper by Mamta Sharma and Monali Mavani for comparison of three algorithms in terms of prediction of students result [32].

In 2009 Fadzilah Siraj and Mansour Ali Abdoulha compared three techniques for understanding undergraduate's student enrolment data and published their work in IEEE [47]. Nbtree was identified as the best classifiers to predict student sequences for course registration planning in paper published by Pathom Pumpuang, Anongnart Srivihok and Prasong Praneetpolgrang in year 2008 in IEEE [62].

Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás concluded that classifier model is appropriate for educational use in terms of accuracy and comprehensibility for decision making in 2008 [64].

Decision tree proved to be consistently 3-12% more accurate than the Bayesian network in predicting academic performance of undergraduate and postgraduate students in a paper titled "A Comparative Analysis of Techniques for Predicting Academic Performance" in 2007 [68], IEEE.

2.4 Correlation among Pre/Post Enrollment Factors and Employability

Recent work published in Emerald Group Publishing Limited, 2014 clearly stated that emotional intelligence, self-management, life experiences are important factors for Employability Development Profile (EDP) [1]. Another work published in Emerald Group Publishing Limited, 2014 described that employability is in strong correlation with competences and dispositions [7].

In 2014, Cairns, Gueni, Fhima, David and Khelifa analyzed employees' profiles during training of a consulting company involved in training of professionals for employability skills and found positive correlation in their jobs/assignments, history etc. [3].

In SA Journal of Industrial Psychology, 2014, authors, Potgieter & Coetzee, observed a number of significant relationships between the participants' personality and employability. Set of eight core employability attributes identified to increase the likelihood of securing and sustaining employment opportunities [8].

David J. Finch, Leah K. Hamilton, Riley Baldwin and Mark Zehner in year 2013 did 30 one-on-one interviews with hiring managers and 115 employers and revealed employers place the highest importance on soft-skills and the lowest importance on academic reputation [12]. In another online survey by Denise Jackson and Elaine Chapman in 2012 reported significant differences between academic and employer skill ratings suggesting prominent skill gap between institutes and corporate [13].

Two papers published by Wiley, authored by V. K. Gokuladas in year 2010 and 2011 respectively. In first paper he reflected that graduates need to possess specific skills beyond general academic education to be employable [33]. In next paper he showcased that GPA and proficiency in English language are important predictors of employability [24].

In 2013, Noor Aieda, Abu Bakar, Aida Mustapha, Kamariah Md. Nasir collected primary data from The Ministry Of Higher Education, Malaysia and found that across industries, graduates have average interpersonal communication skills, lacks in creative and critical thinking, problem solving, analytical skills , and team work [10]. In 2013, Bangsuk Jantawan and Cheng-Fa Tsai, made an effort to design a model for supporting the prediction of the employees' performance in an organization using data mining techniques [9].

2.5 Other areas of Education

In 2012, International Conference on Intelligent Computational Systems published a paper titled "A Classification Model For Edu-Mining" for faculty evaluation based on different parameters [20]. Another paper published in 2009 in IEEE proposed to build evaluation index system and teaching index method based on data mining [49].

International Journal of Computer Science Issues 2011, published a paper titled "A Data Mining View On Class Room Teaching Language", which concluded that mix medium class is more preferred over Hindi and English medium class [27].

Hua-Long Zhao, in his paper titled "Application of OLAP to the Analysis of the Curriculum Chosen by Students" published in IEEE in 2008, analyzed curriculum's establishment from many angles [60]. Prediction of learning disabilities of school-age children was done by Julie M. David and Kannan Balakrishnan in 2010 [43]. In 2007, Vasile Paul Brefelean analyzed students' choice in continuing their education with post University studies (Master degree, PhD) using data mining techniques [67].

Svetlana S. Aksenova, Du Zhang, and Meiliu Lu proposed an approach to predict the enrollment headcount by a predictive model built from new, continued and returned students and published their work in IEEE in 2006 [71]. In 2004, Luis Talavera and Elena Gaudioso described the usage of Data Mining Techniques to support evaluation of collaborative activities [74]. Two papers published by IEEE in 2010 predicted students' drop outs [36] [38].

3. SUMMARY

In summary, all of these researches done previously, reveals some significant areas in education field, where prediction with data mining has reaped benefits; such as finding set of weak students [5], determining student's satisfaction for a particular course [14][50], Faculty Evaluation [20], Comprehensive student evaluation [25][26][40], Class room teaching language selection [27], Predicting students' dropout [34] [36], course registration planning [62], predicting the enrollment headcount [71], evaluation of collaborative activities [74] etc.

Few researches have shown significant relationships between the participants' personality preferences and their employability attributes [8]. It is observed that there are specific skills that graduates need to possess in order to become employable and that these skills are beyond general academic education. [33]. Attributes like emotional intelligence, self-management, and Work & life experience are also important factors for Employability Development Profile (EDP) [1]. Employers also place the highest importance on soft-skills and the lowest importance on academic reputation [12]. Thus, perceived employability is tied to competences and dispositions, rather than only on academic qualification [7]. Significant differences are found between academic and employer skill ratings suggesting prominent skill gap between academia and corporate [13].

4. FUTURE PROSPECTS

One of the most recent and biggest challenge that higher education faces today is making students skillfully employable. Many universities/institutes are not in position to guide their students because of lack of information and assistance from their teaching-learning systems. To better administer and serve student population, the universities/institutions need better assessment, analysis, and prediction tools.

Considerable amount of work is done in analyzing and predicting academic performance, but all of these works are segregated. There is a clear need for unified approach. Other than academic attributes, there are large numbers of factors that play significant role in prediction, which includes non-cognitive factors (set of behaviors, skills, attitudes). Suitable data mining techniques are required to measure, monitor and infer these factors for prediction. Thus enriching the input vector with qualitative values may increase the accuracy rate of prediction as well.

Integrated Models/Frameworks are required for all the stakeholders of an Institution; hence ensuring sustainable growth for all (Management, Teachers, Students and Parents).

5. REFERENCES

[1] Pool, Lorraine Dacre, Pamela Qualter, and Peter J. Sewell. "Exploring the factor structure of the CareerEDGE employability development profile." Education+ Training 56.4 (2014): 303-313.

[2] Saranya, S., R. Ayyappan, and N. Kumar. "Student Progress Analysis and Educational Institutional Growth Prognosis Using Data Mining." International Journal Of Engineering Sciences & Research Technology, 2014

[3] Hicheur Cairns, Awatef, et al. "Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining." IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management. 2014.

[4] Archer, Elizabeth, Yuraisha Bianca Chetty, and Paul Prinsloo. "Benchmarking the habits and behaviors of successful students: A case study of academic-business collaboration." The International Review of Research in Open and Distance Learning 15.1 (2014).

[5] Arora, Rakesh Kumar, and Dharmendra Badal. "Mining Association Rules to Improve Academic Performance." (2014).

[6] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." Expert systems with applications 41.4 (2014): 1432-1462.

[7] Vanhercke, Dorien, Nele De Cuyper, Ellen Peeters, and Hans De Witte. "Defining perceived employability: a psychological approach." Personnel Review 43, no. 4 (2014): 592-605.

[8] Potgieter, Ingrid, and Melinde Coetzee. "Employability attributes and personality preferences of postgraduate business management students." SA Journal of Industrial Psychology 39.1 (2013): 01-10.

[9] Jantawan, Bangsuk, and Cheng-Fa Tsai. "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment." International Journal Of Computer Science And Information Security (2013).

[10] Bakar, Noor Aieda Abu, Aida Mustapha, and Kamariah Md Nasir. "Clustering Analysis for Empowering Skills in Graduate Employability Model." Australian Journal of Basic and Applied Sciences 7.14 (2013): 21-28.

[11] Singh, Samrat, and Vikesh Kumar. "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques." International Journal of Computer Science & Engineering Technology, (2013).

[12] Finch, David J., Leah K. Hamilton, Riley Baldwin, and Mark Zehner. "An exploratory study of factors affecting undergraduate employability." Education+ Training 55, no. 7 (2013): 681-704.

[13] Jackson, Denise, and Elaine Chapman. "Non-technical skill gaps in Australian business graduates." Education+ Training 54.2/3 (2012): 95-113.

[14] Dejaeger, Karel, et al. "Gaining insight into student satisfaction using comprehensible data mining techniques." European Journal of Operational Research 218.2 (2012): 548-562.

[15] Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." Asian Journal Of Computer Science And Information Technology (2012).

[16] Osmanbegović, Edin, and Mirza Suljić. "Data mining approach for predicting student performance." Economic Review 10.1 (2012).

[17] Şen, Baha, Emine Uçar, and Dursun Delen. "Predicting and analyzing secondary education placement-test scores: A data mining approach." Expert Systems with Applications 39.10 (2012): 9468-9476.

[18] Agrewal, S., G. Pandey, and M. Tiwari. "Data mining in education: data classification and decision tree approach." International Journal of e-Education, e-

Business, e-Management and e-Learning, 2 (2) (2012): 140-144.

[19] Pandey, Umesh Kumar, and Brijesh Kumar Bhardwaj. "Data Mining as a Torch Bearer in Education Sector." Technical Journal of LBSIMDS (2012).

[20] Srimani, P. K., and Malini M. Patil. "A Classification Model for Edu-Mining." PSRC-ICICS Conference Proceedings. 2012.

[21] Sukanya, M., S. Biruntha, Dr S. Karthik, and T. Kalaikumaran. "Data mining: Performance improvement in education sector using classification and clustering algorithm." In International conference on computing and control engineering,(ICCCE 2012), vol. 12. 2012.

[22] Yadav, Surjeet Kumar, and Saurabh Pal. "Data Mining Application in Enrollment Management: A Case Study." International Journal of Computer Applications (IJCA) 41.5 (2012): 1-6.

[23] Torenbeek, M., E. P. W. A. Jansen, and W. H. A. Hofman. "Predicting first-year achievement by pedagogy and skill development in the first weeks at university." Teaching in Higher Education 16.6 (2011): 655-668.

[24] Gokuladas, V. K. "Predictors of Employability of Engineering Graduates in Campus Recruitment Drives of Indian Software Services Companies." International Journal of Selection and Assessment 19.3 (2011): 313-319.

[25] Yongqiang, He, and Zhang Shunli. "Application of Data Mining on Students' Quality Evaluation." Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on. IEEE, 2011.

[26] Sakurai, Yoshitaka, Setsuo Tsuruta, and Rainer Knauf. "Success Chances Estimation of University Curricula Based on Educational History, Self-Estimated Intellectual Traits and Vocational Ambitions." Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on. IEEE, 2011.

[27] Pandey, Umesh Kumar, and Saurabh Pal. "A Data mining view on class room teaching language." International Journal Of Computer Science Issues (2011).

[28] Arjariya, Tripti, et al. "Data Mining and It's Approaches towards Higher Education Solutions." International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307.

[29] Aher, Sunita B., and L. M. R. J. Lobo. "Data mining in educational system using Weka." IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT). Vol. 3. 2011.

[30] Suthan, G. Paul, and S. Baboo. "Hybrid chaid a key for mustas framework in educational data mining." IJCSI International Journal of Computer Science Issues 8 (2011): 356-360.

[31] Kumar, Varun, and Anupama Chadha. "An empirical study of the applications of data mining techniques in higher education." International Journal of Advanced Computer Science and Applications 2.3 (2011).

[32] Sharma, Mamta, and Monali Mavani. "Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector." Advances in Computing, Communication and Control. Springer Berlin Heidelberg, 2011. 189-194.

[33] Gokuladas, V. K. "Technical and non- technical education and the employability of engineering graduates: an Indian case study." International Journal of Training and Development 14.2 (2010): 130-143.

[34] Ayesha, Shaeela, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat Khan. "Data mining model for higher education system." Europen Journal of Scientific Research 43, no. 1 (2010): 24-29.

[35] Kovacic, Zlatko. "Early prediction of student success: Mining students' enrolment data." (2010).

[36] Al-shargabi, Asma A., and Ali N. Nusari. "Discovering vital patterns from UST students data by applying data mining techniques." Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on. Vol. 2. IEEE, 2010.

[37] Yan, Zhi-min, Qing Shen, and Bin Shao. "The analysis of student's grade based on Rough Sets." Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on. IEEE, 2010.

[38] Ningning, Gao. "Proposing Data Warehouse and Data Mining in Teaching Management Research." Information Technology and Applications (IFITA), 2010 International Forum on. Vol. 1. IEEE, 2010.

[39] Knauf, Rainer, Yoshitaka Sakurai, Setsuo Tsuruta, and Kouhei Takada. "Empirical evaluation of a data mining method for success chance estimation of university curricula." In Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pp. 1127-1133. IEEE, 2010.

[40] Wu, X., Zhang, H., & Zhang, H. (2010, October). Study of comprehensive evaluation method of undergraduates based on data mining. In Intelligent Computing and Integrated Systems (ICISS), 2010 International Conference on(pp. 541-543). IEEE.

[41] Youping, Bian Xiangjuan Gong. "The application of data mining technology in analysis of college student's performance." Information Science (2010).

[42] Liu, Zhiwu, and Xiuzhi Zhang. "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm." Intelligent Networks and Intelligent Systems (ICINIS), 2010 3rd International Conference on. IEEE, 2010.

[43] Balakrishnan, Julie M. David. "Significance of Classification Techniques In Prediction Of Learning Disabilities." International Journal Of Artificial Intelligence & Applications (2010).

[44] Ranjan, Jayanthi, and Raju Ranjan. "Application of data mining techniques in higher education in India." Journal of Knowledge Management Practice 11.1 (2010): 1-9

[45] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40, no. 6 (2010): 601-618.

[46] Baker, Ryan SJD, and Kalina Yacef. "The state of educational data mining in 2009: A review and future visions." JEDM-Journal of Educational Data Mining1.1 (2009): 3-17.

[47] Siraj, Fadzilah, and Mansour Ali Abdoulha. "Uncovering hidden information within university's student enrollment data using data mining." Modelling & Simulation, 2009. AMS'09. Third Asia International Conference on. IEEE, 2009.

[48] Zhu, Li, Yanli Li, and Xiang Li. "Research on Early-Warning Model of Students' Academic Records Based on Association Rules." Computer Science and Information Engineering, 2009 WRI World Congress on. Vol. 4. IEEE, 2009.

[49] Linjie, Qu, and Lou Lanfang. "Data mining and application of teaching evaluation based on association rules." Computer Science & Education, 2009. ICCSE'09. 4th International Conference on. IEEE, 2009.

[50] Ahmed, Almahdi Mohammed, Norita Md Norwawi, and Wan Hussain Wan Ishak. "Identifying student and organization matching pattern using Apriori algorithm for practicum placement." Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on. Vol. 1. IEEE, 2009.

[51] Nayak, Amar, Jitendra Agarwal, Vinod Kumar Yadav, and Shadab Pasha. "Enterprise architecture for semantic web mining in education." In Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on, vol. 2, pp. 23-26. IEEE, 2009.

[52] Wook, Muslihah, Yuhanim Hani Yahaya, Norshahriah Wahab, Mohd Rizal Mohd Isa, Nor Fatimah Awang, and Hoo Yann Seong. "Predicting NDUM Student's Academic Performance Using Data Mining Techniques." In Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on, vol. 2, pp. 357-361. IEEE, 2009.

[53] Wang, Pei-ji, Lin Shi, Jin-niu Bai, and Yu-lin Zhao. "Mining Association Rules Based on Apriori Algorithm and Application." In Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on, vol. 1, pp. 141-143. IEEE, 2009.

[54] Ramasubramanian, P., K. Iyakutti, and P. Thangavelu. "Enhanced data mining analysis in higher educational system using rough set theory." African Journal of Mathematics and Computer Science Research 2, no. 9 (2009): 184-188.

[55] Selmoune, Nazih, and Zaia Alimazighi. "A decisional tool for quality improvement in higher education." Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on. IEEE, 2008.

[56] Shangping, Dai, and Zhang Ping. "A data mining algorithm in distance learning." Computer Supported Cooperative Work in Design, 2008. CSCWD 2008. 12th International Conference on. IEEE, 2008.

[57] Bresfelean, Vasile Paul, Mihaela Bresfelean, Nicolae Ghisoiu, and C-A. Comes. "Determining students' academic failure profile founded on data mining methods." In Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on, pp. 317-322. IEEE, 2008.

[58] Dimokas, Nikolaos, Nikolaos Mittas, Alexandros Nanopoulos, and Lefteris Angelis. "A prototype system for educational data warehousing and mining."

InInformatics, 2008. PCI'08. Panhellenic Conference on, pp. 199-203. IEEE, 2008.

[59] Zhang, Xiaolong, and Guirong Liu. "Score Data Analysis for Pre-Warning Students in University." Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on. IEEE, 2008.

[60] Zhao, Hua-long. "Application of OLAP to the analysis of the curriculum chosen by students." Anti-counterfeiting, Security and Identification, 2008. ASID 2008. 2nd International Conference on. IEEE, 2008.

[61] Villalon, Jorge J., and Rafael A. Calvo. "Concept Map Mining: A definition and a framework for its evaluation." Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on. Vol. 3. IEEE, 2008.

[62] Pumpuang, Pathom, Anongnart Srivihok, and Prasong Praneetpolgrang. "Comparisons of classifier algorithms: Bayesian network, C4. 5, decision forest and NBTree for Course Registration Planning model of undergraduate students." Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on. IEEE, 2008.

[63] Ranjan, Jayanthi, and Saani Khalil. "Conceptual framework of data mining process in management education in India: an institutional perspective."Information Technology Journal 7.1 (2008): 16-23.

[64] Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás. "Data Mining Algorithms to Classify Students." In EDM, pp. 8-17. 2008.

[65] Ogor, Emmanuel N. "Student academic performance monitoring and evaluation using data mining techniques." Electronics, Robotics and Automotive Mechanics Conference, 2007. CERMA 2007. IEEE, 2007.

[66] Hien, Nguyen Thi Ngoc, and Peter Haddawy. "A decision support system for evaluating international student applications." Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual. IEEE, 2007.

[67] Bresfelean, Vasile Paul. "Analysis and predictions on students' behavior using decision trees in Weka environment." Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on. IEEE, 2007.

[68] Nghe, Nguyen Thai, Paul Janecek, and Peter Haddawy. "A comparative analysis of techniques for predicting academic performance." Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual. IEEE, 2007.

[69] Romero, Cristóbal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." Expert systems with applications 33.1 (2007): 135-146.

[70] Al-Radaideh, Qasem A., Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. "Mining student data using decision trees." In International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan. 2006.

[71] Aksenova, Svetlana S., Du Zhang, and Meiliu Lu. "Enrollment prediction through data mining." Information Reuse and Integration, 2006 IEEE International Conference on. IEEE, 2006.

[72] Rasmani, Khairul A., and Qiang Shen. "Subsethood-based fuzzy rule models and their application to student performance classification." Fuzzy Systems, 2005. FUZZ'05. The 14th IEEE International Conference on. IEEE, 2005.

[73] Delavari, Naeimeh, Mohammad Reza Beikzadeh, and Somnuk Phon-Amnuaisuk. "Application of enhanced analysis model for data mining processes in higher educational system." Information Technology Based Higher Education and Training, 2005. ITHET 2005. 6th International Conference on. IEEE, 2005.

[74] Talavera, Luis, and Elena Gaudioso. "Mining student data to characterize similar behavior groups in unstructured collaboration spaces." Workshop on artificial intelligence in CSCL. 16th European conference on artificial intelligence. 2004.

[75] Salazar, A., J. Gosalbez, I. Bosch, R. Miralles, and L. Vergara. "A case study of knowledge discovery on academic achievement, student desertion and student retention." In Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on, pp. 150-154. IEEE, 2004.

[76] Minaei-Bidgoli, Behrouz, and William F. Punch. "Using genetic algorithms for data mining optimization in an educational web-based system." Genetic and Evolutionary Computation—GECCO 2003. Springer Berlin Heidelberg, 2003.

[77] Kotsiantis, S., C. Pierrakeas, and P. Pintelas. "Efficiency of Machine Learning Techniques in Predicting Students' Performance in Distance Learning Systems." Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece (2002).