



A Novel Educational Datamining Model to Attain Sustainability

Mrs. Anita N Chaware¹

¹Asso.professor, P.G. Department of Computer Science-
MCA, SNDT Womens University,
Mumbai, India
anitadongarwar@gmail.com

Dr. U. A. Lanjewar²

²Institute of Computer Studies & Research,
VMV College of Arts, Commerce and Science
Nagpur, India

Abstract: This work aims at the concept of enhancing the Technical Higher Education to address sustainability and globalization, by doing the research in a novel merged area “Educational data mining” (EDM). Sustainability in education can be achieved by improving the existing education system and by restructuring and reorienting the same. As per the research done in Shandong University, to attain sustainability, more focus should be given in creating student’s opportunities and conducting sustainability related researches in Higher Education. A novel merged area called Educational data mining (EDM) can be used in creating students opportunities by using different DM algorithms. EDM is a study of Inclusion of educational research in data mining and hence addresses two subjects outlined in this paper : The first one corresponds to a review of data mining (DM) techniques , algorithms and its application to Technical Higher Education. Much of the previous work done in EDM area is done is for the online courses or the web-enabled courses. This paper considers the data available with University and mines to model the Higher Education. Hence, the second topic represents a novel Educational Data Mining Model. Finally, We conclude: “Educational data mining (EDM) findings can help the Technical Higher Education system to achieve sustainability and stand firm in the race of Globalization.”

Keywords: Technical Higher Education, University Database, Sustainability, Globalization, Education Data mining Model.

I. INTRODUCTION

The concept of Sustainability Development started in the 80’s era and different meaning and definition were introduced. One among them most frequently quoted definition is from the Brundtland Report[1]:

"Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains two key concepts within it:

First, the concept of needs, in particular the essential needs of the world's poor, to which overriding priority should be given; and Second, the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs."

In other words, sustainability can be defined as something that provides an ability to think in time – to forecast, to think ahead - to plan[2]. Sustainability makes us to move from awareness - to knowledge – to action. It makes people capable to use various process like knowing, inquiring, acting, judging, imagining, questioning and developing an aesthetic response to the environment. Sustainability leads to globalization.

Globalization is defined as economic growth combined together with equality and social peace. Challenges of globalization comes in package, like, one needs to get ready to prepare for a workplace where responsibilities are constantly changing, where vertical management is replaced by networking, where information passes through multiple and informal channels, where initiative-taking is more important than obedience, and where strategies are especially complex because of the expansion of markets beyond national Borders. To meet these challenges, imparting education to each n everyone of all ages is

needed. For this the whole education system needs to be restructured, reorient and refined [3].

In 2010 United Nations University came up with the idea of the decade for education for sustainable development based on a “new vision of education that seeks to empower people of all ages to assume responsibility for creating and enjoying a sustainable future”.

Hence the role of universities is to create highly competent knowledge workers like decision makes, teachers, leaders, entrepreneurs etc. But, to succeed, Universities need to work like corporate house where there is a continuous effort to improve the business. Universities cant be thought of business nut will need to build new business models that are dynamic, modern and fit for the decades ahead. It needs to explore different ways to face and solve the challenges to address sustainability. Universities need to take the responsibility of knowledge transfer activities and the continuous processes of monitoring, centralizing and attempts to manage these activities more efficiently.

One way to do this the “big data” universities have. Our research focuses on this big data the universities have. The wealth of data in higher education if, analysed systematically, can tell the new patterns generating some important knowledge. This knowledge can be used to improve and to enhance the decision in favour of society[4]. Large amounts of data from a variety of sources is collected on daily basis from students admissions, classes, administration, faculty members, programs of study, etc.[5]. Indian Universities needs to look at this big data available and implement new strategies and tools that help them to improve the quality of the students and hence the research. To address sustainability one need to be sustainable. Also to face globalization in education, using knowledge in decision making , can play a significant role in improving and sustaining university competitiveness. Universities need to create new, leaner business models as competition increases

for staff, students, funding and partners. Each of the University stake holders has to take part in this new business model. Each of them need to analyses the data available with them and act accordingly. Our research focus on modelling this “big data” and applying data mining techniques to gain knowledge.

II. DATA MINING

A. Data mining:

Data mining is a process of using different algorithms to find useful patterns or models from data. It is a process of selecting, exploring and modeling large amount of data. Mostly used in technology, also used in areas of real life like finance, marketing, business, different social science methodologies, such as psychology, cognitive science and human behaviour etc.. The ability to continually change and acquire new understanding is a driving force for the application of DM and this allows many new future applications of data mining[6].

B. The DM methodology:

Data mining is a link between massive database and analytical systems. Data mining software analyzes relationships and patterns in stored data based on open-ended user queries [7]. It studies the data and build some relationships among the data. These relationships are used to analyses the data further. The basic relationships are categories as follows:

- a. Classifications: It predicts certain output based on given input. Usually called as supervised learning in which the input/output matches of training data is modeled and tested on test data to find results for new data like in ID3 and C4.5 etc.
- b. Clustering: Data items are grouped according to logical relationships or users requirements. Being Unsupervised learning, In clustering, the algorithms learns by looking at the input and predict the output. Example algorithms are k-means and their advance forms, hierarchical algorithms, density based, etc.
- c. Relationship mining like building Associations: Data can be mined to identify a co-occurrence relationship called associations among the data which may reveal a new knowledge and interesting knowledge discovery.
- d. Sequential patterns: Data is mined to anticipate behaviour patterns and trends.

More study on these relationships can be found in [8,9,10,11,12]. These relationships are then used in many techniques like: Neural networks, fuzzy logic, Intelligence agent systems, Modelling, knowledge-based systems, System optimization and Information systems, together with their applications in nearly all application domains [13] for understanding the patterns or getting more optimized solutions.

To extract (identify) the required / specified knowledge using data mining algorithms, a complete process of selecting the database and preprocessing the data is must. But being educational data most of the data is complete & clean or without any error or missing data and a little preprocessing is required. Hence we can directly consider this for applying the algorithms with little bit of formatting to the data. Most of the data mining algorithms can be viewed as compositions of a few of the above basic

techniques. Also a note of caution Blind application of data mining methods can be a dangerous activity leading to discovery of meaningless patterns.

III. EDUCATION DATA MINING

Educational Data Mining is an emerging interdisciplinary field, concerned with developing data mining methods for exploring the massive data of the educational institutions, and using those methods to better understand students, and the environment in which they learn. The educational data is taken from interactive learning environments, computer-supported collaborative learning environment, or administrative data from the universities. It allows its researchers to use the data mining models to attempt the diversity of research in educational field, working in all the environment like online teaching, classroom or the mixture of both i.e. the blended mode. Researchers study a variety of areas, including individual learning factors, factors that are associated with students failure or non-retention in courses. In particular, the EDM focuses on the prediction, innovative work done on existing models to understand new knowledge of students learning pattern, their emotional and intelligent quotients, the teaching learning process. (“discovery with models”)[4,11].

A. Literature available:

Since 1995 people started working with the Educational Data. But much of the work was on the data available online. Few of the work done are listed below;

- a. Tang and McCalla build a system, presenting a study conducted with simulated students and successful evaluation of the system with real students[9].
- b. The work done by Beikzadesh et.al. suggest that data mining technology can be used by lecturer, student, alumni, manager and other educational staff and is a useful tool for decision making on their educational activities[10].
- c. Romero, C., Ventura, S. (2007) in his survey paper Educational Data Mining: A Survey from 1995 to 2005 has given many such examples which Highlights the importance of data mining in education field and many more [11].
- d. Brijesh Kumar Baradwaj, Saurabh Pal in their paper used the tree method is used the Information like Attendance, Class test, Seminar and Assignment marks were collected from the student's management system, to predict the performance at the end of the semester. [12]
- e. Zaïane in 2001 & 2002 used the online data of students to study on-line courses. This article become one of the milestone to the EDM community proposing and highlighting the EDM's usefulness [16,17].
- f. Aksenova, Zhang, & Lu, in their research has predicted the enrolment of the university using time stamp
- g. Baker, R.S., Corbett, A.T., Koedinger, K.R. in (2004) used data mining techniques build the machine learning technique to Detect Student Misuse of Intelligent Tutoring Systems [18].
- h. Based on the trend fitting model and the linear regression model, Lan Wang, Wei-Wei Zhuang, Yufen Liu tride to predicts the scale of postgraduate education in Hebei Province and points out that the scale of

graduate student enrolment should be compatible with the GDP growth rate. [19]

- i. Muslihah Wook, Yuhanim Hani Yahaya, Norshahriah Wahab, Mohd Rizal Mohd Isa, Nor Fatimah Awang, Hoo Yann Seong compare two data mining techniques Artificial Neural Network (ANN) and the combination of clustering and decision tree classification techniques for predicting and classifying students' academic performance[20]

In this paper, we propose a Novel Higher Educational Data Mining [EDM] model that use the available data from the classroom learning , faculty database , administrative database etc. and modelled it to show how the teaching learning process in higher education can be improved so that our universities can sustain their qualities and produce a quality products in term of students education and research.

The table Educational Data Mining (EDM model) has 3 columns, first Available Database, second The data mining(DM) processes and third the knowledge fro decision making . The DM processes are the jobs /task which generates some patterns leading to knowledge further can be used for decision making / improvements / enhancement to our education systems to sustained itself in the global competitiveness. Once this Acedemic improvement or growth, stakeholders relationship, administrative enhancement is attained then universities become role models in addressing sustainability development. The different stakeholders such as students, academics and administrators, parents, industries with their diverse attitudes, skills, knowledge and experiences makes Universities a multi-stakeholder organization with perfect prerequisites for sustainable development.

IV. EDM MODEL FOR HIGHER EDUCATION

Table: 1

Available Database	Data mining processes	Knowledge to help in decision making.
Students Enrolment data	<ul style="list-style-type: none"> • To predict students learning outcome • To predict enrolment for the course • To predict enrolment of the university from previous enrolment • To predict students selection of course depending on their interest • To predict students dropouts in the initial stage only • To predict students eligibility for higher studies/placements • To form groups/clusters of students depending on their interest, to select the electives • To form groups/clusters of students depending on their parental income , eligible for getting grants. • To find association between mother tongue and medium of instruction • To find association between parental financial condition and student score • To find association between parental education and student score • To find association between distance travelled by the students and their score • Prediction of the numbers of international students and faculties attracting to the University. • Predict the ratio of gender enrolling in university. • Associate learning with gender to predict percentage wise learning 	<ul style="list-style-type: none"> • To improve the enrolment of fresher's • To identify school leavers, continuation education learners, vocational learners domestics and international students, executive learners. • To guide students on their financial status, allocate scholarships etc. • To encourage girls education, higher education in all.
Students Acedemic data	<ul style="list-style-type: none"> • To predict students final year marks from the first yr. marks and some other attributes like distance travelled , parents education, income etc. • To find association between previous students interest n their seminar and projects topics and hence predict the seminar and projects topics for final yr. students. • To predict students eligible for higher studies from students score • To predict the students who should avail extra coaching for their placement • To find associations between attendance and their mark • To group or cluster student according to their interest for placement/training • To Predict the drop-out rate in coming semester • Predicting how many will be selecting by companies as interns • Predicting no. of students recruits during campus. 	<ul style="list-style-type: none"> • To guide students on to improve their performance, Placements and internships, • To drive them for higher studies.
Students feedback data	<ul style="list-style-type: none"> • To find association between feedback and teacher and assist the teacher for his or her improvement- • To predict the course in demand • To predict the resources in demand. 	<ul style="list-style-type: none"> • Help to access teachers, other resources
Co-curricular activities	<ul style="list-style-type: none"> • Associate students interest with co-curricular activities to promote/motivate her/him • To predict the future leaders, academician, innovators, entrepreneur, etc... in students and guide them accordingly. • To group / cluster the students interested in social cause and bring social awareness by motivating them. 	<ul style="list-style-type: none"> • Help in identify leaders, entrepreneur of future and guide them accordingly.
Project and Seminar	<ul style="list-style-type: none"> • To predict seminar topic depending on students interest/ previous yr. marks • Prediction of companies visiting for campus and number of student they will select for projects. 	<ul style="list-style-type: none"> • Academia -Industry collaboration
Faculty database	<ul style="list-style-type: none"> • To associate faculties qualification and research activities. • To predict faculty appraisal depending on different factors • To predict the Acedemic expenses of the faculties • To associate the acedemics of faculties with their interest to update them for their job satisfaction. 	<ul style="list-style-type: none"> • Faculty improvement • Faculty appraisal • Talent management
Syllabus/Course	<ul style="list-style-type: none"> • Predicting what type of students are most likely to take particular type of subjects • Predicting students likeliness to select particular type of subjects as electives • To cluster the subject of similar field for students to select as electives 	<ul style="list-style-type: none"> • Improve in curriculum as per the industry need. • Introduce Need based courses

Library database	<ul style="list-style-type: none"> To associate the students book usage with the number of hours to predict the books , subject, authors in demands To predict whether students wants to paper book or only online material. To optimise the fund spend on book . 	<ul style="list-style-type: none"> To Optimize the Library resources To optimize the financial support to library
Collaboration database	<ul style="list-style-type: none"> Associating the university factors and collaboration for getting good grants Predicting the factors to attract the international students and faculties. Associating student's to the companies and organization for collaboration Predicting the income/grant generation from collaboration Predicting factors responsible for good collaborations. Prediction the good collaboration partners .from different fields. 	<ul style="list-style-type: none"> To make good partnership with industries for getting good grants, faculties support, getting internship program, placing students.
Administration database	<ul style="list-style-type: none"> To predict the financial conditions of university for next year. To predict the enrolment of the university for next yr. To predict investments to be done on different resources for the next year. To predict the our sustainability by measuring the factors for next how many year. To predict the ratio of support staff and academic staff needed. To build expert systems which will help them to select teachers, resources, etc. 	<ul style="list-style-type: none"> Overall improvement of university to achieve sustainability and hence ready for globalization. Teacher appointment procedure

As an example the students enrollment database has all the attributes of the students from their past data from which we can predict the students caliber, intelligence, and interest in subject opting for. Sometimes students select the subject under external pressure, which can be monitored at the entry level and can guide the students to select the subject of his / her own interest. This leads the university to improve the learning and hence the research. Also Factors like previous year results, attendance, financial status of family, parent educational qualification plays an important role in students' education. By applying various DM algorithm to these factors we can predict the students outcome. The faculty database can be helpful to faculties for their research, their appraisal, and overall talent management to the HR

V. ANALYSIS OF THE PROPOSED MODEL

In the proposed model, many of the processes are individually done. Much of the previous work is focused on student performance or faculty management. But to be sustained the overall development is required. In our proposed model there are much more hidden patterns which are necessary to look from the researcher's point of view. Hence every database available with the university can be seen as the assets for research. We tried to take as many database as we can. This work can be extended further with new database found in university.

VI. CONCLUSION

According to the definition of sustainability mentioned in Brundtland report, Sustainability is a blend of knowledge and skills. It's also an ability to work cooperatively with other stakeholders of that environment. Hence we need generations to take part in achieving sustainability and University is the best place to start for sustainability development. But current Environment or Model of University does not have the support to attain sustainability and hence are far behind in the race of the global competitiveness.

The focus of our paper is an attempt to use the available database and model the higher education for making itself sustainable. The university has to take measure for its improvement, build confidence among its all the stake holders by catering their all requirements successfully. A sort of business analytics has to be done on the university business. The University are not business house, they are meant for societal development but it has to run like the business house so as to achieve the goal of sustainability. Finally, We conclude that data mining as used for business analytics, "Educational data mining (EDM)

finding can also help the Technical Higher Education system to achieve sustainability and stand firm in the race of Globalization."

VII. REFERENCES

Journal Papers:

- [1]. V.M. Segovia, Transforming Mindsets Through Education for Sustainable Development, International Encyclopedia of Education (Third Edition), 2010, Pages 746-752
- [2]. Osterwalder k., "The Role of Higher Education Institutions in Promoting Sustainable Development", Editorials and Commentaries, UNU-IAS
- [3]. Jana Dlouhá, Donald Huisingh, Andrew Barton, Learning networks in higher education: universities in search of making effective regional impacts, journal of Cleaner Production, Volume 49, June 2013, Pages 5-10
- [4]. S. ELATIA, D. IPPERCIEL, & A. HAMMAD," Implications and Challenges to Using Data Mining in Educational" , Canadian Journal of Education 35, 2 (2012): 101-119
- [5]. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth "The KDD Process for Extracting Useful Knowledge from Volumes of Data " Communication of the ACM November 1996/Vol. 39, No. 11 .
- [6]. Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011", Expert Systems with Applications 39 (2012) 11303–11311 .
- [7]. Ryan S.J.d. Baker, Kalina Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions, journal of educational data mining, Vol1, No.1, 2009.
- [8]. Sk. Althaf Hussain Basha, A. Govardhan, S. Viswanadha Raju, Nayeem Sultana,"A Comparative Analysis of Prediction Techniques for Predicting Graduate Rate of University", European Journal of Scientific Research ISSN 1450-216X Vol.46 No.2 (2010), pp.186-193
- [9]. Tang, T., McCalla, G. (2005) Smart recommendation for an evolving elearning system: architecture and experiment, International Journal onE-Learning, 4(1), 105–129.,
- [10]. Shu-Hsien Liao , Pei-Hui Chu, Pei-Yuan Hsiao, Data mining techniques and applications – A decade review from 2000 to 2011, Expert Systems with Applications 39 (2012) 11303–11311

- [11]. Romero, C., Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33, 125-146
- [12]. Brijesh Kumar Baradwaj , Saurabh Pal “Mining Educational Data to Analyze Student’s Performance , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011
- [13]. Mrs Anita Chaware, Educational Data Mining: An Emerging Trends in Education, review article in International Journal of Advanced Research in Computer Science, Volume 2, No. 6, Nov-Dec 2011, ISSN No. 0976-5697
- Books, Newsletters and reports:**
- [14]. Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, (The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor , 2nd ed.)
- [15]. Brundtland Commission Report. In world commission on environment and development. *Our Common Future* OUG, (1987) IIEP NEWSLETTER / APRIL-JUNE 1998
- Proceedings Papers:**
- [16]. ZAIANE, O. 2001. Web usage mining for a better web-based learning environment. In *Proceedings of conference on advanced technology for education*, 60-64.
- [17]. ZAIANE, O. 2002. Building a recommender agent for e-learning systems. In *Proceedings of the International Conference on Computers in Education*, 55-59.
- [18]. Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004) Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, 531-540.
- [19]. Lan Wang¹, Wei-Wei Zhuang², Yu-Fen Liu³, — An Empirical study on the prediction model of Postgraduate education in Hebei province, *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, Qingdao, 11-14 July 2010
- [20]. Muslihah Wook, Yuhanim Hani Yahaya, Norshahriah Wahab, Mohd Rizal Mohd Isa, Nor Fatimah Awang, Hoo Yann Seong, Predicting NDUM Student’s Academic Performance Using Data Mining Technique 2009 Second International Conference on Computer and Electrical Engineering, 978-0-7695-3925-6/09 \$26.00 © 2009 IEEE