Descriptive Analysis of Enrollment Data and Adaptive Educational Hypermedia

Nidhi Chopra[1] and **Manohar Lal**[2]

Submitted in November 2012, Accepted in April 2013

Abstract - As the world around is going through a technological revolution with the dawn of digital age, educationists are in some ways compelled to rethink the existing education system and its components. With the tools and the techniques available, nowadays it's imperative to reconsider how they can be used to improve educational institutions and associated bodies. Opportunities for knowledge discovery in educational data have increased tremendously with digital revolution now as compared to the scenario in the past. Educational data is becoming increasingly rich as more and more educational systems are going online and collecting large amounts of data. In this paper a study of an enrollment dataset is presented.

Index Terms - Data Analytics, Educational Data Mining, Enrollment data, Adaptive Educational Hypermedia.

1. INTRODUCTION

In this new digital age, the world of education has also gone under a major transformation. The new technologies and gadgets available help not only enrich and enhance the existing education system but also offer new opportunities and modes which can take the process of learning beyond institutions and allow people to learn on their own time and own terms. These new advances in learning have played a big role in this age of knowledge enhancement via different means and are clearly a sign that there is a need to rethink how the technology potential can be tapped to improve our education system [1, 2]. As of now most of the changes can be seen in the way information is stored, retrieved, distributed or provided to the students such as educational technology, e-learning portals, Learning Management Systems used in distance learning, blended learning & so on. Another emerging and associated area is educational data mining (EDM) where storage, retrieval and analysis of educational data sets can be leveraged to revolutionize education systems [3].

People in all fields and disciplines are becoming more and more informed. They are learning to observe, collect and interpret data trends around them to make better and informed decisions [4, 5]. Analysis of educational data sets is required to understand needs of current society and then also cut down costs in the process [4, 6]. In order to clearly define the framework and the needs for revolutionizing the current education system data have to be analyzed as a first step. In this paper such a study & some solutions are described.

[1, 2] *School of Computer and Information Sciences, Indira Gandhi National Open University, Delhi – 110068, India*
E-mail: [1]nidhichopra@ignou.ac.in and [2]mlal@ignou.ac.in

2. EDUCATIONAL DATA MINING

In the last few years EDM has emerged as a field of its own. The EDM community website [7] defines EDM as follows: "EDM is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." Data mining (DM), or Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data [8]. It finds applications in the fields of artificial intelligence, numeric & combinatorial optimization, business, management, medicine, computer science, engineering etc. [9]. DM largely consists of analyzing available sets of data to interpret, isolate the trends and patterns present in the data i.e. converting raw data into information. The trends obtained can be called as prediction or recommendations [10]. These can be used by educators, educational software developers, teachers, parents or students. However, it is largely understood that EDM methods are often different from standard DM methods. This is because of the non-independence and multilevel hierarchy found in educational data. For the same reason, it is increasingly common to see the psychometrics models being used in EDM [11]. DM is a part of Data Analysis. The outcomes of data based research can be descriptive or actionable, this study includes both.

DM can be visualized as a confluence of multiple disciplines where the background knowledge pertaining to the area of study is processed using tools pertaining to other disciplines such as – information science, database technology, statistics, machine learning & other related fields. Here the 'Area of Study' would be 'Education'. The data can be collected from students' use of interactive learning environments, computer-supported collaborative learning, evaluation, assessment or administrative data (web logs, library usage) from schools and universities. There are various challenges in the field of education like understanding choice of major, appropriate evaluation schemes, student drop out, retention, student unrest and crime, assessment of institution and educationists' goals like quality, access, cost, social and cultural biases. Educational efficacy can be measured and predicted using DM methods [12]. DM is a field which has originated from databases and Artificial Intelligence [13]. Understanding the current trends of our education system could point out towards the underlying issues and help us device an effective plan to address them.

Figure 1 shows broad two possible dimensions of EDM research wherein utilizing the data from point of view of educators and also from those studying the management/administrative aspect of EDM is considered.

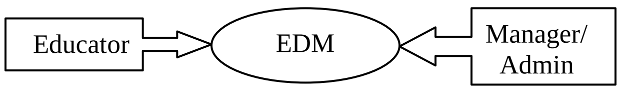


Figure 1: Two possible dimensions of Educational Data Mining research.

2.1 EDM for Educators

An example for such type of EDM is PSLC (Pittsburgh Science of Learning Centre) [14, 15].

2.2 EDM for Administrator and Managers

Education Administrators also use EDM to understand more management related factors such as demographics, enrollment etc. An example of EDM for Admin is the presented case study of enrollment data consisting of 3020 record obtained from SRD (Student Registration Division) of IGNOU (Indira Gandhi National Open University). The data files are in dbf format (figure 2) and can be imported in MS-Excel (figure 3) using FoxPro. In dbf format, one column is shown in an entire screen in figure 2 whereas figure 3 now has a more readable tabular representation.

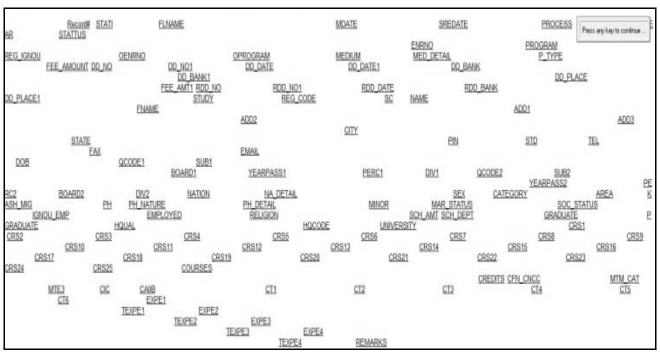
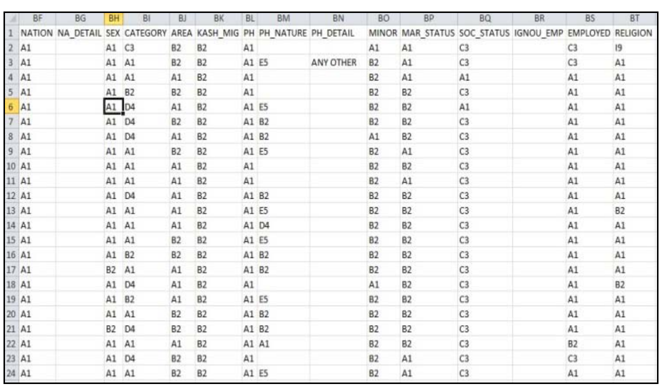


Figure 2: .dbf data file in MS-Visual Studio's Visual FoxPro (raw data file as received)



	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT
1	NATION	NA_DETAIL	SEX	CATEGORY	AREA	KASH_MIG	PH	PH_NATURE	PH_DETAIL	MINOR	MAR_STATUS	SOC_STATUS	IGNOU_EMP	EMPLOYED	RELIGION
2	A1		A1	C3	B2	B2	A1			A1	A1	C3		C3	I9
3	A1		A1	A1	B2	B2	A1	E5	ANY OTHER	B2	A1	C3		C3	A1
4	A1		A1	A1	A1	B2	A1			B2	A1	A1		A1	A1
5	A1		A1	B2	B2	B2	A1			B2	B2	C3		A1	A1
6	A1		A1	D4	A1	B2	A1	E5		B2	B2	A1		A1	A1
7	A1		A1	D4	B2	B2	A1	B2		B2	B2	C3		A1	A1
8	A1		A1	D4	A1	B2	A1	B2		A1	B2	C3		A1	A1
9	A1		A1	A1	B2	B2	A1	E5		B2	A1	C3		A1	A1
10	A1		A1	A1	A1	B2	A1			B2	B2	C3		A1	A1
11	A1		A1	A1	A1	B2	A1			B2	A1	C3		A1	A1
12	A1		A1	D4	A1	B2	A1	B2		B2	B2	C3		A1	A1
13	A1		A1	A1	A1	B2	A1	E5		B2	B2	C3		A1	B2
14	A1		A1	A1	A1	B2	A1	D4		B2	B2	C3		A1	A1
15	A1		A1	A1	B2	B2	A1	E5		B2	B2	C3		A1	A1
16	A1		A1	B2	B2	B2	A1	B2		B2	B2	C3		A1	A1
17	A1		B2	A1	A1	B2	A1	B2		B2	B2	C3		A1	A1
18	A1		A1	D4	A1	B2	A1			A1	B2	C3		A1	B2
19	A1		A1	B2	A1	B2	A1	E5		B2	B2	C3		A1	A1
20	A1		A1	A1	B2	B2	A1	B2		B2	B2	C3		A1	A1
21	A1		B2	D4	B2	B2	A1	B2		B2	B2	C3		A1	A1
22	A1		A1	A1	A1	B2	A1	A1		B2	B2	C3		B2	A1
23	A1		A1	D4	B2	B2	A1			B2	A1	C3		C3	A1
24	A1		A1	A1	B2	B2	A1	E5		B2	A1	C3		A1	A1

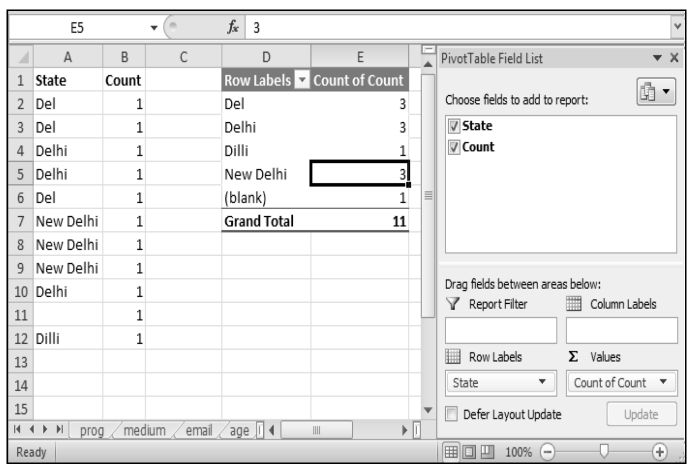
Figure 3: .xlsx data file in MS-Excel (cleaned data file)

For this research, enrollment data of disabled student for an entire year 2009 was obtained from them in January 2010. After data cleaning (using pivot table) some interesting patterns were obtained. The graphs obtained from this analysis are shown below and discussed in the next section. The research methodology has been followed from [16]. A wide variety of DM methods are available such as prediction, clustering, relationship mining, discovery with models, and distillation of data to obtain and present knowledge [17, 18]. Some relevant studies can be found in [19, 20 & 21].

3. STRATEGY FOR DATA CLEANING

Data parsing [22] is easy after importing file from FoxPro to MS-Excel because every column can be viewed separately now. The values are standardized already and discrete verifiable from university website & prospectus. Records were matched to see that there is no repetition of a student's enrollment number which is the primary key. Necessary transformation can be done e.g. to get age from date of birth. So, overall MS-Excel turned out to be a good tool for data cleaning.

'Cold start' is when a data miner has to start from scratch or 'zero' as in this study. Typical real world data sets which are unformatted (raw) need to go through data cleaning steps [22] to be successfully used in a study. After formatting the data appropriately, pivot table feature in excel (a statistical tool in MS-Office package) was used for Data Cleaning. Suppose the variable under consideration is 'State'. Various possible occurrences of State 'Delhi' can be counted as in figure 4. Blanks and wrong fields also got marked ('Del', 'Dilli'). This also explains that why sometimes local understanding of the database can be crucial.



	A	B	C	D	E
1	State	Count		Row Labels	Count of Count
2	Del	1		Del	3
3	Del	1		Delhi	3
4	Delhi	1		Dilli	1
5	Delhi	1		New Delhi	3
6	Del	1		(blank)	1
7	New Delhi	1		Grand Total	11
8	New Delhi	1			
9	New Delhi	1			
10	Delhi	1			
11		1			
12	Dilli	1			
13					
14					
15					

Figure4: Pivot table used to count variables& detect blanks and wrong code.

4. RESULTS

In this section some of the results are presented as obtained using pivot table feature of MS-EXCEL, while doing the data analysis conducted on disabled students of IGNOU who enrolled for various courses in the year 2009.

4.1 54.11% students are of young age group (figure 5).

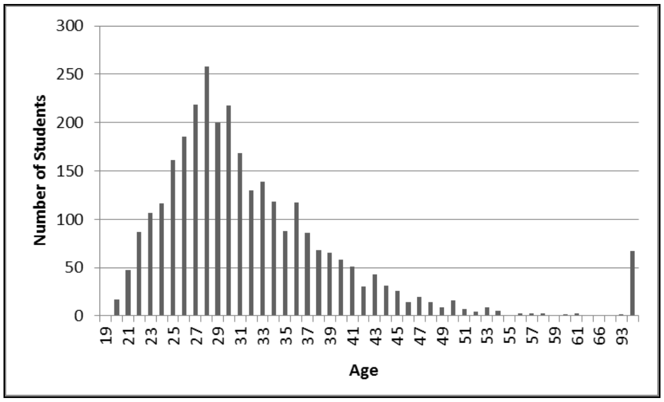


Figure 5: Graph showing age group distribution of students where last column represents 'Wrong code'.

4.2 37.11% students enrolled for Master of Political Sciences and paid an average fee (figure 6).

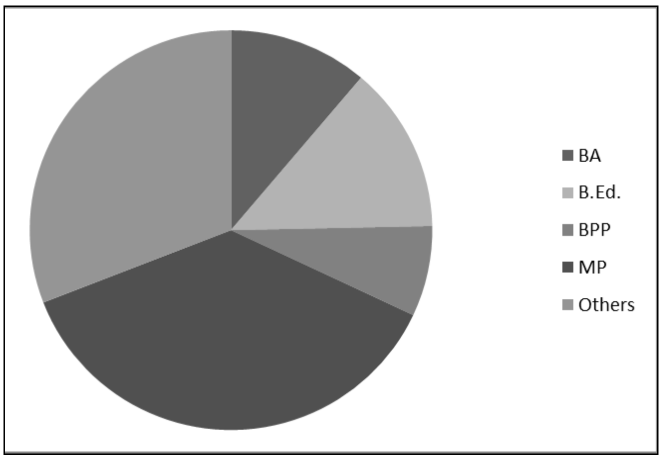


Figure 6: Distribution of courses/programs opted for by the students.

4.3 68. 01% opted for English Medium (figure 7).

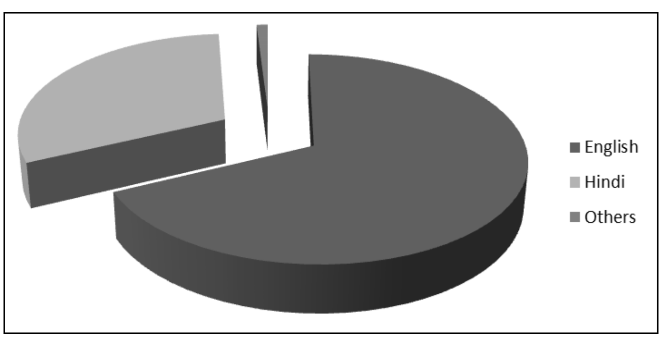


Figure 7: Distribution of medium of instruction as opted by students.

4.4 More than 70% students are male (figure 8).

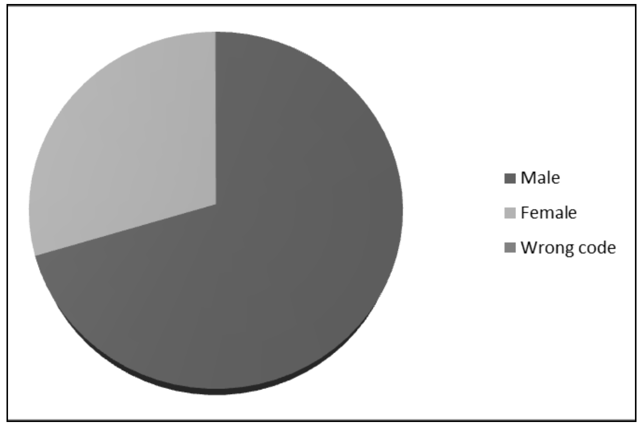


Figure 8: Pie chart showing distribution of gender across student population.

4.5 Most students had finished their previous educational qualification within the past decade as indicated in figure 9 below.

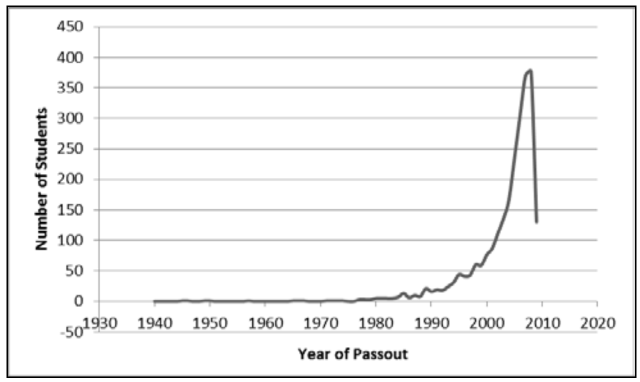


Figure 9: Distribution of students according to the year in which they finished their previous educational qualification.

4.6 54.4% of students are unemployed and 28.1% are employed by IGNOU itself. This shows that number of students who are pursuing education while being employed elsewhere is only 15.3% (figure 10).

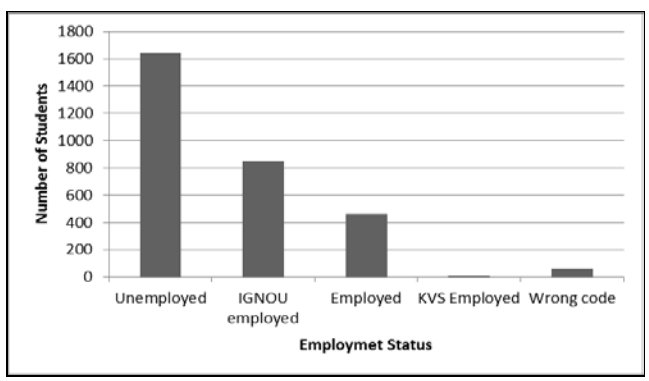


Figure10: Distribution of students accoring to employment status

4.7 Analysis of territory code of students address shows that 60% of students are from urban areas (figure 11).

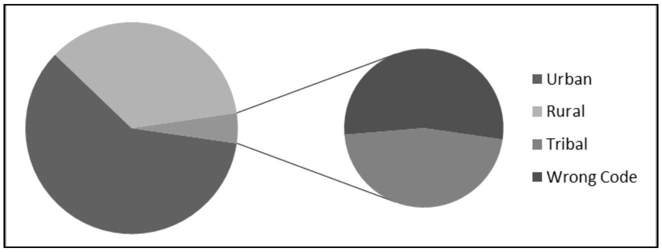


Figure 11: Distribution of student population as per territory code

4.8 To understand the accessibility, we analyzed the email ID field which showed that more than 71% students don't have email ids (figure 12).

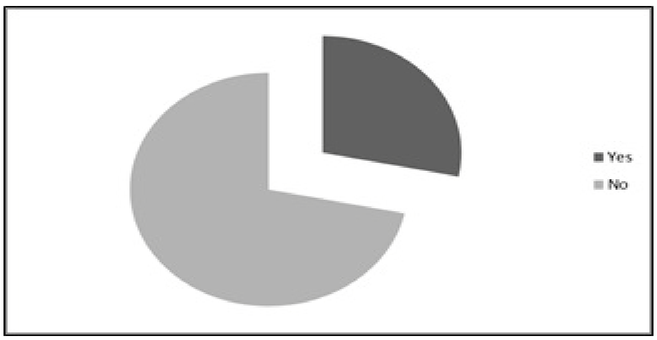


Figure 12: No e-mail ids indicate lack of access to internet and technology.

5. DISCUSSION WITH POSSIBLE ACTIONS

Above results indicate knowledge divide and digital divide. Better methods of increasing outreach are required.

5.1 Result 1 is self-explanatory. Figure 5 has an approximate shape of a normal distribution [23] as often exhibited in biological data [24]. This also resonates well with the model of our current education system where most people like to study or focus on their career value addition in their twenties or early thirties. This kurtosis curve is skewed to the left and a bit slant on the right.

5.2 Result 2 is due to the fact that these students find it easy to do humanities or social sciences courses because there is no help in the form of artificial limbs & training to use them in laboratories (sciences). Science laboratories have no accessibility equipment or area.

5.3 English medium books are comparatively easily available in India at higher education level. IGNOU however plans to launch courses in regional languages. More steps that can be taken are – to encourage translation of books/texts in all subjects and to make them accessible – brail translation, audio books (record and release), video field tours and online repositories of all these educational media.

5.4 Result 4 shows that disability is more common in males in this data sets. But raises more questions about infant & child mortality rate, gender biases or gender divide.

5.5 Result 5 is self-evident. Students are in their 20s, so they have mostly passed out in recent past. Figure 9 is shape of a chi-square distribution [23, 24], with one outlier - 'current' year pass outs.

5.6 Result 6 requires action from Governments to create accessible jobs to increase employment.

5.7 Result 7 indicates the possibility of urban area students having better accessibility to these courses i.e. Knowledge Divide. This needs to be verified by designing a focused future study for the same and improving awareness and accessibility in other areas as well.

5.8 No e-mail ids indicate lack of access to internet and technology for disabled students i.e. Digital Divide.

6. ACCESSIBILITY AND TRACKING

There is need to improve content delivery. It may help in decreasing digital and knowledge gaps. Currently quite a few e-learning and online information delivery platforms are designed with a "One-size-fits-all" approach. Existing distance education system lacks interactivity and can lead to lack of motivation and interest. There is a need for flexible education systems which can also provide guidance as per capacity & learning level [25].

Adaptive Educational Hypermedia (AEH) are flexible and customizable to provide appropriate lesson for each student. Here various views of the same material are created, as desired by the user This can be done by maintaining a student enrollment database combined with user behavior database using tools like link removal (figure 13), stretch text (figure 14) and course monitor (figure 15) for all those students of the university who are using online resources. These proposed tools utilize options set by user & also track and record actions of the user, which media type is chosen most often etc. Combined database form a user model [26] or a student model - goal, previous knowledge, previous performance, background, experience, preferences, stereotypes, user-supplied preferences supplied at run time, analysis of user actions & plan recognition or inference. Providing varying views of the same content is a paradigm shift away from "write once, use once" towards a middleware system [26].

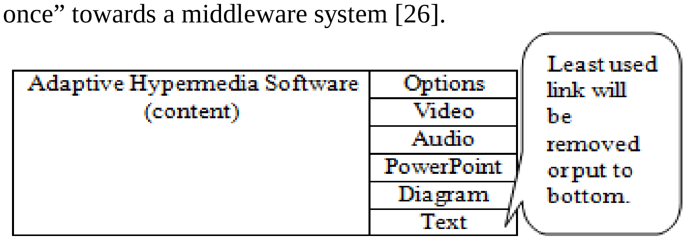


Figure 13: Link removal tool [27]

This link removal tool saves the time spent on looking for most preferred type of media by avoiding confusion. Same is the purpose of stretch text tool and it also makes the interface user friendly. Such tools adapt to the habits & needs of a user (HumanComputer Interface i.e. HCI).

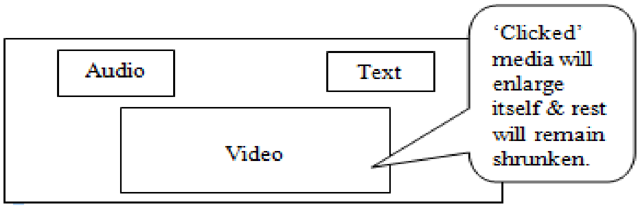


Figure 14: Stretch text tool [28]

Relations in the concept ensure that the student has a study guideline to follow and clearly knows the perquisites and the predecessors for each study. If certain specific perquisites are not fulfilled, the learner will be prompted for the same by course monitor tool. This tool is in accordance with Skinnerian or Linear Approach and can be combined with Programmed Learning. A rule for this can be as below. More components can be added – interest, repetition.

If c1.access = true then set c2.allow_access = true else c2.allow_access = false

If (c1.access = true or c1.test_passed = true) then set c2.allow_access = true else c2.allow_access = false (to include assessment & evaluation options)

Concepts from various disciplines can be combined – "many to many" approach.

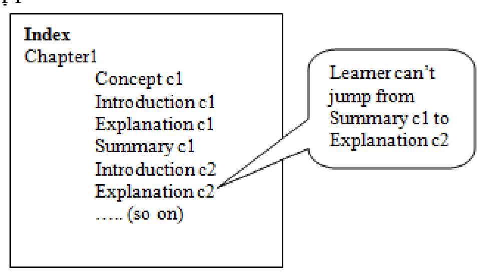


Figure 15: Course monitor tool [29]

7. SCALING IN EDM

A Data Mining problem can be solved through Generalization. To achieve a high degree and accuracy of generalization, a data miner needs a large number of records and more resources in terms of access, time, permissions, teams [30], better software, machines and related facilities as shown in figure 6. Over a longer period of time Agility in Data Analysis can be achieved through Software Re-Engineering. Real world implementations have high complexities [31]. Using tools like WEKA for data mining can give meaningless or useless results. In between steps are not shown as in a calculator so, a required level of understanding may not be obtained. Tools like SPSS require higher system configurations which may not be available to a researcher.

8. CONCLUSION AND FUTURE SCOPE

Analysis of educational data was discussed from the approach of administration. Understanding of the variables provided, exhibited digital divide and knowledge divide. AEH can be used to improve content delivery and may help in decreasing digital and knowledge gaps.It was observed that for developing EDM models, the data obtained should be focused, well organized to achieve effectiveness. At IGNOU where the dataset was collected from an administrative focus and without a preexisting problem statement, more data collection & further studies based on them are required to predict the trends. Studying educational data sets can aid in suggesting pedagogies (teaching methods), site modification, intelligence services & page recommendations in the long run [22, 32].

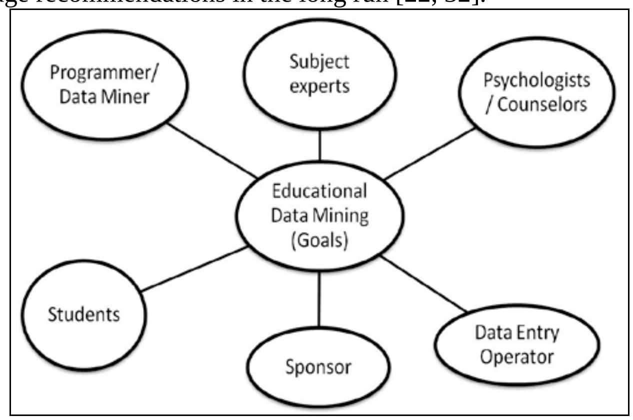


Figure 16: Team work in EDM

To improve the analysis and usability of enrollment data, the enrollment form at universities should be improvised to collect relevant/targeted data fields [33] such as degree of disability, monthly/annual income of the family and other personal data of family, individual/student and assessment. Online enrollment can facilitate collection of data for analysis. Such e-forms can be made adaptive in nature. Background information and performance of every pupil can be assessed throughout the academics and the employment/career to clearly identify any patterns and correlations in the data.

Another possible study can be to analyze the assessment data [34, 35] & log files of these students to find non-independence and multilevel hierarchies in educational data [36]. Such analysis can help us provide more useful insights in the education system of IGNOU for disabled students and to understand the factors affecting students' learning and career path development.

ACKNOWLEDGEMENTS

Authors thank SRD, IGNOU for providing disabled students' enrollment data of the year 2009,for this study.

REFERENCES

[1]. A. Collins and R. Halverson - Rethinking education in the age of technology: the digital revolution and schooling in America. Teachers College Press, 2009.

[2]. C. Romero & S. Ventura. "Educational data mining: A survey from 1995 to 2005". Expert Systems with Applications, 2007, 33(1), 135-146.

[3]. F. Siraj, & M. A. Abdoulha. "Uncovering Hidden Information Within University's Student Enrollment Data Using Data Mining". Third Asia International Conference on Modeling & Simulation, 2009. 413-418. IEEE.

[4]. S. Khan and S. Kumar – "Optimization of Material Procurement Plan – A Database Oriented Decision

Support System", BIJIT - BVICAM's International Journal of Information Technology, 2012.
[5]. Q. A. Al-Radaideh, E. M. Al-Shawakfa & M. I. Al-Najjar.. "Mining student data using decision trees." International Arab Conference on Information Technology, 2006, Yarmouk University, Jordan.
[6]. G. Peters, R. Tagg, and R. Weber. "An application of rough set concepts to workflow management." Proceedings of the 3rd international conference on Rough sets and knowledge technology, 2008, 715-722.
[7]. www.educationaldatamining.org(accessed on 19 December 2011).
[8]. I. H. Witten and E. Frank - Data Mining – Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers, 2005.
[9]. S. V. Chande and M. Sinha – "Genetic Algorithm: A Versatile Optimization Tool", BIJIT - BVICAM's International Journal of Information Technology, 2008.
[10]. C. Vialardi, J. Bravo, L. Shafti, & A. Ortigosa, "Recommendation in higher education using data mining techniques." Proceedings of the 2nd International Conference on Educational Data Mining, 2009, 190-199
[11]. R. S. J. D. Baker, K. Yacef, K. "The state of educational data mining in 2009: A review and future visions." Journal of Educational Data Mining, 2009, 1(1), 3-17.
[12]. T. M. Mitchell - Machine Learning. McGraw-Hill. 1997.
[13]. T. Munakata - Fundamentals of the New Artificial Intelligence, Neural, Evolutionary, Fuzzy and More, Springer, 2008.
[14]. http://learnlab.org/(accessed on 17 August, 2011).
[15]. http://learnlab.org/opportunities/summer/presentations/2011/PSLCSummerSchoolPostersAndFirehoseSlides2011.zip(accessed on 17 August, 2011).
[16]. F. C. Dane - Research Methods, Brooks/Cole Publishing Company, 1990.
[17]. R. S. J. D. Baker, Data mining for education. International Encyclopedia of Education, 2010, 7, 112-118.
[18]. J. Han and M. Kamber - Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2006.
[19]. N. Delavari, M. R. A. Shirazi & M. R. Beikzadeh, "A new model for using data mining technology in higher educational systems." Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training, 2004, 319-324 IEEE.
[20]. N. Delavari, M. R. Beikzadeh, & S. Phon-Amnuaisuk, "Application of enhanced analysis model for data mining processes in higher educational system".6th International conference on Information Technology Based Higher Education and Training, 2005, F4B-1. IEEE.
[21]. C. Vialardi, J. Chue, A. Barrientos, D. Victoria, J. Estrella, A. Ortigosa, J.Peche, "A case study: data mining applied to student enrollment". Proceedings of Third Educational Data Mining Conference, Pennsylvania, USA 2010, 333-335
[22]. M. A. Butt, M. Zaman, Data Quality Tools for Data Warehousing: Enterprise Case Study, IOSR Journal of Engineering, 2013, 3(1), 75-76.
[23]. S. Lipschutz and J. Schiller - Introduction to Probability and Statistics, Tata McGraw Hill, 2005.
[24]. P.S.S.S. Rao and J. Richard - An Introduction to Biostatistics- A Manual for students in Health Sciences, 3rd Edition, PHI Pvt. Ltd. 1999.
[25]. Y. Liao, The Application of Web Mining in Distance Education Problem, Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation, China 2013, Atlantis Press.
[26]. J. Eklund, R. Zeiliger, Navigating the Web: Possibilities and Practicalities for Adaptive Navigational Support. Proceedings of Ausweb'96: The Second Australian World Wide Web Conference. 1996, Southern Cross University Press.
[27]. A. I. Cristea, C. Stewart, Automatic authoring of adaptive educational hypermedia. Web-Based Intelligent e-Learning Systems: Technologies and Applications, 2005,IDEA Publishing group, Zongmin Ma.
[28]. D. Paul, P. Brusilovsky, and G.-J. Houben.Adaptive hypermedia: from systems to framework. ACM Computing Surveys (CSUR) 31.4es , 1999: 12.
[29]. A. I. Cristea, A. D. Mooij. "Designer adaptation in adaptive hypermedia authoring."Proceedings of International conference on Information Technology: Coding and Computing IEEE, 2003.
[30]. A. Cockbum - Agile Software Development, Pearson Education Inc., 2002.
[31]. N. Aggarwal, N. Prakash and S. Sofat – "Mining Techniques for Integrated Multimedia Repositories" BIJIT - BVICAM's International Journal of Information Technology, 2008.
[32]. A. S. Alvi, M. S. Ali., "Revival of Tutor Model: A Domain Independent Intelligent Tutoring System" (ITS)", BIJIT - BVICAM's International Journal of Information Technology, 2010 Vol.2 No.1
[33]. R. Srivastava, I. Srivastava, "Computer and Internet use Among Families: A Case of Botswana", BIJIT - BVICAM's International Journal of Information Technology,2009 Vol.1 No.1
[34]. A. Merceron, K. Yacef, Educational data mining: a case study. Artificial Intelligence in education: supporting learning through Socially Informed Technology.–IOS Press, 2005, 467-474.
[35]. F. Castro, A. Vellido, A. Nebot, F. Mugica, F. Applying data mining techniques to e-learning problems. Evolution of teaching and learning paradigms in intelligent environment, Springer Berlin Heidelberg, 2007, 183-221.
[36]. J. Luan, Data mining and its applications in higher education. New directions for institutional research, 2002, vol. 2002 No. 113, 17-36.