International Journal of Computing Algorithm
Volume: 04 Issue: 01 June 2015 Pages:26-29
ISSN: 2278-2397

EDM: Finding Answers to Reduce Dropout Rates in Schools

Chokanath Hymavathy[1], Viola R.Krishnamani[2], C.P.Sumathi[3]
[1]Research Scholar, Research Department of Computer Science, SDNB Vaishnav College for Women, Chromepet, Chennai, India
[2]Chief Executive Officer, Scholarius Educational Services, Coimbatore, Tamil Nadu, India
[3]Associate Professor & Head, Research Department of Computer Science, SDNB Vaishnav College for Women Chromepet, Chennai Tamil Nadu, India
Email:hyma@qedquest.com, viola@scholarius.net.in, drcpsumathi@gmail.com

Abstract-The focus of this paper is to get a bird's eye view of the various factors that could be analyzed to present schools and governments timely indicators of school dropouts. With the help of EDM, one could have access to enormous amounts of data. The crux of the matter would be to identify the data in line with plans and ideas that would eventually lead to the formulation of policies that enhance the teaching and learning process.

Keywords—EDM; Educational Data Mining; BROMP; Learner; Dropout indicators

I. INTRODUCTION

The education sector has never been in a state of churn as it is right now - this is a statement with specific reference to India, a country that has always prided itself on robust practices and systems in the field of education. On the one hand there are various teaching methodologies, based on both the ancient and the modern; and on the other hand, the scope, depth and variation of subjects (or disciplines) have multiplied many times over. Recent advances in technology have rendered the situation even more complex. This is in spite of the fact that the primary purpose of all technology-driven content is to access information easily and in a timely manner. With data being generated in incredible volumes facility of access has, in many situations, become more virtual than real.

A. What is EDM?

Education Data Mining, or EDM as it is more commonly known, is very akin to finding a needle in a haystack… that is if you are not clear about what you are looking for! Witten and Frank have spoken about this extensively in their well known publication, *Data Mining: Practical Machine Learning Tools and Techniques*. They have refered to the myriad applications that are possible with the enormous quantum of data that is generated every second of the day. The solution however is to find what one wants at the right time from the right source and put it to good and effective use. "As the volume of data increases, inexorably, the proportion of it that people understand decreases, alarmingly. Lying hidden in all this data is information, potentially useful information that is rarely made explicit or taken advantage of."[1]. While on the subject of EDM, it would be relevant to point out the various stages at which the data mined is processed and used. There are stages were the data is 'pre-processed.' "Pre-processing allows transforming the available raw educational data into a suitable format ready to be used by a data mining algorithm for solving a specific educational problem." Romero *et al* went on to describe the ways in which algorithms and learning analytics techniques would be put to use to ensure that the data mined is presented and understood with the maximum amount of clarity.[2] In short, the data mined would be useful only if the EDM techniques used were appropriate and had the bandwidth to accommodate the analyses of the data collected. The purpose of this paper is to highlight the ways in which data collected is made use of in various ways by various key stakeholder groups in the education industry. The case in question here focuses on the government-sponsored education ecosystem in Tamil Nadu, India. Data is collected from various sources, focusing on many specific as well as general areas of interest. The key point however is this: with all this information that is put together over a period of time, what amount of data is made use of, for bettering the state of the whole ecosystem in general and the learning levels of students in particular? More specifically - how does the government education system identify, manage and make use of dropout indicators that are churned out by the vast amount of data collected. It is a question that demands a very detailed answer, based on informed sources.

II. RATIONALE FOR RESEARCH

The 'when', 'what' and the 'how' of data collection is certainly not something that researchers have a problem dealing with. Advances in collection and collation of data have ensured that the methods used are not just tested but also customized to suit the need of specific fields of research. The recipient of all such data collected is able to access humongous amounts of data covering wide-ranging timeframes and study groups. The 'when' and 'what' of data collection could be mildly challenging in situations where the primary question is still unanswered - 'Why'?Why does one need this data? Why does a stakeholder group need it now? Why is this data relevant? Why collect data from a particular source… Far from being rhetorical, these are questions that every researcher asks himself or herself before embarking on that statistical journey of data collection, collation and analysis.

Specifically, in this situation, the need for this research is to understand how to use the various indicators that are available to mitigate dropout rates in the school education system. Therefore, there is an answer one is looking for to the 'why' questions that are raised. Considering the fact that (more than is required) data is available, the answers are not always clearly apparent. Therefore, the key to using data lies in identifying the reason or reasons for collecting data in the first place. Clarity of thought and action at the pre-data collection

International Journal of Computing Algorithm
Volume: 04 Issue: 01 June 2015 Pages:26-29
ISSN: 2278-2397

stage is most likely to provide clear-cut answers and indicators that help educators manage learners in general and dropouts in particular in the public schooling sector.

III. THE SIS STATEMENT

Why collect data? Of course, the task is not an easy one, considering the fact that data collected is primarily sent back to the organization(s) that mandated their collection, with a minimum amount of analysis. Rarely is this data collated to produce wholesome indicators of learning levels in students in the public education system. This paper attempts to list out various indicators that could shed light on performances and non-performances of key stakeholder groups in the sector. These indicators could go a step further and present predictive analytics that would be of great benefit to decision makers in schools as well as in the government. The focal point of this paper is to provide an understanding of how various kinds of data can be assembled and used, following which relevant indicators would help mitigate dropout rates in schools. In other words, the paper tries to provide a glimpse into the sea of indicators available to managerial stakeholders in the education ecosystem in Tamil Nadu. While focusing on specific sources of data, it would be relevant to point out that not all data collected is directly related to the level and range of dropouts in schools. This is where the skill of the analyzer steps in. The key to using the data lies in identifying which data, from which source, could lead to the right kind of information within a specified timeframe. Therefore an understanding of various sources of data generated for and by the government in Tamil Nadu, is crucial to implementing plans for educational restructuring and dropout-mitigation in the public school system.

IV. LITERATURE REVIEW

While focusing on EDM, a lot of emphasis has been placed on the study of the student. Very often one needs to worry about what actually makes a student tick. The answer cannot be given in a couple of words because of the various strands that come together to make a conducive learning environment. Should there be conventional classrooms with rigid structures and curriculums? Should there be student-led inquiry that gives every child a chance to think and learn - where every child matters? Does learner engagement have a direct impact on dropout rates in urban schools? Do ethnic differences play a part in identifying dropout trends?The analysis of the extensive data that is generated helps focus on myriad issues. In the eyes of an educational data analyst, the questions are as varied as the data available. The use of a CHAID prediction model by M. Ramaswami and R. Bhaskaran gives insights into the various indicators that could be of use in a study on dropout rates. [3]Vellido, Castro and Nebot in their essay on Clustering Educational Data point out the close relationship between existing data modeling techniques and new information technologies and database development. This can be further studied in varied learning environments such as e-Learning and any other blended learning situations. [4]Hymavathy, Krishnamani&Sumathi have spoken at length about learner engagement and how it can be factored into the policy-making process on the parts of teachers, managerial stakeholders and governments. Studying learner engagement has tremendous potential because one could get a better insight into whether or not a student has the required level of involvement.

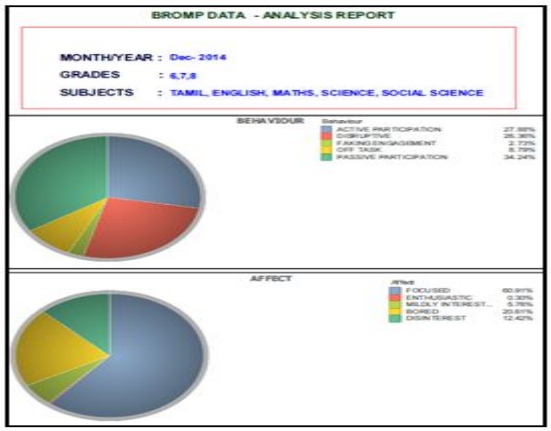

Figure:1

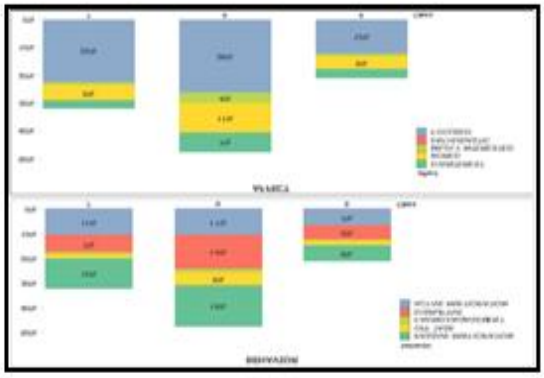

Figure :2

This would in turn indicate whether students had or have a tendency to drop out of school at any specific point of time. [5]Understanding ethnic identities would be yet another important factor as described by Aud, Fox and Ramani. Thought this reference might not be specific to Tamil Nadu as such, the study throws up interesting insight into the role of ethnicity and racial differences on the study profile of students.[6]Thakur and Dev in their study on the Indian educational scenario, made pointed references to the various causes of dropouts in various kinds of schools and sectors. Poverty is probably one of the main causes; different states have shown that basic numerical and linguistic knowledge is grossly lacking at primary and lower secondary levels. [7]

V. METHODOLOGY

Understanding how data can help leads us to two main question points, which can be briefly listed as follows:

1. What are the various kinds of data that are available?
2. What are the dropout indicators that are necessary to focus on?In this paper we take a look at some of the important kinds of data that are generated every day in the public education system. Till very recently, the data available was more or less based on student numbers and other basic demographic details. Individual school heads and sector

International Journal of Computing Algorithm
Volume: 04 Issue: 01 June 2015 Pages:26-29
ISSN: 2278-2397

coordinators had access to assessment details of students. Today, the scenario has changed; from hand-written registers, the government has moved to more analysis-friendly data collected from various stakeholder groups in varying quantum and ranges. Various kinds (sources) of data available: School-based statistics - from the number of students in a class, to the marks they get and the attendance particulars of various classes, schools are able to generate data for every term. Academic data collected focuses on subject-wise achievements of students; one is also able to glean information about the non-performers as well.

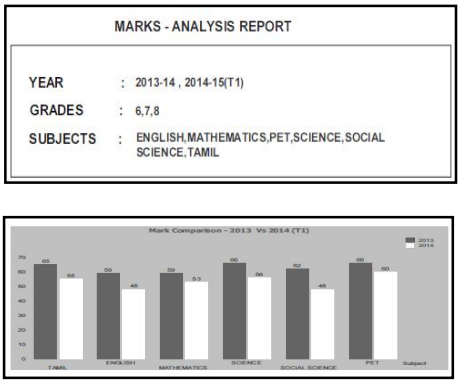

Figure 3: Marks Analysis Report

Learner engagement studies - what students study is governed by syllabuses and subject-requirements; how the same students study what they study is a completely different matrix that has to be studied. This is can be done with a digital application called HART (Human Affect Recording Tool) that is part of an innovative system called BROMP (Baker Rodrigo Observation Method Protocol)[8]- UDISE data - Unified District Information System for Education - developed by India's premier institution NUEPA (National University for Educational Planning and Administration) - provides wide-ranging data that covers information about the various levels in schools, best practices adopted, infrastructure details, programs and protocols introduced, reports and findings of various committees.

Most of the statistical data provided herein points to a few simple facts on how mined data can be used to enhance the teaching and learning experience of students in government-sponsored schools. - Though using dropout indicators could be the best way forward to arrest or mitigate the level of dropouts in government schools; it is important to identify the right indicator - would it be socio-economic indicators or would they be academic indicators? These are questions that would need honest answers before planning the way ahead. Holding up a flag based on an indicator requires both a sense of responsibility as well as a complete understanding of the macro picture. EMIS data - Educational Management Information System - this gives comprehensive information on the socio-economic status of students and their backgrounds. For instance, details of first-generation learners, the income of parents, literacy details etc - these would give decision makers sufficient grounds to conceptualize and implement various kinds of programs. Teacher profile data would help school authorities put together professional development programs and institute beneficial classroom practices to enhance the levels of teaching and learning in an institution.

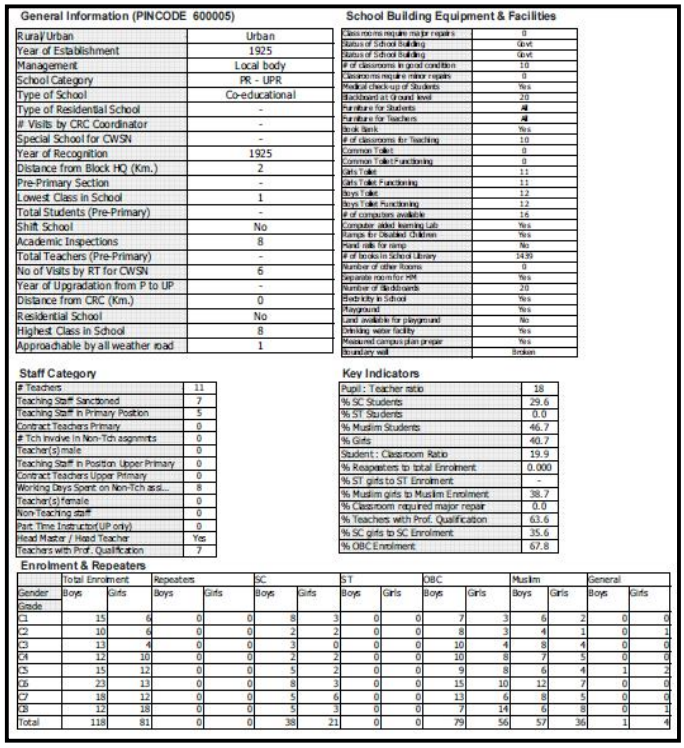

Figure 4: UDISE Data

Dropout indicators to focus onAlex Bowers has presented a comprehensive list of indicators that could help schools collect an enormous amount of data.[9] Before using this list of indicators, it would be relevant to focus on these points first:

- The socio-economic factors - of student, family and community
- Age and grade level of the student
- Curriculum specifications and requirements
- Governmental educational policies
- Trends in student preferences for higher education courses
- Cost factors - at school as well as higher education level

Though the list given above is by no means exhaustive, it serves to understand the range of factors that could impact the continuance of a student in school.

In the case in question, which is the Tamil Nadu school education ecosystem, one of the major factor groups that one needs to watch out for is the socio-economic factors. The dropout rates in schools largely depend on whether or not families are stable. Of course, stability again depends on the income, nature of job and the literacy level of both parents. In addition to this, it is important to remember that external factors such as alcoholism and drug-addiction in parents, could play an important role in study-life of a student.It is also significant to note that dropout indicators are very region-specific and level-specific. To understand this better, it would be good to have a closer look at the statistics presented in this paper. Since the focus is on students from low-income groups, the probability of dropouts is high in certain regions that have high crime rates; similarly students from lower grades remain fairly steadfast and attendance is quite steady. The same however might not happen in the higher classes.

International Journal of Computing Algorithm
Volume: 04 Issue: 01 June 2015 Pages:26-29
ISSN: 2278-2397

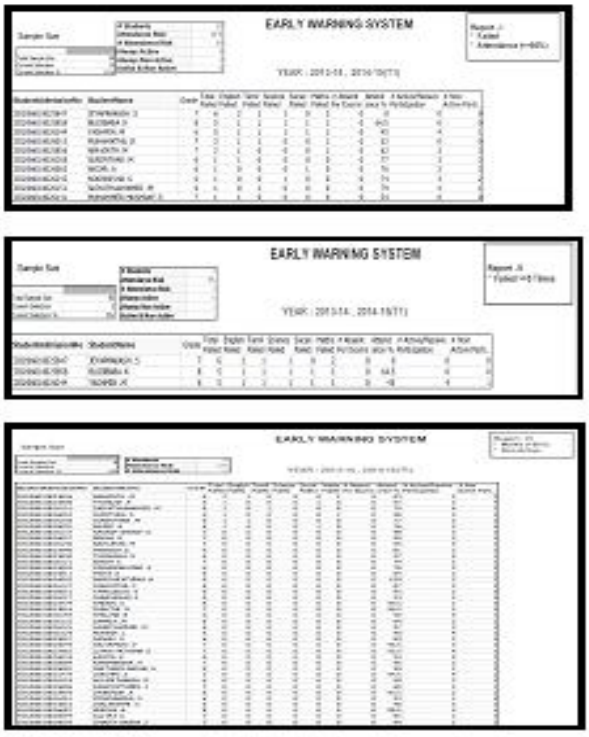

Figure:5

VI. RECOMMENDATIONS AND SUGGESTIONS

One cannot read into statistics without an understanding of the macro picture for one simple reason that there is a tendency to rely on one indicator instead of reading a couple of indicators together.Using dropout indicators to change policy decisions on professional development qualifications for teachers could be a major decision for the government or any stakeholder in school management. Though it is not an easy decision to take, it is certainly one that could use the right indicator(s) to provide insight into the nature, scope, duration and quality of professional development qualifications for teachers.

VII. CONCLUSION

UDISE, BROMP, Learner & Teacher Profiles, Enrolment & Dropout statistics… the list could be customized to suit the aims and objectives of the study. Data thus collected would be sufficient to create new ideas and policies in the field of education, thereby bringing about a change in the way in which public education is actually designed and implemented.Since public education comprises policies for both the teacher as well as the taught, it would be extremely essential to understand the scope of the statistics collected before the same is used in any situation. Early indicators could help devise early intervention programs that would in turn be the foundation for multi-stakeholder-driven initiatives. The important point therefore would be to recognize and monitor early indicators. Goss and Andren have listed out indicators that are more universal than specific to one geographical region. Student engagement, school readiness, behavior, attendance… the list indicates common risk factors that could play the part of early dropout indicators. [10]. In the last decade, a great amount of data that has been generated in the education sector has taken a long time to trickle back into the system. In other words, data is collected, but rarely used effectively. It is hoped that with predictive analytics and modeling, the government and other stakeholders in the public education ecosystem in the country will have sufficient indicators that will lead to the construction of robust programs and sustainable teaching-learning practices.

REFERENCES

[1] I.H. Witten, E. Frank and M.A. Hall. “Data Mining: Practical Machine Learning Tools and Techniques.” Amsterdam: Elsevier, 2011

[2] C. Romero, J. R. Romero & S. Ventura, "A Survey on Pre-Processing Educational Data: Educational Data Mining: Application and Trends". Ed. by Alejandro Pena-Ayala. Switzerland: Springer, 2014

[3] M.Ramaswami and R.Bhaskaran: “A CHAID Based Performance Prediction Model in Educational Data Mining”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No.1, January 2010, ISSN (Online): 1694-0784, ISSN (Print): 1694-0814

[4] A. Velido, F. Castro & A. Nebot, “Clustering Educational Data: Handbook of Educational Data Mining”. Ed by C. Romero, S. Ventura, M. Pechenizkiy, Ryan SJD Baker. Boca Raton: CRC Press, 2010

[5] Hymavathy, C. ; Krishnamani, V.R. ; Sumathi, C.P. : “Analyzing learner engagement to enhance the teaching-learning experience” , IEEE MOOC, Innovation and Technology in Education (MITE), 2014 IEEE International Conference on, 19-20 Dec 2014

[6] S. Aud, M.A. Fox & A. KewalRamani, “Status and Trends in the Education of Racial and Ethnic Groups.” (NCES 2010-015). U.S. Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office

[7] A.K. Thakur & M. Dev. "Education Growth and Development." New Delhi: Deep & Deep Publications Pvt. Ltd. 2010

[8] J. Ocumpaugh, R. S.J.d. Baker, Ma. M.T. Rodrigo, “Baker-Rodrigo Observation Method Protocol (BROMP) 1.0 Training Manual version 1. 0.” Oct. 17, 2012.

[9] Alex J. Bowers, Ryan Sprott & Sherry Taff: “Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity and Specificity”, 2013.

[10] C.L. Goss & K.J. Andren. "Dropout prevention" New York: Guilford Publications. 2014.