

Survey Of Data Mining In Socio-Academic Perspective

Prof. K. Vijayalakshmi M.C.A., M.Phil.,

Abstract : This paper surveys the use of data mining in educational systems by improving the learning and assessment tasks more effectively in order to lower the rate of percentage in unemployment. Learning process has to be focused much in order to attain a better assessment. The optimisation of system to support learning can be predicted easily through the model obtained from Learning Analytics(LA). Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from educational environments. Currently there is an increasing interest in data mining and educational systems, making educational data mining as a new growing research community.

Index Terms: Data Mining(DM), Educational Data Mining(EDM), Learning Analytics (LA), Unemployment.

1 INTRODUCTION

Educational Data Mining (called 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. A key area of EDM is mining student performance and enrollment data. The essential uses of EDM include predicting student performance and to recommend improvements to current educational practice. Data mining techniques can discover useful information that can be used in formative evaluation to assist educators to establish a pedagogical basis for decisions when designing or modifying an environment or teaching approach.

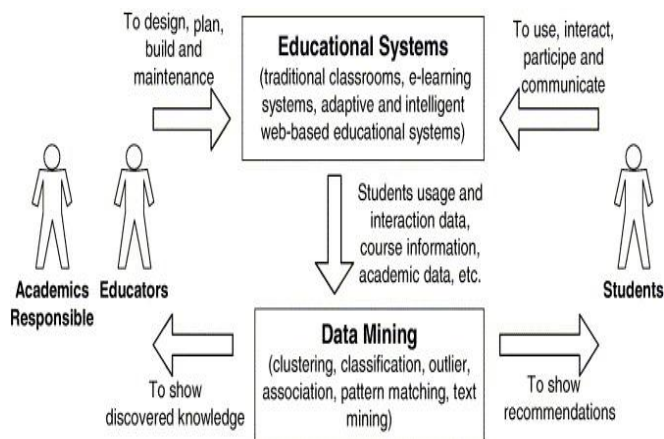


Fig 1. The cycle of applying data mining in educational systems.

The application of data mining in educational systems is an iterative cycle of hypothesis formation, testing, and refinement (Fig. 1). Mined knowledge should enter the loop of the system and guide, facilitate and enhance learning as a whole. Not only turning data into knowledge, but also filtering mined knowledge for decision making. As in Fig. 1, educators and academics are responsible of designing, planning, building and maintaining the educational systems. Students use and interact with them. The discovered knowledge can be used not only by providers (educators) but also by own users (students). So, the application of data mining in educational systems can be oriented to different actors with each particular point of view. Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning. Learning analytics is defined and implemented may vary including:

- To assess their individual achievements and patterns of behaviour in relation to others
- Predicting students required extra support and attention
- to help teaching and non teaching to plan for individuals as well as groups
- to improve current course plans or develop new curriculum offerings as a functional achievement and
- to take decisions on marketing and recruitment by institutional administrators on efficiency and effectiveness measures.

2 LEARNING ANALYTICS Vs EDUCATIONAL DATA MINING

EDM encompasses both learning analytics and academic analytics, the former of which is aimed at governments, funding agencies, and administrators instead of learners and faculty. Learning analytics is an area that combines select institutional data, statistical analysis, and predictive modeling to create intelligence upon which learners, instructors, or administrators can change academic behavior. It can be unambiguous that both can be differentiated based on whether the process is hypothesis driven or not. Regardless of the differences between the LA and EDM communities, the two areas have significant overlap both in the objectives of investigators as well as in the methods and techniques that are used in the investigation. Software used for LA applies to learner interactions with content. LA software tools are:

- Prof. K. Vijayalakshmi is currently working as an Associate Professor in the Department of M.C.A, Reva Institute of Technology & Management, Bangalore India. E-mail: vijaykrishna_2000@yahoo.com

- SNAPP is a learning analytics tool that visualizes the network of interactions resulting from discussion forum posts and replies.
- LOCO-Analyst is a context-aware learning tool for analytics of learning processes taking place in a web-based learning environment
- SAM is a Student Activity Monitor intended for Personal Learning Environments
- BEESTAR INSIGHT is a real-time system that automatically collects student engagement and attendance & provides analytics tools and dashboards for students, teachers & management

2.1 THE CONTEXT OF LEARNING ANALYTICS

- The increasing interest in 'big data' for business intelligence and application of BI techniques to educational data
- The rise of online education focussed an increase in digital data regarding student background and learning log data for the deliverance of high quality affordable education.
- The optimisation of system to support learning and increasing focus on evidencing progress and professional standards for accountability systems
- increasing emphasis was placed on the pedagogic affordances of learning analytics

2.2 TECHNIQUES OF LEARNING ANALYTICS

Learning Analytics draws techniques from:

- Statistics to address hypothesis testing
- Business Intelligence to enable data access and summarise performance indicators.
- Web analytics tools such as Google analytics report on web page visits and references to websites, brands and other keyterms across the internet.
- Operational research aims at highlighting design optimisation for maximising objectives and these techniques are implicated in LA to create models of real world behaviour for practical application.
- Artificial intelligence and Data mining methods are capable of detecting patterns in data.
- Social Network Analysis used to explore clusters of networks, influence networks, engagement and disengagement, and has been deployed for these purposes in learning analytic contexts.
- Information visualization is to sense making around the data provided.

2.3 METHODS OF LEARNING ANALYTICS

Methods for learning analytics include:

- Social network analysis (SNA) provides both a visual and a mathematical analysis of human relationships.
- Behavioral trust analysis uses instances of conversation and propagation as an indicator of trust.
- Influence and passivity measure assess the influence of people and information by measuring the number of times it is passed on, cited, or retweeted.
- Discourse Analytics aims to capture meaningful data on student interactions which aims to explore the

properties of the language used, as opposed to just the network of interactions, or forum-post counts, etc.

- Social Learning Analytics explores the role of social interaction in learning, the importance of learning networks, discourse used to sensemake, etc.
- Disposition Analytics seeks to capture data regarding student's dispositions to their own learning, and the relationship of these to their learning.
- Impact of interaction
 - ✓ Prediction
 - ✓ Personalization & Adaptation
- Intervention is the information visualization in the form of learning dashboards.

3 WHAT IS ASSESSMENT?

Assessment involves systematically gathers, analyze, and interpret the evidence to determine how well performance matches those expectations and standards, and using the resulting information improve performance (Fig 2). Classroom Assessment is a simple method to provide faculty and students with information and insights needed to improve teaching effectiveness and learning quality. It is an approach designed to help teachers find out what students are learning in the classroom and how well they are learning it.

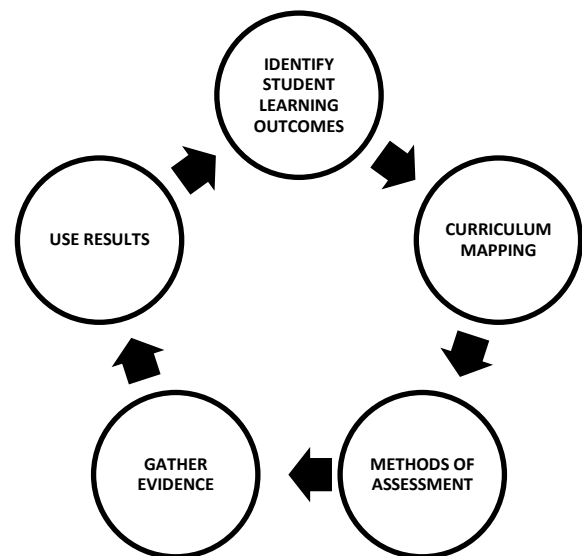


Fig 2. Assessment cycle

Educational assessment is the process of documenting in measurable terms such as knowledge, skills, attitudes, and beliefs. Assessment can focus on the individual learner, the learning community, the institution, or the educational system as a whole. The final purposes of assessment practices in education depends on the theoretical framework of the practitioners and researchers, their assumptions and beliefs about the nature of human mind, the origin of knowledge and the process of learning.

3.1 TYPES OF ASSESSMENT

Assessment is divided using the following distinctions:

- initial, formative, and summative
- objective and subjective
- referencing
- formal and informal.

3.1.1 INITIAL, FORMATIVE AND SUMMATIVE

Assessment is often divided into initial, formative, and summative categories for the purpose of considering different objectives for assessment practices.

- Initial assessment is a pre-assessment or diagnostic assessment conducted prior to instruction from which individual student growth can be measured.
- Formative assessment is an educative assessment used to aid learning.
- Summative assessment is carried out at the end of a course or project.

Assessment of learning is generally summative in nature and intended to measure learning outcomes and report those outcomes to students, parents and administrators.

3.1.2 OBJECTIVE AND SUBJECTIVE

Objective assessment is a form of questioning which has a single correct answer. Subjective assessment is a form of questioning which may have more than one correct answer. Objective assessment is well suited to the increasingly popular computerized or online assessment format.

3.1.3 REFERENCING

Referencing is the process of acknowledging your sources. They are classified as follows:

- Criterion-referenced assessment occurs when candidates are measured against defined criteria.
- Norm-referenced assessment is known as grading on the curve which is effectively a way of comparing students.
- Ipsative assessment is self comparison that is measuring yourself against yourself either in the same domain over time, or comparative to other domains within the same student.

3.1.4 FORMAL AND INFORMAL

Formal assessment usually a written document, such as a test, quiz, or paper. It is given a numerical score or grade based on student performance. Informal assessment does not contribute to a student's final grade. It usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performance and portfolio assessments, participation, peer and self-evaluation, and discussion.

INTERNAL AND EXTERNAL

Assessment may be either Internal or external in nature. Internal assessment is set and marked by the school (i.e. teachers). External assessment is set by the governing body, and is marked by non-biased personnel.

3.2 STANDARDS OF QUALITY ASSESSMENT PARAMETERS

High-quality assessments are considered those with a high level of reliability and validity.

3.2.1 RELIABILITY

Reliability relates to the consistency of an assessment. A reliable assessment is one which consistently achieves the same results with the same cohort of students. The reliability of an assessment is based on the following:

- Temporal stability is the performance on a test is comparable on two or more separate occasions.
- Form equivalence is the performance among examinees is equivalent on different forms of a test based on the same content.
- Internal consistency is the responses on a test are consistent across questions.

3.2.2 VALIDITY

A valid assessment is one which measures what it is intended to measure. Basically on

- Does the content of the test measure stated objectives?
- Do scores correlate to an outside reference?
- Does the assessment correspond to other significant variables?
- Does the item or theory make sense, and is it seemingly correct to the expert reader?

A good assessment has both validity and reliability. The subject-matter validity used widely in education and predictive validity used widely in the workplace, predicts performance. Thus, a subject-matter valid test of knowledge of student admission rules is appropriate while a predictively valid test would assess whether the student could follow those rules.

3.2.3 EVALUATION STANDARDS

In the field of evaluation, and in particular educational evaluation, the Joint Committee on Standards for Educational Evaluation has published three sets of standards for evaluations.

- The Personnel Evaluation Standards was published in 1988
- The Program Evaluation Standards (2nd edition) was published in 1994, and
- The Student Evaluation Standards was published in 2003.

Each publication presents and elaborates a set of standards for use in a variety of educational settings. The standards provide guidelines for designing, implementing, assessing and improving the identified form of evaluation. Each of the standards has been placed in one of four fundamental categories to promote educational evaluations that are proper, useful, feasible, and accurate.

4 EDM METHODS

The types of EDM method are related to those found in data mining in general, but with some differences based on the unique features of educational data. Ryan Baker classifies the areas of EDM as follows:

- Prediction
 - ✓ Classification
 - ✓ Regression
 - ✓ Density estimation
- Clustering
- Relationship mining
 - ✓ Association rule mining
 - ✓ Correlation mining
 - ✓ Sequential pattern mining
 - ✓ Causal data mining
- Distillation of data for human judgment
- Discovery with models

Baker and Kalina Yacef claim that discovery with models is particularly prominent in EDM, as compared to data mining in general. In discovery with models, a model of a phenomenon is developed through any process that can be validated in some fashion and this model is then used as a component in another analysis, such as prediction or relationship mining.

Tool name	Authors	Mining task
Mining tool	Za€iane and Luo (2001)	Association and patterns
MultiStar	Silva and Vieira (2002)	Association and classification
Data Analysis Center	Shen et al. (2002)	Association and classification
EPRules	Romero et al. (2003)	Association
KAON	Tane et al. (2004)	Text mining and clustering
TADA-ED	Merceron and Yacef (2005)	Classification and association
O3R	Becker et al. (2005)	Sequential patterns
Synergo/ColAT	Avouris et al. (2005)	Statistics and visualization
GISMO/CourseVis	Mazza and Milani (2005)	Visualization
Listen tool	Mostow et al. (2005)	Visualization
TAFPA	Damez et al. (2005)	Classification
iPDF_Analyzer	Bari and Benzater (2005)	Text mining

Table 1: EDM tools

4. 1 PROCESS MINING ON EDUCATIONAL DATA

The process of mining on educational data consists of three subfields:

- Conformance checking
- Model discovery
- Model extension

Process Mining Educational Data Set consists of:

- Data Preparation
- Visual Mining with Dotted Chart Analysis
 - ✓ Conformance Analysis
 - ✓ Conformance Checking
- LTL Analysis
- Process Discovery with Fuzzy Miner

In modern education various information systems are used to support educational processes. In the majority of cases these systems have logging capabilities to audit and monitor the processes they support. These data can be analyzed from various levels and perspectives, showing different aspects of organization, and giving us more insight in the overall educational system (Fig 3) focus on integrating domain knowledge in the mining process.

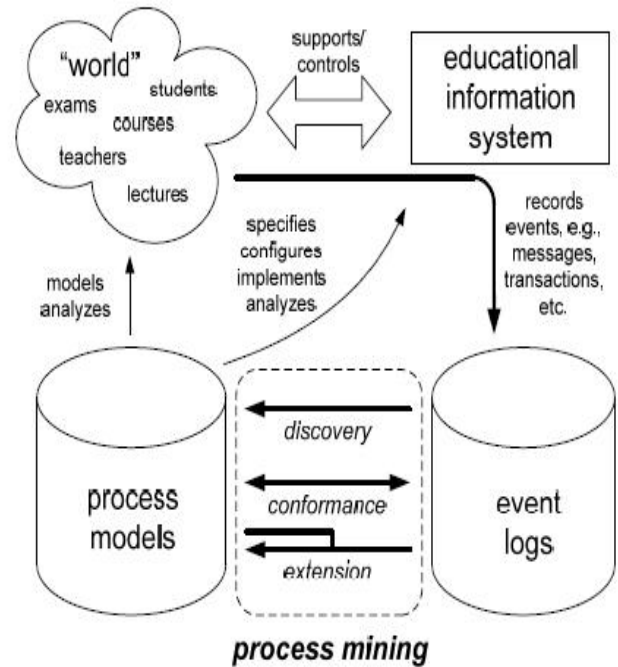


Fig 3. Process Mining on Educational Data

However, with the development and increasing popularity of blended learning and e-learning, information systems enable us to capture activities also at different levels of granularity.

4.2 APPLICATIONS

The primary applications of EDM are:

- Analysis and visualization of data
- Providing feedback for supporting instructors
- Recommendations for students
- Predicting student performance
- Student modeling
- Detecting undesirable student behaviors
- Grouping students
- Social network analysis
- Developing concept maps
- Constructing courseware
- Planning and scheduling

5. GLOBAL CHALLENGES IN EDUCATION

Most important is the challenge of the change of mindset. Various challenging parameters are access, quantity and quality & pedagogy. Impart quality education to increase numbers in order to develop learning metrics. The art of teaching should be move to the science of learning and improve the capacity of the students in how to learn. Adopt 21st century communication tools for enhanced learning experience. Even after applying all in the learning and teaching process, why students still fail to learn?. Probably the reasons are:

- Technology mislead them in wrong direction.
- They are learned helplessness
- Sometimes language used in the course make them that learning is very difficult.
- Their learning goals are not articulated.
- Lack of adequate reading and comprehensive rate.
- Not devoting time required for the learning tasks.
- Inadequacy of available learning technologies.
- Not ready for the present learning intervention.
- Choice of education is taken by parents rather than children.

Learning styles are different ways that a person can learn. It's commonly believed that most people favor some particular method of interacting with, taking in and processing stimuli or information. The basic fundamental learning styles are

- visual learners have a preference for seeing : pictures, visual aids such as overhead slides, diagrams, handouts, etc..
- auditory learners best learn through listening lectures, discussions, tapes, etc..
- reading-writing preference learners;
- kinesthetic learners or tactile learners prefer to learn via experience : moving, touching, and doing active exploration of the world; science projects; experiments, etc..

Also Joy Reid defined learning styles according to perceptual preferences:

- visual - written or visual information
- auditory - verbal
- kinesthetic - physical activity
- tactile - working with materials
- social group
- social individual

Benfits of changing the learning & teaching protocols are listed below:

- Make strategies to keep participants on task.
- Ensure an equal exchange of ideas, feedback, and expertise even from less-verbal participants.
- Use a dialogue and reflection before decision making.
- Learn a collaborative process for developing ground rules.
- Promote a safe and supportive structure for discussing challenges.

Now it is required to upgrade education to include skills. The board should have both academic and industry experts to raise the standards of the syllabus. Institutions may choose to include an appropriate combination of the following in their assessment documentation

- to ensure that assessment is efficient, effective, and purposeful, and supported with appropriate resources and that results are used appropriately.
- providing information needed to carry out assessment practices

Twenty-first century learning is about the process of integrating and using knowledge, not just the acquisition of facts and procedures. Hence, educators need to build assessments for learning, rather than solely of learning. If we are to help students succeed in a 21st-century economy and society, we must find ways to measure their ability to apply knowledge to complex and challenging tasks, and to behave in other ways that predict successful engagement in the world as it is now.

5. 1 ASSESSMENT IN THE AGE OF INNOVATION

Within the past 50 years, we've seen our country move from an industrial economy to an information-based economy. Now, early in the 21st century, it appears we are shifting to an innovation-based economy, one that requires what the psychologist Robert J. Sternberg calls successful intelligence, a three-point foundation of analytical, practical, and creative skills. The Internet enables some of the best teaching minds to bond together in powerful professional learning communities. Teachers are expected to adhere to leadership as traditionally defined in hierarchical organizations. Most teacher leadership is still not viewed as a means to effect meaningful change in education

6. INDIA UNEMPLOYMENT RATE

Unemployment Rate in India decreased to 3.8 % in 2012 from 9.40% in 2010. Historically, from 1983 until 2011, India Unemployment Rate averaged 7.57% reaching an all time high of 9.4% in December of 2010 and a record low of 3.8% in December of 2011. In India, the unemployment rate measures the number of people actively looking for a job as a percentage of the labour force.

6.1 UNEMPLOYMENT

The unemployment rate can be defined as the number of people actively looking for a job divided by the labour force.

$$\text{Rate of UE} = \frac{\text{No. of people looking for a job}}{\text{labour force}}$$

The percentage of UE is less only when labour force is more than the number of people looking for a job. Engineering roles are at the top of the pay scale for fresh hires in India, and saw the highest salary increase this financial year, outperforming other jobs like sales, HR, administration and support services.

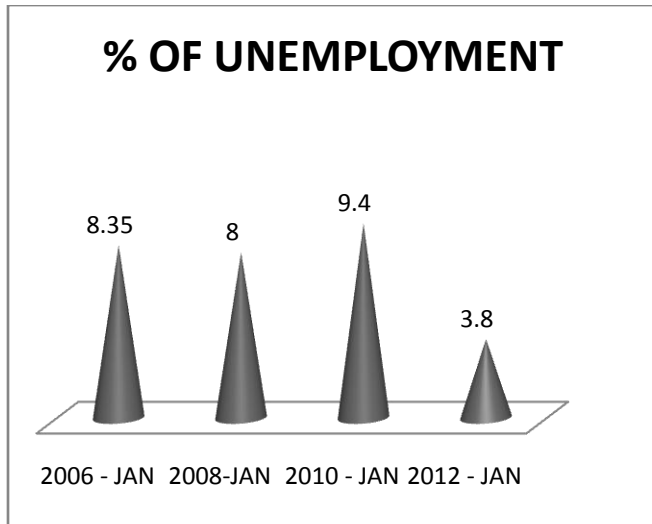


Fig4. Unemployment percentage in India

The median monthly salary of graduate engineers in engineering roles has increased to Rs 25,000 in the current fiscal, from Rs 22,500 last year, a 11% increase, says a survey by global management consultancy Hay Group. Median salaries for HR professionals went up 10.1% to Rs 20,923, while those for administration and support roles increased 8.2% to Rs 18,500. The entry-level salaries in non-IT engineering companies are comparatively higher because of the supply-demand mismatch. Most companies are fighting for the same pool of talent. The war for talent is not about headcount, but based on skillsets.

7. Conclusion

Educational data mining is an upcoming field related to several well-established areas of research including e-learning, adaptive hypermedia, intelligent tutoring systems, web mining, data mining, etc. The application of data mining in educational systems has specific requirements not present in other domains, mainly the need to take into account pedagogical aspects of the learner and the system. Educational data mining is a young research area and it is necessary more specialized and oriented work educational domain in order to obtain a similar application success level to other areas, such as medical data mining, mining e-commerce data, etc. Once if we tune well the learning and assessment process, the performance of the students in turn will be increased implies the rate of unemployment will get decreased. From the fig 4, comparing Jan 2010 to Jan 2012, there is a immense reduction in the rate of unemployment by 5.6 that shows that there is a possibility in future that the rate of unemployment could be probably 0. Since

9.4 -----> 3.8 in the year 2012
3.8-----> 0 in 2020 atleast.

Some future recommendations are:

- Mining tools must have a more intuitive and easy to use interface, with parameter-free data mining algorithms to simplify the configuration and execution, and with good visualization facilities to make their results meaningful to educators and e-learning designers.
- The data mining tool has to be integrated into the e-learning environment as another author tool. Feedback and results obtained with data mining can be directly applied to the e-learning environment.
- Education-specific mining techniques can help much better to improve the instructional design and pedagogical decisions.

ACKNOWLEDGMENT

I take this opportunity to thank all the people who involved for completing this work more comfortably and successfully. Also I extend my sincere thanks to my management for the extended support and motivation they have given ever for the academic promotional activities.

REFERENCES:

- [1]. Leite, Walter L.; Svinicki, Marilla; and Shi, Yuying: Attempted Validation of the Scores of the VARK: Learning Styles Inventory With Multitrait–Multimethod Confirmatory Factor Analysis Models, pg. 2. SAGE Publications, 2009.
- [2]. Thomas F. Hawk, Amit J. Shah (2007) "Using Learning Style Instruments to Enhance Student Learning" Decision Sciences Journal of Innovative Educationdoi:10.1111/j.1540-4609.2007.00125.x
- [3]. Black, Paul, & William, Dylan (October 1998). "Inside the Black Box: Raising Standards Through Classroom Assessment." Phi Beta Kappan. Available at PDKintl.org. Retrieved January 28, 2009.
- [4]. Earl, Loma (2003). Assessment as Learning: Using Classroom Assessment to Maximise Student Learning. Thousand Oaks, CA, Corwin Press. ISBN 0-7619-4626-8. Available at [1], Accessed January 23, 2009.
- [5]. Reed, Daniel. "Diagnostic Assessment in Language Teaching and Learning." Center for Language Education and Research, available at Google.com. Retrieved January 28, 2009.
- [6]. Educational Technologies at Virginia Tech. "Assessment Purposes." VirginiaTech DesignShop: Lessons in Effective Teaching, available at Edtech.vt.edu. Retrieved January 29, 2009.
- [7]. Yu, Chong Ho (2005). "Reliability and Validity." Educational Assessment. Available at Creative-wisdom.com. Retrieved January 29, 2009.
- [8]. Moskal, Barbara M., & Leydens, Jon A (2000). "Scoring Rubric Development: Validity and Reliability." Practical Assessment, Research & Evaluation, 7(10). Retrieved January 30, 2009 from [3]

- [9]. Committee on Standards for Educational Evaluation. (2003). *The Student Evaluation Standards: How to Improve Evaluations of Students*. Newbury Park, CA: Corwin Press.
- [10]. Fadel, Charles, Honey, Margaret, & Pasnik, Shelley (May 18, 2007). "Assessment in the Age of Innovation." *Education Week*. Retrieved January 29, 2009

Mrs. K. Vijayalakshmi obtained her M.C.A from Bharathidasan University & M.Phil from Manonmaniam Sundaranar University, Tamil Nadu, India. She is currently an Associate Professor in the Department of M.C.A, Reva Institute of Technology & Mangement, Affiliated to V.T.U, Bangalore, Karnataka, India. She has 12 years of teaching experience & presented papers in national conferences and published papers in journals.