International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.111.08

A Study of Easy Educational Data Mining for E-Learning Log Data from complex and large Dataset

S. Arumugam[1], A. Kovalan[2], A.E.Narayanan[3]
[1]*Department of Computer Science and Applications,*
Periyar Maniammai Institute of Science and Technology, Thanjavur, India.
[2]*Professor, Department of Computer Science, Theivanai Ammal College for Women, Villupuram, India*
[3]*Asst.Prof.(SG), Department of Computer Science Engineering,*
Periyar Maniammai Institute of Science and Technology, Thanjavur, India.

Abstract - Datamining(DM) techniques are widely adopted for finding out new knowledge from huge and complex data sets. The use of interactive learning environments and computer aided learning techniques have resulted in a huge collection of data. This is because it provides opportunities to log on toall user actions during the process of learning. But applying the same techniques of DM on huge time series session data is a challenging task. By applying data mining techniques on to these 'e-learning session log data', one could learn and assess the behavior of on-line system of learning. It could also study the computer based self learning environments. The mined knowledge and information data can be effectively used to remodel the e-learning system. Moreover it also helps to construct steps to identify "poor learners" even before conducting the examination. Time series data stores the same measurements that are recorded on a regular interval. But, the UCI(University of California, Irvine) EPM(Educational Process Mining) data is a kind of Non Linear Time Series Data. The periodicity of the students' process are not exactly resembled with the regular processes. In this work, instead of treating it as time series data, they are converted into simple multidimensional data for ease of data mining. The purpose of this research is to see explore the possibilities of applying Data Mining techniques on the same dataset to envisage the result that can be applied to improve e- Learning, Newer Educational Models.
Keywords: E-learning, Learning Analytics, Technology Enhanced Learning, Educational Data Mining.

I. INTRODUCTION

E-Learning is a term used to represent any form of computer, web or electronics aided based learning and teaching both inside and outside the class room [1]. Learning Management Systems (LMSs) are the modern day learning that has totally changed the way of learning. It focuses on a lot of research interest since these LMSs encounter plenty of data during the process of learning. An LMS holds the students' academic actions and interactions are stored in files. It is available as rough data [2]. Application of Data mining (DM) on such a huge academic dataset and domain is named as educational data mining (EDM).

Data mining (DM) "is a process that applies statistical, mathematical, AI and machine learning techniques to extract, classify and analyze information and knowledge from huge databases"[3]. Embedded information stored in the database are explored using e-learning systems to extract knowledge by Data Mining. It can be done by the analysis of the information available in the database. The objective of e-learning is to find the guidelines of the usage of the system by teachers and students. It helps to identify the students' learning behaviors[4]. EDM converts such data in the Educational system into an meaningful information which can be used by students, teachers, parents, educational researchers[4].

After the preprocessing the huge data and by applying statistical, data mining and visualization techniques the unknown knowledge could be retrieved from the data set. This could be done by any statistical techniques like as correlation, regression, anova etc., or complex data mining techniques such as clustering, text mining etc.,[4].

The behavior of real-world-information can be represented dynamically by measurements along temporal dimension, known as time-series. It is collected over a long period of time and this time series is a source of interesting behaviors of the system and huge in size. Time dependent, high dimensionality of the time-series data makes it complex to understand the system. This also makes it difficult to apply any conventional data mining algorithm directly. If it is possible to remove such time related dynamics of this complex time-series data, without loosing any information, then the resultant multidimensional data can readily be used to mine by applying any conventional data mining algorithms.

II. E-LEARNING

E-Learning is electronic learning in this learning type, no face-to-face corporal presents between the faculty and the students. It is also called as web based learning. It employs Internet technology and delivers digital contents, provides a new environment for faculties and students [5].

The following standardized e-learning processes are available for adoption[5]

- Metadata
- Content Wrapping
- User Profile
- Student Registration
- Content Delivery

E-learning reduces learning costs, encourages both faculty and student to go for virtual teaching-learning, improves standards of course delivery with expanded and expedited knowledge development [6]. Applying the data mining techniques in e-learning educational system helps to predict the outcome of course management systems, student characteristics, performance[7].

Knowledge Discovery in Database (KDD) is a method of collecting and analyzing a large volume of data from various point of view. It helps to discover patterns or relations and meaningful information. The KDD process that includes data selection, transformation, cleaning, pattern searching, finding presentation, finding interpretation and finding evaluation [8].

In [9], the authors display the implementation of association rules applied in to an intelligent tutoring system as a recognized proof. They adopted this data mining technique to highlight the faults frequently encountered when solving the problem.

III. DATA MINING TECHNIQUES

Data Mining is used to extract knowledge also from education systems. It helps to discover the students' learning behavior. It also helps the students in choosing and using educational software at various levels such as keystroke event, question answer, session oriented, student level, classroom observation, and school wise. In the study of educational data, some critical factors such as time, sequence play important roles10].

The following DM methods are also used in e-learning applications:

- Clustering
- Prediction techniques
- Visualization techniques

Association rules mining is a well known data handling tasks. It finds out relationships among attributes in databases[11]. It is very useful for educational Institutions to become familiar with their weak areas and drawbacks so that they could work on betterment of their learning and working skills[12].

In [13], to forecast the students division based on the previous database and the classification task is applied. The decision tree method is used for classification. The various academic data like percentage of Attendance, Class Test and Seminar performance and Assignment Marks were composed from the database, this helps to predict the performance of the end semester results. This assists to know the students who requires a additional attention to reduce the failure in the upcoming semester examination[13].

Data mining relies on some of the techniques like classification, categorization, estimation, and visualization[14]. An algorithm is a specific, mathematically determined data mining method. Some of the DM techniques mentioned in [14] are

- Clustering
- Classification
- Association rule
- Regression
- Neural Networks
- Decision Trees
- Nearest Neighbor Method

Some of the EDM approaches that are mentioned in[14] are

- Text Mining
- Social Network Analysis
- Discovery with models
- Data for Human Judgment distillation

Application of data mining techniques on the educational data has led to new research domains areas. The existing techniques are used to model the learning behaviors of the students. It is identified through Learning Analytics(LA). Most of the available system-implementations are focused on providing abilities to the teachers and other knowledge providers to discover the students with learning difficulties. It also helps the teachers in offering necessary guidance to the students to improve their capabilities or learning capacities[15].

DM is used to analyze the existing work, to identify the existing fissure and to help to improve further. The data produced by the execution of data mining technique is useful for faculties and the students[16] in enhancing their performance.

Complexity in programming languages refers to a large number of interacting components in a system. It relates to a code of a program and the path that the programs codes performs. Cyclomatic complexity is a software metric. The measurement of the quantity of the logical complexity of a program can be found[17]

Association rules are less extensive compared to the other DM techniques, because only one algorithm is used to extract association rules from data. The collection of items and analyzing the data remains instinctive. It gives different results for the different classification task of the same set of data [18].

Random forest is a blend of tree predictors. Here tree depends on the random vector sample values. It is distributed with the all trees in the forest[19]. Multi Layer Perception(MLP) classifies and implements back-propagation algorithm[20].

Classification methods are categorized the email messages into supervised and unsupervised ones. The authors used Naïve Bayes classifier, J48 Decision Tree algorithm. They used Support Vector Machines for supervised methods and K-means clustering algorithm for unsupervised methods[21].

In[22], authors implemented neural network method known as Radial Basis Function Network (RBFN). An RBFN carried out classifies using the measurement of the similar input samples in the dataset.

IV. EDUCATIONAL DATA MINING(EDM)

The Educational Data Mining Society plans to improve the scientific development of the educational institutions using DM techniques[23].

According to [18], Educational Data Mining (EDM) provides opportunities for researchers and academicians for the betterment of teaching/learning process. This EDM helps academic institutions to improve institutional effectiveness as well as student performance. Data mining is a powerful tool for better decision making and analyzing of an organization.

In[24], the authors compared two types research approaches, they are Educational Data Mining (EDM) and Learning Analytics and Knowledge (LAK). They have applied a method to solve the learning research problems. They highlight the communication and relations between EDM and LAK that helps the analysis of the data.

Clustering is a task preprocessing task to improve Educational Process Mining(EPM). One approach employs all event log-data to reveal the nature of a student's behavior. Another approach used two clustering methods, namely manual clustering and automatic clustering in interactions with the Moodle course[25].

In [26], the authors selected WEKA (Waikato Environment for Knowledge Analysis) software which was developed at the University of Waikato in New Zealand, as an EDM tool. WEKA is a open source software. This tool contains techniques to find various data mining techniques like preprocessing of data, classification, regression, clustering, association rule. They identified the classifications with J48 algorithm.

Learning Analytics(LA) is collecting, analyzing and reporting about the learner and the learning environment.. In [27], the authors mentioned the common EDM and LA methods and their application.

In [28], the authors say that EDM intends to address many characteristics of educational system. With the objective of solving the following problems.

- Student Modeling
- Predictive Modeling
- Generating Recommendations
- Analyzing learner's behavior:
- Maintaining and improving courses:
- Learners
- Educators and
- Administrators

Every complexity within the environment can be found and classified using any computational methods into significant prototypes. Various types of interaction can be noted as behavioral schemes and decoded for decision

making. The students learning science, psychology, pedagogy and computer science interconnect each other[29]. The application of data mining methods to learning data is reasonable [30]

4.1 Dataset and Algorithms used in the Educational Data Mining Task

Most of the existing research speak to troubles of classification and clustering[31]. Classification problems define the existing associations with a set of multivariate data sets. The following classification methods are used in a Data Mining process are

Fuzzy logic

- Artificial Neural Networks and Evolutionary Computation
- Graphs, Trees
- Association Rule Mining
- Multi-agent systems(MAS)

In [32], the authors based their study on the data collected through six laboratory sessions of first-year students of Computer Engineering at the University of Genoa, using digital electronics simulator. They investigated and compared the learning processes of students using Process Mining(PM) techniques. They correlated the various clusters of students based on their academic performance. The results revealed that the measured the complexity has positive correlation with the final grades of students and negative correlation with the difficulty of the laboratory sessions.

A comprehensive experiment was designed with 115 students[33]. A new design of the experiment to log on the interaction data for comparing to the learning outcomes was carried out. They aimed to extract valuable knowledge through LA methods from gathered data using LA Data Collector (LADC).

The authors explained[34] that LA and EDM are both two rising fields, even though they have dissimilarities in their origins and applications. LA is a multi-disciplinary field that involves Machine Learning(ML), artificial intelligence, information retrieval, statistics, and visualization.

4.2 Time Series Data and Multidimensional Data

The Original UCI DPM data set is a kind of non-liner time-series data. The features in this dataset are selected and presented in a appropriate format catering to the complex Process Mining tasks. But to apply a clustering and classification algorithm, it is required to process the data further.

In this work, it is proposed to apply a simple data transformation technique to convert the non-linear time- series data into a simple multidimensional data so that more data mining algorithms could be directly applied to the data set to discover new knowledge and information.

4.3 About the Educational Process Mining (EPM) Dataset of UCI

A Learning Analytics Data Set has been provided to UC Irvine (UCI) machine learning repository by the researchers of Non-Linear Complex Systems Laboratory, Italy and Department of Industrial Design, Eindhoven University of Technology, Netherlands.

It was done based on the experiments that have been carried out with a 115 students of first-year, undergraduate Engineering graduates of the University of Genoa under a simulation environment named DEEDS (Digital Electronics Education and Design Suite) and was used for e-learning process. This ambiance presents learning materials through browsers for the students, and informs them to solve various assignments with different levels of complexity.

The data set contains the students' time-series of activities in six laboratory sessions of the course of digital electronics. The students information of 6 sessions are stored in files. The session folder that contains up to 99 CSV files, each is allotted to note the student log during that session. Each file contains 13 attributes.

V. DATASET AND ALGORITHMS USED IN THIS EDUCATIONAL DATA MINING TASK

5.1 Features of the EPM Dataset

"The features selected for this data set come from preprocessing of data collected through a logging program. Due to ethical reasons and to ensure the anonymity of the users, the original log files could not be shared, instead, the data transformed and cleaned in an appropriate format are shared. The original logs contain the logging data of client system for approximately every second, while the features are calculated in order to be allocated for a particular activity". They are displayed for Process Mining. The data is presented for each session, per student, and for every exercise. The extension .CSV file contains data on specific session of a specific student with their IDs. The files

have many exercises of that session available in 'exercise' feature. The exercises hold start time, end time and other features.
Process Mining is a process management technique that allows to extract valuable knowledge from the event logs. In Process Mining, normally events / activities are linked together in a process instance or case. The potential cases in this data set contains students laboratory session(Vahdat et al., 2014; Vahdat et al 2015). The following is the detailed list of features:

Table-1: Session Attributes

Sl. No	Feature	Description
1	Session	Number of lab sessions from 1 to 6.
2	Student-ID	115 students' ID number
3	Exercise	It shows the working Ex.ID of the student. (Es_2_1 represents Session 2, Exercise 1).
4	Activity	The activities are grouped into 15 categories related to Exercise, Using Deeds Simulator, Using Text Editor, Working on Diagram, Working on Properties Window, Viewing Study Materials, Using Finite State Machine Simulator, Using Aulaweb, Blank and other irrelevant Activities
5	Start-time	Starting date and time of a specific activity.
6	End-time	Ending date and time of a specific activity.
7	Idle-time	The duration of idle time between the start and end time.
8	Mouse-wheel	Volume of mouse wheel operations during an activity.
9	Mouse-wheel-click	Number of mouse wheel clicks during an activity.
10	Mouse-click-left	Number of left mouse clicks during an activity.
11	Mouse-click-right	Number of right mouse clicks during an activity.
12	Mouse-movement	Distance moved by the mouse movements during an activity.
13	Keystroke	Number of key presses during an activity.

5.2 Grades Data of EMP Dataset
The grades data contain two Excel files. The grades are given to every student Id.
Final grades are the ones got in the final exam. Some Ids are missing in final grades, because those students did not attend the final exams.
Intermediate grades are obtained from the evaluation of students' assignments for each session (exercises). The grades are given to the students' assignments from session 2 to 6. There is no grade for Session 1. The students are required to work during each laboratory session and submit their work afterwards. The intermediate grades are given after reviewing their works.
The features in this dataset are chosen and modified in a suitable format for complex Process Mining tasks.

5.3 J48 Decision Tree Classifier
The J48 Decision tree classifier supports the algorithm given below. A decision tree of the training data available is created using the attributes. It is used to classify a new item. It encounters the training set items whenever it identifies a feature, and then it differentiates the various instances. This feature can apply to most of the data instances helping to classify them better. Amongst the available values of this feature, without ambiguity, the data instances lessening inside its group have the same value for the target variable, then the branch is finished and allotted to it the target value found.

VI. THE PROPOSED MODEL OF EDM SYSTEM

6.1 The Data Conversion Algorithm
In section 5, the Educational Process Mining (EPM) Dataset of UCI was described. The following pseudo code explains the data conversion algorithm that converts the Time Series text Data Set into multidimensional numeric data set.
Let the Time Series Data Set contains data from N sessions and there are n log files corresponding to n student IDs in each session. Each log file may have m number of activity logs belonging to any one of 15 categories of activities.
Input : $N \times n \times m$ number of text files of Time Series Data Set
Output: One, Simple, Multi Dimensional Numeric Data se

```
Let Record Index R=0
for S=1:N   {
  for ID=1 to n {
   for each Activity log L in LogFile(ID) {
       Features=Extract13Features(L)
       SessionID=Feature(1)
       Student_ID=Feature(2)
       // the third feature is not used
       //Exercise=Feature(3)
       ActivityID=aID=Feature(4)
       Start_time=Feature(5)
       End_time=Feature(6)
        Idle_time= Feature(7)
      //MKFs= MouseAndKeyFeatures
       MKFs= Feature(8:13)
      // ActivityCounts AC
       AC(aID)= AC (aID)+1
       // ActivityTimeCount ATC
        ATC(aID)= End_time-Start_time
       Idle_time_Count(aID)= Idle_time_Count(aID)+ Idle_time
       For f=1 to 6 {
          //Average MouseAndKeyFeatureCounts MKFCs
          MKFsC(f)=( MKFsC (f) + MKFs(f))/2;
       }
     R=R+1;
     MDDataSet(R)={S,ID, AC (1:15), ATC(1:15),  MKFC(1:6) }
    }
}
}
 Save(FeatureDataSet)
EPMFeaturesDataset=Features;
save('EPMFeaturesDataset.mat','EPMFeaturesDataset');t
```

VII. THE IMPLEMENTATION RESULTS AND DISCUSSION

7.1 About the Implementation

The proposed time series data conversation system and using Matlab version 7.4.0 (R2007a) was developed. Even after such transformation, the data is still very complex in nature making the data mining task a challenging one. To test the suitability of the converted data, J48 Classifier is used as a that measures the suitability of the data using different accuracy metrics of classification Precision, Recall and F_Score and used k-fold validation.

7.2 The Metrics

The following are the metrics considered for measuring the performance of the classifier of the EDM system. The results are interpreted based on the arrived values of these metrics.

7.3 Precision

Precision can be defined as a measure of a classifiers exactness. A low precision can also indicate the presence of large number of false positives in the classification results. Precision is also known as Positive Predictive Value. Precision is calculated using the following relation :

$$\text{Precision} = \frac{\text{True Positive(TP)}}{\text{True Positive(TP)} + \text{False Positive(FP)}}$$

That is, high precision represents that an algorithm resulted considerably more relevant results than irrelevant ones.

7.4 Recall

Recall represents the proportion of actual positives that are correctly identified as identified. It is calculated using the following relation :

$$\text{Recall} = \frac{\text{True Positive(TP)}}{\text{True Positive(TP)} + \text{False Negative(FN)}}$$

Recall is used to measure of a classifiers completeness. High recall signifies that an algorithm resulted the most of the relevant results. A low recall points to many False Negatives.

7.5 F-Score

F-Score is known as the harmonic mean of precision and recall. The F-score or f-measure is calculated using the following relation :

$$\text{F-Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

7.6 The Numerical Results as Precision, Recall and F-Measure

The original EPM data set is a kind of time series data in text format files, on which the conventional J48 like algorithm can not be directly applied. Hence it is converted as a multidimensional numeric data. If the data is really classifiable, then it would produce meaningful result in terms of accuracy metrics.

The following results show the classification in terms of Precision, Recall and F-Measure. It is obtained by applying the J48 algorithm on the combined, transformed data from session-2 to session-6 since the class grades information for session-1 is not available with the original UCI EPM data set

Table 2 - Classification Accuracy Achieved by J48Algorithms

Grades	Precision	Recall	F-Measure
Average	0.609	0.683	0.601

In this experiment, a combined data set from session-2 to Session-6 data is formed and used for training and testing.

7.7 Predictability of Grades in Terms of Different Metrics

The following chart shows the predictability of the proposed EDM system in terms of Precision, Recall, F-Score, TP-Rate and FP-Rate.

The following chart shows the predictability of grades. The results of the following chart prove that the data that is transformed is classifiable.

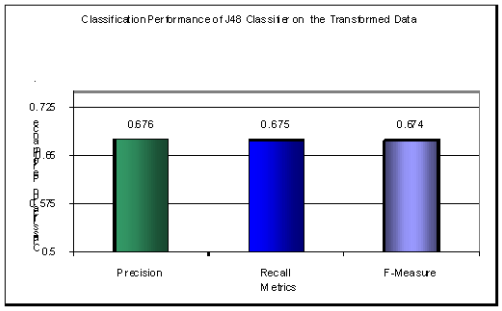

Figure 1: Performance in terms of Precision

The results and contributions of this work are being summarized as follows:

It is proposed that a simple data transformation algorithm which can transform text logs based time series data in to simple, numerical multidimensional data.

A conventional classification algorithm is applied on the transformed data and proved that the converted data is classifiable.

VIII. CONCLUSION

The proposed data transformation algorithm in MATLAB7 was implemented and transformed the text format time-series data of five e-learning sessions into a simple, numerical multi dimensional data set. It has applied a conventional J48 classification algorithm. The average accuracy measured in terms of different metrics is higher than 0.6.

This does not provide a total solution to all the problems. This measures only the average performance of J48 in terms of precision, recall and f-score. But, it is also required to know what would be the result of poor grade students identified from their e-learning behavior, as well as and what would be the result of high grade students. Class wise measurements of accuracy only can answer these questions. In this work, it has used the combined data from all the sessions (2 to 6). The data from a single session is classified, then the results may vary. In other words, it is necessary to find session-wise performance to forecast the student behavior in a forthcoming session. This may be addressed by the future researches.

IX. REFERENCES

[1] https://docs.opencv.org/2.4/modules/ml/doc/random_trees.html
[2] Cristóbal Romero, Rebeca Cerezo, Alejandro Bogarín1 and Miguel Sánchez-Santillán, [Chapter-1] Data Mining and Learning Analytics: Applications in Educational Research, John Wiley & Sons pub, Online published book, 15 SEP 2016
[3] Jean Baptiste Batware" Real-Time Data Mining For E-Learning",(2007)
[4] Shimaa Abd Elkader Abd Elaal, "E-learning using data mining", Chinese-Egyptian Research Journal
[5] Dr. Mohamed F. AlAjmi, Shakir Khan, Dr. Arun Sharma : Studying Data Mining and Data Warehousing with Different E-Learning System in : International Journal of Advanced Computer Science and Applications, Vol. 4, No.1, 2013
[6] Priyanka R. Pradhan, Mr. R. B. Kulkarni : Review on Secure E-Learning using Data Mining Techniques and Concepts in : International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS070930 www.ijert.org (This work is licensed under a Creative Commons Attribution 4.0 International License.) Vol. 4 Issue 07, July-2015
[7] Jyoti Upadhyay1 , Pratima Gautam : Forecasting Some Aspects of E-Learning Educational System Using Data Mining Techniques in : International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 3 Issue 5, May 2014
[8] V.Vijayalakshmi,i.Shamili,K.Kaarguzhali:survey on e-learning:comparison of techniques and algorithms in : International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 12 Issue 4 FEBRUARY 2015.
[9] Richard A. Huebner Norwich University : A Survey of Educational Data-Mining Research in : Research in Higher Educational Research????
[10] Ryan S.J.d. Baker, Pittsburgh : Data Mining for Education in : International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.
[11] Enrique García , Cristóbal Romero , Sebastián Ventura , Toon Calders : Drawbacks and solutions of applying association rule mining in learning management systems in : Proceedings of the International Workshop on Applying Data Mining in e-Learning 2007
[12] Dušan Perović : data mining influence on e-learning in : The Sixth International Conference on e-Learning (eLearning-2015), 24- 25 September 2015
[13] Brijesh Kumar Baradwaj, Saurabh Pal : Mining Educational Data to Analyze Students' Performance in : International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011
[14] Jiechao Cheng : Data-Mining Research in Education in : arXiv : 1703.10117 [stat.OT],2017
[15] Thilina Ranbaduge : Use of Data Mining Methodologies in Evaluating Educational Data in : International Journal of Scientific and Research Publications, Volume 3, Issue 11, November 2013
[16] Sunil, Prof. M. N. Doja : Data Mining Techniques to Discover Students Visiting Patterns in E-learning Resources in : l, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.6, June- 2017
[17] Noraini Mohameda , Raja Fitriyah Raja Sulaimanb, Wan Rohana Wan Endut : The Use of Cyclomatic Complexity Metrics in Programming Performance's Assessment in : Science Direct, Procedia - Social and Behavioral Sciences 90,2013
[18] "Revisiting interestingness of strong symmetric association rules in educational data" by Merceron and Yacef , International Workshop on Applying Data Mining in e-Learning (ADML'07), Second European Conference on Technology Enhanced Learning (EC-TEL07)
[19] http://www.stat.berkeley.u/users/breiman/RandomForests/
[20] http://scikit-learn.org/stable/modules/neural_networks_supervised.html
[21] http://www.d.umn.edu/~padhy005/Chapter5.html
[22] http://mccormickml.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/
[23] "EducationalDataMining.org", (2010)
[24] George Siemens , Ryan S J.d. Baker:Learning Analytics and Educational Data Mining Towards Communication and Collaboration in Conference'10, Month 1–2, ACM,2010
[25] Samira Elatia, Donald Ipperciel, Osmar R. Zaïane : Data Mining and Learning Analytics in : Educational Process mining: a Tutorial and case study using moodle data sets John Wiley & Sons, Inc., Hoboken, New Jersey,2016
[26] S. Lakshmi Prabha, Dr.A.R.Mohamed Shanavas : Application of Educational Data mining techniques in e-Learning- A Case Study in :) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015
[27] Calvet Liñán, L., & Juan Pérez, Á. A.: Educational Data Mining and Learning Analytics: differences,similarities, and time evolution. RUSC. Universities and Knowledge Society Journal, 12(3). pp. 98-112, 2015
[28] Geeta Kashyap, Ekta Chauhan : REVIEW ON EDUCATIONAL DATA MINING TECHNIQUES in :2nd International conference on Recent Innovation in Science, Engineering and Management, 2015
[29] Zacharoula Papamitsiou and Anastasios A. Economides : Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence in : Educational Technology & Society, 17 (4), 49–64, 2014

[30] "Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes" by Anjewierden et al, International Workshop on Applying Data Mining in e-Learning (ADML'07), Second European Conference on Technology Enhanced Learning (EC-TEL07)

[31] Félix Castro1, Alfredo Vellido, Àngela Nebot, and Francisco Mugica : Applying Data Mining Techniques to e-Learning Problems. In: Jain L.C., Tedman R.A., Tedman D.K. (eds) Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, vol 62. Springer, Berlin, Heidelberg,2007

[32] Mehrnoosh Vahdat, Luca Oneto, Davide Anguita, Mathias Funk, and Matthias Rauterberg:A Learning Analytics Approach to Correlate the Academic Achievements of Students with Interaction Data from an Educational Simulator in : Conole et al. (eds.): EC-TEL 2015, LNCS 9307, pp. 352-366. Springer (2015). DOI: 10.1007/978-3-319-24258-3 26

[33] Mehrnoosh Vahdat, Luca Oneto, Alessandro Ghio, Giuliano Donzellini, Davide Anguita, Mathias Funk, and Matthias Rauterberg : A Learning Analytics Methodology to Profile Students Behavior and Explore Interactions with a Digital Electronics Simulator In: de Freitas, S., Rensing, C., Ley, T., Munoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 596–597. Springer (2014).

[34] Mehrnoosh Vahdat, Alessandro Ghio, Luca Oneto, Davide Anguita, Mathias Funk and Matthias Rauterberg in : Advances in Learning Analytics and Educational Data Mining in : European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (2015).

X. AUTHORS PROFILE

Mr. S. Arumugam. received his M.C.A., and M.Phil. in computer Science from Bharathidasan University. Now working as Assistant Professor at the Department of Computer Science and Application, Periyar Maniammai
Institute of Science and Technology, Vallam, Thanjavur, Tamilnadu India. He authored several books/course materials for postgraduate students of PMU and published papers in national and international Journals. He is a Life Member of Indian Society for Technical Education (ISTE) and Computer Society of India(CSI). He organized several workshops on different topics. Currently he is pursuing Ph.D. in Computer Science a PMIST, Tamilnadu, India

.

Dr. A. Kovalan received his MCA, and PhD in Computer Science from Bharathiyar University. He has presented more than 20 papers at National and International Conferences. He has published more than 20 papers at National and International Journals. He has contributed in the development of Software on Online Entrance Examination system(PMU EES v1.0) at PMU. His area of interest includes Image Processing, Signal Processing, Datamining. He is providing research guidance for Ph.D scholars from different areas of research and has presented several invited talks in this areas of research.

Dr. A. E. Narayanan received his M.Tech (Computer and Information Technology) from Manonmaniam Sundaranar University and Ph.D in Computer Science Engineering from Periyar Maniammai University. He has presented more than 10 papers at National and International Conferences. He has published at National and International Journals. He organizes various National Conferences such as RITIDS in association with Ministry of Earth Sciences, Govt. of India New Delhi. He also conducted various National Level workshops in collaboration with IIT Bombay, ISTE and MHRD and guided various M.Tech students for their project and research work. His areas of research are Digital Image Processing and Fault Tolerant Systems. He is the Member of ISTE, CSI and IEI. He is providing research guidance for Ph.D scholars from different areas of research and has presented several invited talks in this areas of research.