ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2015.0145

APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS – A LITERATURE REVIEW

Katrina Sin[1] and Loganathan Muthu[2]
[1]*Faculty of Education and Languages, Open University Malaysia, Malaysia*
E-mail: coachkatrina@oum.edu.my
[2]*Department of Computer Science, Bharathiyar University, India*
E-mail: vmmlog@gmail.com

Abstract
The usage of learning management systems in education has been increasing in the last few years. Students have started using mobile phones, primarily smart phones that have become a part of their daily life, to access online content. Student's online activities generate enormous amount of unused data that are wasted as traditional learning analytics are not capable of processing them. This has resulted in the penetration of Big Data technologies and tools into education, to process the large amount of data involved. This study looks into the recent applications of Big Data technologies in education and presents a review of literature available on Educational Data Mining and Learning Analytics.

Keywords:
Big Data, Learning Analytics, LMS, Educational Data Mining

1. INTRODUCTION

Learning that initially started in the class room was based on three models namely behavioral, cognitive and constructivist models [2] The behavioral models rely on observable changes in the behavior of the student to assess the learning outcome. The cognitive models are based on the active involvement of teacher in the learning which helps in guided learning. In the constructivist models, the students have to learn on their own from the knowledge available to them.

Siemens (2004) [4] proposed a new model termed “Connectivism” which was characterized as the “amplification of learning, knowledge and understanding through the extension of personal network”. According to this model, learning is no longer an internal activity [5] Connectivism proposed learning in a network of nodes which improved the learning experience of students and reduced the need for the direct involvement of an instructor. Since then, traditional learning environments have gradually mutated into community based learning environments.

2. EMERGENCE OF BIG DATA IN LEARNING

Research in education has resulted in several new pedagogical improvements. Community based learning environments have increased in number. In the current learning environments, users learn in online communities like discussion forums, online chats, instant messaging clients and various Learning Management Systems like Moodle. Recent learning methods like Flipped Classroom [6] greatly depend on online activities. Several frameworks [7] and models have been proposed for online learning management systems to improve the learning experience. Entry of open source projects in mobile computing has led to low cost smartphones and smartphones have penetrated much. Students have started using smart phones to access learning content. As the learning environments have become accessible anywhere through the internet, students access their courses anywhere and indulge in learning activities. Students’ activities through learning management systems create large amount of data that can be utilized in developing the learning environment, helping the students in learning and improving the overall learning experience.

In addition to the data available from student activities, data are also created by educational institutions which use applications to manage courses, classes and students. The amount of data made available in the above scenarios is so enormous [1] that traditional processing techniques cannot be used to process them. Due to the limitations of the conventional data processing applications, the educational institutions have started exploring “Big Data” technologies to process the educational data.

3. BIG DATA

The term “Big Data” refers to any set of data [3] that is so large or so complex that conventional applications are not adequate to process them. The term also refers to the tools and technologies used to handle “Big Data”. Examples of Big Data include the amount of data shared in the internet everyday, YouTube videos viewed, twitter feeds and mobile phone location data. In the recent years, data produced by learning environments have also started to get big enough raising the need for Big Data technologies and tools to handle them.

3.1 CHALLENGES IN HANDLING BIG DATA

Several challenges need to be addressed while handling Big Data. Those challenges include

3.1.1 Storage:

While the common capacity of hard disks nowadays is in the range of terabytes, the amount of data generated through internet everyday is in the order of exabytes. Though the data generated in education is not as large as all the data generated through internet, it is large enough, and would get much larger in future. The traditional RDBMS tools will be unable to store or process such Big Data. To overcome this challenge, databases that don't use traditional SQL based queries are used. Compression technology is used to compress the data at rest and in memory.

3.1.2 Analysis:

As data generated to several types of online learning sites differ in structure and the size of the data is also huge, analysis of the data may consume a lot of time and resources. To overcome this, scaled out architectures are used to process the data in a distributed manner. Data are split into smaller pieces and processed in a vast number of computers available throughout the network and the processed data is aggregated.

3.1.3 Reporting:

Traditional reports involve display of statistical data in the form of numbers. When large amount of data are involved, traditional reports become difficult to interpret by human beings. In those cases the reports need to be represented in a form that can be easily recognized by looking into them.

The Big Data technologies overcome these challenges using various techniques.

3.2 TOOLS AND TECHNIQUES

3.2.1 Techniques:

The challenges faced in processing Big Data technologies are overcome by using various techniques. The most popular techniques used in educational data mining are listed below.

- **Regression** – Regression is used in predicting values of a dependant variable by estimating the relationship among variables using statistical analysis
- **Nearest Neighbor** – In this technique the values are predicted based on the predicted values of the records that are nearest to the record tha needs to be predicted.
- **Clustering** – Clustering involves grouping of records that are similar by identifying the distance between them in an n-dimensional space where n is the number of variables.
- **Classification** – Classification is the identification of the category/class to which a value belongs to, on the basis of previously categorized values.

3.2.2 Open Source Tools:

Several Open source tools exist which help in taming Big Data [9] Some of the top tools are listed below.

- **MongoDB** is a cross platform document oriented database management system. It uses JSON like douments instead of a table based architecture.
- **Hadoop** is a framework that allows distributed processing of big datasets across clusters of networked computers using simple programming models.
- **MapReduce** is a programming model and framework used by hadoop. It enables processing huge amount of data in parallel on large clusters of compute nodes.
- **Orange** is a python based tool for processing and mining big data. It has an easy to use interface with drag & drop functionalities with variety of add-ons.
- **Weka** is a java based tool for processing large amount of data. It has a vast selection of algorithms that can be used in mining data.

3.2.3 Proprietary Tools:

- **SAP HANA** is a proprietary in-memory RDBMS capable of handling large amount of data. It uses Parallel InMemory relational query techniques, Columnar stores and Compression technology to overcome the challenges faced in handling Big Data.

4. APPLICATIONS IN LEARNING

Big Data techniques can be used in a variety of ways in learning analytics as listed below [8]:

- Performance Prediction
  - o Student's performance can be predicted by analyzing student's interaction in a learning environment with other students and teachers
- Attrition Risk Detection
  - o By analyzing the student's behavior, risk of students dropping out from courses can be detected and measures can be implemented in the beginning of the course to retain students.
- Data Visualization
  - o Reports on educational data become more and more complex as educational data grow in size. Data can be visualized using data visualization techniques to easily identify the trends and relations in the data just by looking on the visual reports.
- Intelligent feedback
  - o Learning systems can provide intelligent and immediate feedback to students in reponse to their inputs which will improve student interaction and performance.
- Course Recommendation
  - o New courses can be recommended to students based on the interests of the students identified by analyzing their activities. That will ensure that students are not misguided in choosing fields in which they may not have interest.
- Student skill estimation
  - o Estimation of the skills acquired by the student
- Behavior Detection
  - o Detection of student behaviors in community based activities or games which help in developing a student model
- Grouping & collaboration of students
- Social network analysis
- Developing concept maps
- Constructing courseware
- Planning and scheduling

4.1 PERFORMANCE PREDICTION

Predictive Analytics enables prediction of student's behavior, skill and performance by analyzing various activities performed by the student while interacting with the Learning Management System or with fellow students. Based on the activities of the students, the performance of the students can be predicted using the data mining techniques that can be used in identifying the underperforming students so that the instructors can focus on developing them.

Vince Kellen (2013) [10] in his case study titled “Applying Big Data in Higher Education: A Case Study”, describes the successful implementation of a Big Data analysis tool: “SAP's HANA”, in the University of Kentucky. By monitoring and analyzing the student's background data, their system calculates a scored termed “K-Score” for each student. The score depicts the involvement of students in the learning activities. A low score represents an underperforming student who needs to be taken care of.

4.2 SKILL ESTIMATION

Skill Estimation refers to the estimation of the skills of the students so that the learning environment can be adjusted to suit the student's skills. Skills were calculated based on the interaction of the student with the system or in the message boards or discussion forums.

Paulo Blikstein in his paper “Using learning analytics to assess students' behavior in open-ended programming tasks” [11] notes the use of a tool named “NetLogo”. He logs the mouse inputs of students into the lab machines through the software and with help of the data logged finds the error rates and progress rates of the students.

Beheshti [12] , in his paper titled “Predictive performance of prevailing approaches to skills assessment techniques: Insights from real vs. synthetic data sets”, uses synthetic as well as real data sets for assessing the skills of learners. He compares the differences between the details of the skills obtained by the different techniques and analyses his methodology. The results show that the real data provides more accurate results.

4.3 BEHAVIOR DETECTION

Joseph Grafsgaard [13], in his paper “Predicting Learning and Affect from Multimodal Data Streams in Task-Oriented Tutorial Dialogue”, presents a system for recognizing the facial expressions of the students to predict the engagement, frustration and learning outcomes of students after the learning session. He also uses gesture detection and posture tracking algorithms to capture non-verbal behaviors of students and associates them with the learning patterns.

Seong Jae Lee [14], in his paper “Learning Individual Behavior in an Educational Game: A Data-Driven Approach”, describes a framework for modeling user's behavior. The proposed system learns individual policies from the movement of the players in the game and builds a cognitive model. He states that this type of modeling will help in understanding learning processes of the user who interacts with the system and in adapting the learning environment to the user.

4.4 ATTRITION RISK PREDICTION

Researchers have also used the big data techniques in predicting the risk of attrition associated with students. In educational institutions where the students are likely to drop out of courses, the student's activities are monitored and the student's engagement score is predicted. The predicted score was found to be dependent on the attrition rates.

Lalitha Agnohotri, in her paper titled “Building a Student At-Risk Model: An End-to-End Perspective From User to Data Scientist”, proposes a new model named “Student At Risk” that is capable of calculating risk ratings for new joinees. She uses historical data to model the student's behavior and uses the created model in the system to calculate the attrition risks of new joinees. The ratings can be used to identify the students at risk and take needed measures to retain those students in advance.

4.5 DATA VISUALIZATION

Paulo Blikstein also points to a tool named “SNAPP” (Social Networks Adapting Pedagogical Practice) that is used by instructors to visualize the interaction of students in online blogs and find the course in which they are interested the most.

Vince Kellen (2013) uses the data obtained on the student activities in classrooms and builds a visualization of classroom utilization for each classroom on a weekly and hourly basis. The visualization clearly shows the patterns of the data that can be easily recognized compared to the traditional way of reporting the data. Also the data are not aggregated and are directly processed with the help of HANA to produce the visualization.

5. LITERATURE REVIEW

Since 2011, the annual International Conference on Educational Data Mining (EDM) and the annual International Conference on Learning Analytics and Knowledge (LAK) have seen many papers submitted and presented to showcase the emerging and fast developing field of Educational Data Mining and Learning Analytics (refer Fig.1.). The trend in the numbers of articles submitted/published in the two journals in the past 5 years clearly shows the growing interest in the field.

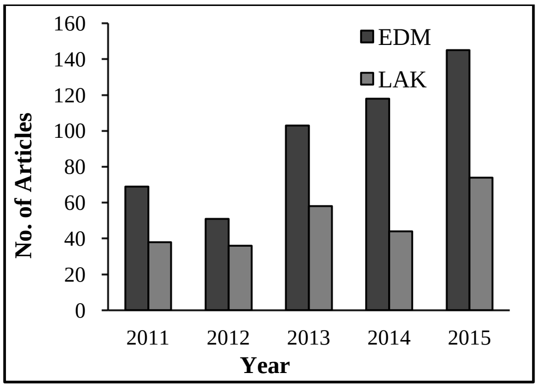

Fig.1. Articles published/submitted in EDM & LAK

A total of 119 papers have been submitted in the recent international conference on Educational Data Mining held in 2014 which shows dramatic increase in the number of papers submitted when compared to the 103 and 51 papers submitted in the same conference in 2013 and 2012 respectively. The graph shown below clearly shows the growth of research in this field.

The research papers submitted in EDM 2014 covered a variety of topics under Education Data Mining and Big Data. But around 54% of the submitted papers were addressing the top 8 topics.

5.1 MAJOR TOPICS OF INTEREST

The major topics in which the researchers have concentrated in the EDM 2014 conference are listed below in the order of descending popularity:

- Behavior Detection
- Skill Estimation
- Game-based Learning
- Student Modeling

- Performance Prediction
- Q-Matrix
- Adaptive Learning
- Attrition Risk Prediction

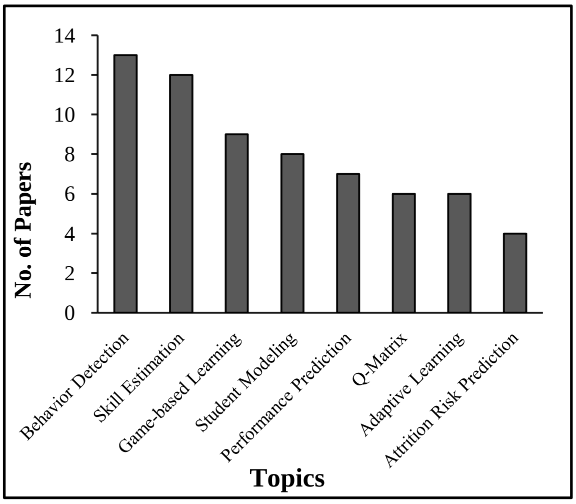

Fig.2. Popular Topics in EDM 2014

13 out of the 119 papers that were submitted in the 2014 edition of the EDM conference were associated with Behavior Detection techniques. Most of these papers were involved in studying and detecting the behavior and engagement of students in educational games. In Game-Based Learning environments, the student's were allowed to play some game, while the system monitored the student's activities. Using the data available on the activities of the student while playing the game, the system detects the behavior of the student and applies it to adapt to the student or for modeling the student behavior.

There are also many articles on learning analytics published in various journal such as Journal of Learning Analytics, Journal of Educational Technology & Society, International Review of Research in Open and Distance Learning (IRRODL), International Journal of Technology Enhanced Learning, International Journal of E-Learning and Distance Education (JDE), Australasian Journal of Educational Technology, International Journal of Learning Technology and British Journal of Educational Technology

However, for the purpose of this study, we only review searched literature on Google Scholar sorted by relevance relating to Educational Data Mining and Learning Analytics.

5.2 METHODS

5.2.1 Data Collection:

This study reviews literature chosen with the primary focus in Educational Data Mining and Learning Analytics and its implications to higher education, educational technology and instructional design.

Google Scholar was used to search and locate academic papers from journals, conference proceedings and professional magazines with the keywords “educational data mining” and “learning analytics”. The search period was set from 2010 to 2015 and the papers reviewed include both qualitative and quantitative studies from researchers in the field of educational data mining and learning analytics worldwide. The search for the keywords in Google Scholar when sorted by relevance yielded the below results.

Table.1. Search results for each Keyword

Keyword	Search Results
Educational Data Mining	5290
Learning Analytics	5890
Educational Data Mining and Learning Analytics	1370

For the purpose of this study, the data collection process resulted in the identification of a total of 90 distinct articles. 45 articles were selected for the search term “Educational Data Mining” [15-59] and another 45 articles were selected for the search term “Learning Analytics” [11; 60-103]. From the search results, it can be seen that all these articles have been frequently cited by other researchers. As these articles are frequently cited by researchers, they can be selected as a good representative sample of the literature in the field.

5.2.2 Data Classification/Analysis

Articles were classified both quantitatively and qualitatively. The quantitative analysis was used to classify the papers according to the publication year and the type of publication in which the articles appeared. Papers were qualitatively classified using open coded content analysis whereby each paper was reviewed to identify themes and trends in the literature.

5.3 RESULTS

5.3.1 Quantitative Details:

From the 45 articles selected on “Educational Data Mining” , 17 were published in 2010, 6 in 2011, 14 in 2012, 5 in 2013 and 3 in 2014 (Fig.3.). From the 45 articles selected on “Learning Analytics”, 2 were published in 2010, 8 articles in 2011, 20 articles in 2012, 14 articles in 2013, 1 article in 2014 and none in 2015 (Fig.4.). As Google Scholar searched on relevance of the article based on frequency of views and citation, articles from 2014 and 2015 were not listed at the top results of the search, therefore not reviewed in this study.

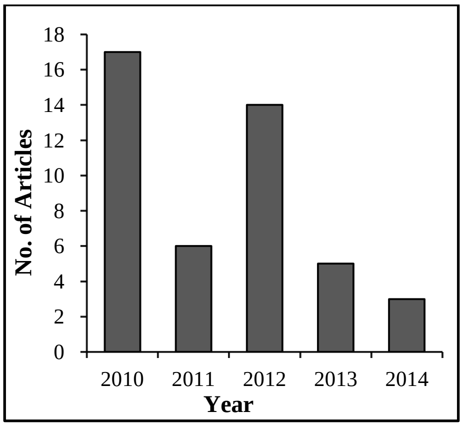

Fig.3. Articles on "Educational Data Mining" classified by Year

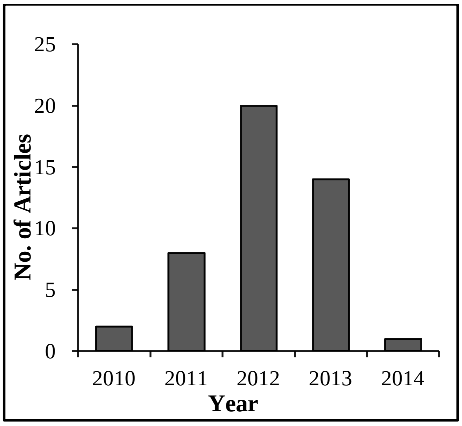

Fig.4. Articles on "Learning Analytics" classified by Year

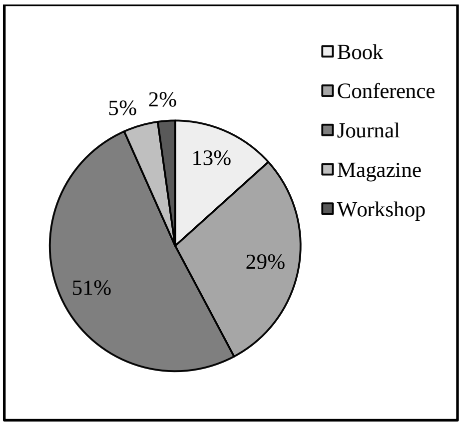

Fig.5. Articles on "Educational Data Mining" classified by Publication Type

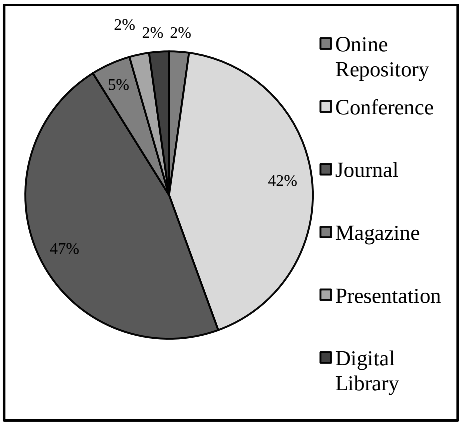

Fig.6. Articles on "Learning Analytics" classified by Publication Type

The articles were also classified by publication year. In “Educational Data Mining”, 23 articles were from journals, 13 from conference proceedings, 6 from books, 2 from magazines and 1 from workshop presentation. In “Learning Analytics”, there were 21 articles from journals, 19 conference publications, 2 academic magazine articles, 1 from online repository, 1 from a scientific digital library and 1 workshop presentation. Fig.5 and Fig.6 show the percentage of articles in each publication type. Journals and Conferences seem to contribute much to the field.

The Fig.7 and Fig.8 represent the classification of articles by both publication type and year published. The chart shows that many articles from the year 2010 on “Educational Data Mining” are much referenced. Conference articles on “Learning Analytics” were mainly from the International Conference on Learning Analytics and Knowledge (LAK) 2011-2013, and from 2011, there was an increase in journals too.

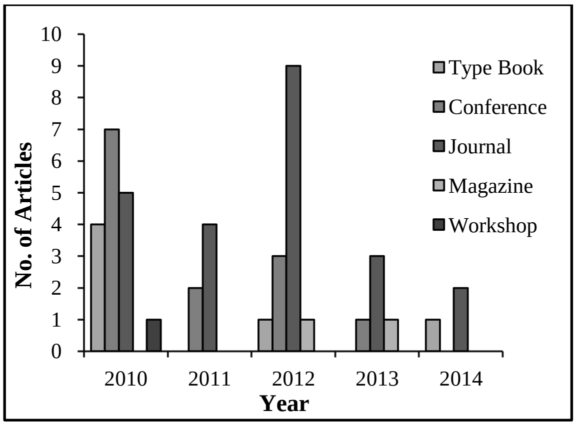

Fig.7. Articles on “Educational Data Mining” classified by Publication Type and Year

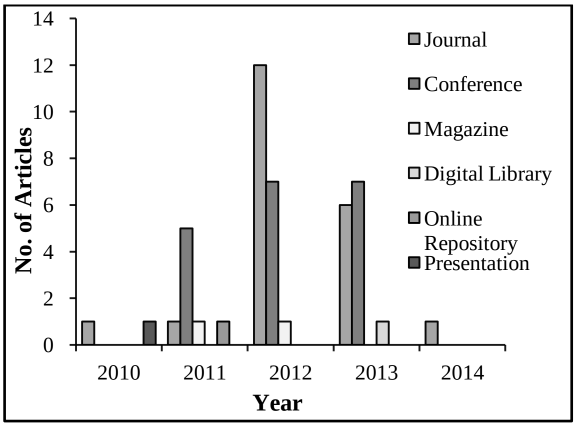

Fig.8. Articles on “Learning Analytics” classified by Publication Type and Year

A list of authors with more than one article being reviewed here are presented in Table.2 and Table.3. Pal had the highest number of publications with 7 articles in “Educational Data Mining” followed by Romero and Ventura. In “Learning Analytics”, Siemens had the highest number of publications with 6 articles and Ferguson and Shum had 4 articles each. Author with multiple publications have worked in teams and several such teams can be identified from co-authorship, such as Siemens & Gasevic, Ferguson & Shum, Prinsloo & Slade and Drachsler & Greller.

Table.2. Authors with multiple publications in the review - “Educational Data Mining”

Author	No. of Articles	Details
S Pal	7	Bharadwaj & Pal 2012 (2); Pal 2012; Pandey & Pal 2011 (2);

		Yadav, Bharadwaj & Pal 2012; Yadav & Pal 2012
C Romero	6	Lopez, Luna, Romero & Ventura 2012; Marquez-Vera, Romero & Ventura 2010; Perez, Romero & Ventura 2010; Romero, Romero (Jose), Luna & Ventura 2010; Romero & Ventura 2010; Romero & Ventura 2013
S Ventura	6	Lopez, Luna, Romero &Ventura 2012; Marquez-Vera, Romero & Ventura 2010; Perez, Romero &Ventura 2010; Romero, Romero (Jose), Luna &Ventura 2010; Romero & Ventura 2010; Romero & Ventura 2013
R Baker	5	Baker 2010; Koedinger, Baker, Cunningham, Skogsholm, Leber & Stamper 2010; Gobert, Sao Pedro, Baker, Toto & Montalvo 2012; Gobert, Sao Pedro, Raziuddin & Baker 2013; Winne & Baker 2013
B Bharadwaj	3	Bharadwaj & Pal 2012 (2); Yadav, Bharadwaj & Pal 2012
D Garcia-Seiz	2	Garcia-Saiz, Palazuelos & Zorrilla 2014; Zorrilla, Garcia-Saiz & Balcazar 2010
J Gobert	2	Gobert, Sao Pedro, Baker, Toto & Montalvo 2012; Gobert, Sao Pedro, Raziuddin & Baker 2013
N Heffernan	2	Pardos, Heffernan, Anderson & Heffernan (Cristina) 2010; Trivedi, Pardos, Sarkozy & Heffernan 2010
U Pandey	2	Pandey &Pal 2011 (2)
Z Pardos	2	Pardos, Heffernan, Anderson & Heffernan (Cristina) 2010; Trivedi, Pardos, Sarkozy & Heffernan 2010
M Sao Pedro	2	Gobert, Sao Pedro, Baker, Toto & Montalvo 2012; Gobert, Sao Pedro, Raziuddin & Baker 2013;
S Yadav	2	Yadav, Bharadwaj &Pal 2012; Yadav &Pal 2012
M Zorrilla	2	Garcia-Saiz, Palazuelos & Zorrilla 2014; Zorrilla, Garcia-Saiz & Balcazar 2010

Table.3. Authors with multiple publications in the review

Author	No. of Articles	Details
D Clow	3	Clow 2012; Clow 2013
D Gasevic	2	Siemens & Gasevic 2011; Siemens & Gasevic 2012;
E Duval	3	Duval 2011; Verbert, Duval, Klerkx & Govaerts 2013; Verbert, Manouselis, Drachsler, Duval 2012;
G Siemens	6	Siemens 2010; Siemens 2012; Siemens & Gasevic 2011; Siemens & Gasevic 2012; Siemens & Long 2011; Siemens &Baker 2012;
H Drachsler	3	Drachsler & Greller 2012; Greller & Drachsler 2012; Verbert, Manouselis, Drachsler, Duval 2012;
K Verbert	2	Verbert, Duval, Klerkx & Govaerts 2013; Verbert, Manouselis, Drachsler, Duval 2012;
P Blikstein	3	Blikstein 2011; Blikstein 2013; Worsley & Blikstein 2010;
P Prinsloo	2	Prinsloo & Slade 2013; Slade & Prinsloo 2013;
R Ferguson	4	Ferguson 2012; Ferguson & Shum 2011; Ferguson & Shum 2012; Shum &Ferguson2012;
S Dawson		Lockyer, Heathcote & Dawson 2013; Macfadyen & Dawson 2012;
S Slade	2	Prinsloo & Slade 2013; Slade & Prinsloo 2013;
SB Shum	4	Ferguson &Shum 2011; Ferguson & Shum 2012; Shum & Ferguson2012; Shum & Crick 2012;
W Greller	2	Drachsler & Greller 2012; Greller & Drachsler 2012;

5.3.2 Qualitative Details – Topics & Themes:

The articles reviewed covered a wide range of topics and themes relating to Educational Data Mining and learning analytics. The keyword from the articles on Educational Data Mining are as follows:

Apriori Algorithm, Bayesian Classifier, Bayesian Networks, Bootstrap Aggregating, Classification, Clustering, Collaborative Filtering, Critical Relative Support, Data Preparation, Discovery with Models, Dropout Management, Education Analytics, Ensemble Learning, Evidence-centered Design, Fine-Grained Skill Models, ID3 Algorithm, Inference, Intelligent Tutoring Systems, Knowledge Discovery in Database, K-means Clustering, Least Association Rules, Lecture Capturing, Machine Learning, Metacognition, Moodle, Motivation, Online Interaction, Performance Improvement, Prediction, Psychometrics, Recommender Systems, Relationship Mining, Rule Mining, School Failure, Self-regulated Learning, Sequence Rules, Session Identification, Social Network Analysis, SoTL, Special Clustering, Student Profiling, Text Replay Tagging, Video Lectures, Visualization, Web Usage Mining, Weka.

The keywords from the articles on “Learning Analytics” are as follows:

Analytics, Academic analytics, Action analytics, Automated assessment, Big data, Change management, Collaboration, College student success, Constructionism, Content analysis, Data integration, Data for learning, Datasets, Documentation, Discourse analysis, Domain design, Education, Education data mining, Early intervention, Educational dialogue, Educational games, E-learning standards, Ethics, Framework, Feedback, Higher education, Information visualization, Learning analytics, Learning design, Learning dashboards, Learning management systems, Multi-modal interaction, MOOCs, Pedagogy, Practise, Policy, Qualitative evaluations, Reference model, Research, Retention, Semantic web, Social learning analytics, Theory and Virtual learning environments.

The articles on “Educational Data Mining” were categorised in major general themes as follows:

1. **Introductory –** These articles explain the application of data mining techniques and various data mining algorithms in education in general. They also review the other articles availabe in the same context.
2. **Student Performance** – These articles explain the design, development or implementation of new models, frameworks and algorithms for predicting or measuring student's performance. Some articles also focus on predicting student's failure rates.
3. **Data Mining** – These articles focus on applying specific algorithms to mine educational data and extract information from it that can be used in improving the learning environment.
4. **Pedagogy** – These articles discuss the impact of different learning environments and pedagogical models on learners.

Table 4 lists the articles on “Educational Data Mining” based on the categories.

Table.4. Articles by Category, Title, Author and Year published - “Educational Data Mining”

Category	No. of Articles	Title	Author/Year
Introductory	18	Educational data mining	Scheuer & McLaren 2012
		Data mining for education	Baker 2010
		Classifiers for educational data mining	Hamalainen & Vinni 2010
		A Java desktop tool for mining Moodle data	Perez, Romero &Ventura 2010
		A survey and future vision of data mining in educational field	Sachin &Vijay 2012
		Educational data mining: A review	Mohamad & Tasir 2013
		Academic analytics and data mining in higher education	Baepler & Murdoch 2010
		Importance of data mining in higher education system	Bhise, Thorat & Supekar 2013
		An empirical study of the applications of data mining techniques in higher education	Kumar & Chadha, 2011
		Design and discovery in educational assessment: evidence-centered design, psychometrics, and educational data mining	Mislevy, Behrens, Dicerbo & Levy 2012
		The survey of data mining applications and feature scope	Padhy, Mishra & Panigrahi 2012
		Educational data mining: A survey and a data mining-based analysis of recent works	Pena-Ayala 2014
		Educational data mining: a review of the state of the art	Romero &Ventura 2010
		Mining educational data to improve students’ performance: a case study	Tair & El-Halees 2012
		The potentials of educational data mining for researching metacognition, motivation and self-regulated learning	Winne &Baker 2013
		Data mining applications: A comparative study for predicting student's performance	Yadav, Bharadwaj & Pal 2012
		Introduction to the special section on educational data mining	Calders & Pechenizkiy 2012
		Data mining in education	Romero & Ventura 2013
Student Performance	18	Using fine-grained skill models to fit student performance with Bayesian networks	Pardos, Heffernan, Anderson & Heffernan (Cristina) 2010
		Mining educational data to analyze students'performance	Bharadwaj & Pal 2012

		Data Mining: A prediction for performance improvement using classification	Bharadwaj & Pal 2012
		Application of data mining in educational databases for predicting academic trends and patterns	Parack, Zahid & Merchant 2012
		Spectral clustering in educational data mining	Trivedi, Pardos, Sarkozy & Heffernan 2010
		Collaborative filtering applied to educational data mining	Toscher & Jahrer 2010
		Classification via Clustering for Predicting Final Marks Based on Student Participation in Forums.	Lopez, Luna, Romero & Ventura 2012
		Data Mining: A prediction for Student's Performance Using Classification Method	Ahmed & Elaraby 2014
		Data Mining: A prediction of performer or underperformer using classification	Pandey & Pal 2011
		A CHAID based performance prediction model in educational data mining	Ramaswami & Bhaskaran 2010
		Data mining: A prediction for performance improvement of engineering students using classification	Yadav & Pal 2012
		Recommender system for predicting student performance	Thai-Nghe, Drumond, Krohn-Grimberghe & Schmidt-Thieme 2010
		Knowledge mining from student data	Chandra & Nandhini 2010
		Predicting school failure using data mining	Marquez-Vera, Romero & Ventura 2010
		Mining educational data to reduce dropout rates of engineering students	Pal 2012
		Impact of different pre-processing tasks on effective identification of users' behavioral patterns in web-based educational system	Munk & Drlik 2011
		Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds	Gobert, Sao Pedro, Baker, Toto & Montalvo 2012
		From log files to assessment metrics: Measuring students'science inquiry skills using educational data mining	Gobert, Sao Pedro, Raziuddin & Baker 2013
Data Mining	6	A data repository for the EDM community: The PSLC DataShop	Koedinger, Baker, Cunningham, Skogsholm, Leber & Stamper 2010
		A data model to ease analysis and mining of educational data	Kruger, Merceron & Wolf 2010
		Mining significant association rules from educational data using critical relative support approach	Abdullah, Herawan, Ahmad & Deris 2011
		Usage reporting on recorded lectures using educational data mining	Gorissen, Bruggen & Jochems 2012
		Mining rare association rules from e-learning data	Romero, Romero (Jose), Luna & Ventura 2010
		Towards parameter-free data mining: Mining educational data with yacaree	Zorrilla, Garcia-Saiz & Balcazar 2010
Pedagogy	3	Using educational data mining methods to study the impact of virtual classroom in e-learning	Falakmasir & Habibi 2010
		A Data mining view on class room teaching language	Pandey & Pal 2011
		Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users	Garcia-Saiz, Palazuelos & Zorrilla 2014

The articles on “Learning Analytics” were categorised in a few general themes as follows:

1. Introductory – general explanation of definitions, concepts, potentials and challenges in the emerging field of learning analytics
2. Development - design, implementation, reference model and evolution of learning analytics in higher education and e-learning, improvement of student performance.
3. Pedagogy, learning theory and design – relates learning analytics to pedagogical intent and learning design and theory
4. Data mining and datasets – explores research and methods of educational data mining and learning analytics
5. Assessment – learning analytics as a tool for assessment
6. Frameworks & Dashboards – learning analytics information visualization
7. Research & Ethics – research practises and ethical consideration
8. Social learning analytics – design and implementation
9. Discourse analytics – design and implementation
10. MOOCs – learning analytics use for tracking in MOOCs

The Table.5 shows the number of articles, titles, author and year published according to 10 categories listed above.

Table.5. Articles by Category, Title, Author and Year published - “Learning Analytics”

Category	No. of Articles	Article Titles	Author/Year
Introductory	5	1. What are learning analytics	Siemens 2010
		2. Penetrating the Fog: Analytics in Learning and Education	Siemens &Long 2011
		3. Guest editorial-learning and knowledge analytics	Siemens &Gasevic, 2012
		4. An overview of learning analytics	Clow, 2013
		5.Analytics in higher education: Establishing a common language	van Barneveld, Arnold & Camplbell 2012
Development	10	1.A qualitative evaluation of evolution of a learning analytics tool	Ali,Hatala, Gašević & Jovanović, 2012
		2.The pulse of learning analytics understandings and expectations from the stakeholders	Drachsler & Greller,2012
		3.Learning analytics: drivers, developments and challenges	Ferguson, 2012
		4.A reference model for learning analytics	Chatti, Dyckhoff,Schroeder & Thus, 2012
		5.Design and implementation of a learning analytics toolkit for teachers	Dyckhoff, Zielke, Bültmann, Chatti & Shroeder, 2012
		6. Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan	Macfadyen &Dawson, 2012
		7. E-Learning standards and learning analytics. Can data collection be improved by using standard data models?	Del Blanco, Serrano,Freire, Martinez-Ortiz &Fernández-Manjón 2013
		8.Learning analytics as a middle space	Suthers &Verbert 2013
		9.Using learning analytics to predict (and improve) student success: A faculty perspective	Dietz-Uhler &Hurn 2013
		10.Multimodal learning analytics	Blikstein 2013
Pedagogy, learning theory and design	4	1.Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics	Shum &Crick, 2012
		2.The learning analytics cycle: closing the loop effectively	Clow, 2012
		3.Informing pedagogical action: Aligning learning analytics with learning design	Lockyer, Heathcote &Dawson, 2013
		4. Learning analytics for online discussions: a pedagogical model for intervention with	Wise, Zhao &Hausknecht, 2013

		embedded and extracted analytics	
Data mining and datasets	6	1.Learning analytics and educational data mining: towards communication and collaboration	Siemens &Baker, 2012
		2.The Evolution of Big Data and Learning Analytics in American Higher Education.	Picciano, 2012
		3.Dataset-driven research to support learning and knowledge analytics	Verbert, Manouselis, Drachsler &Duval, 2012
		4.Fostering analytics on learning analytics research: the LAK dataset	Taibi &Dietze, 2013
		5.Interpreting data mining results with linked data for learning analytics: motivation, case study and directions	d'Aquin &Jay, 2013
		6.Educational data mining and learning analytics	Baker &Inventado, 2014
Assessment	6	1.What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis	Worsley &Blikstein, 2010
		2.Using learning analytics to assess students'behavior in open-ended programming tasks	Blikstein, 2011
		3. Course signals at Purdue: using learning analytics to increase student success	Arnold &Pistilli, 2012
		4.Learning analytics as a tool for closing the assessment loop in higher education	Mattingly, Rice &Berge, 2012
		5.Tracing a little for big improvements: Application of learning analytics and videogames for student assessment	Serrano-Laguna, Torrente, Moreno-Ger &Fernández-Manjón, 2012
		6.Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics	Ellis, 2013
Frameworks & Dashboards	5	1.Stepping out of the box: towards analytics outside the learning management system	Pardo&Kloos, 2011
		2.Open Learning Analytics: an integrated &modularized platform	Siemen, Gasevic, Haythornthwaite, Dawson, Shum Ferguson…&Baker,2011
		3.Attention please!: learning analytics for visualization and recommendation	Duval, 2011
		4.Translating learning into numbers: A generic framework for learning analytics	Greller &Drachsler, 2012 Verbert, Duval, Klerkx,
		5.Learning analytics dashboard applications	Govaerts &Santos, 2013
Research & Ethics	3	1.Learning analytics: envisioning a research discipline and a domain of practice	Siemens, 2012
		2.Learning analytics ethical issues and dilemmas	Slade &Prinsloo, 2013
		3.An evaluation of policy frameworks for addressing ethical considerations in learning analytics	Prinsloo &Slade, 2013
Social learning analytics	2	1.Social learning analytics: five approaches	Ferguson &Shum,2012
		2.Social learning analytics	Shum &Ferguson, 2012
Discourse analytics	2	1.Discourse-centric learning analytics	De Liddo, Shum, Quinto, Bachler &Cannavacciuolo, 2011
		2.Learning analytics to identify exploratory dialogue within synchronous text chat	Ferguson &Shum,2011
MOOCs	2	1.The value of learning analytics to networked learning on a personal learning environment	Fournier, Kop &Sitlia 2011
		2.MOOCs and the funnel of participation	Clow, 2013

5.4 LIMITATIONS

Due to time constraint, only 90 articles were reviewed, which represents only 1% of the search results for the terms “Educational Data Mining” and “Learning Analytics” in Google Scholar. Also due to Google Scholar search algorithm, articles from year 2014 and 2015 were not listed at the top of the search results, thus not selected for the review in this study.

Another limitation is that articles published in languages other than English were not considered for this review. As the search was conducted only on Google Scholar, articles from journals and academic databases were also not included.

5.5 DISCUSSION

The growth in the emerging fields of educational data mining and learning analytics can be seen from the availability of literature from 2010 until present. In the 45 articles selected for review in “Educational Data Mining”, 18 were focusing on exploring the ways in which the data mining techniques can be applied in education, while another 18 focused on evaluating or predicting student's performance. A few articles focused on specific data mining algorithms and pedagogical analysis. This clearly shows that researchers are focusing mainly on the top two themes, the application of data mining techniques in education and prediction of student performance. The 4 themes were selected as they clearly distinguish the articles into 4 groups and help identify the theme most used by the researchers.

Even though only 45 articles on “Learning Analytics” were reviewed in this study, it can be seen from the theme categories that researchers are focusing their efforts on 3 major trends in the field of learning analytics.

First, there are articles focusing on the development of academic analytics, and introduction of learning analytics, its concepts, implications and impact to higher education and e-learning, the importance of aligning with pedagogy, learning theory and design and the research frameworks and ethics policies.

Second, from the technical view point, researchers are looking at educational data mining and the use of datasets to improve learning analytics, especially through communication and collaboration between educational data mining and learning analytics communities. To help teachers and students visualise learning traces, researchers are designing and developing frameworks and dashboards for information visualization. The use of learning analytics in assessments enabling automated, real time feedback using multiple modalities and videogames to students, will impact student performance and success.

Third, the use of learning analytics in social learning and MOOCs. Learners build knowledge together in their cultural and social settings through discourse, from online forums, synchronous text to social networks, supporting the constructivist learning theory.

The study also revealed less number of articles focusing on evaluating learning outcomes by analyzing natural language text. Researchers can focus on predicting student performance in learning environments where students interact through forums. As essay answers and forum content are currently manually evaluated by teachers, it has greater scope for the application of big data techniques in the future.

6. CONCLUSION

As the data involved in education becomes larger, the applications of Big Data techniques become more and more necessary in learning environments. MOOCs are good examples of learning environments that were resource hungry and raised the need for data mining in education. The recent trends in the published papers in EDM indicate the growth in data mining in education field. Apart from EDM which we saw in this study, other communities are also involved in researching this field. Exploring those communities will provide greater insights in the field. Educational Data Mining is sure to reshape the way in which the forthcoming generations would learn.

This study revealed two major trends from articles located from top 45 search results of a Google scholar search on “Educational Data Mining”. The articles were mainly from journals and conferences related to Educational Data Mining. The major trends were identified as “Introduction to the concepts of applying Data Mining in Education” and “Prediction or measurement of Student Performance using Data Mining”. As the former trend also involved articles on performance prediction, we can conclude that the biggest focus is on “Performance Prediction using Data Mining”.

This study also highlighted 3 major trends and 10 themes from articles located from the top 45 search results of a Google Scholar search on “Learning Analytics”. The articles were mainly from Educational Technology journals and conference proceedings. There were 22 articles on the first trend of articles introducing the concept and development of the field of learning analytics to higher education and e-learning, 17 articles on the second trend of technical development of learning analytics framework and tools, and 6 articles on the third trend of learning analytics use in social learning.

As this study has reviewed only a tiny portion of the available articles, there remains a need for a systematic study of published literature on the fast growing field of application of big data in education and learning.

REFERENCES

[1] Saptarshi Ray, “Big Data in Education”, Gravity, the Great Lakes Magazine, pp. 8-10, 2013.

[2] Peggy A. Ertmer and Timothy J. Newby, “Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective”, *Performance improvement quarterly*, Vol. 6, No. 4, pp. 50-72, 1993.

[3] Wikipedia, “Big data --- Wikipedia, The Free Encyclopedia”, https://en.wikipedia.org/w/index.php?title=Big_data&oldid=669888993. Accessed 2015.

[4] George Siemens, “Connectivism: A learning theory for the digital age”, *International Journal of Instructional Technology & Distance Learning*, Vol. 2, No. 1, 2005.

[5] R. Shriram and Steve Carlise Warner, “Connectivism and the impact of Web 2.0 technologies on education”, *Asian*

Journal for Distance Education, Vol. 8, No. 2, pp. 4-17, 2010.

[6] Bill Tucker, "The flipped classroom", *Education Next*, Vol. 12, No. 1, pp. 82-83, 2012.

[7] Shriram Raghunathan and Abtar Kaur, "Assessment of online interaction pattern using the Q-4R framework", *The International Lifelong Learning Conference*, 2011.

[8] Manolis Mavrikis, Patricia Charlton and Demetra Katsifli, "The Potential of Learning Analytics and Big Data", http://www.ariadne.ac.uk/issue71/charlton-etal. Accessed 2013.

[9] Cynthia Harvey, "50 Top Open Source Tools for Big Data", http://www.datamation.com/data-center/50- top-open-source-tools-for-big-data-1.html. Accessed 2012.

[10] Vince Kellen, Cutter Consortium, Adam Recktenwald and Stephen Burr, "Applying Big Data in Higher Education: A Case Study", *Cutter Consortium*, Vol. 13, No. 8, pp. 1-39, 2013.

[11] Paulo Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks", *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 110-116, 2011.

[12] Behzad Beheshti and Michel Desmarais, "Predictive performance of prevailing approaches to skills assessment techniques: Insights from real vs. synthetic data sets", *Proceedings of the 7th International Conference on Educational Data Mining*, pp. 409-410, 2014.

[13] Joseph Grafsgaard, Joseph Wiggins, Kristy Elizabeth Boyer, Eric Wiebe and James Lester, "Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue", *Proceedings of the 7th International Conference on Educational Data Mining*, 2014.

[14] Seong Jae Lee, Yun-En Liu and Zoran Popovic, "Learning Individual Behavior in an Educational Game: A Data Driven Approach", *Proceedings of the 7th International Conference on Educational Data Mining*, 2014.

[15] Paul Baepler and Cynthia James Murdoch, "Academic analytics and data mining in higher education", *International Journal for the Scholarship of Teaching and Learning*, Vol. 4, No. 2, pp. 17, 2010.

[16] Ryan S. J. D. Baker and others, "Data mining for education", International encyclopedia of education, Vol. 7, pp. 112-118, 2010.

[17] E Chandra and K Nandhini, "Knowledge mining from student data", *European journal of scientific research*, Vol. 47, No. 1, pp. 156-163, 2010.

[18] Mohammad Hassan Falakmasir and Jafar Habibi, "Using educational data mining methods to study the impact of virtual classroom in e-learning", *Proceedings of the 3rd International Conference on Educational Data Mining*, pp. 241- 248, 2010.

[19] Wilhelmiina Hämäläinen and Mikko Vinni, "Classifiers for educational data mining", *Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, pp. 57-71, 2010.

[20] Kenneth R Koedinger, Ryan SJd Baker, Kyle Cunningham, Alida Skogsholm, Brett Leber and John Stamper, "A data repository for the EDM community: The PSLC DataShop", *Handbook of educational data mining*, Vol. 43, pp. 1-21, 2010.

[21] André Krüger, Agathe Merceron and Benjamin Wolf, "A data model to ease analysis and mining of educational data", *Proceedings of the 3rd International Conference on Educational Data Mining*, pp. 132-140, 2010.

[22] Carlos Marquez-Vera, Cristobal Romero and Sebastián Ventura, "Predicting school failure using data mining", *Proceedings of the 4th International Conference on Educational Data Mining*, pp. 271-276, 2011.

[23] Zachary A Pardos, Neil T Heffernan, Brigham Anderson, Cristina L Heffernan and Worcester Public Schools, "Using fine-grained skill models to fit student performance with Bayesian networks", Chapman \& Hall/CRC Press, 2010.

[24] Rafael Pedraza Perez, Cristobal Romero and Sebastián Ventura, "A Java desktop tool for mining Moodle data", *Proceedings of the 4th International Conference on Educational Data Mining*, pp. 1-2, 2011.

[25] M Ramaswami and R Bhaskaran, "A CHAID based performance prediction model in educational data mining", *International Journal of Computer Science Issues (IJCSI)*, Vol. 7, No. 1, pp. 10-18, 2010.

[26] Cristóbal Romero, José Raúl Romero, Jose Maria Luna and Sebastián Ventura, "Mining rare association rules from elearning data", *Proceedings of the 3rd International Conference on Educational Data Mining*, pp. 1-10, 2010.

[27] Cristóbal Romero and Sebastián Ventura, "Educational data mining: a review of the state of the art", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, Vol. 40, No. 6, pp. 601-618, 2010.

[28] Nguyen Thai-Nghe, Lucas Drumond, Artus KrohnGrimberghe and Lars Schmidt-Thieme, "Recommender system for predicting student performance", *Procedia Computer Science*, Vol. 1, No. 2, pp. 2811-2819, 2010.

[29] A. Toscher and Michael Jahrer, "Collaborative filtering applied to educational data mining", *Journal of Machine Learning Research*, pp. 1-11, 2010.

[30] Shubhendu Trivedi, Zachary Pardos, Gábor Sárközy and Neil Heffernan, "Spectral clustering in educational data mining", *Proceedings of the 4th International Conference on Educational Data Mining*, pp. 129-138, 2010.

[31] Marta E Zorilla, Diego Garcia-Saiz and José L Balcázar, "Towards parameter-free data mining: Mining educational data with yacaree", *Proceedings of the 4th International Conference on Educational Data Mining*, pp. 363-364, 2010.

[32] Zailani Abdullah, Tutut Herawan, Noraziah Ahmad and Mustafa Mat Deris, "Mining significant association rules from educational data using critical relative support approach", *Procedia-Social and Behavioral Sciences*, Vol. 28, pp. 97-101, 2011.

[33] Brijesh Kumar Baradwaj and Saurabh Pal, "Mining educational data to analyze students' performance", *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6, pp. 63-69, 2011.

[34] Varun Kumar and Anupama Chadha, "An empirical study of the applications of data mining techniques in higher education", *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 3, pp.80-84, 2011.

[35] Michal Munk and Martin Drlik, "Impact of different preprocessing tasks on effective identification of users'

behavioral patterns in web-based educational system", *Procedia Computer Science*, Vol. 4, pp. 1640-1649, 2011.

[36] Umesh Kumar Pandey and Saurabh Pal, "A Data mining view on class room teaching language", *International Journal of Computer Science Issues*, Vol. 8, No. 4, pp. 277-282, 2011.

[37] Umesh Kumar Pandey and Saurabh Pal, "Data Mining: A prediction of performer or underperformer using classification", *International Journal of Computer Science and Information Technologies*, Vol. 2, No. 2, pp. 686-690, 2011.

[38] Brijesh Kumar Bhardwaj and Saurabh Pal, "Data Mining: A prediction for performance improvement using classification", *International Journal of Computer Science and Information Security*, Vol. 9, No. 4, 2012.

[39] Toon Calders and Mykola Pechenizkiy, "Introduction to the special section on educational data mining", *ACM SIGKDD Explorations Newsletter*, Vol. 13, No. 2, pp. 3-6, 2012.

[40] Janice D Gobert, Michael A Sao Pedro, Ryan SJd Baker, Ermal Toto and Orlando Montalvo, "Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds", *JEDM-Journal of Educational Data Mining*, Vol. 4, No. 1, pp. 111-143, 2012.

[41] Pierre Gorissen, Jan Van Bruggen and Wim Jochems, "Usage reporting on recorded lectures using educational data mining", *International Journal of Learning Technology*, Vol. 7, No. 1, pp. 23-40, 2012.

[42] Manuel Ignacio Lopez, J.M. Luna, C. Romero and S. Ventura, "Classification via Clustering for Predicting Final Marks Based on Student Participation in Forums.", *Proceedings of the 5th International Conference on Educational Data Mining*, pp. 148-151, 2012.

[43] Robert J Mislevy, John T Behrens, Kristen E Dicerbo and Roy Levy, "Design and discovery in educational assessment: evidence-centered design, psychometrics, and educational data mining", *JEDM-Journal of Educational Data Mining*, Vol. 4, No. 1, pp. 11-48, 2012.

[44] Neelamadhab Padhy, Mishra, Rasmita Panigrahi and others, "The survey of data mining applications and feature scope", *International Journal of Computer Science, Engineering and Information Technology*, Vol. 2, No. 3, 2012.

[45] Saurabh Pal, "Mining educational data to reduce dropout rates of engineering students", *International Journal of Information Engineering and Electronic Business (IJIEEB)*, Vol. 4, No. 2, pp. 1, 2012.

[46] Suhem Parack, Zain Zahid and Fatima Merchant, "Application of data mining in educational databases for predicting academic trends and patterns", *International Conference on Technology Enhanced Education (ICTEE), IEEE*, pp. 1-4, 2012.

[47] R Barahate Sachin and M Shelake Vijay, "A survey and future vision of data mining in educational field*", Proceedings of the 2nd International Conference on Advanced Computing & Communication Technologies (ACCT)*, pp. 96-100, 2012.

[48] Oliver Scheuer and Bruce M McLaren, "Educational data mining", *Encyclopedia of the Sciences of Learning*, pp. 1075-1079, 2012.

[49] Mohammed M Abu Tair and Alaa M El-Halees, "Mining educational data to improve students' performance: a case study", *International Journal of Information and Communication Technology Research*, Vol. 2, No. 2, pp. 140-146, 2012.

[50] Surjeet Kumar Yadav, Brijesh Bharadwaj and Saurabh Pal, "Data mining applications: A comparative study for predicting student's performance", *International Journal of Innovative Technology & Creative Engineering*, Vol. 1, No. 12, pp. 13-19, 2012.

[51] Surjeet Kumar Yadav and Saurabh Pal, "Data mining: A prediction for performance improvement of engineering students using classification", *World of Computer Science and Information Technology Journal*, Vol. 2, No. 2, pp. 51-56, 2012.

[52] R.B Bhise, S.S Thorat and A.K Supekar, "Importance of data mining in higher education system", *IOSR Journal Of Humanities And Social Science*, Vol. 6, No. 6, pp.18-21, 2013.

[53] Janice D Gobert, Michael Sao Pedro, Juelaila Raziuddin and Ryan S Baker, "From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining", *Journal of the Learning Sciences*, Vol. 22, No. 4, pp. 521-563, 2013.

[54] Siti Khadijah Mohamad and Zaidatun Tasir, "Educational data mining: A review", *Procedia-Social and Behavioral Sciences*, Vol. 97, pp. 320-324, 2013.

[55] Cristobal Romero and Sebastian Ventura, "Data mining in education", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 3, No. 1, pp. 12-27, 2013.

[56] Philip H. Winne and Ryan S.J.d Baker, "The potentials of educational data mining for researching metacognition, motivation and self-regulated learning", *Journal of Educational Data Mining*, Vol. 5, No. 1, pp. 1-8, 2013.

[57] Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method", *World Journal of Computer Application and Technology*, Vol. 2, No. 2, pp. 43-47, 2014.

[58] Diego Garcia-Saiz, Camilo Palazuelos and Marta Zorrilla, "Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users", *Educational Data Mining Studies in Computational Intelligence*, Vol. 524, pp. 411-439, 2014.

[59] Alejandro Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works", *Expert systems with applications*, Vol. 41, No. 4, pp. 1432-1462, 2014.

[60] George Siemens, "What are learning analytics", *Retrieved March*, Vol. 10, pp. 2011, 2010.

[61] Marcelo Worsley and Paulo Blikstein, "What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis", *Educational Data Mining*, pp. 235-240, 2011.

[62] Anna De Liddo, Simon Buckingham Shum, Ivana Quinto, Michelle Bachler and Lorella Cannavacciuolo, "Discourse centric learning analytics", *Proceedings of the 1st*

International Conference on Learning Analytics and Knowledge, pp. 23-33, 2011.

[63] Erik Duval, "Attention please!: learning analytics for visualization and recommendation", *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 9-17, 2011.

[64] Rebecca Ferguson and Simon Buckingham Shum, "Learning analytics to identify exploratory dialogue within synchronous text chat", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 99-103, 2011.

[65] Hélène Fournier, Rita Kop and Hanan Sitlia, "The value of learning analytics to networked learning on a personal learning environment", *1st Learning Analytics Conference*, pp. 104-109, 2011.

[66] Abelardo Pardo and Carlos Delgado Kloos, "Stepping out of the box: towards analytics outside the learning management system", *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 163-167, 2011.

[67] George Siemens, et. al., "Open Learning Analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques, http://solaresearch.org/OpenLearningAnalytics.pdf

[68] George Siemens and Phil Long, "Penetrating the Fog: Analytics in Learning and Education", *EDUCAUSE review*, Vol. 46, No. 5, pp. 31-40, 2011.

[69] Liaqat Ali, Marek Hatala, Dragan Gašević and Jelena Jovanović, "A qualitative evaluation of evolution of a learning analytics tool", *Computers & Education*, Vol. 58, No. 1, pp. 470-489, 2012.

[70] Kimberly E. Arnold and Matthew D. Pistilli, "Course signals at Purdue: using learning analytics to increase student success", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* pp. 267-270, 2012.

[71] Angela van Barneveld, Kimberly Arnold and John Campbell, "Analytics in higher education: Establishing a common language", *EDUCAUSE learning initiative*, 2012.

[72] M. A. Chatti, A. L. Dyckhoff, U. Schroeder and H. Thüs, "A reference model for learning analytics", *International Journal of Technology Enhanced Learning*, Vol. 4, No. 5, pp. 318-331, 2012.

[73] Doug Clow, "The learning analytics cycle: closing the loop effectively", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 134-138, 2012.

[74] Hendrik Drachsler and Wolfgang Greller, "The pulse of learning analytics understandings and expectations from the stakeholders", *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 120-129, 2012.

[75] Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti and Ulrik Schroeder, "Design and implementation of a learning analytics toolkit for teachers", *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 58-76, 2012.

[76] Rebecca Ferguson, "Learning analytics: drivers, developments and challenges", *International Journal of Technology Enhanced Learning*, Vol. 4, No. 5/6, pp. 304-317, 2012.

[77] Rebecca Ferguson and Simon Buckingham Shum, "Social learning analytics: five approaches", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 23-33, 2012.

[78] Wolfgang Greller and Hendrik Drachsler, "Translating learning into numbers: A generic framework for learning analytics", *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 42-57, 2012.

[79] Leah P. Macfadyen and Shane Dawson, "Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan", *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 149-163, 2012.

[80] Karen D. Mattingly, Margaret C. Rice and Zane L. Berge, "Learning analytics as a tool for closing the assessment loop in higher education", *Knowledge Management & E-Learning*, Vol. 4, No. 3, pp. 236-247, 2012.

[81] Anthony G. Picciano, "The evolution of big data and learning analytics in American higher education", *Journal of Asynchronous Learning Networks*, Vol. 16, No. 3, pp. 9-20, 2012.

[82] Ángel Serrano-Laguna, Javier Torrente, Pablo Moreno-Ger and Baltasar Fernández-Manjón, "Tracing a little for big improvements: Application of learning analytics and videogames for student assessment" *Procedia Computer Science*, Vol. 15, pp. 203-209, 2012.

[83] Simon Buckingham Shum and Ruth Deakin Crick, "Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 92-101, 2012.

[84] Simon Buckingham Shum and Rebecca Ferguson, "Social learning analytics", Journal of educational technology & society, Vol. 15, No. 3, pp. 3-26, 2012.

[85] George Siemens, "Learning analytics: envisioning a research discipline and a domain of practice", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 4-8, 2012.

[86] George Siemens and Ryan S. J. D. Baker, "Learning analytics and educational data mining: towards communication and collaboration", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 252-254, 2012.

[87] George Siemens and Dragan Gasevic, "Guest editorial - Learning and Knowledge Analytics", *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 1-2, 2012.

[88] Katrien Verbert, Nikos Manouselis, Hendrik Drachsler and Erik Duval, "Dataset-driven research to support learning and knowledge analytics", *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 133-148, 2012.

[89] Mathieu D. Aquin and Nicolas Jay, "Interpreting data mining results with linked data for learning analytics: motivation, case study and directions", *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 155-164, 2013.

[90] A. del Blanco, A. Serrano, M. Freire, I. Martínez-Ortiz and B. Fernández-Manjón, " E-Learning standards and learning analytics. Can data collection be improved by using

standard data models", *Global Engineering Education Conference (EDUCON) IEEE*, pp. 1255-1261, 2013.

[91] Paulo Blikstein, "Multimodal learning analytics", *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* , pp. 102-106, 2013.

[92] Doug Clow, "An overview of learning analytics", *Teaching in Higher Education*, Vol. 18, No. 6, pp. 683-695, 2013.

[93] Doug Clow, "MOOCs and the funnel of participation", *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 185-189, 2013.

[94] Beth Dietz-Uhler and Janet E Hurn, "Using learning analytics to predict (and improve) student success: A faculty perspective", *Journal of Interactive Online Learning*, Vol. 12, No. 1, pp. 17-26, 2013.

[95] Cath Ellis, "Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology*, 44(4), 662-664, 2013.

[96] Lori Lockyer, Elizabeth Heathcote and Shane Dawson, "Informing pedagogical action: Aligning learning analytics with learning design", *American Behavioral Scientist*, 2013.

[97] Paul Prinsloo and Sharon Slade, "An evaluation of policy frameworks for addressing ethical considerations in learning analytics", *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 240-244, 2013.

[98] Sharon Slade and Paul Prinsloo, "Learning analytics ethical issues and dilemmas", *American Behavioral Scientist*, Vol. 57, No. 10, pp. 1510-1529, 2013.

[99] Dan Suthers and Katrien Verbert, "Learning analytics as a middle space", *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 1-4, 2013.

[100] Davide Taibi and Stefan Dietze, "Fostering analytics on learning analytics research: the LAK dataset", http://ceur-ws.org/Vol-974/lakdatachallenge2013_preface.pdf. Accessed in 2013.

[101] Katrien Verbert, Erik Duval , Joris Klerkx, Sten Govaerts and José Luis Santos, "Learning analytics dashboard applications", *American Behavioral Scientist*, Vol. 57, No. 10, pp. 1500-1509, 2013.

[102] Alyssa Friend Wise, Yuting Zhao and Simone Nicole Hausknecht, "Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics", *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 48-56, 2013.

[103] Ryan Shaun Baker and Paul Salvador Inventado, "Educational data mining and learning analytics", *Learning Analytics*, pp. 61-75, 2014.