A Research Travelogue Towards Educational Data Mining

Bernard Ugalde
AMA University Online Education
Quezon City, Philippines

R. Venkateswaran
Salalah College of Technology
Salalah, OMAN

ABSTRACT
In the recent era of technology, educational institutions innovate itself to find new ways to serve its educational community efficiently and effectively. Information systems have been there for quite some time as the backbone of education institutions to support its daily operations. At this point, educational databases have much information but remain utilized. In order to make benefit from such big data, a power tool is required like data mining for analysis and prediction. Data mining has been proven useful in various aspects of our lives like in advertising, marketing, loans and now a new frontier in the field of education. It has been noted that there is no unified approach among researchers in educational data mining and a considerable amount of work is required towards this field. This research presents a comprehensive travelogue (2010- 2017) in educational data mining with respect to related international journals available from various sources, and secondary data collected from the organization in the form of survey reports.

Keywords
Educational Data Mining, Trends in EMD, Future vision of EDM

1. INTRODUCTION
In the last decade, educational institutions have been continuously innovating itself to cope up with the fast-changing industry needs. They employ qualified staff, create state-of-the-art laboratories for practice and simulation, develop and employ a tool that will help in the administration and management of schools. However, with all of these measures in place, schools are still confronted with problems of student's dropout, poor academic performance, and unemployment of graduates.

Although educational institutions collect and keep student records, it has been noted that most of the time these data and information are not utilized properly. Data are just collected every academic year, and are kept in filing cabinets in school vaults, school's information systems or dump somewhere.

Analyzing and understanding factors that affect the number of dropouts, monitoring academic performance, keeping track of alumni employment in education institution entails complex, tedious and endless. Academic marks are maintained by these organizations for purpose of printing student's transcript later. Data on a number of enrollments and analysis on student's demographics and background are unutilized and does not support any decision or policy-making within institutions. Thus, in order for these problems to be addressed, a very useful tool for an educational institution is needed. There must be a way to analyze, understand and predict performance needs of students scientifically, and a way to inform and support the whole educational community.

Only if universities could get a clear picture of the current standing of the entire student populations of the institution that will help them make pro-active measures to form of policies and programs that will align and address the needs of the students and the whole school community. This will promote efficient and effective learning on the side of the student. On the teacher's side, they can plan ahead with their delivery inside the classroom. School management will get its benefit by attracting a great number of enrollees. As a whole getting the quality education that each deserved. In the first place, the genuine goal of each educational institution is to let the student learn and let them add value to the workforce later.

Educational data mining have shown noteworthy advantages to education institutions It is a very powerful tool to reveal the pattern and precious knowledge, which otherwise may not be identified. It also helps to respond to educational question and problems that might be needed data analysis and interpretation. Substantial work has been done towards the usage of data mining techniques in education, but still, there are many untouched areas and no unified approach is followed.

This paper presents a comprehensive literature review of relevant researches done in educational data mining from 2010 to 2017. Section 2 gives a short note about educational data mining. Section 3 presents the review of related researchers in a form of a table. Section 4, 5 and 6 presents the research objectives, methodology, and findings of this study respectively, followed by conclusions and recommendation, and references.

2. A BRIEF NOTE ABOUT EDUCATIONAL DATA MINING
The following discusses educational data mining in several areas as follows:

2.1 What is Educational Data Mining?
According to International Educational Data Mining Society [1], Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in. EDM focuses on developing new tools and algorithm for discovering data patterns [2]. Particularly in education settings, it can answer questions like (1) what sequences of topics is most effective for a specific student? (2) Which student action is associated with better learning and higher grades? (3) Which action indicates satisfaction and engagement? (4) What features of an online learning environment lead to better learning?

2.2 Goals of Educational Data Mining
Table 1 below presents the goal of educational data mining. EDM can predict student future learning behavior and can

determine the optimal instructional sequences to support the student's learning style. It provides a clear picture of the academic performance of a student and determines the needed support. Finally, EDM test learning theories and inform its stakeholders about the effectiveness and efficiency of the whole teaching-learning process. All of these relates to the direct participation of the student in teaching and learning process which falls under EDM's student-oriented goal.

Table 1 Goals of EDM

SN	Goal
1	Predict student's future learning behavior.
2	Determine the optimal instructional sequences to support the student's learning style.
3	Provide a clear picture of the academic performance of a student.
4	Determine the needed academic support of the student, and assess if the support given is effective
5	Provide feedback and adapt learning recommendations for behavior
6	Test and evaluate the learning theories and learning materials.
7	Evaluates the whole teaching-learning process

A. *Educational Data Mining Process*

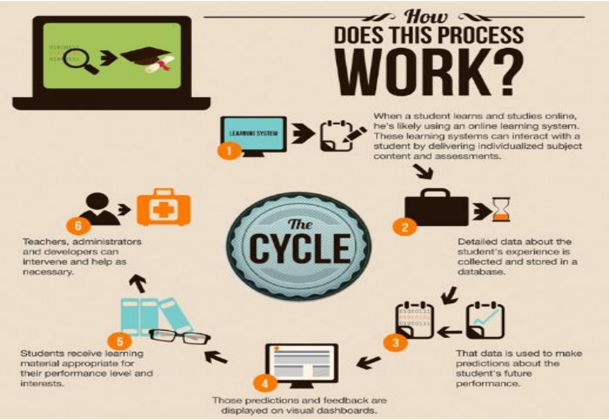

Figure 1. EDM Process

According to [3], EDM is viewed as a cycle. It starts with data collection and pre-processing. Students are directed to use the online learning system. Delivery of the course is based on individualized subject content and assessment metrics. Detailed student's performance and experience are recorded and stored in the database. The student information is transformed into models that are used to predict the future learning behavior and feedback using dashboard. A student will receive learning materials based on their performance level and learning style. Along the ways, other stakeholders can intervene to mitigate any issues or difficulties encountered. The whole EDM process is illustrated in figure 1.

2.3 Educational Data Mining Stakeholders

There are generally three stakeholders in EDM but can change based on the size of the organization. Teachers prepare and develop the course materials and set the assessment metric. They mentor the students by monitoring student's academic performance and assessing their needs. They facilitate the learning process of the student. In a smaller organization, teachers are also tasked to prepare the course objectives and goals. They will also analyze the whole student's experience in the database and provide necessary actions. Students will undergo the learning process by accessing the learning materials and assessments. At the end, he then gives feedback and recommendation to the learning process as a whole. Administrators and management set priorities, approve plans, support the cost incurred in from the conceptualization of the EDM to its implementation and maintenance. Table 2 summarizes the stakeholders of EDM.

Table 2. Stakeholders of EDM

Stakeholder	Role
Teachers	Prepares and develop the course materials and sets the assessment metric. Mentors the students by monitoring student's academic performance and assessing their needs. Facilitate the learning process of the student. Prepare the course objectives and goals. Analyze the whole student's experience in the database and provide necessary actions.
Students	Undergo the learning process by accessing the learning materials and assessments. Give feedback and recommendation to the learning process as a whole.
Management and Administrator	Set priorities; approve plans support the cost incurred in from the conceptualization of the EDM to its implementation and maintenance.

2.4 EMD Methods and Techniques

To discover knowledge from databases, various algorithms, methods, and techniques are used. The following DM methods are popular with the EDM research community.

1. **Prediction** is a technique which goal is to develop a model that can infer a single aspect of data from some combination of other aspects of data [4]. Generally, it has three types: classification, regression, and density estimation. Classification uses can use either categorical or binary input variables, while regression uses continuous input variables. Density estimation can be done with the help of various kernel functions.
2. **Clustering** divides data set into various groups, known as clusters [5]. It is best used if the data set is already specified. As per clustering phenomenon, the data point of one cluster and should be more similar to other data points of the same cluster and more dissimilar to data points of another cluster. There are two ways of initiation of a clustering algorithm. Firstly, start the clustering algorithm with no prior assumption and second is to start clustering algorithm with a prior postulate.
3. **Relationship mining** generally refers to contrive new relationships between variables, done on a large data set, having a no of variables [6]. It attempts to discover the variable which is most closely associated with the specified variable. There are four types of relationship mining: association rule mining, correlation mining, and sequential pattern mining and causal data mining. Association data mining is based on an if-then rule that is if some particular set of variable value appears then it generally has a specified value. In correlation mining, the linear correlations are discovered between variables. The aim of

sequential pattern mining is to extract temporal relationships between variables.

4. **Discovery** with models includes the designing of model based on some of the basic concepts like prediction, clustering and knowledge engineering used to discover new predicted variables [7].

5. **Distillation of data for human judgment** by identifying and classifying the data in a way that human can easily recognize the pattern [8].

2.5 EMD Tools

The following table presents the popular tools use for educational data mining though there are many tools available right now on the market.

Table 3. Tools of EDM

Ref	Name of Tool and Developer	License	Functions and Features	Techniques and Tools	Operating System
[9]	MSSQL Server (Microsoft)	Commercial	Provides DM functions both in relational db system and Data Warehouse (DWH) system environment.	Integrates the algorithms developed by third party vendors and application users.	Windows, Linux
[10]	Oracle Data Mining (Oracle Corp)	Commercial	Provides an embedded DWH infrastructure for multidimensional data analysis	Association Mining, Classification, Prediction, Regression, Clustering, Sequence similarity search and analysis.	Windows, Mac, Linux
[11]	SPSS Clementine (IBM)	Commercial	Provides an integrated data mining development environment for end users and developers.	Association Mining, Clustering, Classification, Prediction and visualization tools	Windows, Solaris, Linux
[12]	WEKA (University of Waikato, New Zealand)	Open/ free	Provides machine learning algorithms for data mining tasks. Well-suited for developing new machine learning schemes.	Data pre-processing, classification, regression, clustering, association rules, and visualization.	Windows, Linux
[13]	ORANGE (University of Ljubljana, Slovenia)	Open/ free	Provides open source data visualization and analysis for novice and experts	Text mining and Bioinformatics add-on	Windows, Linux
[14]	Carrot	Open/ free	Provides ready-to-use components for fetching search results from various sources	Clustering	Windows, Linux

All these concepts and theories in EDM mentioned in this section will be used in the following section of the study.

3. REVIEW OF RELATED LITERATURE

A comprehensive literature survey of various significant researchers in the area of Educational Data Mining ranging from the year 2010 to 2017 is presented below (Table 4). The leverage points of this survey are trends of DM methods and respective key findings.

Table 4. Educational Data Mining Papers

Reference	Year, Author(s)	Methodology	Key Findings
[15]	2017, Ms. Ganesan Kavitha, and Dr. Lawrance Raj	Gain ratio	Analyzed students' performance in their assessment to discover the students at risk of failing the final exam
[16]	2017, Jirˇí Rˇihák and Radek Pelánek	Clustering techniques, matrix factorizations and neural networks	Measured similarity of educational items using data on learners' performance
[17]	2017, Aaron Bauer, Jeff Flatten and Zoran	Iterative visualization-based methodolog	Analyzed problem-solving behavior in

	Popovi´c	y	open-ended scientific-discovery game challenges
[18]	2017, Zhiyun Ren, Xia Ning, Huzefa Rangwala	factorization-based approach called Matrix Factorization with Temporal Course-wise Influence	Predicted Grade with Temporal Course-wise Influence
[19]	2017, Angela Stewart, Nigel Bosch and Sidney K. D'Mello	Variety of machine learning techniques	Generalizes Face-Based Mind Wandering Detection Across Task Contexts
[20]	2017, Andrew M. Olney, Dariush Bakhtiari, Daphne Greenberg and Art Graesser	Q-MATRIX & LOGISTIC MIXED MODELS	Assessed computer literacy of adults with low literacy skills
[21]	2017, Renu Balyan, Kathryn S. McCarthy and Danielle S. McNamara	Random Forests, Trees and Support Vector Machines	Assessed Literary Text Comprehension
[22]	2016, Ankita Katar and Shubha Dubey	Classification algorithms	Evaluated student performance
[23]	2016, Amjad Abu Saa	Multiple classification methods	Predicted the students' grade at the end of the semester
[24]	2016, Alexandria K. Vail, Joseph B. Wiggins, Joseph F. Grafsgaard, Kristy Elizabeth Boyer, Eric N. Wiebe and James C. Lester	Regression	Analyzed of student affective response, as evidenced by multimodal data streams, immediately following tutor questions.
[25]	2016, Martin Stapel, Zhilin Zheng and Niels Pinkwart	Specialized scope classifiers	Predict student performance
[26]	2016, Mirka Saarela and Tommi K¨arkk¨ainen	Robust clustering technique	Analyzed students' overall performance

			based on learning skill and core studies
[27]	2016, Arjun Sharma, Arijit Biswas, Ankit Gandhi, Sonal Patil, Om Deshmukh	Neural network	predicted liveliness in educational videos
[28]	2016, Ke Niu, Zhendong Niu, Xiangyu Zhao, Can Wang, Kai Kang, Min Ye	Clustering algorithms	Analyzed users' learning behaviors and help to provide personalized learning guides in traditional Web-based learning systems.
[29]	2015, Engin Bumbacher, Shima Salehi, Miriam Wierzchula and Paulo Blikstein	Cluster analysis	How learning environment and inquiry behaviors affects the development of conceptual understanding in Physics
[30]	2015, Juraj Nižnan, Radek Pelánek, Jirˇí Rˇ ihák	Bayesian extension, a hierarchical model, and a networked model	Analyzed Student Models for Prior Knowledge Estimation
[31]	2015, Yang Chen, Pierre-Henri Wuillemin, Jean-Marc Labat	Probabilistic Association Rules	Discovering Prerequisite Structure of Skills
[32]	2015, Aysu Ezen-Can and Kristy Elizabeth Boyer	Multiple regression models	Identified the factors that may be influential in students' levels of interaction with the system
[33]	2015, Mirka Saarela, Tommi Kärkkäinen	Clustering Approach	Determine if country stereotypes exist in Programme for International Student Assessment (PISA)
[34]	2014, Peña-Ayala, Alejandro	Statistical and Clustering Processes	Identified kinds of educational systems,

			disciplines, tasks, methods, and algorithms.
[35]	2014, Saranya, S., R. Ayyappan, and N. Kumar	Naive Bayes Algorithm	Graphically represented institutional growth prognosis and students' progress analysis.
[36]	2014, Archer, Elizabeth, Yuraisha Bianca Chetty, and Paul Prinsloo	Experimental Usage of Employee Profiling Software	Experimented the usage of a commercial product generally used for employee profiling in corporate, for higher education environment.
[37]	2014, Hicheur Cairns, Awatef, et al.	Clustering Technique	Professionals' data was analyzed during training of a consulting company.
[38]	2014, Arora, Rakesh and Dharmendra Badal	Association Analysis Algorithm	Found set of weak students based on graduation and post-graduation marks.
[39]	2013, Stephen E. Fancsali, Tristan Nixon, and Steven Ritter	Bayesian Knowledge Tracing	Optimal and Worst-Case Performance of Mastery Learning Assessment with Bayesian Knowledge Tracing
[40]	2013, Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe, James C. Lester	Bayesian Information Criterion	Automatically Recognizing Facial Expression: Predicting Engagement and Frustration
[41]	2013, William Hawkins, Neil Heffernan, Yutao Wang, Ryan S.J.d. Baker	Bayesian networks with Knowledge Tracing	Extending the Assistance Model: Analyzing the Use of Assistance over Time
[42]	2012, Osmanbegović, Edin, and Mirza Suljić	Chi-Square Test, One RTest, Info Gain and	Found predicting model for academic

		Ratio Test, Naive Bayes, DTree	performance that is user friendly for professors or non-expert users.
[43]	2012, Sukanya, M., S. Biruntha, Dr S. Karthik, and T. Kalaikumaran	Bayesian Classification Method	Analyzed and assisted the low academic achievers in higher education.
[44]	2011, Torenbeek, M., E. P. W. A. Jansen, and W. H. A. Hofman	Structural Equations Modeling, Correlation Matrix	Examined two variables, Pedagogical approach and skill development in the first 10 weeks of enrollment
[45]	2011, Yongqiang, He, and Zhang Shunli	Association Rules Analysis	Guidance provided for scientific management and comprehensive evaluation of students.
[46]	2011, Sakurai, Yoshitaka, Tsuruta, and Rainer Knauf	Decision Tree	Estimated success chances of curricula by implementing student profiling with storyboard system
[47]	2011, Aher, Sunita B., and L. M. R. J. Lobo	Classification and Clustering	Analyzed the performance of final year students.
[48]	2011, Sharma, Mamta, and Monali Mavani	Decision Tree, Sota, Naïve Bayes	Comparison of three algorithms in terms of prediction of students result.
[49]	2010, Ayesha, Shaeela, Tasleem, Ahsan, Inayat	K-Means Clustering	Analyzed students' learning behavior to check the performance of students and predicted weak students.
[50]	2010, Kovacic, Zlatko	Classification Tree Models	Investigated enrolment attributes to pre-identify success of students.
[51]	2010, Al-shargabi,	Clustering, Association	Analyzed students'

	Asma A., and Ali N. Nusari	Rules and Decision Trees	academic achievement, students' drop out, and students' financial behavior.
[52]	2010, Yan, Zhi-min, Qing Shen, and Bin Shao	Rough Set Theory	Students' grades were analyzed.
[53]	2010, Ningning, Gao	Neural Network, Rough Set Theory	Predicted drop outs from course
[54]	2010, Knauf, Rainer, Yoshitaka Sakurai, Setsuo Tsuruta, and Kouhei Takada	Decision Tree	Analyzed successful Storyboard (e-learning system) success paths for students.
[55]	2010, Wu, X., Zhang, H., & Zhang, H.	Decision Tree	Suggested comprehensive evaluation method of that can objectively distinguish the grades of students.
[56]	2010, Youping, Bian Xiangjuan Gong	Decision Tree	Evaluated the high school students and studying effectiveness.
[57]	2010, Liu, Zhiwu, and Xiuzhi Zhang	Decision Tree	Built forecasting model for students' marks to identify negative learning habits or behaviors of students.

Research Objectives

The main objective of the paper is to identify related researches in the field of EDM ranging from 2010-2017. Specifically, it sought to identify the:

1. Trends and future vision of EDM methodologies and techniques.
2. Trends and future vision of EDM tools.
3. Trends and future vision of EDM educational outcomes.

4. RESEARCH METHODOLOGY

This paper is a descriptive study using survey analysis of researches available from journals, and secondary data from various sources like application websites and website articles.

5. RESULTS AND FINDINGS

Past studies reveal the interesting field of educational data mining as a topic for research. The salient features of the features can be summarized below:

5.1 Trends and Future Vison of EDM Methodologies and Techniques

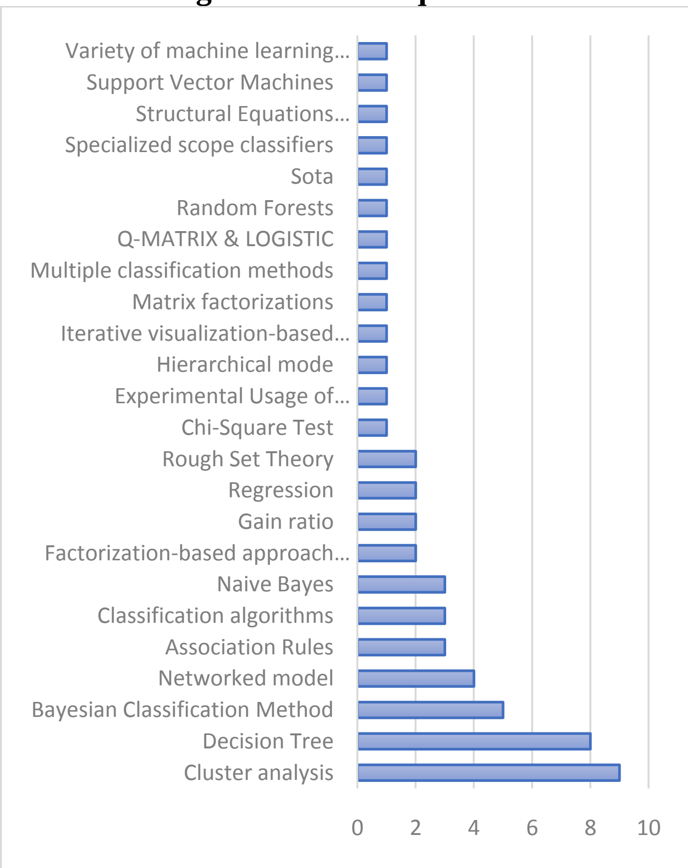

Figure 2. Trends of EDM Methodologies and Techniques

During 2010-2011 as presented in figure 2, the highest research involves decision trees. Changes in data mining methods and technique are seen during the periods of 2012-2015. During this period, highest research involve is clustering, however, it worth noting that methods used were in a combination of more than one technique.

This trend continues to be seen from the periods of 2015 to 2017, however, this time, more complex implementation was made like further factoring and iterating the basic methods.

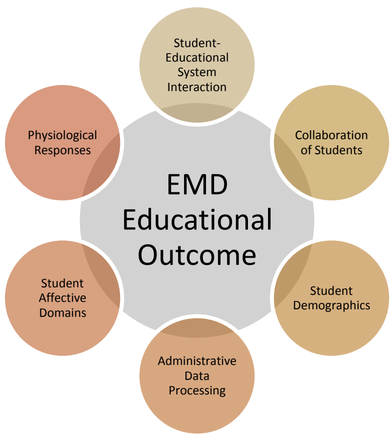

Figure 3. Journey of EDM Education Outcomes

Now and in the future data of interest as presented in figure 3 is not restricted to interactions of individual students with an educational system alone(navigation behavior, input to quizzes and interactive exercises) but might also include data from collaborating students (chat, forums, messages), administrative data (school, district, teacher, campus), and demographic data (gender, age, school grades). Data on student affective domain (motivation, emotional states) and physiological responses (facial expression, seat posture, blinking of the eyes, perspiration, and heartbeat) are also interesting areas of research.

5.2 Trends and Future Vision of EDM Educational Outcomes

During 2010-2015, topic themes revolve around forecasting and analysis students' behavior, evaluation of student academic performance, estimation of success and effectiveness of program or undertakings, and comparison of predictions.

The paradigm shift of the EDM research has been seen during 2016. During this period, noteworthy researches were prediction and forecasting patterns in educational videos, analysis of problem-solving behavior in games, and effective responses of learning using data generated from face detection devices.

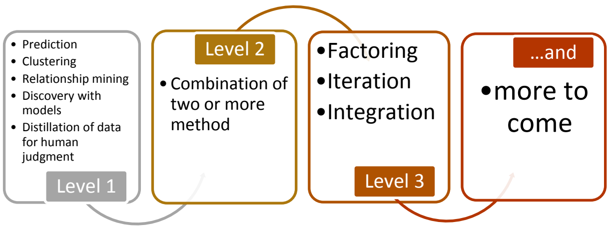

Figure 4. Journey of EDM Education Outcomes

The basic methodologies are still there as the foundation of research as shown in figure 4. However as shown earlier, researchers have combined the basic techniques to come up with meaningful research. And, it does not stop from there, as presented in the paper, researchers are applying further statistical computations like factoring, iteration, integration and the that goes along with complex education goals as presented in the earlier slides.

5.3 Trends and Future Vision of EDM Tools

DM tools are required to validate the large set of data collected from heterogeneous environments. During 2010-2017 it is found that researcher mostly preferred open source tool like WEKA and then a commercial tool such as SPSS Clementine to validate their dataset.

As mentioned, these tools will not be enough to support future EDM studies and research thus the law will find its new limits. New tools need to be developed to address the growing needs in the following areas:

1. Network Computing
2. Cloud Computing
3. Stream and Distributed Data Processing
4. Internet of Things

Furthermore, user interface designs will be more complicated as PC will fade out. The text will be replaced by the new form of data like videos and images.

6. CONCLUSION AND RECOMMENDATION

It is always a constant challenge to educational institutions to make the learning process effective allowing full learning on the side of the students and effective delivery on the side of the teachers. Thus educational institutions make its way to innovate itself and come forward to provide the best and quality learning environment, and one of them is to evaluate and analyze existing data of students find similarity, look for differences and patterns, and predict learning performance and behavior of students.

This research is a valuable paper that contains related researches in the field of educational data mining. It presents the journey of research and practice from the year 2010 to 2017. This work focuses on research trends in data mining tools, techniques, and outcomes in an educational context and its future vision.

As an extension of this work in the future is to build an EDM framework taking into consideration the specific application of tools and techniques that will use by any college to strengthen the utilization of existing resources.

7. ACKNOWLEDGEMENT

The authors wish to thank Salalah College of Technology for the continuous support on this undertaking.

8. REFERENCES

[1]"educationaldatamining.org", *Educationaldatamining.org*, 2017. [Online]. Available: http://educationaldatamining.org/. [Accessed: 13- Dec- 2017].

[2]"Mining educational data to analyze learning and teaching methods, the case of medicine | Persyval-Lab", *Persyval-lab.org*, 2017. [Online]. Available: https://persyval-lab.org/en/exploratory-project/Med_LT_med. [Accessed: 13- Dec- 2017].

[3]A. Francisco, "Realizing the Opportunity for Big Data in Education - Digital Promise", *Digital Promise*, 2017. [Online]. Available: http://digitalpromise.org/2014/04/17/realizing-the-opportunity-for-big-data-in-education/. [Accessed: 13- Dec- 2017].

[4]"Data Mining Classification & Prediction", *www.tutorialspoint.com*, 2017. [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm. [Accessed: 14- Dec- 2017].

[5]"What is Clustering in Data Mining?", *Big Data Made Simple - One source. Many perspectives.*, 2017. [Online]. Available: http://bigdata-madesimple.com/what-is-clustering-in-data-mining/. [Accessed: 14- Dec- 2017].

[6]C. Rygielski, J. Wang and D. Yen, "Data mining techniques for customer relationship management", 2017. .

[7]"Educational data mining - EduTech Wiki", *Edutechwiki.unige.ch*, 2017. [Online]. Available:

http://edutechwiki.unige.ch/en/Educational_data_mining. [Accessed: 14- Dec- 2017].

[8]D. Thanamani, "An Overview of Knowledge Discovery Databaseand Data mining Techniques", *Rroij.com*, 2017. [Online]. Available: http://www.rroij.com/open-access/an-overview-of-knowledge-discovery-databaseand-data-mining-techniques.php?aid=48833. [Accessed: 14- Dec- 2017].

[9]"MS SQL Server Tutorial", *www.tutorialspoint.com*, 2017. [Online]. Available: https://www.tutorialspoint.com/ms_sql_server/. [Accessed: 14- Dec- 2017].

[10]"Oracle Data Mining", *Oracle.com*, 2017. [Online]. Available: http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html. [Accessed: 14- Dec- 2017].

[11]D. science and P. analytics, "IBM SPSS Software | IBM Analytics", *Ibm.com*, 2017. [Online]. Available: https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software. [Accessed: 14- Dec- 2017].

[12]"Weka 3 - Data Mining with Open Source Machine Learning Software in Java", *Cs.waikato.ac.nz*, 2017. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 14- Dec- 2017].

[13]"Orange – Data Mining Fruitful & Fun", *Orange.biolab.si*, 2017. [Online]. Available: https://orange.biolab.si/. [Accessed: 14- Dec- 2017].

[14]D. Weiss, "Carrot2 - Open Source Search Results Clustering Engine", *Project.carrot2.org*, 2017. [Online]. Available: https://project.carrot2.org/. [Accessed: 14- Dec- 2017].

[15]G. Kavitha and L. Raj, "Educational Data Mining and Learning Analytics Educational Assistance for Teaching and Learning", *International Journal of Computer & organization Trends*, vol. 41, no. 1, pp. 21-25, 2017.

[16]J. Rˇ ihák and R. Pelánek, "Measuring Similarity of Educational Items Using Data on Learners' Performance", *10th International Conference on Educational Data Mining*, China, 2017.

[17]J. Rˇ ihák and R. Pelánek, "Measuring Similarity of Educational Items Using Data on Learners' Performance", *10th International Conference on Educational Data Mining*, China, 2017.

[18]Z. Ren, X. Ning and H. Rangwala, "Grade Prediction with Temporal Course-wise Influence", in *10th International Conference on Educational Data Mining*, China, 2017.

[19]A. Stewart, N. Bosch and S. D'Mello, "Generalizability of Face-Based Mind Wandering Detection Across Task Contexts", *10th International Conference on Educational Data Mining*, China, 2017.

[20]A. Olney, D. Bakhtiari, D. Greenberg and A. Graesser, "Assessing Computer Literacy of Adults with Low Literacy Skills", in *10th International Conference on Educational Data Mining*, China, 2017.

[21]R. Balyan, K. McCarthy and D. McNamara, "Combining Machine Learning and Natural Language Processing to Assess Literary Text Comprehension", in *10th International Conference on Educational Data Mining*, China, 2017.

[22]A. Katare and S. Dubey, "A Study of various Techniques for Predicting student Performance under Educational Data Mining", International Journal of Electrical, Electronics and Computer Engineering, 2016.

[23]A. Abu, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications, vol. 7, no. 5, 2016.

[24]A. Vail, J. Wiggins, J. Grafsgaard, K. Boyer, E. Wiebe and J. Lester, "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement", 9th International Conference on Educational Data Mining, 2016.

[25]M. Stapel, Z. Zheng and N. Pinkwart, "An Ensemble Method to Predict Student Performance in an Online Math Learning Environment", in 9th International Conference on Educational Data Mining, 2016.

[26]M. Saarela and T. K ̈arkk ̈ainen, "Analysing Student Performance using Sparse Data of Core Bachelor Courses", Journal of Educational Data Mining, vol. 7, no. 1, 2016.

[27]A. Sharma, A. Biswas, A. Gandhi, S. Patil and O. Deshmukh, "LIVELINET: A Multimodal Deep Recurrent Neural Network to Predict Liveliness in Educational Videos", in 9th International Conference on Educational Data Mining, 2016.

[28]K. Niu, Z. Niu, X. Zhao, C. Wang, K. Kang and M. Ye, "A Coupled User Clustering Algorithm for Web-based Learning Systems", in 9th International Conference on Educational Data Mining, 2016.

[29]E. Bumbacher, S. Salehi, M. Wierzchula and P. Blikstein, "Learning Environments and Inquiry Behaviors in Science Inquiry Learning: How their Interplay Affects the Development of Conceptual Understanding in Physics", in 8th International Conference on Educational Data Mining, Spain, 2015.

[30]J. Nižnan, R. Pelánek and J. Rˇ ihák, "Student Models for Prior Knowledge Estimation", in 8th International Conference on Educational Data Mining, Spain, 2015.

[31]Y. Chen, P. Wuillemin and J. Labat, "Discovering Prerequisite Structure of Skills through Probabilistic Association Rules Mining", in 8th International Conference on Educational Data Mining, Spain, 2015.

[32]A. Ezen-Can and K. Boyer, "Choosing to Interact: Exploring the Relationship Between Learner Personality, Attitudes, and Tutorial Dialogue Participation", in 8th International Conference on Educational Data Mining, Spain, 2015.

[33]M. Saarela and T. Kärkkäinen, "Do Country Stereotypes Exist in PISA? A Clustering Approach for Large, Sparse, and Weighted Data.", in 8th International Conference on Educational Data Mining, Spain, 2015.

[34] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent

works." Expert systems with applications 41.4 (2014): 1432-1462.

[35] Saranya, S., R. Ayyappan, and N. Kumar. "Student Progress Analysis and Educational Institutional Growth Prognosis Using Data Mining." International Journal Of Engineering Sciences & Research Technology, 2014

[36] Archer, Elizabeth, Yuraisha Bianca Chetty, and Paul Prinsloo. "Benchmarking the habits and behaviors of successful students: A case study of academic-business collaboration." The International Review of Research in Open and Distance Learning 15.1 (2014).

[37] Hicheur Cairns, Awatef, et al. "Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining." IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management. 2014.

[38] Arora, Rakesh Kumar, and Dharmendra Badal. "Mining Association Rules to Improve Academic Performance." (2014).

[39]S. Fancsali, T. Nixon and S. Ritter, "Optimal and Worst-Case Performance of Mastery Learning Assessment with Bayesian Knowledge Tracing", in International Conference on Educational Data Mining, USA, 2013.

[40]J. Grafsgaard, J. Wiggins, K. Boyer, E. Wiebe and J. Lester, "Automatically Recognizing Facial Expression: Predicting Engagement and Frustration", in International Conference on Educational Data Mining, USA, 2013.

[41]W. Hawkins, N. Heffernan, Y. Wang and R. Baker, "Extending the Assistance Model: Analyzing the Use of Assistance over Time", in International Conference on Educational Data Mining, USA, 2013.

[42] Osmanbegović, Edin, and Mirza Suljić. "Data mining approach for predicting student performance." Economic Review 10.1 (2012).

[43] Sukanya, M., S. Biruntha, Dr S. Karthik, and T. Kalaikumaran. "Data mining: Performance improvement in education sector using classification and clustering algorithm." In International conference on computing and control engineering,(ICCCE 2012), vol. 12. 2012.

[44] Torenbeek, M., E. P. W. A. Jansen, and W. H. A. Hofman. "Predicting first-year achievement by pedagogy and skill development in the first weeks at university." Teaching in Higher Education 16.6 (2011): 655-668.

[45] Yongqiang, He, and Zhang Shunli. "Application of Data Mining on Students' Quality Evaluation." Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on. IEEE, 2011.

[46] Sakurai, Yoshitaka, Setsuo Tsuruta, and Rainer Knauf. "Success Chances Estimation of University Curricula Based on Educational History, Self-Estimated Intellectual Traits and Vocational Ambitions." Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on. IEEE, 2011.

[47] Aher, Sunita B., and L. M. R. J. Lobo. "Data mining in educational system using Weka." IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT). Vol. 3. 2011.

[48] Sharma, Mamta, and Monali Mavani. "Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector." Advances in Computing, Communication and Control. Springer Berlin Heidelberg, 2011. 189-194.

[49] Ayesha, Shaeela, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat Khan. "Data mining model for higher education system." Europen Journal of Scientific Research 43, no. 1 (2010): 24-29.

[50] Kovacic, Zlatko. "Early prediction of student success: Mining students' enrolment data." (2010).

[51] Al-shargabi, Asma A., and Ali N. Nusari. "Discovering vital patterns from UST students data by applying data mining techniques." Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on. Vol. 2. IEEE, 2010.

[52] Yan, Zhi-min, Qing Shen, and Bin Shao. "The analysis of student's grade based on Rough Sets." Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on. IEEE, 2010.

[53] Ningning, Gao. "Proposing Data Warehouse and Data Mining in Teaching Management Research." Information Technology and Applications (IFITA), 2010 International Forum on. Vol. 1. IEEE, 2010.

[54] Knauf, Rainer, Yoshitaka Sakurai, Setsuo Tsuruta, and Kouhei Takada. "Empirical evaluation of a data mining method for success chance estimation of university curricula." In Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pp. 1127-1133. IEEE, 2010.

[55] Wu, X., Zhang, H., & Zhang, H. (2010, October). Study of comprehensive evaluation method of undergraduates based on data mining. In Intelligent Computing and Integrated Systems (ICISS), 2010 International Conference on(pp. 541-543). IEEE.

[56] Youping, Bian Xiangjuan Gong. "The application of data mining technology in analysis of college student's performance." Information Science (2010).

[57] Liu, Zhiwu, and Xiuzhi Zhang. "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm." Intelligent Networks and Intelligent Systems (ICINIS), 2010 3rd International Conference on. IEEE, 2010.

9. AUTHOR PROFILE

Mr. Bernard Ugalde is a graduate of Bachelor of Science in Computer Science, and master degree holders in Information Technology and MA in College Teaching.

Dr. R. Venkateswaran received his professional degree MCA from Bharathiar University, Tamilnadu, India, He received his M.Phil in computer science from Bharathidasan University, Tamilnadu, India, and Ph.D in Computer Science with Informaiton Security Specialization from Karpagam Academy of Higher Education, Karpagam University, Tamilnadu, India in the field of Cryptography and Network Security. Presently he is the faculty of Networking and Information Security, IT Department, Salalah College of

Technology, Salalah, OMAN from the Year 2012. He has 15 years of teaching experience and 3 years of research experience. He has participated in many national level conferences and workshops, published papers in several international conferences and proceedings. He has Published 8 papers in international refereed journals. He is a member of CSI, IAENG, IACSIT, CSTA and many online forums. He has completed Oracle 9i at Oracle University and He is Cisco Certified CCNA Professional. His research interests are in cryptography and network security, information security, software engineering and database management systems.