Data Mining for Education Decision Support: A Review

http://dx.doi.org/10.3991/ijet.v9i6.3950

Suhirman[1,2], Jasni Mohamad Zain[1], Haruna Chiroma[3], and Tutut Herawan[3]
[1]University Technology of Yogyakarta, Yogyakarta, Indonesia
[2]Universiti Malaysia Pahang, Gambang, Kuantan Pahang, Malaysia
[3]University of Malaya, Kuala Lumpur, Malaysia

***Abstract*—Management of higher education must continue to evaluate on an ongoing basis in order to improve the quality of institutions. This will be able to do the necessary evaluation of various data, information, and knowledge of both internal and external institutions. They plan to use more efficiently the collected data, develop tools so that to collect and direct management information, in order to support managerial decision making. The collected data could be utilized to evaluate quality, perform analyses and diagnoses, evaluate dependability to the standards and practices of curricula and syllabi, and suggest alternatives in decision processes. Data minings to support decision making are well suited methods to provide decision support in the education environments, by generating and presenting relevant information and knowledge towards quality improvement of education processes. In educational domain, this information is very useful since it can be used as a base for investigating and enhancing the current educational standards and managements. In this paper, a review on data mining for academic decision support in education field is presented. The details of this paper will review on recent data mining in educational field and outlines future researches in educational data mining.**

***Index Terms*—Data mining; Decision making; Education; Review.**

I. INTRODUCTION

Higher education institutions are overwhelmed with huge amounts of information regarding student's enrollment, number of courses completed, achievement in each course, performance indicators and other data. This has led to an increasingly complex analysis process of the growing volume of data and to the incapability to take decisions regarding curricula reform and restructuring. On the other side, educational data mining is a growing field aiming at discovering knowledge from student's data in order to thoroughly understand the learning process and take appropriate actions to improve the student's performance and the quality of the courses delivery [1].

The biggest challenge is how to predict college potential challenges and opportunities in the future, including the quality of inputs, processes and outputs. Anticipate the flow of information, the data that is in college to be optimized utilization. The significance of computer science for economics and society is undisputed. In particular, computer science is acknowledged to play a key role in schools (e.g., by opening multiple career paths) [2]. So that predictions can answer required. Prediction is very useful in management decision making colleges. One way to predict the future challenges this college is to analyze the data using data mining techniques. Data mining has been widely demonstrated success in predicting profits in the business world, but is rarely used to predict the gains in education. Results indicate a significant relationship between students use of technology for academic purposes and self-reported educational gains as well as technological gains. These results add support to claims of the need for a more symbiotic relationship between technology and pedagogy [3].

Data mining and knowledge discovery in databases are treated as synonyms, but data mining is actually a step in the process of knowledge discovery. The sequences of steps indentified in extracting knowledge from data are shown in Figure 1.

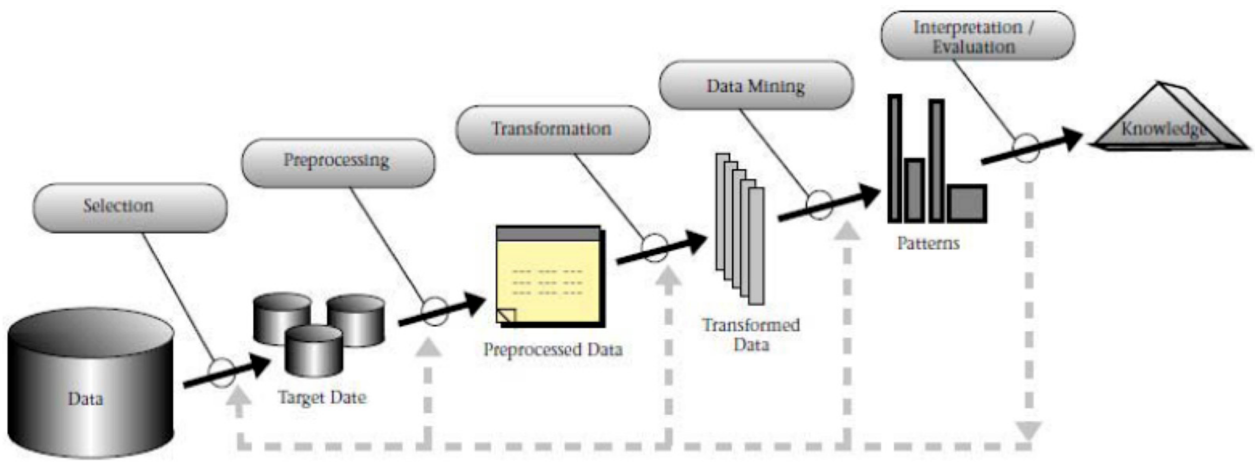

Figure 1. The steps of extracting knowledge from data [4]

The main functionality of data mining techniques is applying various methods and algorithms in order to discover and extract patterns of stored data. These interesting patterns are presented to the user and may be stored as new knowledge in knowledge base. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making [5]. Data mining (DM) and Decision Support Systems (DSS) are well suited technologies to provide decision support in the higher education environments, by generating and presenting relevant information and knowledge towards quality improvement of education processes and management [6].

The main objective of higher education institutions is to provide quality education. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Data mining techniques in context of higher education by offering a data mining model for higher education system in the university [7]. In today's competitive situation, institution need to use discovery knowledge techniques to make better, more informed decisions. Its main advantage is that by simply indicating where the data file is, the service itself is able to perform all the process [8].

In this paper, a review on data mining for academic decision support in higher education is presented. The paper reviews on recent data mining in educational field and outlines future researches in educational data mining. The rest of this paper is organized as follow. Section 2 describes educational data mining definition and its methods. Section 3 reviews on existing recent works on educational data mining. Finally, the conclusion of this work and future research direction are described in Section 4.

II. EDUCATIONAL DATA MINING

A. *Definition*

Educational datamining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. This paper surveys the most relevant studies carried out in this field to date. First, it introduces EDM and describes the different groups of user, types of educational environments, and the data they provide. It then goes on to list the most typical/common tasks in the educational environment that have been resolved through data-mining techniques, and finally, some of the most promising future lines of research are discussed. The field of educational data mining is a ripe for explosive growth. Machine learning and data mining have developed a vast array of tools and techniques that have been wellstudied and examined in myriad context.

Educational data mining can be applied to wide areas of research including elearning, intelligent tutoring systems, text mining, social network mining etc. In education, EDM can function as a replacement for less accurate but more established psychometric techniques. Educational data mining is an interactive cycle of hypothesis formation, testing and refinements that alternates between two complementary types of activities. One type of activity is qualitative analysis, focuses on understanding individual tutorial events. Other type involve, knowledge tracing analyses the growth curve by aggregating over successive opportunities to apply skills [5].

The emerging fields of academic analytics and educational data mining are rapidly producing new possibilities for gathering, analyzing, and presenting student data. University might soon be able to use these new data sources as guides for course redesign and as evidence for implementing new assessments and lines of communication between instructors and students. This essay links the concepts of academic analytics, data mining in higher education, and course management system audits and suggests how these techniques and the data they produce might be useful to those who practice the scholarship of teaching and learning [9].

The EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. This process does not differ much from other application areas of DM, like business, genetics, medicine, etc., because it follows the same steps as the general DM process [10]. Even so, there are some important issues that differentiate the application of DM, specifically to education, from howit is applied in other domains :

a) *Objective:* The objective of DM in each application area is different. For example, in EDM, there are both applied research objectives, such as improving the learning process and guiding students' learning, as well as pure research objectives, such as achieving a deeper understanding of educational phenomena. These goals are sometimes difficult to quantify and require their own special set of measurement techniques.
b) Data: In educational environments, there are many different types of data available for mining. These data are specific to the educational area, and therefore have intrinsic semantic information, relationships with other data, and multiple levels of meaningful hierarchy.
c) Techniques: Educational data and problems have some special characteristics that require the issue ofmining to be treated in a different way. Although most of the traditional DM techniques can be applied directly, others cannot and have to be adapted to the specific educational problem.

There are actually more groups involved with many more objectives, namely :

a) Learners/students
b) Educators/lecturers
c) Course/educational/researchers
d) Organizations/learning/providers/universities/privat training companies
e) Administrators

Lecturers and academics section are in charge of planning, designing, building and maintaining the educational systems. Students use and interact with them. Starting from all the available information about courses, students, usage and interaction, different data mining techniques can be applied in order to discover useful knowledge that helps to improve the learning process. The discovered knowledge can be used not only by providers (lecturers) but also by own users (students). So, the application of data mining in educational systems can be oriented to different actors with each particular point of view (figure 2). An existing learning management system is improved by using data mining techniques and increasing the efficiency of the courses using custom modules [11].

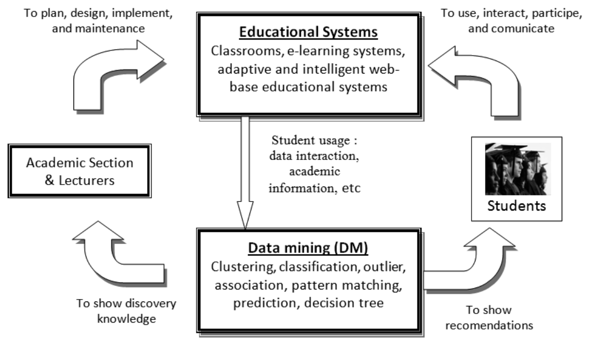

Figure 2. The cycle of applying data mining in educational systems

B. Methods

Baradwaj and Pal [7] Categorize methods in educational data mining into the following general categories. Viewpoint is focused on applications of educational data mining to data. These methods are listed as web mining methods, and are quite prominent in mining web data and in mining other forms of educational data. These categories of educational data mining methods are largely acknowledged to be universal across types of data mining.

1) Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

2) Clustering

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

3) Prediction

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

4) Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

Association Rules Mining is one of the popular techniques used in data mining. Positive association rules are very useful in correlation analysis and decision making processes. In educational context, determine a "right" program to the students is very unclear especially when their chosen programs are not selected. In this case, normally they will be offered to other programs based on the programs‟‟ availability and not according to their program‟‟s field interests. The main concern is, by assigning inappropriate program which is not reflected their overall interest; it may create serious problems such as poorly in academic commitment and academic achievement [12]. Least association rules are the association rules that consist of the least item. These rules are very important and critical since they can be used to detect the infrequent events and exceptional cases. However, the formulation of measurement to efficiently discover least association rules is quite intricate and not really straight forward [13].

Sequential rule mining is an important data mining task used in a wide range of applications. However, current algorithms for discovering sequential rules common to several sequences use very restrictive definitions of sequential rules, which make them unable to recognize that similar rules can describe a same phenomena [14].

5) Neural networks

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

6) Decision Trees

Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

7) Nearest Neighbor Method

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k is greater than or equal to 1). Sometimes called the k-nearest neighbor technique.

III. A REVIEW ON DATA MINING FOR ACADEMIC DECISION SUPPORT IN HIGHER EDUCATION

A. Providing Information for Supporting Educators

The objective is to provide feedback to support teachers/administrators in decision making (about how to improve students' learning, organize instructional resources more efficiently, and enable them to take appropriate proactive action. It is important to point out that this task is different than data analyzing and visualizing tasks, which only provide basic information directly from data (reports, statistics, etc.). Moreover, providing feedback completely new and interesting information. Several DM techniques have been used in this task, although association-rule mining has been the most common. Association-rule mining reveals interesting relationships among variables in large databases and presents them in the form of strong rules, according to the different degrees of interest.

There are many studies that apply/compare several DM models that provide feedback. Association rules, clustering, classification, sequential pattern analysis, dependency modeling, and prediction have been used to enhance web-based learning environments to improve the degree to which the educator can evaluate the learning process [15]. Association-rule mining has been used to confront the problem of continuous feedback in the educational process. to provide feedback to the course author about how to improve courseware [10]. to help the teacher to discover beneficial or detrimental relationships between the use of web-based educational resources and student's learning [16].

Other different DM techniques have been applied to provide feedback such as: domain-specific interactive DM to find the relationships between log data and student's behavior in an educational hypermedia system [17]. A special type of feedback is when data come specifically from tests, questions, assessments, etc. In this case, the objective is to analyze it in order to improve the questionnaires and to answer questions such as: what items / questions test the same information, and which are of the most use for predicting course/test results, etc. Several DM approaches and techniques (clustering, classification, and association analysis) have been proposed for joint use in the mining of student's assessment data [18].

Finally, another special type of feedback involves the use of text data. In this case, the objective of applying text/DM to educational data is to analyze educational contents, to summarize/analyze the learner's discussion process, etc., in order to provide instructor feedback. Automatic text analysis, content analysis, and text mining have been used to extract and identify the opinions found on Web pages in e-learning systems[19]. The applicability data mining techniques to identify the main drivers of student satisfaction in education institutions. In the end, the resulting models are to be used by the management to support the strategic decision making process [20].

B. Educational Recommendation Systems

Educational recommender system is very important. Systems in Learning Networks in order to provide learners advice on the suitable learning activities to follow. Learning networks target lifelong learners in any learning situation, at all educational levels and in all national contexts. They are community-driven because every member is able to contribute to the learning material. Existing Recommender Systems and recommendation techniques used for consumer products and other contexts are assessed on their suitability for providing navigational support. The similarities and differences are translated into specific requirements for learning and specific requirements for recommendation techniques. On the use of memory-based recommendation techniques, which calculate recommendations based on the current data set. To need is proposed a combination of memory-based recommendation techniques that appear suitable to realise personalised recommendation on learning activities in the context of e-learning.

Recommendation systems may assist learners in identifying potentially helpful information objects. Online Information Searching Strategies Inventory (OISSI) was applied to examine the participants' perceptions of how students applied information searching strategies [21]. An initial model for the design of such systems in Learning networks and a roadmap for their further development are presented. Future research should further analyze which attributes of learners and learning activities and techniques perform best recommendations [22]. Student data was mined, using clustering, association rules and numerical analysis, to find common patterns affecting the learners performance that used as a basis for providing hints to the students. Students who were provided hints achieved higher average marks [23].

The objective is to be able to make recommendations directly to the students, teachers, and administrators with respect to their activities, links to visits, the next task or problem to be done, etc., and also to be able to adapt learning contents, interfaces, and sequences to each particular student. Several DM techniques have been used for this task, but the most common are association-rule mining, clustering, and sequential pattern mining. Sequence/sequential pattern mining aims to discover the relationships between occurrences of sequential events to find if there exists any specific order in the occurrences.

C. Prediction of student academic performance

The objective of prediction is to estimate the unknown value of a variable that describes the student. In education, the values normally predicted are performance, knowledge, and score. This value can be numerical/continuous value (regression task) or categorical/discrete value (classification task). Regression analysis finds the relationship between a dependent variable and one or more independent variables. Classification is a procedure in which individual items are placed into groups based on quantitative information regarding one or more characteristics inherent in the items and based on a training set of previously labeled items. Prediction of a student's performance is one of the oldest and most popular applications of DM in education, and different techniques and models have been applied (neural networks, Bayesian networks, rule-based systems, regression, and correlation analysis). Modeling and prediction of student success is a critical task in education.Using the trained model enables interpretation of how different courses affect performance on a specific course in the future [24].

In the field of academics, data mining can be very useful in discovering valuable information which can be used for profiling students based on their academic record [25]. The classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination [7].

The various data mining techniques like classification, clustering and relationship mining can be applied on educational data to predict the performance of a student in the examination and bring out betterment in his academic performance. Rule based classification techniques can be used to predict the result of the students in the final semester based on the marks obtained by them in the previous semesters [26]. A computational method that can efficiently estimate the ability of students from of a Web-based learning environment capturing their problem solving processes [27].

An approach based on grammar guided genetic programming, which classifies students in order to predict their final grade based on features extracted from logged data in a web based education system. This approach could be quite useful for early identification of students at risk, especially in very large classes, and allows the instructor to provide information about the most relevant activities to help students have a better chance to pass a course [28]. Predicting student failure at school has become a difficult challenge due to the high number of factors that can affect the low performance of students and the imbalanced nature of these types of datasets. A genetic programming algorithm and different data mining approaches are proposed for solving these problems using real data. Firstly, we select the best attributes in order to resolve the problem of high dimensionality. Then, rebalancing of data and cost sensitive classification have been applied in order to resolve the problem of classifying imbalanced data [29].

The factors that lead to success or failure of students at placement tests is an interesting and challenging problem. Since the centralized placement tests and future academic achievements are considered to be related concepts, analysis of the success factors behind placement tests may help understand and potentially improve academic achievement. The sensitivity analysis revealed that previous test experience, whether a student has a scholarship, student's number of siblings, previous years' grade point average are among the most important predictors of the placement test scores [30]. Teachers can also benefit from the use of adaptive educational systems enabling them to detect situations in which students experience problems [31]. The interactions that students have with each other, with the instructors, and with educational resources are valuable indicators of the effectiveness of a learning experience. It is shown to have significant correlation with student academic achievement thus validating the approach to be used as a prediction mechanism [32]. Prediction scores of the courses is one of the effective approaches which helps the students to select their courses intelligently. Bayesian Network model for predicting the student course scores based of the student's educational history [33].

D. Cognitive Modeling of Student Learning

The objective of student modeling is to develop cognitive models of users/students, including a modeling of their skills and declarative knowledge. DM has been applied to automatically consider user characteristics (motivation, satisfaction, learning styles, affective status, etc.) and learning behavior in order to automate the construction of student models. The ability of an adaptive hypermedia system to create tailored environments depends mainly on the amount and accuracy of information stored in each user model. Some of the difficulties that user modeling faces are the amount of data available to create user models, the adequacy of the data, the noise within that data, and the necessity of capturing the imprecise nature of human behavior. Data mining and machine learning techniques have the ability to handle large amounts of data and to process uncertainty. These characteristicsmake these techniques suitable for automatic generation of user models that simulate decision making. So it needs surveyed different data mining techniques that can be used to efficiently and accurately capture user behavior. Hence the need for representation guidelines that show which techniques may be used more efficiently according to the task implemented by the application [34].

Student modeling is one of the key factors that affects automated tutoring systems in making instructional decisions. A student model is a model to predict the probability of a student making errors on given problems. A good student model that matches with student behavior patterns often provides useful information on learning task difficulty and transfer of learning between related problems, and thus often yields better instruction. Manual construction of such models usually requires substantial human effort, and may still miss distinctions in content and learning that have important instructional implications. In here, proposed an approach that automatically discovers student models using a state-of-art machine learning agent. The discovered model is of higher quality than human-generated models, and demonstrate how the discovered model can be used to improve a tutoring system's instruction strategy [35]. Student modeling is widely used in educational data mining and intelligent systems for making scientific discoveries guidance and to guide instruction. For both these purposes, the model has a high accuracy is essential, and researchers have incorporated various features into a model student. However, due to the different techniques using a variety of features, when evaluating approaches, not easy to figure out what is the key to the high prediction accuracy: models or features. To build such knowledge, so it takes a variety of empirical studies that show models are considered as goods, skills, and transfer model. Difficulty items better predictor than the difficulty skill or expertise of students in the addition model. Previous work has shown that considering the overall student skills better [36].

A basic question of instructional interventions is how effective it is in promoting student learning. This study to determine the relative efficacy of different instructional strategies by applying an educational data mining technique and learning decomposition. Using logistic regression to determine how much learning is caused by different methods of teaching the same skill, relative to each other. Comparing of results with a previous study, which used classical analysis techniques and reported no main effect. Our results show that there is a marginal difference, suggesting giving students scaffolding questions is less effective at promoting student learning than providing them delayed feedback. This study utilizes learning decomposition, an easier and quicker approach of evaluating the quality of ITS interventions than experimental studies. The usage of computer-intensive approach, bootstrapping, for hypothesis testing in educational data mining area [37].

E. Detecting Behaviour of Student Learning

The objective of detecting student behavior is to discover/detect those students who have some type of problem or unusual behavior such as: erroneous actions, low motivation, misuse, cheating, dropping out, academic failure, etc. Several DM techniques mainly classification, and clustering have been used to reveal these types of students to provide them with appropriate help everyday.

Student dropout occurs quite often in universities. Subsequently, an attempt was made to identifying the most appropriate learning algorithm for the prediction of students dropout. A number of experiments have taken place with data provided. interesting conclusion is that the Naive Bayes algorithm can be successfully used. A prototype web based support tool, which can automatically recognize students with high probability of dropout, has been constructed by implementing this algorithm. It was proved that the learning algorithms predict dropout of new students with satisfying accuracy and thus become a useful tool in an attempt to prevent and therefore reduce dropouts. The comparison of the six algorithms showed that the Naive Bayes algorithm is the most appropriate. A prototype web based support tool, which can recognize students with high probability of dropout. In a future work, in order to achieve the highest possible prediction accuracy with the usage of the fewest attributes (collecting student data is often expensive and time consuming and the classifier becomes more complicated), a wrapper attribute selection methodology along with the Naive Bayes algorithm will be used [38].

In the emerging field of educational data mining, a strong bias towards data-rich digital learning environments. However, in many educational institutes a lot of regular course data will probably be more readily available. This data may also be used to support and advise students in various ways, for the better of the student as well as the institute. Based on experience, the department claims to be able to distinguish the potentially successful students from the first year before the end semester. To do this in an early stage is important for the student as well as for the university, but the selection is only loosely based on assumed student similarities over the years. There is no thorough analysis. Data mining techniques may corroborate and improve the accuracy of this prediction. Furthermore, data mining techniques may point out indicators of academic success that are missed until now. The techniques are applied on data that is readily available in the institution's database [39].

Exams failure among university students has long fed a large number of debates, many education experts seeking to comprehend and explicate it, and many statisticians have tried to predict it. Understanding, predicting and preventing the academic failure are complex and continuous processes anchored in past and present information collected from scholastic situations and students' surveys, but also on scientific research based on data mining technologies. The experiments in the educational area, based on classification learning and data clustering techniques, made in order to draw up the students profile for exam failure/success. The results presented are a part of a larger research which is to be used to make numerous correlations, analysis and to be presented to the higher education institution managers, to offer a better knowledge of students present scholastic situations, their opinions regarding the each component of the educational process, and to predict some important aspects of their future scholastic situation. The purpose is to contribute to optimal managerial decision taking, in preventing students' exams failure, improving learning abilities and scholastic results [40].

The incorporates virtual reality and artificial intelligence to simulate virtual autonomous characters and their cognitive processes in dangerous working situations. It generates behaviour-based errors to support learning and risk prevention. It uses new mechanisms taking into account human factors with respect to cognitive modelling of human behaviour regarding risky situations. In the simulated environment the trainee can visualize the risks incurred during his work with the virtual agents. The emergent risks depend on the cognitive characteristics of the virtual operators and on the expertise of the trainee. The multi-agent system to support the control of virtual operators represented by virtual ognitive agents. The cognitive agents are enriched with a planner for selecting actions according to goals, the environment and to the personal characteristics of the agents. The system developed to model virtual autonomous characters and their activities in risky situations to support learning, decision-making and risk prevention. Because human-factors are essential in such a training system we based our work on a cognitive model. Our multi-agent system is based up on action selection and cognitive planning. The decisions and create a plan depending on their physical and cognitive characteristics. The control mode delimitates the choice of an action. Depending on this parameter the agent plans broadly and chooses the actions more adapted to the situation or plans to a more limited degree and compromises on safety aspects to gain productivity [41].

Analyzing data from existing system and databases developed has allowed for system enhancements and an improved ability to meet student priorities with both system developments and support services. Error data provides insight into not only system performance issues but also the system usability and user skills, which are then translated into supporting the users through design improvements and staff development and support mechanisms. Data on usage patterns provides insight for design teams into new developments and improvements required. Support request data highlights common requests and issues, allowing technical support services and even staff development activities to be tailored accordingly [42]. The availability and use of computers in teaching has seen an increase in the rate of plagiarism among students because of the wide availability of electronic texts online. A classification of types of plagiarism is presented, and an analysis is provided of the most promising technologies that have the potential of dealing with the limitations of current state-of-the-art systems. Furthermore, the article concludes with a discussion on legal and ethical issues related to the use of plagiarism detection software [43].

Presentation of framework that provides valuable knowledge to teachers and students, mainly based on fuzzy logic methodologies. the proposed framework is applied to the Didactic Planning course of Centre of Studies in Communication and Educational Technologies virtual campus. The application shows it usefulness, improving the course understanding and providing valuable knowledge to teachers about the course performance [44]. Classification methods lke Bayesian network, rule mining and decision trees can be used to extract the hidden knowledge about the students behavior. These methods can be applied on the educational data to identify the weak students and can also be used to predct the students behavior and performance in the examination [45].

F. Student Learning Groups

The objective is to create groups of students according to their customized features, personal characteristics, etc. Then, the clusters/groups of students obtained can be used by the instructor/developer to build a personalized learning system, to promote effective group learning, to provide adaptive contents, etc. The DMtechniques used in this task are classification (supervised learning) and clustering (unsupervised learning).

The adoption of Learning Management Systems to create virtual learning communities is a unstructured form of allowing collaboration that is rapidly growing. Compared to other systems that structure interactions, these environments provide data of the interaction performed at a very low level. For assessment purposes, this fact poses some difficulties to derive higher lever indicators of collaboration. So the need is proposed to shape the analysis problem as a data mining task. The typical data mining cycle bears many resemblances with proposed models for collaboration management. Some preliminary experiments using clustering to discover patterns reflecting user behaviors. Results are very encouraging and suggest several research directions.

A lesson learned from the analysis of this type of data is that data collection needs to be carefully designed and tuned to include all the possibly useful information. Finally, focusing on clustering methods, at least two additional research opportunities. First, building clusters from low level features to provide some guide to instructors about how higher level features can be derived for further analysis. A larger number of clusters would be probably more useful for this task. Secondly, clustering can be also directly applied to more elaborated data obtained in semi-structured workspaces, so that patterns can be automatically obtained instead of manually exploring individual or global reports. The data mining cycle, widely used for modeling business problems, fits very well into the recent line of research characterizing and classifying analysis methods systems. A lot of promising research directions combining aspects from both views. Data mining can be a valuable source for data processing and model building techniques. It can provide methods to represent and integrate richer domain knowledge which, in fact, is still an open problem in data mining research [46]. Because, not all profiles may be present in the population. Combining a flexible of kmeans and determine efficient starting centers based on the -matrix substantially improves the clustering results and allows for analysis of data sets previously thought impossible [47].

There has been a proliferation of web-based learning programs. Unlike traditional computerbased learning programs, It is used by a population of learners who have diverse background. How different learners access It has been investigated by several studies, which indicate that cognitive style is an important factor that influences learners' preferences. However, these studies mainly use statistical methods to analyze learners' preferences. Findings in this study show that Field Independent learners frequently use backward/forward buttons and spent less time for navigation. On the other hand, Field Dependent learners often use main menu and have more repeated visiting. The cognitive style is an important factor that determines students learning behavior. Further work needs to be undertaken with a larger sample to provide additional evidence. The decision trees, is a useful tool for classifying students' cognitive styles. The advantage of using data mining rather than statistical methods is that it is not necessary to make any assumptions. Further work can analyse students learning patterns using other classification methods, such as k-nearest neighbor or support vector machines. It would be interesting to see whether similar results would be found by using these classification methods [48].

The efficacy of online learning programs is tied to the suitability of the program in relation to the target audience. Based on the dataset that provides information on student enrollment, academic performance, and demographics extracted from a data warehouse, the factors that could distinguish students who tend to take online courses from those who do not. To address this issue, data mining methods, including classification trees and multivariate adaptive regressive splines, were employed. Unlike parametric methods that tend to return a long list of predictors, data mining methods suggest that only a few variables are relevant, namely, age and discipline. Previous research suggests that older students prefer online courses and thus a conservative approach in adopting new technology is more suitable to this audience. However, younger students have a stronger tendency to take online classes than older students. These findings can help policymakers prioritize resources for online course development and also help institutional researchers, faculty members, and instructional designers customize instructional design strategies for specific audiences [49].

Specifying the criteria of a rubric to assess an activity establishing the different quality level of proficiency of development and defining weight for every criterion is not easy. Besides, the complexity increases when the involve more than one lecturer. Reaching an agreement about the criteria and the level of proficiency might be easier taking into account the abilities student must achive according to the purpose of the subject. However, the disagreement about the weight of every criterion in an assesment rubric might easily appear. This focus on the automatic weight adjustment for the criteria of a rubric. So, It can be considered as a global perception that the whole group of lecturers have about the accuracy of solving an activity. Each lecturer makes a proposal from a set of student globally expresses who of each pair has solved better an activity for which the rubric was designed. Approach base on the pairwise learning is proposed to obtain adequate weights for the criteria of rubric. The system commits fewer errors than the lecturer and make them improve and reconsider some aspects of the rubric [50]. A path model is used to investigate how two online peer-assessment activities rubric based assessment and peer feedback affected the learning performance of assessors and assessees. Several path models were tested and found that the original mom peers and student exam scores in a prior Humanities course were removed. model did not fit when the variable of cognitive feedback from peers was included. The best fit model was the one in which direct paths from cognitive feedback from peers and student exam scores in a prior Humanities course were removed [51].

The growing demand in e-learning, numerous research have been done to enhance teaching quality in e-learning environments. The researchers have indicated that adaptive learning is a critical requirement for promoting the learning performance of students. Adaptive learning provides adaptive learning materials, learning strategies and/or courses according to a students learning style. Hence, to achieving adaptive learning environments is to identify students learning styles. A learning style classification mechanism to classify and then identify students learning styles. The proposed mechanism improves k-nearest neighbor (k-NN) classification and combines it with genetic algorithms (GA). To demonstrate the viability of the proposed mechanism, the proposed mechanism is implemented on an open-learning management system. The learning behavioral students are collected and then classified by the proposed mechanism. The experimental results indicate that the proposed classification mechanism can effectively classify and identify students learning styles [52].

If any, studies have investigated the factors that might contribute to the integration or implementation of e-learning portals in universities. The support that by eff ctively developing higher levels of Groupware systems, teachers empower students to make better, more informed decisions and facilitate the utilization of e-learning portals [53]. the extraction of rare association rules when gathering student usage data from a Moodle system. This type of rule is more difficult to find when applying traditional data mining algorithms. Some relevant results obtained when comparing several frequent and rare association rule mining algorithms [54]. Co-Clustering simultaneously measures the degree of homogeneity in both data instances and features, thus also achieving clustering and dimensionality reduction simultaneously. Students and features could be modelled as a bipartite graph and a simultaneous clustering could be posed as a bipartite graph partitioning problem [55].

G. The Analysis of Social Networks

The Analysis of Social Networks, or structural analysis, aims at studying relationships between individuals, instead of individual individual attributes or properties. A social network is considered to be a group of people, an organization or social individuals who are connected by social relationships like friendship, cooperative relations, or informations exchange. In Web-based cooperative learning environments, peer-to-peer interaction often suffers from the difficulty due to lack of exploring useful social interaction information, so that peers cannot find appropriate learning partners to make an effective cooperative learning. This problem easily results in poor learning outcomes in Web-based cooperative learning environments. Generally, learning partners assigned by instructors cannot ensure to compose suitable learning groups for individual learners in cooperative learning environments. Inappropriate learning partners not only easily lead to poor learning interaction and achievement, but also lose the meaning of cooperative learning. As a result, present a novel scheme of mining social interactive networks for recommending appropriate learning partners for individual learners in a cooperative problem-based learning environment. The experimental results reveal that the proposed scheme provides likely benefits in terms of promoting learners learning interaction and learning performance in cooperative learning environments [56].

The basic and obvious benefit of the system to the students is as a course management system that keeps information about courses they have taken and facilitates communication with their advisors. Providing social navigation support and community-based recommendation provides more benefit and encouragement to use the system. However, to encourage students to evaluate the courses they have taken, The Career section of the system is very important. The results suggest that the “do-it-for-yourself” approach succeeds in providing more course recommendations. Observing progress toward each career goal is an important motivation for the students to use the system while also providing more explicit and implicit feedback to the system [57]. The presentation a model which can automatically detect a variety of student dialoque acts as students collaborate within a computer supported collaborative learning environment. In addition, an analysis is presented which gives substantial insight as to how students learning is associated with students speech acts, knowledge that will significantly influence how this model is utilized by running learning software. Within Piagetian theory, the cognitive conflict of ideas between students is seen as beneficial for learning. Which sorts of interpersonal behaviors lead to most effective learning, however, is open to debate, with some researchers arguing that cooperation is most effective and others arguing that interpersonal conflict is a natural part of collaborative learning. In fact, interpersonal conflict is associated with positive learning, a finding that must be taken into account, in designing interventions that rely upon detectors of students speech acts [58].

Reffay and Chanier [59] have been argued that cohesion plays a central role in collaborative learning. In face-to-face classes, it can be reckoned from several visual or oral cues. In a Learning Management System environment, such cues are absent. The Social Network Analysis concepts, adapted to the collaborative distance-learning context, can help measuring the cohesion of small groups. That processing embodied in monitoring tools, can display global properties both at individual level and at group level and efficiently assist the tutor in following the collaboration within the group. It seems to be more appropriate than the long and detailed textual analysis of messages and the statistical distribution of participants contributions. The diversified social and cultural backgrounds of online learners and instructors complicate the conceptualization of online social presence and pose challenges to instructors in course design. Multiple-group confirmatory factor analysis of the scores from the Computer-Mediated Communication Questionnaire (CMCQ), using Structural Equation Modeling, to assess the equality of the underlying factor structure across the low-context culture (LCC) and the high-context culture (HCC) groups. The cultural groups perceived online social presence in a slightly different manner [60].

H. Concept Map Assessment of Classroom Learning

The objective of constructing concept maps is to help instructors/educators in the automatic process of developing/constructing concept maps. Some DM techniques (mainly, association rules, and text mining) have been used to construct concept maps. [61], Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line, referred to as linking words or linking phrases, specify the relationship between the two concepts. Concept as a perceived regularity in events or objects, or records of events or objects, designated by a label. Propositions are statements about some object or event in the universe, either naturally occurring or constructed. Propositions contain two or more concepts connected using linking words or phrases to form a meaningful statement.

For achieving the adaptive learning, a predefined concept map of a course is often used to provide adaptive learning guidance for learners. However, it is difficult and time consuming to create the concept map of a course. Thus, how to automatically create a concept map of a course becomes an interesting issue. There are Two-Phase Concept Map Construction approach to automatically construct the concept map by learners historical testing records. Phase 1 is used to preprocess the testing records; i.e., transform the numeric grade data, refine the testing records, and mine the association rules from input data. Phase 2 is used to transform the mined association rules into prerequisite relationships among learning concepts for creating the concept map. Therefore, Set Theory to transform the numeric testing records of learners into symbolic data, apply Education Theory to further refine it, and apply Data Mining approach to find its grade fuzzy association rules. Then, in Phase 2, based upon observation in real learning situation are used multiple rule [62].

Recent researches have demonstrated the importance of concept map and its versatile applications especially in e-Learning. The designing adaptive learning materials, designers need to refer to the concept map of a subject domain. Moreover, concept maps can show the whole picture and core knowledge about a subject domain. Research from literature also suggests that graphical representation of domain knowledge can reduce the problems of information overload and learning disorientation for learners. However, construction of concept maps typically relied upon domain experts in the past; it is a time consuming and high cost task. Concept maps creation for emerging new domains such as e-Learning is even more challenging due to its ongoing development nature. The constructed concept maps can provide a useful reference for researchers, who are new to the e-Learning field, to study related issues, for teachers to design adaptive learning materials, and for learners to understand the whole picture of e-Learning domain knowledge [63].

To make learning process more effective, the educational systems deliver content adapted to specific user needs. Adequate personalization requires the domain of learning to be described explicitly in a particular detail, involving relationships between knowledge elements referred to as concepts. Manual creation of necessary annotations is in the case of larger courses a demanding task. A concept relationship discovery problem that is a step in adaptive e-course authoring process. A method of automatic concept relationship discovery for an adaptive e-course. An approaches based on domain model graph analysis. The further advantage of this method is that although the variants are targeted at the e-learning domain,not limited to the presented computations are also applicable to different environments. Concept maps constructed over the Web pages should in first step serve as backbone for development of richer semantic descriptions. Involvement of social annotations or folksonomies shifts method applicability even further [64].

The subject materials of most enterprise e-training programs were mainly developed by employees of the enterprises; therefore, it becomes a challenging issue to efficiently and effectively translate the knowledge and experiences of the employees to computerized subject materials, especially for those who are not an experienced teacher. In addition, to develop an etraining course, personal ignorance or incorrect concepts might significantly affect the quality of the course if only a single employee is asked to develop the subject materials. To cope with this problem, a multiexpert e-training course design model is proposed. Accordingly, an e-training course development system has been implemented. Moreover, a practical application has showed that of the novel approach not only can improve the quality, but also help the experts to organize their domain knowledge. Nowadays, the notation of business management has changed from the emphasis of lower cost and high efficiency to the achievement of intellectual properties and innovation. Thus, how to maximize the benefits of intellectual properties via the assistance of new technologies has become an important and challenging issue. The development of e-training courses is one of the effective ways to preserve and promote the intellectual properties of an enterprise. In this paper, we propose a Multiple-Expert approach to cooperatively developing e-training courses. In the novel approach, the method is employed to elicit and integrate the course design knowledge from multiple experts [65].

I. Courseware Construction Based on Data Mining

The objective of constructing courseware is to help instructors and developers to carry out the construction/development process of courseware and learning contents automatically. On the other hand, it also tries to promote the reuse/exchange of existing learning resources among different users and systems.

Different DM techniques and models have been used to develop coursewareRought set theory try to provide a possible solution to improve the learning performance of e-Learning system. The clustering method is used to construct a clustering concept hierarchy. The rough set theory can help solve uncertain problem well [66]. Several DM techniques have been applied to reuse learning resources. Hybrid unsupervised DM techniques have been employed to facilitate LO reuse and retrieval from the Web or from different LO repositories. Comparison tests between OntoDNA and other ontology mapping and merging tools have also indicated that OntoDNA outperforms most tools in terms of precision, recall and f-measure. In the future work, The enhancing of mapping and merging algorithm on many-to-one mapping and many-to-many mapping at conceptual level and instance level for constraint mapping and merging [67].

Due to massive information overload on the web it's hard to index and reuse existing learning resources. Classifying learning resources according to domain specific concept hierarchies could address the problem of indexing and reusability. Manual classification is a tedious task and, as a result, automatic classifiers are in high demand. For this task we present an automated approach based on machine learning technique to exploit hierarchal knowledge in order to classify learning resources in a given hierarchy of concepts. The experiment that using hierarchical information and content of unclassified documents provides better accuracy [68]. Most currently used e-Learning Systems do not often offer search functionality. Even if methods are provided to search for Learning Objects (LOs), they don't usually utilize information about users' interests stored in their profiles. Moreover, most of the search engines are only use query conformity to order the result list. User profiling methods are usually absent, as behaviour analysis methods are difficult to implement in specialised e-Learning systems. In this paper, a new approach for profiled search, which enables better adjustment of the order of results for end-users' expectations is proposed. It is related to the situation when both LOs and users profile descriptions are standardized [69].

J. *Effective Process Planning and Scheduling*

The objective of planning and scheduling is to enhance the traditional educational process by planning future courses, helping with student course scheduling, planning resource allocation, helping in the admission and counseling processes, developing curriculum, etc. Different DM techniques have been used for this task mainly association rules.

Data mining techniques to analyze the course preferences and course completion rates of enrollees in extension education courses at a university. First, extension courses were classified into five broad groups. Records of enrollees in extension courses were then analyzed by three data mining algorithms: Decision Tree, Link Analysis, and Decision Forest. Decision tree was used to find enrollee course preferences, Link Analysis found the correlation between course category and enrollee profession, and Decision Forest found the probability of enrollees completing preferred courses. Results will be used as a reference for curriculum development in the extension program [70]. The main contributions of this study discusses on how the various data mining techniques can be applied to the set of educational data and what new explicit knowledge or models are discovered. The models are classified based on the type of techniques used, including predictive and descriptive. The obtained rules from each model are translated into plain English as a factor to be considered by the managerial system to either support their current decision makings or help them to set new strategies and plan to improve their decision making procedures. The final results have been analyzed and validated with real situations in a university. The factors affecting the anomalies have been discusses in detail. The final result from each model using various techniques [71].

Increasing the quality of personnel by cultivating talents for the future becomes an extremely important issue. With the growth of firms and the increase in their needs, the database is also growing. therefore determine how to recognize and extract the useful information contained in this database in order to apply it in such a way that assists institution in meeting their increasing and changing needs. Data collecting of personnel educational training by cluster analysis, decision tree algorithm and back-propagation neural networks for mining analysis and classification. The key factors essential to the success of educational training. Once identified, this information can then serve as the basis for other firms future planning of educational training strategies with regard to innovation and breakthrough [72].

The role of management education offers great opportunity for many interesting and challenging data mining applications. The meaningful knowledge and potentially useful patterns extracted through data mining can assist in improving the quality of education and performance of students. The conceptual frame work of data mining process in management education. Management institutions will find larger and wider applications for data mining as these institutes carry research and teachings that relates to creation, transformation and utilization of knowledge. The framework helps management institutes to explore the effects of probable changes in recruitments, admissions and courses and ensures efficiency in the quality of students, student assessments, evaluations and allocations. The data mining process in management education in general and academic aspects of admission and counseling process in particular. The patterns that can be generated using data mining techniques are also suggested [73]. In the actual economical context, characterized by the competition pressures, the use of decisional tools become a valuable advantage to make the difference. The concepts of data warehouse, multidimensional model and data mining, are essentials in the design and the deployment of such tools. The design and implementation of a decisional tool, dedicated to a no lucrative goal institution. An university that hopes to improve the quality of his service, by analyzing the pedagogical results, to discover the success and failure factors, and attempt to increase success chances of students. In this perspective is used OLAP (Online Analytical Process) and association rules discovery techniques [74].

A data mining technique and a genetic algorithm are applied to an automatic course scheduling system to produce course timetables that best suit students and teachers' needs. A practical automatic course scheduling system based on students needs, wherein the course scheduling process is divided into two stages. In the first stage, students needs in course selection are considered and an association among courses selected by students is mined using the association mining technique; while in the second stage, the genetic algorithm is used to arrange the course timetable. The experiment results that the automatic course scheduling system proposed in this study not only can efficiently replace the onerous operation of conventional manual course scheduling, but also produce course timetables that truly fulfill users needs and increase students and teachers satisfaction, thereby providing a win-win-win solution for students, teachers and the school [75].

K. Data Visualization

The objective of data visualization is to highlight useful information and support decision making. In the educational environment, for example, it can help educators and course administrators to analyze the students' course activities and usage information to get a general view of a student's learning. Statistics and visualization information are the two main techniques that have been most widely used for this task. While statistics is a mathematical science used concerning the collection, analysis, interpretation or explanation, and presentation of data. It is relatively easy to get descriptive statistics from statistical software. This descriptive analysis can provide such global data characteristics as summaries and reports about learner. Statistical analysis is also very useful to obtain reports assessing, how many minutes the student has worked, how many minutes he has worked today, how many problems he has resolved, and his correct percentage, our prediction of his score, and his performance level.

Information visualization uses graphic techniques to help people to understand and analyze data. Visual representations and interaction techniques take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once. There are several studies oriented toward visualizing different educational data such as: patterns of annual, seasonal, daily, and hourly user behavior on online forums [76]. Presentation of unified data framework that allows the aggregation of high demand data sources into a single useful research resource that is relevant to research in higher education. The unified data framework guides the aggregation of existing and new data sets, and provides the option of connecting and automatically, or semi-automatically, updating data from the original sources. The unified data framework presents to researchers of higher education a robust suite of analytic tools for data mining and visualization of combined and complex data sources [77].

TABLE I.
A REVIEW SUMMARY OF DATA MINING TECHNIQUES IN EDUCATION FIELDS

Authors	Year	Method	Dataset	Result	Advantages	Disadvantages
Fayyad *et al.*	1996	DM & KDD	A simple dataset with two classes	Direction of current and future research	There are a better understanding	Limited in scope
Zaiane *et al*	2001	Learning Activity Evaluation	Students in web-based learning environment	Improve the quality of web-based learning environment	Can help students and educators	The results have not been optimal
Kotsiantis *et al.*	2003	Machine Learning Techniques	Student dataset from the Hellenic Open University	Students can identify potential dropouts	Automatically	Collecting student data is often expensive and time consuming
Reffay *et al.*	2003	Collaborative distance-learning	Data extracted from a 10-week distance-learning experiment	Efficiently assist the tutor in following the collaboration within the group	More appropriate than the long and detailed textual analysis	There is no monitoring system
Romero *et al.*	2004	grammar-based genetic programming, prediction rules	Students Usage Information	Improving methodology of adaptive systems for web-based education	Useful knowledge for teachers	This method may not be applicable to common problem
Tavalera *et al.*	2004	Clustering to discover patterns	Student dataset	The patterns reflecting user behaviors.	Widely used for modeling business problem	Still need more research
Burr *et al.*	2004	Time use analysis	Student dataset from University Online Forums	Clearly demonstrate that the available technology does not influence study habits	There is little work done on how and when students make use of such facilities	Required surcharge
Saini *et al.*	2005	Document clustering	Document dataset	Provides better accuracy	Expectation Maximization	The algorithms presently classify documents on only the leaf nodes
Farzan *et al.*	2006	Adaptive community-based hypermedia system	Student course dataset	Recommendations based on students assessment of course relevance to their career goals	There are explicit student feedback and then evaluates	Not taking into account implicit feedback
Frias-Martinez *et al.*	2006	Adaptive hypermedia	Relevant information about the behavior of a user (or set of users)	The ability to handle large amounts of data and to process uncertainty	Efficiently and accurately capture user behavior	The system has not been integrated
Novak *et al.*	2006	Concept Maps	The concepts represented by the words, and the propositions or ideas	The basic theory and the origins of the concept map	Simple but deeply meaningful	Still need to be revised periodically
Hubscher *et al.*	2007	Clustering	Students dataset and the educational hypermedia system	Finding of some interesting patterns	More general approaches	Complex algorithm and costly
Orzechowski *et al*	2007	Collaborative filtering	Data from e-learning systems	Search results in a way as close to the users	Computational complexity kept at minimum	Very time-consuming
Chu *et al.*	2007	a multiexpert e-training course	Employees of the company	Can improve the quality of the etraining course	Help the experts to organize their domain knowledge	Compares several features of traditional course design system
Heathcote *et al*	2007	Transaction log analysis	Data harvesting from various databases	More proactive support services	Usefulness of the system increases	The design of the system is not good

Kiu *et al*.	2007	OntoDNA, which employs hybrid unsupervised data mining techniques	Statistic on paired ontologies	The OntoDNA is able to resolve semantic, lexical and structural heterogeneity	OntoDNA outperforms most tools in terms of precision	Mapping is still low
Song *et al*.	2007	Conditional Random Fields	Subject content value dictionaries	Support for e-learning system	Can help teachers and users	Not been applied to a real e-learning
Tseng *et al*.	2007	Fuzzy association rules	Student's score dataset	Building a concept map automatically	Can evaluate the results	Experimental testing rather than real learning
Chiu *et al*.	2008	Rough set theory	Frequently asked questions (FAQ) from Univ.	The relevant FAQs for the user query are generated	Improvement of learning performance	Not applying the methods in natural language
Drachsler *et al*.	2008	Collaborative Filtering	Student dataset from university	Support to make a personal decision	Can be used for lifelong learning	Limited attribute
Wang *et al*.	2008	Association mining	Student, course, lecturer, and class	An automatic course scheduling system	Win-win solution	Complex algorithm and costly
Pechenizkiy *et al*.	2008	Clustering, classification and association	The online assessment of students	With a modest size dataset and well-defined problem	It's easier to get results	Less focus on the discovery of patterns
Bresfelean *et al*.	2008	classification learning, clustering	Student's surveys dataset	Framework to profile students who failed / successful	Can make predictions	accuracy of less
Chen *et al*.	2008	Social network analysis	72 learners	Presents a novel scheme of interactive mining social networks	Interaction and learning performance in cooperative learning environments	Not to be applied real
Chen *et al*.	2008	Text-mining techniques	Conference papers in e-Learning domain as data sources	Provide a useful reference for researchers	Learning can be done better	Narrow data range
Delavari *et al*.	2008	Classification, prediction, association rule analysis	Historical and operational data that reside in the educational organization	Decision support system	Ability of data mining is better	Only applied to higher education
Edward *et al*.	2008	Multi-agent system	Human factors with respect to cognitive modelling of behaviour	A multi-agent system to support the control of virtual operators	Cognitive agents are enriched with a planner for selecting actions according to goals	Limitations of system by integration of knowledge
Hsia *et al*.	2008	Decision Tree, Link Analysis, and Decision Forest	Enrollees in extension education courses at a university in Taiwan	Reference for curriculum development in the extension program	as a reference for curriculum development and marketing in the field of higher education	Just extension education
Ranjan *et al*.	2008	Data mining, pattern	Student course dataset	The framework helps management to explore the information needed	Improving the quality of learning	Intelligence system is still lacking
Selmoune *et al*.	2008	Multidimensional model	Student dataset from university	Quality of care improved	Can increase the chances of student success	Lack of data
Wang *et al*.	2008	Association mining	Students, teachers, space, and lessons dataset	Optimal course schedule	Lectures run better	The systems only analyze the results of course selection
Yu *et al*.	2008	Classification trees and multivariate adaptive regressive splines	Students of online courses	That younger students have a stronger tendency to take online classes	can help policy makers prioritize resources	The variables used are limited
Bresfelean *et al*.	2009	DM and DSS	Academic dataset from high education	Presenting relevant information	Time efficiency	High cost
Dekker	2009	The binary classication	Student dataset from institutes	An accuracy of 75% to 80%	The techniques are applied on data that is readily available	No additional data
Feng *et al*.	2009	Decomposition, bootstrapping randomization	Response data of student	That delayed feedback tutoring strategy is more effective	An easier, quicker, and low cost	Computationally intensive techniques are less powerful
Huang *et al*.	2009	The algorithm classification	Personnel educational training	Basic planning agency stretegis	Better planning	It took extra time for planning
Lee *et al*.	2009	Statistical methods	Learners	Learners frequently use less time for navigation	Time efficiency	Learners often repeated visiting
Prata *et al*.	2009	The Bayes classification algorithm	Students from elementary school near Pittsburgh	There are correlations between pre and post-test learning	Learning gains were positively	Have not been able to detect conflicts
Quevedo *et al*.	2009	Pairwise learning model	The dataset from a core course	Obtaining weights for the criteria of rubric	It reduces error	Lack of criteria
Romero *et al*.	2009	Fuzzy rules	Moodle dataset from University Cordoba	Subgroup discovery	Better results	Slightly of rules which are understandable
Simko *et al*	2009	Automatic concept relationship	Student dataset from institute	PageRank-based variant achieves better results	Not limited to it	A course authoring process

Chang *et al*.	2009	k-Nearest neighbor classification	117 elementary school students	The classification mechanism can effectively classify and identify students learning styles	Students can be classified and identified	Data only for elementary school students
Gong *et al*.	2009	Knowledge Tracing (KT) model	Student dataset	high self-discipline students had significantly higher initial knowledge	Expand the flow of information	There can be problems of both over and under reporting
Baepler *et al*.	2010	academic analytics, data mining	Academic dataset from higher education	recommendations for action	It can begin to sift through the noise and provide researchers with a new set of tools	Yet really effective
Dominguez *et al*.	2010	Clustering, association rules and numerical analysis	Student data in 2008	Users who were provided with hints achieved higher average marks	The hinting system had greatly helped users	Need to intelligent systems development
Mozgovoy *et al*	2010	A classification of types of plagiarism	Student course dataset	Detection can be done better	Plagiarism can be reduced	Belum dipertimbangkan aspek hukum dalam sistem
Nugent *et al*.	2010	The Empty K-Means Algorithm with Automatic Specification	Student course dataset	Improves the clustering results	Allows for analysis of data sets that previously thought impossible	Some natural clusters were not present
Romero *et al*.	2010	Association Rules	Student usage data from a Moodle system	Relevant results and good performance	Packed with illustrative examples	Not compared to the previous algorithm
Abdullah *et al*.	2011	Positive association rules	Student dataset from UMT	Association rules with high correlation	Can help decision making	No other datasets
Baradwaj *et al*.	2011	Decision tree	End of the semester students test scores	Describes students performance	It helps in identifying the dropouts students	Complex algorithm
Li *et al*.	2011	A state-of-art machine learning agent	Students who study Algebra	Improvement a tutoring systems	Could be used for such cross-datase	Limited to one class
Abdullah *et al*.	2011	Least association rules	Students examination result dataset	Reduce up to 98% of uninterested association rules	Can be used to reveal the significant rules	Not using a real dataset
Yen *et al*.	2011	Structural equation modeling	Online learners and instructors	That cultural groups perceived online social presence in a slightly different manner	Learners can be grouped according to the social	It may lead to social inequality
Sachin *et al*.	2012	EDM, KDD	Student dataset from institutes	EDM has proven to be more successfull	EDM can be applied to wide areas	Research is still limited
Alper *et al*.	2012	Machine learning methods	Student dataset	Can predict students' future	Improving the prediction results	Measurements have not been detailed
Anderson *et al*.	2012	symbiotic relationship between technology and pedagogy	administrators, instructors, and students	a significant relationship between students' use of technology for academic purposes	self-reported educational gains as well as technological gains	the gap between high and low technology users
Dejaeger *et al*.	2012	Multi class classification	Students two business education institutions	The strategic decision process support	Decision making faster	In the context of the other less effective
Fournier-Viger *et al*.	2012	Association rule mining	three real-life public datasets	CMRules is faster and has a better scalability for low support thresholds	a successful application of the algorithm in a tutoring agent	The algorithm is still slow
Gaudioso *et al*.	2012	Predictive models, Descriptive models	Secondary education students datasets	The predictive modeling in the area of supporting teachers in adaptive educational systems	Teachers can detect situations in which students experience problems	The models are still very simple
Kumar *et al*.	2012	Rule based classification	The students in the final semester	Predicting the performance of students	The accuracy is quite high	The algorithm used is rather complicated
Lee	2012	Log file analysis	407 college students	Significantly enhance the learning efficiency of students	To estimate the ability of students	There is still uncertainty factor
Liu *et al*.	2012	Recommendation systems on Internet-based learning	Student course dataset	The recommendations may influence the behaviors	Focused searchers and broad searchers	Required wider data
Lu *et al*.	2012	Path Model	180 high school students	The best fit model was direct paths from cognitive feedback	More effective	Not tested in college
Blagojevic *et al*.	2012	PDCA (Plan, Do, Check, Act)	Student dataset from University of Kragujevac	The design, implementation, and evaluation of the system	Increasing the efficiency of the courses	There is no comparative analysis
Mishra *et al*.	2012	Classification, ID3 decision tree	Student course dataset	It helps the faculties to improve and bring out betterment in the result of students.	A fully automated	Accuracy of data is rather doubtful
Navarro *et al*.	2012	Multinomial logistic regression	210 students in a Spanish university	Teachers empower students to make better	More informed decisions	Not yet developed to the level of higher
Nebot *et al*.	2012	Fuzzy inductive reasoning (FIR)	Teachers and students	Providing valuable knowledge to teachers about the course performance.	Can provide feedback for further action	Applied in the virtual campus

Ngo *et al.*	2012	Unified Data Framework	The Unified Data Framework guides the aggregation of existing and new data sets	Analytic tools for data mining and visualization	Automatically updating data from the original sources	The algorithm is rather complex
Parack *et al.*	2012	Apriori Algorithm, K-means Clustering	Student grades dataset	The implemented algorithms offer an effective way in educational systems	Can find important information	Number paremeter less
Romero-Zaldivar *et al.*	2012	Virtual appliances	Student course dataset	Significant correlation with student academic achievemen	Being able to provide predictions	Analysis of quantitative data only
Sen *et al.*	2012	Classification, Prediction	A large and feature rich dataset from Secondary Education Transition System in Turkey	Decision tree algorithm is the best predictor with 95% accuracy	High accuracy	Efficiency is less
Torabi *et al.*	2012	Bayesian networks	Student scores dataset	Provides predictions that improve the quality of student learning	The possibility of failure can be minimized	Not shown in the percentage rate of success
Trandafili *et al.*	2012	Clustering	Students of higher education institutions dataset	Improving the quality and performance of students	Reveal interesting and important students profiles	Complex process and it may take longer
Zafra *et al.*	2012	Classification	Student in web based educational environments dataset	The proposed method achieved better accuracy	Useful for early identification of students at risk	Qualitative data have not been taken into account
Zendler *et al.*	2012	The assessment of content and process concepts relevant	Teachers of computer science in schools	Computer science professors attach more importance to content concepts	Computer science will continue to evolve	Computer science teachers will be left
Zorrilla *et al.*	2012	Service-oriented architecture	profile for the multimedia data set	Quality of service improved	The service itself is able to perform all the process	Users are less specific
Trivedi *et al.*	2012	Spectral Coclustering	Student dataset	The proposed method is very useful for institutions	Enhance better performance	The dataset is not suitable as co-occurrence tables and believe
Marquez-Vera *et al.*	2012	Classification, Grammar-based genetic programming	670 high school students from Zacatecas, Mexico	Higher accuracy	Students who may fail immediately in anticipation	Complex algorithm

IV. CONCLUSION

A. Conclusion

The database is owned by a higher education need to be explored more in the data mining to obtain very valuable information and knowledge beneficial to the development of higher education in the future. Patterns data contained in the database is then converted into a model and used to predict the trend of the data with accuracy high. As a result, the agency is expected to more easily manage their resources appropriately and wisely.

EDM has been introduced as an area of future research related to several well-established area of research. Therefore, it can be said that EDM is no longer in the initial study, but has yet to be an area of research. In fact, we have outlined some of the front line were interesting, but to be more mature areas, is also necessary for researchers to develop a more cohesive and collaborative research. Thus, the full integration of the DM in the educational environment will become reality, and fully implementing the operation can be made available not only to researchers and developers, but also for external users.

B. Directions and Opportunities for future research

There is a lot of future work to be considered in EDM, indicate in continuation what arguably are the most interesting and influential. In fact, a few initial studies on some of these points have already begun to appear, namely :

a) EDM Equipment should be designed to be easier for educators or novice user in DM. DMtools usually designed more for strength, flexibility, and simplicity. Most current DM tools are too complex for educators to use and their features go beyond the scope of what educators might want to do. One solutionis possible development using standard algorithms for each task and the free parameters DM algorithms to simplify the configuration and execution for the novice user. EDM tool should also have a more intuitive interface that is easy to use.
b) DM tool should be integrated into the learning environment. All data mining tasks (preprocessing, data mining, and postprocessing) should be done in a single application with a standard interface. DM tool should be integrated into the learning environment. All data mining tasks (preprocessing, data mining, and postprocessing) should be done in a single application with a standard interface.
c) Standardization of data and models. No re-use of common tools or tools that can be applied to any system of education. Therefore, standardization of data input, processing and output of the model is required, as along with preprocessing, find, and postprocessing tasks.
d) Traditional mining algorithms need to be adjusted to take into account the context of education. DM techniques to use semantic information when applied to education data. Special education in mining engineering can greatly improve instructional design, managerial and pedagogical decisions, and the purpose of the Semantic Web is to facilitate data management in an educational environment.

REFERENCES

[1] Trandafili, E., Allkoçi, A., Kajo, E. and Xhuvani, A. *Discovery and evaluation of student's profiles with machine learning*. City, 2012.

[2] Zendler, A. and Klaudt, D. Central Computer Science Concepts to Research-Based Teacher Training in Computer Science: An Experimental Study. *Journal of Educational Computing Research*, 46, 2 2012), 153-172. http://dx.doi.org/10.2190/EC.46.2.c

[3] Anderson, B. and Horn, R. Community Colleges in the Information Age: Gains Associated with Students' Use of Computer Technology. *Journal of Educational Computing Research*, 47, 1 2012), 51-65. http://dx.doi.org/10.2190/EC.47.1.c

[4] Fayyad, U., Shapiro, G. P. and Smyth, P. *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence, City, 1996.

[5] Sachin, B. R. and Vijay, S. M. A Survey and Future Vision of Data mining in Educational Field. In *Proceedings of the Second International Conference on Advanced Computing & Communication Technologies* (2012). IEEE Computer Society, [insert City of Publication],[insert 2012 of Publication].

[6] Bresfelean, V. P., Ghisoiu, N., Lacurezeanu, R. and DA, S.-T. *Towards the Development of Decision Support in Academic Environments*. City, 2009.

[7] Baradwaj, B. K. and Pal, S. Mining Educational Data to Analyze Student Performance. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 2, 6 2011), 63-69.

[8] Zorrilla, M. and García-Saiz, D. A service oriented architecture to provide data mining services for non-expert data miners. *Decision Support Systems*2012), 1-13.

[9] Baepler, P. and Murdoch, C. J. Academic Analitics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, 4, 2 2010), 1-9.

[10] Romero, C., Ventura, S. and Bra, P. D. *Knowledge discovery with genetic programming for providing feedback to courseware author*. Kluwer Academic Publisher, City, 2004.

[11] Blagojevic, M. and Micic, Z. A web-based intelligent report e-learning system using data mining techniques. *Computers and Electrical Engineering*2012), 1-10.

[12] Abdullah, Z., Herawan, T., Ahmad, N. and Deris, M. M. Extracting highly positive association rules from students‟ enrollment data. *Procedia - Social and Behavioral Sciences*, 282011), 107-111.

[13] Abdullah, Z., Herawan, T., Ahmad, N. and Deris, M. M. Mining significant association rules from educational data using critical relative support approach. *Procedia - Social and Behavioral Sciences*, 282011), 97-101.

[14] Fournier-Viger, P., Faghihi, U., Nkambou, R. and Nguifo, E. M. CMRules: Mining sequential rules common to several sequences. *Knowledge-Based Systems*, 252012), 63–76.

[15] Zaiane, O. *Web Usage Mining for a BetterWeb-Based Learning Environment*. Banff, AB, City, 2001.

[16] Romero, C., González, P., Ventura, S., Jesus, M. J. d. and Herrera, F. Evolutionary algorithms for subgroup discovery in e-learning : A practical application using Moodle data. *Expert System Application Journal*, 362009), 1632-1644.

[17] Hubscher, R., Puntambekar, S. and Nye, A. *Domain Speci¯c Interactive Data Mining*. Corfu Greece, City, 2007.

[18] Pechenizkiy, M., Calders, T., Vasilyeva, E. and Bra, P. D. *Mining the Student Assessment Data : Lessons Drawn from a Small Scale Case Study*. City, 2008.

[19] Song, D., Lin, H. and Yang, Z. *Opinion Mining in e-Learning System*. IEEE Computer Society, City, 2007.

[20] Dejaeger, K., Goethals, F., Giangreco, A., Mola, L. and Baesens, B. Gaining insight into student satisfaction using comprehensible data mining techniques. *European Journal of Operational Research*, 2182012), 548–562.

[21] Liu, C. C., Chang, C. J. and Tseng, J. M. The effect of recommendation systems on Internet-based learning for different learners : A data mining analysis. *British Journal of Educational Technology*2012).

[22] Drachsler, H., Hummel, H. G. and Koper, R. Personal recommender systems for learners in lifelong learning networks: The requirements, techniques and model. *Int. Journal. Learning Technology*, 3, 4 2008), 404-423. http://dx.doi.org/10.1504/IJLT.2008.019376

[23] Dominguez, A. K., Yacef, K. and Curran, J. R. *Data Mining for Individualised Hints in eLearning*. City, 2010.

[24] Alper, M. E. and Çataltepe, Z. *Improving course success prediction using ABET course outcomes and grades* City, 2012.

[25] Parack, S., Zahid, Z. and Merchant, F. *Application of data mining in educational databases for predicting academic trends and patterns*. City, 2012.

[26] Kumar, S. A. and Vijayalakshmi, M. N. *Mining of student academic evaluation records in higher education* City, 2012.

[27] Lee, Y. J. Developing an efficient computational method that estimates the ability of students in a Web-based learning environment. *Computers & Education*, 582012), 579-589.

[28] Zafra, A. and Ventura, S. Multi-instance genetic programming for predicting student performance in web based educational environments. *Applied Soft Computing Journal*, 12, 8 2012), 2693-2706. http://dx.doi.org/10.1016/j.asoc.2012.03.054

[29] Márquez-Vera, C., Cano, A., Romero, C. and Ventura, S. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied Intelligence*2012), 1-16.

[30] Sen, B., Uçar, E. and Delen, D. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 392012), 9468-9476.

[31] Gaudioso, E., Montero, M. and Hernandez-del-Olmo, F. Supporting teachers in adaptive educational systems through predictive models : A proof of concept. *Expert Systems with Applications*, 392012), 621–625.

[32] Romero-Zaldivar, V. A., Pardo, A., Burgos, D. and Kloos, C. D. Monitoring student progress using virtual appliances: A case study. *Computers & Education*, 582012), 1058-1067.

[33] Torabi, R., Moradi, P. and Khantaimoor, A. R. *Predict Student Scores Using Bayesian Networks*. City, 2012.

[34] Frias-Martinez, E., Chen, S. and Liu, X. Survey of Data Mining Approaches to User Modeling for Adaptive Hypermedia. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 36, 6 2006), 734-749. http://dx.doi.org/10.1109/TSMCC.2006.879391

[35] Li, N., Matsuda, N., Cohen, W. and K.R. Koedinger *A Machine Learning Approach for Automatic Student Model Discovery*. Forbes Ave, Pittsburgh, PA 15232, City, 2011.

[36] Gong, Y., Rai, D., Beck, J. E. and Heffernan, N. T. *Does self-discipline impact students' knowledge and learning?* , City, 2009.

[37] Feng, M., Beck, J. E. and Heffernan, N. T. *Using Learning Decomposition and Bootstrapping with Randomization to Compare the Impact of Different Educational Interventions on Learning*. City, 2009.

[38] Kotsiantis, S. B., Pierrakeas, C. J. and Pintelas, P. E. *Preventing Student Dropout in Distance Learning Using Machine Learning Techniques*. Springer-Verlag Berlin Heidelberg, City, 2003.

[39] Dekker, G. W. *Predicting students drop out: a case study*. City, 2009.

[40] Bresfelean, V. P., Bresfelean, M. and Ghisoiu, N. *Determining Students' Academic Failure Profile Founded on Data Mining Methods*. City, 2008.

[41] Edward, L., Lourdeaux, D., Barthès, J., Lenne, D. and Burkhardt, J. Modelling Autonomous Virtual Agent Behaviours in a Virtual Environment for Risk. *The International Journal of Virtual Reality*, 7, 3 2008), 13-22.

[42] Heathcote, E. and Prakash, S. *What your learning management system is telling you about supporting your teachers: Monitoring system information to improve support for teachers using educational technologies at Queensland University of Technology*. City, 2007.

[43] Mozgovoy, M., Kakkonen, T. and Cosma, G. Automatic Student Plagiarism Detection: Future Perspectives. *Journal of Educational Computing Research* 43, 4 2010), 511531.

[44] Nebot, A., Mugica, F. and Castro, F. *A framework to provide real time useful knowledge in e-learning environments* City, 2012.

[45] Mishra, K. K., Tripathi, A. and Misra, A. K. A performance evaluation tool for behavioral analysis of students. *Journal of Ap-*

plied Sciences, 12, 15 2012), 1572-1579. http://dx.doi.org/10.3923/jas.2012.1572.1579
[46] Talavera, L. and Gaudioso, E. *Mining Student Data To Characterize Similar Behavior Groups In Unstructured Collaboration Spaces*. City, 2004.
[47] Nugent, R., Dean, N. and Ayers, E. *Skill Set Profile Clustering: The Empty K-Means Algorithm with Automatic Specification of Starting Cluster Centers*. City, 2010.
[48] Lee, M. W., Chen, S. Y., Chrysostomou, K. and Liu, X. Mining student's behavior in web-based learning programs. *Expert Systems with Applications Journal*, 362009), 3459–3464.
[49] Yu, C. H., Digangi, S., Jannasch-pennell, A. K. and Kaprolet, C. Profiling students who take online courses using data mining methods. *Online Journal Distance Learning Administration*, 11, 2 2008), 1-14.
[50] Quevedo, J. R. and Motanes, E. *Obtaining rubric weitghts for assessments by more than one lecturer using a pairwise learning model*. City, 2009.
[51] Lu, J. and Zhang, Z. Understanding the Effectiveness of Online Peer Assessment : A Path Model. *Journal of Educational Computing Research*, 46, 3 2012), 313-333. http://dx.doi.org/10.2190/EC.46.3.f
[52] Chang, Y. C., Kao, W. Y., Chu, C. P. and Chiu, C. H. A learning style classification mechanism for e-learning. *Computer & Education Journal*, 53, 2 2009), 273–285. http://dx.doi.org/10.1016/j.compedu.2009.02.008
[53] Navarro, J. G. C. and Rodríguez, F. J. C. Factors Affecting the Use of an E-Learning Portal at University. *Journal of Educational Computing Research*, 46, 1 2012), 85-103. http://dx.doi.org/10.2190/EC.46.1.d
[54] Romero, C., Romero, J. R., Luna, J. M. and Ventura, S. *Mining Rare Association Rules from e-Learning Data*. City, 2010.
[55] S. Trivedi, Z. A. P., G.N. Sarkozy, N.T. Heffernan *CoClustering by Bipartite Spectral Graph Partitioning for OutofTutor Prediction*. City, 2012.
[56] Chen, C. H., Hong, C. M. and Chang, C. C. *Mining Interactive Social Network for Recommending Appropriate Learning Partners in a Web-based Cooperative Learning Environment*. City, 2008.
[57] Farzan, R. and Brusilovsky, P. *Social Navigation Support in a Course Recommendation System*. City, 2006.
[58] Prata, D., Baker, R., Costa, E., Rose, C. and Cui, Y. *Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments*. City, 2009.
[59] Reffay, C. and Chanier, T. *How social network analysis can help to measure cohesion in collaborative distance-learning*. City, 2003.
[60] Yen, C. J. and Tu, C. H. A Multiple-Group Confirmatory Factor Analysis of the Scores for Online Social Presence: Do They Measure the Same Thing Across Cultural Groups? *Journal of Educational Computing Research*, 44, 2 2011), 219 - 242. http://dx.doi.org/10.2190/EC.44.2.e
[61] Novak, J. D. and A. J. Canas The theory underlying concept maps and how to construct and use them. Florida Institute for Human and Machine Cognition (IHMC), City, 2006.
[62] Tseng, S. S., Sue, P. C., Su, J. M. and Weng, J. F. A new approach for constructing the concept map. *Computer Education Journal*, 492007), 961-707.
[63] Chen, N. S., Kinshuk, Wei, C. W. and Chen, H. J. Mining e-Learning domain concept map from academic articles. *Computer Education Journal*, 502008), 1009–1021.
[64] Simko, M. and Bielikova, M. *Automatic Concept Relationships Discovery for an Adaptive E-course*. City, 2009.
[65] Chu, H. C., Hwang, G. J., H.Wu, P. and Chen, J. M. *A computer-assisted collaborative approach for E-training course design*. City, 2007.
[66] Chiu, D. Y., Pan, Y. C. and Chang, W. C. *Using Rough Set Theory to Construct E-Learning FAQ Retrieval Infrastructure*. City, 2008.
[67] Kiu, C. C. and Lee, C. S. *Learning objects reusability and retrieval through ontological sharing: A hybrid unsupervised data mining approach*. City, 2007.
[68] Saini, P. S., Sona, D., Veeramachaneni, S. and Ronchetti, M. *Making E-Learning Better Through Machine Learning*. City, 2005.
[69] Orzechowski, T., Ernst, S. and Dziech, A. *Profiled Search Methods for e-Learning Systems*. City, 2007.
[70] Hsia, T., Shie, A. and Chen, L. Course planning of extension education to meet market demand by using data mining techniques -An example of Chinkuo technology university in Taiwan. *Expert System Application Journal*, 34, 1 2008), 596-602. http://dx.doi.org/10.1016/j.eswa.2006.09.025
[71] Delavari, N., Phon-amnuaisuk, S. and Beikzadeh, M. Data mining application in higher learning institutions. *Informatics in Education Journal*, 7, 1 2008), 31-54.
[72] Huang, C., Lin, W., Wang, W. and Wang, W. Planning of educational training courses by data mining: Using China Motor Corporation as an example. *Expert System Application Journal*, 36, 3 2009), 7199-7209. http://dx.doi.org/10.1016/j.eswa.2008.09.009
[73] Ranjan, J. and Khalil, S. Conceptual framework of data mining process in management education in India: An institutional perspective. *Information Technology Journal*, 7, 1 2008), 16-23. http://dx.doi.org/10.3923/itj.2008.16.23
[74] Selmoune, N. and Alimazighi, Z. *A decisional tool for quality improvement in higher education*. City, 2008.
[75] Wang, Y., Cheng, Y., Chang, T. and Jen, S. M. *On the application of data mining technique and genetic algorithm to an automatic course scheduling system*. City, 2008.
[76] Burr, L. and Spennemann, D. H. Pattern of user behaviour in University online forum *International Journal Instruct. Techno. Distance Learn*, 1, 10 2004), 11-28.
[77] Ngo, L. B., Dantuluri, V. B., Stealey, M. B., Ahalt, S. B. and Apon, A. *An architecture for mining and visualization of U.S. higher educational data*. City, 2012.

AUTHORS

Suhirman is with University Technology of Yogyakarta, Yogyakarta, Indonesia and Universiti Malaysia Pahang, Gambang, Kuantan Pahang, Malaysia.

Jasni Mohamad Zain is with University Technology of Yogyakarta, Yogyakarta, Indonesia.

Haruna Chiroma is with the Department of Artificial Intelligence, University of Malaya, Kuala Lumpur, Malaysia.

Tutut Herawan is with the Department of Information System, University of Malaya, Kuala Lumpur, Malaysia.

Submitted 08 June 2014. Published as resubmitted by the authors 08 December 2014.