

Data Mining in Education

Abdulmohsen Algarni
College of Computer Science
King Khalid University Abha
61421, Saudi Arabia

Abstract—Data mining techniques are used to extract useful knowledge from raw data. The extracted knowledge is valuable and significantly affects the decision maker. Educational data mining (EDM) is a method for extracting useful information that could potentially affect an organization. The increase of technology use in educational systems has led to the storage of large amounts of student data, which makes it important to use EDM to improve teaching and learning processes. EDM is useful in many different areas including identifying at-risk students, identifying priority learning needs for different groups of students, increasing graduation rates, effectively assessing institutional performance, maximizing campus resources, and optimizing subject curriculum renewal. This paper surveys the relevant studies in the EDM field and includes the data and methodologies used in those studies.

Index Terms—Data mining, Educational Data Mining (EDM), Knowledge extraction.

I. INTRODUCTION

One of the primary goals of any educational system is to equip students with the knowledge and skills needed to transition into successful careers within a specified period. How effectively global educational systems meet this goal is a major determinant of both economic and social progress.

Some countries provide free education for all citizens from grade one through the university years. Therefore, a large number of students enter universities every year. For example, King Khalid University (KKU) accepted approximately 23,000 students in 2013. It has become difficult to provide high quality teaching and guidance to such a large number of students. As a result, many students fail to complete their degrees within the required periods. EDM can present universities with a clear picture of specific hindrances to student learning. For example, students can fail in advanced subjects because they did not learn the basic information from the prerequisite subjects. Using data mining (DM) techniques to analyze student information can help identify possible reasons for student failures.

Data mining provides many techniques for data analysis. The large amount of data currently in student databases exceeds the human ability to analyze and extract the most useful information without help from automated analysis techniques. Knowledge discovery (KD) is the process of nontrivial extraction of implicit, unknown, and potentially useful information from a large database. Data mining has been used in KD to discover patterns with respect to a users needs. The pattern definition is an expression in language that describes a subset of data. An example of a KD pattern definition appears in [1].

The increasing use of technology in educational systems has made a large amount of data available. EDM provides a significant amount of relevant information [2] and offers a clearer picture of learners and their learning processes. It uses DM techniques to analyze educational data and solve educational issues. Similar to other DM techniques extraction processes, EDM extracts interesting, interpretable, useful, and novel information from educational data. However, EDM is specifically aimed at developing methods that use unique types of data in educational systems [3]. Such methods are then used to enhance knowledge about educational phenomena, students, and the settings in which they learn [4]. Developing computational approaches that combine data and theory will help improve the quality of T& L processes.

From a practical point of view, EDM allows users to extract knowledge from student data. This knowledge can be used in different ways such as to validate and evaluate an educational system, improve the quality of T& L processes, and lay the groundwork for a more effective learning process [5]. Similar ideas have been applied successfully, especially in business data, in different datasets, such as e-commerce systems, to increase sales profits [6]. Thus, the success of applying DM techniques in business data encourages its adoption in different domains of knowledge. Notably, DM has been applied to educational data for research objectives such as improving the learning process and guiding students learning or acquiring a deeper understanding of educational phenomena. However, while EDM has made comparatively less progress in this direction than other fields, this situation is changing due to increased interest in the use of DM in the educational environment [7].

Many tasks or problems in educational environments have been managed or resolved through EDM. Baker [8], [4] suggested four key areas of EDM application: improving student models, improving domain models, studying the pedagogical support provided by learning software, and conducting scientific research on learning and learners. Five approaches/methods are available: prediction, clustering, relationship mining, distillation of data for human judgment, and discovery with models. Castro [9] categorized EDM tasks into four different areas: applications that deal with the assessment of students learning performance, course adaptation and learning recommendations to customize students learning based on individual students behaviors, developing a method to evaluate materials in online courses, approaches that use

feedback from students and teachers in e-learning courses, and detection models for uncovering student learning behaviors.

II. DATA MINING

DM is a powerful artificial intelligence (AI) tool, which can discover useful information by analyzing data from many angles or dimensions, categorize that information, and summarize the relationships identified in the database. Subsequently, this information helps make or improve decisions. In DM solutions, algorithms can be used either independently or together to achieve the desired results. Some algorithms can explore data; others extract a specific outcome based on that data. For example, clustering algorithms, which recognize patterns, can group data into different n-groups. The data in each group are more or less consistent, and the results can help create a better decision model. Multiple algorithms, when applied to one solution, can perform separate tasks. For example, by using a regression tree method, they can obtain financial forecasts or association rules to perform a market analysis.

A large amount of data in databases today exceeds the human ability to analyze and extract the most useful information without help from automated analysis techniques. Knowledge discovery is the process of nontrivial extraction of implicit, unknown, and potentially useful information from a large database. Data mining used in KD has discovered patterns with respect to a users needs. The pattern definition is an expression in the language that describes a subset of data; an example is shown in [1].

The accurate discovery of patterns through DM is influenced by several factors, such as sample size, data integrity, and support from domain knowledge, all of which affect the degree of certainty needed to identify patterns. Typically, DM uncovers a number of patterns in a database; however, only some of them are interesting. Useful knowledge constitutes the patterns of interest to the user. It is important for users to consider the degree of confidence in a given pattern when evaluating its validity.

The KD process is interactive and examines many decisions made by the user. Loops can occur between any two steps in the process, which are needed for further iteration.

First, it is important to develop an understanding of the application domain, including relevant prior knowledge, and identify the end users goal. Second, choose a target dataset and focus on the subset of variables or data samples targeted for examination. Third, clean and preprocess the data by reducing noise, designing strategies for dealing with missing data, and accounting for time-sequence information and known changes. Fourth (the data reduction and projection phase), find useful features to represent the data such as dimensionality reduction or transformation methods. Fifth, use the goals of the KD to choose the appropriate DM strategy. Sixth, match the dataset with DM algorithms to search for patterns. Seventh, extract interesting patterns from a particular representational form or set. Eighth, interpret these mined patterns and/or return to any previous steps for an additional iteration. Finally, use the

discovered knowledge by taking action and documenting or reporting the knowledge [10].

III. EDUCATIONAL DATA MINING

Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings and using those methods to better understand students and the settings which they learn in [3]. Different from data mining methods, EDM, when used explicitly, accounts for (and avail of opportunities to exploit) the multilevel hierarchy and lacks independent educational data [3].

IV. EDM METHODS

Educational data mining methods come from different literature sources including data mining, machine learning, psychometrics, and other areas of computational modelling, statistics, and information visualization. Work in EDM can be divided into two main categories: 1) web mining and 2) statistics and visualization [11]. The category of statistics and visualization has received a prominent place in theoretical discussions and research in EDM [8], [7], [12]. Another point of view, proposed by Baker [3], classifies the work in EDM as follows:

- 1) Prediction.
 - Classification.
 - Regression.
 - Density estimation.
- 2) Clustering.
- 3) Relationship mining.
 - Association rule mining.
 - Correlation mining.
 - Sequential pattern mining.
 - Causal DM.
- 4) Distillation of data for human judgment.
- 5) Discovery with models.

Most of the above mentioned items are considered DM categories. However, the distillation of data for human judgment is not universally regarded as DM. Historically, relationship mining approaches of various types have been the most noticeable category in EDM research.

Discovery with models is perhaps the most unusual category in Bakers EDM taxonomy, from a classical DM perspective. It has been used widely to model a phenomenon through any process that can be validated in some way. That model is then used as a component in another model such as relationship mining or prediction. This category (discovery with models) has become one of the lesser-known methods in the research area of educational data mining. It seeks to determine which learning material subcategories provide students with the most benefits [13], how specific students behavior affects students learning in different ways [14], and how tutorial design affects students learning [15]. Historically, relationship mining methods have been the most used in educational data mining research in the last few years.

Other EDM methodologies, which have not been used widely, include the following:

- Outlier detections discover data points that significantly differ from the rest of the data [16]. In EDM, they can detect students with learning problems and irregular learning processes by using the learners response time data for e-learning data [17]. Moreover, they can also detect atypical behavior via clusters of students in a virtual campus. Outlier detection can also detect irregularities and deviations in the learners or educators actions with others [18].
- Text mining can work with semi-structured or unstructured datasets such as text documents, HTML files, emails, etc. It has been used in the area of EDM to analyze data in the discussion board with evaluation between peers in an ILMS [19], [20]. It has also been proposed for use in text mining to construct textbooks automatically via web content mining [21]. Use of text mining for the clustering of documents based on similarity and topic has been proposed [22], [23].
- Social Network Analysis (SNA) is a field of study that attempts to understand and measure relationships between entities in networked information. Data mining approaches can be used with network information to study online interactions [24]. In EDM, the approaches can be used for mining group activities [25].

A. Prediction

Prediction aims to predict unknown variables based on history data for the same variable. However, the input variables (predictor variables) can be classified or continue as variables. The effectiveness of the prediction model depends on the type of input variables. The prediction model is required to have limited labelled data for the output variable. The labelled data offers some prior knowledge regarding the variables that we need to predict. However, it is important to consider the effects of quality of the training data in order to achieve the prediction model.

There are three general types of predictions:

- Classification uses prior knowledge to build a learning model and then uses that model as a binary or categorical variable for the new data. Many models have been developed and used as classifiers such as logistic regression and support vector machines (SVM).
- Regression is a model used to predict variables. Different from classification, regression models predict continuous variables. Different methods of regression, such as linear regression and neural networks, have been used widely in the area of EDM to predict which students should be classified as at-risk.
- Density estimation is based on a variety of kernel functions including Gaussian functions.

Prediction methodology in EDM is used in different ways. Most commonly, it studies features used for prediction and uses those features in the underlying construct, which predicts student educational outcomes [26]. While different

approaches try to predict the expected output value based on hidden variables in the data, the obtained output is not clearly defined in the labels data.

For example, if a researcher aims to identify the students most likely to drop out of school, with the large number of schools and students involved, it is difficult to achieve using traditional research methods such as questionnaires. The EDM method, with its limited amount of sample data, can help achieve that aim. It must start by defining at-risk students and follow with defining the variables that affect the students such as their parents educational backgrounds. The relation between variables and dropping out of school can be used to build a prediction model, which can then predict at-risk students. Making these predictions early can help organizations avoid problems or reduce the effects of specific issues.

Different methods have been developed to evaluate the quality of a predictor including accuracy of linear correlation, Cohens Kappa, and A [27]. However, accuracy is not recommended for evaluating the classification method because it is dependent on the base rates of different classes. In some cases, it is easy to get high accuracy by classifying all data based on the large group of classes sample data. It is also important to calculate the number of missed classifications from the data to measure the sensitivity of the classifier using recall [28]. A combined method, such as an F-measure, considers both true and false classification results, which are based on precision and recall, to give an overall evaluation of the classifier.

B. Clustering

Clustering is a method used to separate data into different groups based on certain common features. Different from the classification method, in clustering, the data labels are unknown. The clustering method gives the user a broad view of what is happening in that dataset. Clustering is sometimes known as an unsupervised classification because class labels are unknown [10].

In clustering, we have started to find data points that naturally group together to split the dataset into different groups. The number of groups can be predefined in the clustering method. Generally, the clustering method is used when the most common group in the dataset is unknown. It is also used to reduce the size of the study area. For example, different schools can be grouped together based on similarities and differences between them [29], [30].

C. Relationship mining

Relationship mining aims to find relationships between different variables in data sets with a large number of variables. This entails finding out which variables are most strongly associated with a specific variable of particular interest. Relationship mining also measures the strength of the relationships between different variables. Relationships found through relationship mining must satisfy two criteria: statistical significance and interestingness. Large amounts of data contain many variables and hence have many associated rules. Therefore, the measure of interestingness determines the

most important rules supported by data for specific interests. Different interestingness measures have been developed over the years by researchers including support and confidence. However, some research has concluded that lift and cosine are the most relevant used in educational data mining[31].

Many types of relationship mining can be used such as association rule mining, sequential pattern mining, and frequent pattern mining. Association rule mining is the most common EDM method. The relationship found in association rule mining is $\text{if} \rightarrow \text{then}$ rules. For example, if {Student GPA is less than two, and the student has a job} \rightarrow {, the student is going to drop out of school}. The main goal of relationship mining is to determine whether or not one event causes another event by studying the coverage of the two events in the data set, such as TETRAD [32], or by studying how an event is triggered.

D. Discovery with Models

In discovery, models are generally based on clustering, prediction, or knowledge engineering using human reasoning rather than automated methods. The developed model is then used as part of other comprehensive models such as relationship mining.

E. Distillation of data for human judgement

Distillation of data for human judgment aims to make data understandable. Presenting the data in different ways helps the human brain discover new knowledge. Different kinds of data require specific methods to visualize it. However, the visualization methods used in educational data mining are different from those used in different data sets [33], [34] in that they consider the structure of the education data and the hidden meaning within it.

Distillation of data for human judgment is applied in educational data for two purposes: classification and/or identification. Data distillation for classification can be a preparation process for building a prediction model [35]; identification aims to display data such that it is easily identifiable via well known patterns that cannot be formalized [36].

As mentioned previously, there is a wide variety of methods used in educational data mining. These methods have been divided by Rayn [37] into five categories: clustering, prediction, relationship mining, discovery with models, and distillation of data for human judgement are illustrated in Table I.

V. EDUCATIONAL DATA MINING DATA AND APPLICATIONS

The main goal of EDM is to extract useful knowledge from educational data including student records, student usage data, inelegant tutre, and LMS systems. The extracted knowledge can improve the process of teaching and learning in the educational system[38]. It can also lead to the development of new teaching processes. Similar ideas have been applied successfully in different domains of knowledge. For example, e-commerce systems and basket analysis are popular applications in data mining [39]. They increase sales by analyzing

users shopping behaviors. While it is clear that data mining methods in education have not progressed as far as they have in business [40], in the last few years, EDM has drawn more attention from researchers. Applying DM to educational data is different than it is in other domains, as defined below:

- 1) Objective: Applying DM methods to any specific data is led by the objectives. The main objective for using EDM is to improve teaching and learning processes. Research objectives, such as gaining a deeper understanding of the teaching and learning phenomena, occasionally influence the objectives. Applying traditional research methods to achieve goals is sometimes difficult.
- 2) Data: Using technology in education has led to increased data in educational systems, which differs from basic information, such as student information, because it includes more information, which is generated by different systems such as the LMS system. Applying EDM methods to educational data can make extracting specific knowledge either quite simple or more complicated such as in applying relational mining. One example would be applying relational mining to find the relation between students success in courses that contain several chapters organized into lessons, with each lesson including several concepts.
- 3) Techniques: The application of DM to any problem is driven by the objectives of the research and the type of data at hand. Therefore, applying data mining successfully to educational data requires specific adoption. The adoption can be for either the DM methods or pre-processing of the data. Some DM methods can be applied directly, without any modifications, and some cannot. Moreover, some DM techniques are used for specific problems in the educational domain. However, choosing certain techniques depends on the researchers perspective of the problem and the objectives of the research [41]. For example, EDM methods can improve the teaching and learning processes in the classroom, identify at-risk students, customize teaching processes, and provide recommendations to teachers and students. Most current research involves only teachers and students. However, more groups can be involved in research that has other objectives such as course development [42].

A. Data used in EDM

EDM offers a clear picture and a better understanding of learners and their learning processes. It uses DM techniques to analyze educational data and solve educational issues. Similar to other DM techniques extraction processes, EDM extracts interesting, interpretable, useful, and novel information from educational data. However, EDM is specifically concerned with developing methods to explore the unique types of data in educational settings [3]. Such methods are used to enhance knowledge about educational phenomena, students, and the settings in which they learn [4]. Developing computational

TABLE I: Educational data mining methodology categories.

Category	objectives	Key applications
prediction	Develop a model to predict some variables base on other variables. The predictor variables can be constant or extract from the data set.	Identify at-risk students. Understand student educational outcomes
Clustering	Group specific amount of data to different clusters based on the characteristics of the data. The number of clusters can be different based on the model and the objectives of the clustering process.	Find similarities and differences between students or schools. Categorized new student behavior
Relationship Mining	Extract the relationship between two or more variables in the data set.	Find the relationship between parent education level and students drooping out from school. Discovery of curricular associations in course sequences; Discovering which pedagogical strategies lead to more effective/robust learning
Discovery with Models	It aims to develop a model of a phenomenon using clustering, prediction, or knowledge engineering, as a component in more comprehensive model of prediction or relationship mining.	Discovery of relationships between student behaviours, and student characteristics or contextual variables; Analysis of research question across wide variety of contexts
Distillation of Data for Human Judgement	The main aim of this model to find a new way to enable researchers to identify or classify features in the data easily.	Human identification of patterns in student learning, behaviour, or collaboration; Labelling data for use in later development of prediction model

approaches that combine data and theory will help improve the quality of T& L processes.

The increasing use of technology in educational systems has made a large amount of data available. Educational data mining (EDM) provides a significant amount of relevant information [2]. Therefore, the main source of data used in EDM to date can be categorized as follows:

- Offline education, also known as traditional education, is where knowledge transfers to learners based on face-to-face contact. Data can be collected by traditional methods such as observation and questionnaires. It studies the cognitive skills of students and determines how they learn. Therefore, the statistical technique and psychometrics can be applied to the data.
- E-learning and learning management systems (LMS) provide students with materials, instruction, communication, and reporting tools that allow them to learn by themselves. Data mining techniques can be applied to the data stored by the systems in the databases.
- Intelligent tutoring systems (ITS) and adaptive educational hypermedia systems (AEHS) try to customize the data provided to students based on student profiles. As a result, applying data mining techniques is important for building user profiles. The data generated by that system can then assist in further research.

Based on the three categories established by Romero etl [26], we can group EDM research according to the type of data used: traditional education, web-based education (e-learning), learning management systems, intelligent tutoring systems, adaptive educational systems, tests questionnaires, texts contents, and others.

B. EDM Application

Many studies have been developed in the area of EDM. A framework for examining learners behaviors in online education videos was recommended by Alexandro & Georgios [43].

The proposed framework consisted of capturing learner performance data, designing a data model for storing the activity data, and creating modules to monitor and visualize learner viewing behavior using captured data. Researchers relied on most of the students to watch videos in the few days prior to exams or an assignment due date. Moreover, pausing and resuming was mainly observed in videos associated with an assignment. One lamentation was that the author did not study what affected learner viewing behavior or why some learners refrained from viewing online videos altogether.

In other research, Saurabh Pal [44] built a model using data mining methodologies to predict which students would likely drop out during their first year in a university program. That study used the Nave Bayes classification algorithm to build the prediction model based on the current data. The result of the system was promising for identifying students who needed special attention to reducing the dropout rate. Leila Dadkhahan [45] tried to justify what was needed for student retention in higher education institutions to reduce the number of dropouts. As a result, using data mining techniques led to increased student retention and graduation rates.

VI. CONCLUSIONS

The increased use of technology in education is generating a large amount of data every day, which has become a target for many researchers around the world; the field of educational data mining is growing quickly and has the advantage of containing new algorithms and techniques developed in different data mining areas and machine learning. The data mining of educational data (EDM) is helping create development methods for the extraction of interesting, interpretable, useful, and novel information, which can lead to better understanding of students and the settings in which they learn.

EDM can be used in many different areas including identifying at-risk students, identifying priorities for the learning needs of different groups of students, increasing graduation rates, effectively assessing institutional performance, maximizing

campus resources, and optimizing subject curriculum renewal. This paper surveyed the most relevant studies carried out in the field of EDM including data used in certain studies and the methodologies employed. It also defined the most common tasks used in EDM as well as those that are the most promising for the future.

REFERENCES

- [1] S.-T. Wu, "Knowledge discovery using pattern taxonomy model in text mining," 2007.
- [2] J. Mostow and J. Beck, "Some useful tactics to modify, map and mine data from intelligent tutors," *Natural Language Engineering*, vol. 12, no. 02, pp. 195–208, 2006.
- [3] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013.
- [4] R. Baker *et al.*, "Data mining for education," *International encyclopedia of education*, vol. 7, pp. 112–118, 2010.
- [5] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors," *User Modeling and User-Adapted Interaction*, vol. 14, no. 5, pp. 425–464, 2004.
- [6] N. S. Raghavan, "Data mining in e-commerce: A survey," *Sadhana*, vol. 30, no. 2-3, pp. 275–289, 2005.
- [7] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*. CRC Press, 2010.
- [8] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- [9] F. Castro, A. Vellido, A. Nebot, and F. Mugica, "Applying data mining techniques to e-learning problems," in *Evolution of teaching and learning paradigms in intelligent environment*, pp. 183–221, Springer, 2007.
- [10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The kdd process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.
- [11] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, in *Proc. Educational Data Mining 2009: 2nd International Conf*, 2009.
- [12] S. L. Tanimoto, "Improving the prospects for educational data mining," in *Track on Educational Data Mining, at the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling*, pp. 1–6, 2007.
- [13] J. E. Beck and J. Mostow, "How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students," in *Intelligent tutoring systems*, pp. 353–362, Springer, 2008.
- [14] M. Cocea, A. Hershkovitz, and R. S. Baker, "The impact of off-task and gaming behaviors on learning: immediate or aggregate?," 2009.
- [15] H. Jeong and G. Biswas, "Mining student behavior models in learning-by-teaching environments," in *EDM*, pp. 127–136, Citeseer, 2008.
- [16] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.
- [17] C. C. Chan, "A framework for assessing usage of web-based e-learning systems," in *Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on*, pp. 147–147, Sept 2007.
- [18] M. Muehlenbrock, "Automatic action analysis in an interactive learning environment," in *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, pp. 73–80, 2005.
- [19] M. Ueno, "Data mining and text mining technologies for collaborative learning in an ilms" ssamura," in *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on*, pp. 1052–1053, IEEE, 2004.
- [20] L. P. Dringus and T. Ellis, "Using data mining as a strategy for assessing asynchronous discussion forums," *Computers & Education*, vol. 45, no. 1, pp. 141–160, 2005.
- [21] J. Chen, Q. Li, L. Wang, and W. Jia, "Automatically generating an e-textbook on the web," in *Advances in Web-Based Learning-ICWL 2004*, pp. 35–42, Springer, 2004.
- [22] J. Tane, C. Schmitz, and G. Stumme, "Semantic resource management for the web: an e-learning application," in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pp. 1–10, ACM, 2004.
- [23] C. Tang, R. W. Lau, Q. Li, H. Yin, T. Li, and D. Kilis, "Personalized courseware construction based on web data mining," in *Web Information Systems Engineering, 2000. Proceedings of the First International Conference on*, vol. 2, pp. 204–211, IEEE, 2000.
- [24] J. Scott, *Social network analysis*. Sage, 2012.
- [25] P. Reyes and P. Tchounikine, "Mining learning groups' activities in forum-type tools," in *Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pp. 509–513, International Society of the Learning Sciences, 2005.
- [26] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students," in *EDM*, pp. 8–17, 2008.
- [27] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [28] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.
- [29] C. R. Beal, L. Qu, and H. Lee, "Classifying learner engagement through integration of multiple data sources," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 21, p. 151, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.
- [30] S. Amershi and C. Conati, "Automatic recognition of learner groups in exploratory learning environments," in *Intelligent Tutoring Systems*, pp. 463–472, Springer, 2006.
- [31] A. Merceron and K. Yacef, "Interestingness measures for associations rules in educational data," *EDM*, vol. 8, pp. 57–66, 2008.
- [32] C. Wallace, K. B. Korb, and H. Dai, "Causal discovery via mml," in *ICML*, vol. 96, pp. 516–524, Citeseer, 1996.
- [33] A. Hershkovitz and R. Nachmias, "Developing a log-based motivation measuring tool," in *EDM*, pp. 226–233, Citeseer, 2008.
- [34] J. Kay, N. Maisonneuve, K. Yacef, and P. Reimann, "The big five and visualisations of team work activity," in *Intelligent tutoring systems*, pp. 197–206, Springer, 2006.
- [35] R. S. d Baker, A. T. Corbett, and V. Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing," in *Intelligent Tutoring Systems*, pp. 406–415, Springer, 2008.
- [36] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction*, vol. 4, no. 4, pp. 253–278, 1994.
- [37] R. Baker *et al.*, "Data mining for education," *International encyclopedia of education*, vol. 7, pp. 112–118, 2010.
- [38] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors," *User Modeling and User-Adapted Interaction*, vol. 14, no. 5, pp. 425–464, 2004.
- [39] N. S. Raghavan, "Data mining in e-commerce: A survey," *Sadhana*, vol. 30, no. 2-3, pp. 275–289, 2005.
- [40] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*. CRC Press, 2010.
- [41] M. Hanna, "Data mining in the e-learning domain," *Campus-wide information systems*, vol. 21, no. 1, pp. 29–34, 2004.
- [42] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no. 6, pp. 601–618, 2010.
- [43] K. Alexandros and E. Georgios, "A framework for recording, monitoring and analyzing learner behavior while watching and interacting with online educational videos," in *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*, pp. 20–22, IEEE, 2013.
- [44] S. Pal, "Mining educational data using classification to decrease dropout rate of students," *arXiv preprint arXiv:1206.3078*, 2012.
- [45] L. Dadkhahan and M. A. Al Azmeh, "Critical appraisal of data mining as an approach to improve student retention rate," *International Journal of Engineering and Innovative Technology (IJEIT) Volume*, vol. 2.