Analysis of Data in School Level Education System by using Datamining

[1]Prof Arif Hakeem, [2]Dr. Tryambak Hiwarkar
[1]Research Scholar in SSSUTMS, Sehore
[2]Professor, Department of Computer Science and Engineering, SSSUTMS, Sehore

ABSTRACT

Data mining techniques are used to extract useful knowledge from raw data. The knowledge gathered is valuable and greatly influences the decision maker. School Educational Data Mining (SEDM) is a method of gathering useful information that affects an organization. The increasing use of technology in education systems has led to the storage of large amounts of student data, which is important for the use of SEDM to improve teaching and learning processes. SEDM is useful in many areas such as identifying vulnerable students, identifying priority learning needs for different groups of students, increasing 12th grade rates, effectively assessing organizational performance, maximizing campus resources, and optimizing subject course renewal. This paper surveys the relevant studies in the SEDM field and contains the data and methodology used in those studies.

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

I. INTRODUCTION

One of the primary goals of any education system is to equip students with the knowledge and skills needed to move into a successful career in a given period of time. How effectively the global education system achieves this goal is a major determinant of economic and social progress.

Some countries offer free education to all citizens from grade one to graduation. Therefore, a large number of students enter schools every year. It is difficult to give high quality teaching and guidance to such a large number of students. As a result, many students fail to complete their middle school in the required period. SEDM can be presented with a clear picture of specific barriers to student learning. For example, students may fail in advanced subjects because they did not take basic information from pre-existing subjects. Using data mining (DM) methods to analyze student information can help identify the causes of student failures. Data mining offers several methods for data analysis. The large amount of data currently in student databases exceeds the human capacity to analyze and gather highly useful information without the aid of automated analysis methods. Knowledge Discovery (KD) is the nontrivial extraction process of vague, unknown and useful information from a large database. Data mining in KD was used to find samples related to customer needs. Sample definition is the expression in language that describes a subset of data. An example of the KD model definition appears in [1].

Growing technology in education systems has made large amounts of data available. SEDM provides significant relevant information [2] and provides a clear picture of practitioners and their learning processes. It uses DM methods to analyze educational data and solve educational problems. Like the extraction methods of other DM methods, SEDM captures interesting, detailed, useful, and novel information from educational data. However, SEDM specifically aims to develop methods that use specialized data in education systems [3]. Such techniques are used to increase knowledge about educational programs, students and the settings they learn in [4]. Developing computational methods that combine data and theory can help improve the quality of T&L processes.

From a practical point of view, SEDM allows users to extract knowledge from student data. This knowledge can be used in a variety of ways, such as validating and evaluating the education system, improving the quality of T&L processes, and laying the groundwork for a more effective learning process [5]. Similar ideas have been successfully applied to increase sales profits, especially in business datasets and various datasets such as e-commerce systems. Therefore, the success of applying DM methods in business data encourages its adoption in various fields of knowledge. In particular, DM is applied to educational data for research purposes such as improving learning processes and guiding learning students or gaining an in-depth understanding of academic topics. However, although SEDM has made little progress in this direction compared to other sectors, this situation is changing, with increasing interest in the use of DM in the academic environment [7].

Many tasks or problems in the educational environment are handled or solved by SEDM. Baker [8], [4] suggested four main areas of SEDM application: improving student design, improving domain design, studying and collaborating on the educational support provided by software learning, and minimizing scientific research by learners. Five methods / techniques are available: prediction, clustering, relationship mining, data distillation for human judgment, and search with models. Castro [9] categorizes SEDM tasks into four distinct categories: applications with learning recommendations to optimize student learning based on student performance evaluation, course adaptation, and individual student behavior. Work, develop a method for assessing content in online courses, use feedback from students and teachers in e-learning courses, and identify patterns to highlight student learning behavior.

II. DATA MINING

DM is a powerful Artificial Intelligence (AI) tool that can search for useful information by analyzing data from multiple angles or dimensions, classifying that information, and capturing the relationships found in the database. Subsequently, this information can help in decision making or improvement. In DM solutions, the algorithm can be used independently or together to achieve the desired result. Some algorithms can detect data; Others get a specific result based on that data. For example, clustering algorithms that identify patterns can convert data into separate n-groups. The data in each group is more or less consistent and the results help to create better decision models. Multiple algorithms, when applied to a single solution, can do different things. For example, by using the regression tree method, they can obtain financial references or association rules for conducting market analyzes.

Today the vast amount of data in databases goes beyond the ability of HA-MAN to analyze and collect highly useful information without the aid of automated analysis methods. Knowledge Discovery is the process of extracting vague, unknown and useful information from a large database. Data mining used in KD explored patterns related to customer needs. Sample definition is the expression in language that describes a subset of data; An example is shown in [1].

Accurate innovation of models by DM is influenced by a number of factors such as sample size, data integrity and support from domain knowledge, all of which affect the level of certainty required to identify samples. Typically, a DM opens several models in one database; However, only a few of them are interesting. Useful knowledge makes the user interested. It is important for consumers to consider the level of trust in a given model when evaluating its authenticity.

The KD process is interactive and examines many decisions made by the user. There may be loops between any two stages in this process, which require further repetition.

First, it is important to raise awareness about the application domain, which includes relevant prior knowledge and identifying the end user target. Second, select the target dataset and focus on a subset of the target variable or data models for testing. Third is the process of minimizing noise, creating strategies to deal with missing data, and cleaning and pre-processing data by calculating time-order information and known changes. Fourth (data reduction and projection phase), find useful features to indicate data such as dimensionality reduction or transition methods. Fifth, use KD's objectives to select the appropriate DM strategy. Sixth, compare the dataset with the DM algorithm to search for patterns. Seventh, collect interesting patterns from a specific represented form or set. Eighth, describe these mining patterns and / or go back to the previous steps for additional repetition. Finally, take action using the knowledge found and document or report 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, SEDM, when used explicitly, accounts for (and avail of opportunities to exploit) the multilevel hierarchy and lacks independent educational data [3].

IV. SEDM METHODS

The methods of educational mining come from various literary sources, including data mining, machine learning, psychometrics and computational modeling, statistics and other fields of data viewing. Work in SEDM can be divided into two main categories: 1) web mining and 2) statistics and visualization [11]. The category of statistics and visualization has found a prominent place in theoretical discussion and research in SEDM [8], [7], [12]. Another approach proposed by Baker [3] categorizes work in SEDM as follows:

1) Prediction.
   Classification.
   Regression.
   Density estimates.

2) Clustering.

3) Relationship mining.
   Association Governance Mining.
   Re correlation mining.
   Sequential pattern mining.
   Reason DM.

4) Distillation of data for human determination.

5) Discovery with model.

Most of the above are considered DM cat- e.g. However, the distillation of data is not universally regarded as DM for human decision. Historically, a wide variety of relationship mining practices have been very significant in SEDM research.

Discovery with models from the perspective of classical DM is probably the most unusual category in the Bakers SEDM classification. It has been widely used to model an event through any process that can be verified in some way. That model can be used as a component of another model, such as relation mining or prediction. This category (searching with models) has become one of the lesser known methods in the research field of educational data mining. It seeks to find out which learning content subcategories are most beneficial to students [13], how specific student behavior affects students learning in different ways [14] and how tutorial design [15] for students affects learning. Historically, relation mining methods have been increasingly used in educational data mining research over the past few years.

Other SEDM policies that are not widely used include:

Det More detectors search for data points that are significantly different from the rest of the data [16]. At SEDM, they can identify learning problems and students with irregular learning processes using learning reaction time data from e-learning data [17]. In addition, they can also eliminate unusual behavior by groups of students on the virtual campus. External identification can also detect manipulations and differences in the actions of learners or teachers [18].

Mining Text Mining can work with semi-structured or stretched-dataset datasets such as text documents, HTML files, emails and more. It has been used in the SEDM field in discussion boards. ILMS [19], [20]. It is also proposed to automatically build textbooks through web content mining [21] for use in text content. The use of text mining for clustering of documents based on similarity and content has been proposed [22], [23].

Network Social Network Analysis (SNA) is a study that seeks to understand and measure the relationships between entities in network information. Data mining can be used with network information to study online interactions [24]. In SEDM, this approach can be used for mining group operations [25].

An estimate

The goal of estimation is to estimate unknown variables based on history data for the same variable. However, input variables (predictive variables) can be classified or continued as variables. The effect of the model attendance pattern depends on the type of input variable. The estimated model requires limited label data for the output variable. Labeled data provides some foreknowledge of the variables we need to follow. However, in order to obtain a model attendance model, it is important to consider the effects of the quality of the training data.

There are three common types of estimates:

Ification taxonomy uses prior knowledge to construct a learning model and uses that model as a binary or hierarchical variable for new data. Most models are used as classifiers such as de-liquidated and logistic regression and support vector machines (SVMs).
Regression is a model used to estimate variables. In contrast to classification, regression models predict continuous variables. Various methods of regression, such as linear regression and neural networks, are widely used in the SEDM field to help students categorize risk.
Density estimates are based on a variety of kernel fungi, including Gaussian functions.
The estimation method in SEDM is used in a variety of ways. In general, it studies the features used for assessment and uses those features in the underlying structure to evaluate students' academic outcomes [26]. While different Procedures attempt to estimate the output output value based on hidden variables in the data, the output label obtained is not clearly defined in the data.
For example, if a researcher’s intention is to exclude students from school, it is difficult to use traditional research methods such as questionnaires, with a large number of schools and students participating. The SEDM method, with its limited sample data, can help achieve that goal. It should begin with defining students at risk and following the definition of variables that affect students, such as their parents' educational background. The relationship between variables and the school dropout can be used to construct a model attendance model that can assess students at risk. Doing these tips in advance can help organizations avoid problems or reduce the impact of specific problems.
Various methods have been developed to assess the quality of the predictor, including the linear correlation, the accuracy of Cohen's Kappa and A [27]. However, accuracy is not recommended for the evaluation of the classification method because it depends on the base rates of different classes. In some cases, it is easy to achieve high accuracy by classifying all data based on large groups of data. It is also important to calculate the number of taxonomies missing from the data to measure the sensitivity of the taxonomy using recall [28]. To give an overall evaluation of the classification, a mixed method, such as F-measurement, considers the true and false classification results to be accurate and recall based.

B. Clustering
Clustering is a method used to separate data into different groups based on certain characteristics. Unlike the classification method, in clustering, the data labels are unknown. The clustering method gives the user a comprehensive view of what is going on in that dataset. Clustering is sometimes called an unsupervised classification because class labels are unknown [10].
In clustering, we start searching for naturally grouped data points to divide the dataset into separate groups. The number of clusters can be predefined in the clustering method. In general, 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 classified based on their similarities and differences [29], [30].

C. Relationship Mining
The purpose of relationship mining is to find relationships between different variables in a data set consisting of a large number of variables. It helps to identify which variables are strongly related to a particular variable of particular interest. Relationship mining also measures the consistency of relationships between different variables. Relationships discovered through relationship mining must meet two criteria: statistical significance and interest. Large amounts of data contain many variables and therefore many related rules. Therefore, the measure of interest determines the most important rules that data support for specific interests. Researchers have developed a number of interesting activities over the years, including support and confidence. However, some research has shown that elevators and cations are very relevant in educational data mining [31].
A variety of relationship mining can be used, such as association rule mining, sequential pattern mining and often sample mining. Association rule mining is the most common SEDM method. The rule found in Association Rule Mining fıf → then the rule. For example, {If the student's GPA is less than two, and the student has a job, the student leaves school. The main goal of relationship mining is to determine whether the coverage of two events in a data set, such as Tetrad [32], or the induction of one event causes another event.

3D Discovery with Model

In search, models are usually based on clustering, prediction or knowledge engineering, using human logic rather than automated methods. The developed model will be used as part of other comprehensive models such as linkage-ship mining.

E. Distillation of data for human determination

The purpose of distilling data for human decision making is to make the data understandable. Presenting data in different ways helps the human brain discover new knowledge. Specific methods are required to visualize different types of data. However, the methods of visualization used in educational data mining differ from those used in different data sets [33], [34] in that they examine the structure and hidden meaning of educational data. Huh.

Data distillation for human judgment is applied in educational data for two purposes: taxonomy and / or identification. Data distillation model for classification A manufacturing process to create a hazardous model [35]; The purpose of identification is to present easily identifiable data through models that cannot be formalized [36].

As mentioned earlier, a wide variety of methods are used in educational data mining. These methods are divided into five categories by Ray [37]: clustering, prediction, relationship mining, search with models, and data distillation for human determination as described in Table I.

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

TABLE I: Educational data mining methodology categories.

V. Educational data mining data and Applications

The main objective of SEDM is to gather useful knowledge from academic data, including student data, student usage data, and LMS systems. Acquired knowledge improves the teaching and learning process in the education system [38]. It also leads to the development of new learning processes. Similar ideas have been successfully applied in the fields of science. For example, popular applications in e-commerce systems and basket analysis data mining [39]. They increase sales by analyzing consumers' shopping behavior. It is clear that data mining methods in education have not yet progressed as they are in business [40], and over the years, SEDM has attracted much attention from researchers. Applying DM to educational data is different in other domains, as defined below:

1) Objective: By applying DM methods to any specific data. The main purpose of using SEDM is to improve teaching and learning processes. Research goals such as gaining a deeper understanding of teaching and learning phenomena sometimes affect goals. Applying traditional research methods to achieve goals is sometimes difficult.

2) Data: The use of technology in education has led to an increase in data in education systems, which is different from basic information such as student information because it contains more information that is generated by different systems such as LMS systems. Applying SEDM methods to educational data can be very simple or complex in gathering specific knowledge such as implementing relational mining. An example is the application of relational mining to establish relationships between student success in courses that have multiple chapters in school.

3) Techniques: The application of DM to any problem is motivated by the objectives of the research and the type of data at hand. Therefore, specific adoption of educational data is required for successful implementation of data mining. Adoption can be for DM methods or pre-processing of data. Some DM methods can be applied directly without modification and some will not. In addition, some DM methods are used for specific problems in the field of education. However, the choice of certain methods depends on the researcher's approach to the problem and the objectives of the research [41]. For example, SEDM techniques can improve teaching and learning processes in the classroom, identify students at risk, optimize learning processes, and make recommendations to teachers and students. Only teachers and students were involved in the current research. However, more groups may be involved in rediscovery with other goals, such as curriculum development [42].

A. Data used in SEDM

SEDM provides a clear picture and a better understanding of learners and their learning processes. It uses DM methods to analyze educational data and solve educational problems. Like the extraction methods of other DM methods, SEDM captures interesting, detailed, useful, and novel information from educational data. However, SEDM is particularly relevant for development 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 approaches that combine data and theory will help improve the quality of T& L processes.

Growing technology in education systems has made large amounts of data available. Educational Data Mining (SEDM) provides significant relevant information [2]. Therefore, the main sources of data used so far in SEDM can be classified as follows:

Offline learning, also known as traditional traditional education, where knowledge is transferred to learners based on face-to-face interaction. Data can be collected through traditional methods such as observations and questionnaires. It studies students 'cog-local skills and determines how they learn. Therefore, statistical methods and psychometrics can be applied to data.

-E-Learning and Learning Management Systems (LMS) loses convenient materials, teaching, communication and reporting tools from students. Data mining methods can be applied to data stored by the system in a database.

Intelligent Tutoring System (ITS) and Adaptive Educational Hypermedia System (AEHS) seek to customize the data provided to students based on student profiles. As a result, it is important to apply data mining techniques to create user profiles. The data generated by that system will aid in further research.

Based on the three categories founded by Romero Etl [26], we can group SEDM research into the data type used: traditional education, web-based learning (e-learning), learning management systems, intelligent tutoring systems, adaptive education system, and test questionnaire. Others.

B. SEDM Applications

Several studies have been developed in the field of SEDM. Alexandro and Georgios recommended a framework to examine the behavior of learners in online education videos [43].

The proposed framework includes the capture of practitioner performance data, the creation of a data model for storing activity data, and the creation of modules to monitor and visualize practitioner behavior using the extracted data. Researchers relied heavily on students to watch the video a few days before the exam or deadline. In addition, stagnation and the onset of relapse appeared primarily in videos associated with an appointment. One lament is that the author did not study the factors that influenced the learner's behavior or why some practitioners completely avoided watching online videos.

Among other researches, Saurabh Paul [4] created a model using data mining techniques to estimate that students will drop out of the university program within their first year. That study used the Nave Base classification algorithm to build a prediction model based on current data. The result of the system is to identify students who need special attention to reduce the drop out rate. Leela Dodkhan [45] sought to justify dropping out of higher education institutions what was needed to retain students. As a result, the use of data mining methods has increased student retention and graduation rates.

End of VI

The increasing use of technology in education is generating large amounts of data every day, which has become a target for many researchers around the world; The field of academic data mining is growing rapidly and has the advantage of utilizing new algorithms and methods developed in various data mining areas and in machine learning. Data Mining of Educational Data (SEDM) helps to develop development methods for extracting interesting, detailed, useful and novel information, enabling students to better understand and have the settings they are learning. SEDM can be used in many areas such as identifying students, identifying students at risk, identifying the learning needs of different groups of students, raising graduation rates, and effectively evaluating organizational performance.

Optimization of campus resources and revival of subject curriculum. This paper surveys the most relevant studies conducted in the field of SEDM, including the data used and the methodology used in some studies. It defines the most common functions used in SEDM, as well as 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," JSEDM-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 min- ing 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. Educa- tional 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 SEDM, 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 12 th 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" ssamurai"," 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 Informa- tion 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 th 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 SEDM, 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.," SEDM, 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 SEDM, 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.