Review on Educational Data Mining

Ms. Twinkle[1], Mr. Gurpreet Singh[2]
Research Scholar[1], Assistant Professor[2]
Department of Computer Engineering, Punjabi University, Patiala

Abstract- As we have extensive measure of information in industry so it is important to investigate that data and extract the useful information by applying distinctive data mining techniques. With the help of technique like association rule mining, we can find different pattern of student performance by taking different measures of student like their course, their study of hour, monthly income of parents, and their background sources and many others. In this survey paper we have mainly two goals, the first is to advance development of education, the second is to organize the data and improve the student performance and their learning behavior by applying different data mining techniques. It has been seen that methods of education data mining are different from standard data mining methods. In educational data mining we mainly need to focus on data that are related to student learning and their performance.

Keywords: *Data Mining, Knowledge Discovery Database, Methods*

I. INTRODUCTION

Data Mining: It is the process of extracting previously unknown data from large databases. It is also called knowledge discovery database (KDD) or knowledge mining from data. Its main objective is to take information from a dataset and converted into understandable form. It is also used to solve problems by analyzing the data which is present in the database system. And the various techniques help in decision making. Information mining is not a system, yet rather a gathering of factual methodology that, as indicated by [3], empower the client to set up how intense the separation procedure is while distinguishing factors that can anticipate execution. In easier terms, it is a technique for misusing information and removing profitable data. Education (Information) Data Mining: As data mining is used in education research therefore it is named as "Educational Data Mining". As there is a growth in education so we have large amount of data and therefore mining of that data is important. The mined data can help in planning, teaching, learning and etc. As Educational Data Mining is concerned with developing methods to explore the student learning behavior and their performance and also their grades in examination. Indeed, the term instructive mining (EDM) has picked up money as a worldview equipped towards information investigation in training, the point of which is to look for, break down, and extricate data to create prescient models that assistance to upgrade instructive procedures. DM might be directed or unsupervised. As per [4], in unsupervised DM the calculation consequently finds designs information, with no earlier details about the kind of information. Then again, regulated calculations work with developed models and decide the kind of information being dissected. With the help of various methods educators will predict the score of students in examination and then compared the results with their original scored obtained in examination. Along these lines we can undoubtedly dissect the execution of week understudies and after that guide them that how they can build their execution. Educational data mining attempt to building a model of individual's objectives and their insight. In this way teachers can take input to take feedback on student learning experience with the model. For more effective learning we can apply the techniques of data mining such as decision tree, classification, association rule mining and neural networks. EDM can discover diverse examples with the help of association rule mining and anticipate the execution of understudies.

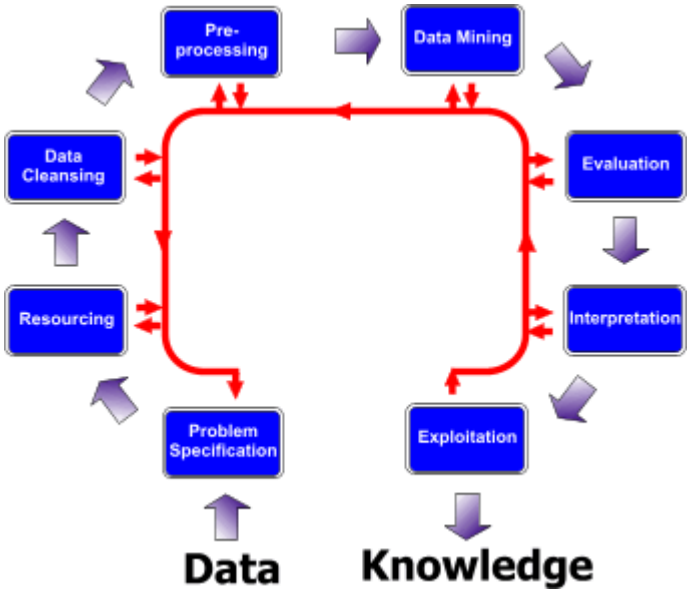

Fig 1 How data is mined?

The performance of student is very important in building their future. Our principle target is to distinguish the moderate learner execution. We will examine the performance of all types of students with the help of Clustering technique. Many creators connected a no of strategies on the information we have gathered and discover the best system which gives better outcome to dissect the attributes of understudies.

II. RELATED WORK

Data Mining: It is used in educational field to increase our understanding of learning process. Its main objective is to identifying, extracting and evaluating the variables related to students. It is described by Alaa el-Halees. Data Mining in educational environment is also known as Educational Data Mining.

Sr. no.	Year	Authors	Topic	Methods adopted (feature selection and classification algorithms)	Merits/limitations
1.	2014	C. Anuradha, T. Velmurugan	A Data Mining based Survey on Student Performance Evaluation System	In this, It is identified that to enhance the nature of higher education, the curriculum setup is the key point. The investigators are given new methods and patterns to maximize the student's performance through EDM techniques. Moreover classification algorithms ID3 and C4.5 are used to identify the various categories of students' performances.	The foremost intention of this research is to traverse the data mining techniques which are used for the improvement of student's performance and also identify the best suited structure of curriculum for the current environment.
2.	2015	Allan H.K. Yuen, Vincent Tam, Edmund Y. Lam, S.T. Fung and W.W.T. Fok	Enhancing Educational Data Mining Techniques on Online Educational Resources with A Semi-Supervised Learning Approach	an effective and systematic framework of a semi-supervised learning approach in which a concept-based classifier using the bag-of-word (BOW) approach is co-trained with an explicit semantic analysis (ESA) classifier to derive a common set of precedence rules based on a diverse set of online educational resources.	Their proposal of semi-supervised learning approach sheds light on many possible directions including the integration with other sophisticated optimizers such as the evolutionary algorithms to formulate more optimized learning paths for personalized learning, the pedagogical and other impacts of their proposals to sophisticated e-Learning systems for future exploration
3.	2015	Bo Guo, Rui Zhang, Guang Xu, Chuangming Shi and Li Yang	Predicting Students Performance in Educational Data Mining	In this study a prediction system, called Students Performance Prediction Network (SPPN), is proposed to predict student performance using emerging trend Deep Learning approach.	A deep learning architecture for predicting students performance for unlabeled data by automatically learning multiple levels of representation, can train model on a relatively large real world students dataset.
4.	2015	Kamaljit Kaur, Kuljit Kaur	Analyzing the Effect of Difficulty Level of a Course on Students Performance Prediction using Data Mining	The study investigates the possibility to predict the subject wise success rate of students in CBCEGS with the help of contemporary tools of data mining. Weka 3.6.12 has been used. decision tree model (CART) supplemented by ensemble classifier AdaBoost provide high accuracy in prediction of students' grades in a course . Another tree model J48 , Non-linear regression model (MSP), Non-linear regression model (MSP) help management, teachers, students to take useful decisions related to learning behavior of students.	As such problems are hard to solve manually to find out hidden patterns and knowledge, Weka serves as a free and open source tool for the analysis and evaluation of small data sets to solve various problems.
5.	2015	Kartika Maharani, Teguh Bharata Adji, Noor Akhmad Setiawan, Indriana Hidayah	Comparison Analysis of Data Mining Methodology and Student Performance Improvement Influence Factors in Small Data Set	Variations of feature selections have been accomplished (Gain Ratio, Principal Component Analysis, Classifier Subset Evaluator). Each selected features are then tested by classifiers (Naive Bayes) and being validated (Cross Validation). Best accuracy and smallest variance are achieved by CSE's selected features.	Overcome the problem of small dataset by preventing imbalanced class phenomenon. Imbalanced class is usually happened in limited number of observed data. SMOTE is one of various methods for handling the fault classification of minority class.
6.	2015	Kartika Maharani, Teguh Bharata Adji, Noor Akhmad Setiawan, Indriana Hidayah	Comparison Analysis of Data Mining Methodology and Student Performance Improvement Influence Factors in Small Data Set	Variations of feature selections have been accomplished (Gain Ratio, Principal Component Analysis, Classifier Subset Evaluator). Each selected features are then tested by classifiers (Naive Bayes) and being validated (Cross Validation). Best accuracy and smallest variance are achieved by CSE's selected features.	Overcome the problem of small dataset by preventing imbalanced class phenomenon. Imbalanced class is usually happened in limited number of observed data. SMOTE is one of various methods for handling the fault classification of minority class.
7.	2015	Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad	Educational Data Mining for Prediction and Classification of Engineering Students Achievement	used Neuro-Fuzzy classification for academic achievement for electrical engineering students. Classification method would allow more flexibility to judge on a single/group of student performance, and neuro-fuzzy linguistic is a value showing the probability of students to achieve excellent grade even if the student achieved weak in certain	In this, those techniques are used in which tells how to obtain knowledge from databases such as large arrays of student data from academic Institution databases.

				course/subject.	
8	2016	Anoopkumar M, Dr. A. M. J. Md. Zubair Rahman	A Review on Data Mining Techniques and Factors Used in Educational Data Mining to Predict Student Amelioration	give a comprehensive survey towards the research papers which would have discussed different Data Mining Methods especially the mostly used and trendy algorithms applied to EDM context.	This survey is very helpful for achieving good overview of educational data mining methods and tools which is used presently to bring about improvements in teaching and predicting the performance of Students so as to predict Academic Performance in Learning Programming.
9	2016	Mohammed Hussain, Mohamed Al-Mourad, Sujith Mathew, Abdullah Hussein	Mining Educational Data for Academic Accreditation: Aligning Assessment with Outcomes	The paper provided the detailed design of one of the framework processes, that is, the process of aligning assessments with student learning outcomes.	Framework to collect scope and verify the large amount of data. The analysis of the data is used to evaluate the institution against a standard set by an accreditation body, for the purpose of the academic accreditation of higher education programs. Therefore, the framework reduces human involvement in accreditation.
10	2016	Shaymaa E. Sorour, Tsunenori Mine	Building an Interpretable Model of Predicting Student Performance Using Comment Data Mining	established the main six attributes that represent relationships among sets of interrelated variables of students' comments towards learning subject. Finally, rules are extracted based on the prediction results, which will have great effect for eliciting more accurate decisions. This finding helps an instructor to prepare his lesson and focus on the attributes of each group.	The proposed method has two limitations: first, 'black box' and 'white box' models are not antipodes. If one can combine both approaches in an overall strategy, the advantages of both paradigms will be combined, and the disadvantages are eliminated. Second, they built the dictionary of attributes manually which extracted the shared factors from all the students and the words related to each attribute.

III. EDUCATIONAL DATA MINING METHODS

There are such a variety of techniques for educational data mining which are:

Prediction: In predictive modelling we will investigate the current information set and frame a model of its important attributes. In this we have two phases: the first one is training phase which involves building a model utilizing a huge specimen of information as training set. The second one is testing phase, which includes trying out the model on new data. Predictive modelling is associated with the procedures of classification and value prediction. The model is developed using a supervised learning approach. Once the model is defined clearly it can be used for forecast purposes. Both training and testing of the model should be performed. Training requires vast information whereas testing is done on little information.

Clustering: Clustering is a educational data mining technique that makes significant group of items that have comparative characteristics of students. By using clustering technique in educational data mining we can find different patterns of students and find connections between them. In this way we have a gathering of clusters like week students, normal students type and after that we can easily guide them how to enhance their learning behaviour.

Discovery with models: In this we will built up a model with the help of prediction, clustering and afterward it will be used as a segment in another investigation of information namely in prediction and relationship mining. In the prediction method the model was made and used to foresee a new variable and in relationship mining the made model will examinations the new predictions which are turned out as well as additional variables which are related to our study. The utilizations of this method include finding the relationships between student behaviours, their qualities and logical factors in the learning environment.

Distillation of data for human judgement: Humans can make different inferences about the information. I educational data mining, information is refined for human judgement for the most part for two purposes: the first one is identification and the second one is characterization. The purpose of this method is to compress the critical information in an intuitive way and understand the large measure of data which are utilized as a part of EDM to take decisions. Therefore we can say that this technique is very exceptionally valuable to teachers to comprehend the data.

IV. SOME OF THE CHALLENGES OF EDUCATIONAL DATA MINING

Educational data is incremental in nature: As the information is becoming quick therefore maintain the data in the data warehouse system is getting to be distinctly troublesome. Checking the operational data sources, student interest for their specific course, and its effect in particular organisation is the principle issue. The other issue in EDM is arrangement and

interpretation of the incremental data. We have to focus on time, content and their arrangement. In this HR is another issue in EDM.

Incorporation of background knowledge: The Background learning, or basic data with respect to the space under review, can be used to direct the revelation procedure and permit the found examples to be communicated in brief terms. Area learning that are identified with databases, such as integrity constraints and deduction rules and help into accelerate an information mining process.

Possibility of Uncertainty: Due to the presence of uncertain errors in the data, no construct model can foresee exact outcomes in terms of student learning and their performance or general scholarly arranging.

V. APPLICATIONS OF DATA MINING IN EDUCATION FIELD

Analysis and Visualization of Data: EDM is utilized to highlight imperative information and help in decision making. In the educational sector, it can be helpful for course administrators and teachers for analysing the basic data and students activities during their course to get a brief thought of an understudies taking in, their behaviour and their performance. Visualization information is the primary method that has been used for this strategy. Visualization uses some of the graphical methods to help people in understanding and examining the gathered information. There are number of studies related to visualization and analysis of different educational data such as patterns of hourly worked every day and yearly client conduct online gatherings.

Predicting Student Performance: In student performance, we will foresee the estimation value of some variable that defines the student. In educational data mining sector, the predicted values are student's performance, their imprints, learning or score got in exams. Classification technique is used to combine individual data items based upon quantitative qualities or based after preparing training set of previously unknown items. Different techniques of EDM and models are applied for prediction of student's performance by utilizing strategies such as decision trees, neural networks, rule based systems, Bayesian networks and so forth.

Enrolment Management: Enrolment management is basically utilized in higher education to clarify the very much arranged procedures and approaches to change the enrolment of student to meet arranged objectives. It is an organizational idea and also a systematic set of activities are designed to allow educational organisations to exert more influence over student's enrolments. Such practices are regularly incorporate maintenance projects, advertising, and confirmation strategies.

Grouping Students: In this case the gathering of students are made according to their customized features, their personal characteristics, their scores and so forth. These bunches/gathering of understudies can be used by the instructor or engineer to build a learning framework which can easily promote effective group in learning. Some of the DM techniques are used in this task are classification and clustering. Many clustering algorithms are used to group students are hierarchical agglomerative clustering, K-means and model-based clustering.

Planning and scheduling: Both Planning and scheduling are utilized to upgrade the traditional educational data mining process by arranging future courses, planning of their resource allocation which helps in the admission and counselling procedures, and so forth. Numerous DM techniques are used for this task are classification, categorization, estimation, and decision trees, link analysis.

VI. CONCLUSION

As Data mining is an extremely helpful tool for a wide variety of real-world problems where tremendous measure of information is put away and gathered. In Educational Data Mining we will first highlights the principle qualities that are used and the patterns found out by applying data mining techniques. When we apply association rule mining technique we will discovered many interesting relationship among data and from that we can anticipate the execution of understudy and their learning conduct. As every system of educational data mining has different focal points and impediments so it turn out to be extremely hard to discover which strategy is ideal but at the same time is rely on the reason for which instructive information is mined. In this manner we can state that this paper give a survey on procedures on instructive information mining and additionally their applications.

VII. REFERENCES

[1]. Lee, J., & Shute, V. J.," Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. Educational Psychologist, (2010), 45, 185–202.

[2]. González, C., Caso, J., Díaz, K., & López, M.," Academic performance and associated factors " Contributions from some large-scale assessments. Bordón (2012) , 64(2), 51–68

[3]. Romero, C., Ventura, S.,"Data mining in education" Wiley Interdisc. Rev.: Data Min. Knowl. Discovery 3(1), 12–27 (2013)

[4]. C. Anuradha, T. Velmurugan, " A Data Mining based Survey on Student Performance Evaluation System" Published in: Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on Date of Conference: 18-20 Dec. 2014

[5]. Allan H.K. Yuen, Vincent Tam, Edmund Y. Lam, S.T. Fung and W.W.T. Fok," Enhancing Educational Data Mining Techniques on Online Educational Resources with A Semi-Supervised Learning Approach" Published in: Teaching, Assessment, and Learning for Engineering (TALE), 2015 IEEE International Conference on: 10-12 Dec. 2015

[6]. Kamaljit Kaur, Kuljit Kaur," Analyzing the Effect of Difficulty Level of a Course on Students Performance Prediction using Data Mining" Published in: 2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015) Dehradun, India, 4-5 September 2015

[7]. Kartika Maharani, Teguh Bharata Adji, Noor Akhmad Setiawan, Indriana Hidayah," Comparison Analysis of Data Mining Methodology and Student Performance Improvement Influence Factors in Small Data Set" Published in: 2015 International Conference on Science in Information Technology (ICSITech)

[8]. Ms.Tismy Devasia, Ms.Vinushree T P, Mr.Vinayak Hegde,"Prediction of Students Performance using Educational Data Mining" Published in: The Third Information Systems International Conference 2015 Volume 72, 2015, Pages 414-422

[9]. Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement" Published in: Engineering Education (ICEED), 2015 IEEE 7th International Conference on Date of Conference: 17-18 Nov. 2015

[10]. Parneet Kaur, Manpreet Singh, Gurpreet Singh Josanc, " Classification and prediction based data mining algorithms to predict slow learners in education sector" Published in: 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015) Volume 57, 2015, Pages 500-508

[11]. Wil M. P. van der Aalst , Shengnan Guo, Pierre Gorissen, " Comparative Process Mining in Education: An Approach Based on Process Cubes" Published in: IFIP International Federation for Information Processing 2015 SIMPDA 2013, LNBIP 203, pp. 110–134, 2015.

[12]. Anoopkumar M, Dr. A. M. J. Md. Zubair Rahman," A Review on Data Mining Techniques and Factors Used in Educational Data Mining to Predict Student Amelioration" Published in: Data Mining and Advanced Computing (SAPIENCE), International Conference on Date of Conference: 16-18 March 2016`

[13]. J.Naren, Kirthika Ashokkumar," A Semantic Feedback on Student's performance with Data Mining Techniques: State of the Art Survey" Published in: Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on Date of Conference: 16-18 March 2016

[14]. Mohammed Hussain, Mohamed Al-Mourad, Sujith Mathew, Abdullah Hussein," Mining Educational Data for Academic Accreditation: Aligning Assessment with Outcomes" Published in: Global Journal of Flexible Systems Management pp 1–10

[15]. Sapan H Mankad, " Predicting learning behaviour of students: Strategies for making the course journey interesting" Published in: Intelligent Systems and Control (ISCO), 2016 10th International Conference on: 7-8 Jan. 2016

[16]. Shaymaa E. Sorour, Tsunenori Mine, " Building an Interpretable Model of Predicting Student Performance Using Comment Data Mining" Published in: Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on Date of Conference: 10-14 July 2016

[17]. Steven Lehr, Hong Liu, Sean Klinglesmith, Alex Konyha, " Use Educational Data Mining to Predict Undergraduate Retention" Published in: 2016 IEEE 16th International Conference on Advanced Learning Technologies