International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

Survey on Student's Performance Prediction Techniques

Niyati Lashkari
Department of Computer Science and Engineering, Sushila Devi Bansal College of Technology, Indore (M.P), India

Submitted: 25-06-2021 Revised: 01-07-2021 Accepted: 03-07-2021

ABSTRACT: The purport and objective of this survey paper is to review, data mining in the educational sector.
Accurate feedback on student's learning performance can help to improve the productivity of students. This feedback can use prediction techniques to predict the student's upcoming performance. In this subject, significant research has been carried out. Thus to understand the utility of student's performance prediction a survey on existing student's performance prediction system has been done. The work includes the study of educational data mining techniques, methods, applications, and recent research and development in student performance prediction. It also has a brief discussion about student performance prediction and the challenges in the existing prediction model. Finally based on the collected literature conclusion has been made and a future extension of the work has been proposed.
Keywords: Data mining, Educational data mining, Student performance prediction, Data mining algorithms, applications of EDM, methods of EDM.

I. INTRODUCTION

There is a huge amount of data available in the different Industries. This data is of no use until it is converted into information. Thus it is necessary to analyze this data and extract information from it. In the extraction of information, data mining played an essential role by involving a set of processes like Data Cleaning, Integration, Transformation, Mining, Evaluation, and Presentation. After these processes, we would be able to use this information in many applications like Academics, Fraud Detection, Market Analysis, Production, Science, customer retention, and others.

1.1 Data Mining

Data mining is a technique that takes information as input and outputs knowledge. One of the earliest and most cited data mining definitions is given by Fayyad, Piatetsky-Shapiro, and Smyth (1996), who defines it as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."[1]

Data Mining is the technique of extracting and studying massive data repositories to discover patterns withinside the data sets to offer data with the aid of using statistical and artificial intelligence methods.[2]. Data mining is the analysis stage in the process of "knowledge discovery in databases"(KDD). In addition, it also includes data and database management, data preprocessing, model and inference analysis, complexity considerations, post-processing of identified structures, online visualization and update, etc. There are various applications of data mining in the area of healthcare, telecommunication, market-based analysis, financial services, manufacturing, banking, food industries, road traffic, motor industries, retail sales, bioinformatics, and counter-terrorism, educational field, scientific and engineering, and business and many more other.[3]

1.2 History of Data Mining

The term "data mining" originated in the 1990s, however, data mining is the evolution of a field with a long history. The roots of data mining can be traced back to three families: classic statistics, artificial intelligence, and machine learning. The origin of data mining can be derived back to the development of artificial intelligence in the 1950s. The development of statistics mining is shown in Figure 1. [4]

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

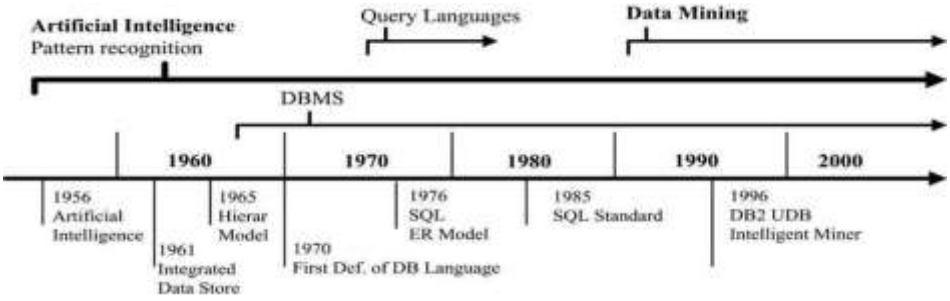

Figure 1: History of data mining development

1.3 Data Mining Process

The first three steps in Figure 2 involve preparing the data for mining, relevant data must be selected from a large and diverse data set, and then necessary preprocessing must be performed. Finally, the data transformed into a suitable representation. The fourth stage is data mining in which specialized computer algorithms are applied to preprocessed data to identify patterns. The patterns that are generated may take various forms (graphs, decision trees, etc.). Finally, the results of data mining are carefully evaluated and interpreted. [5]

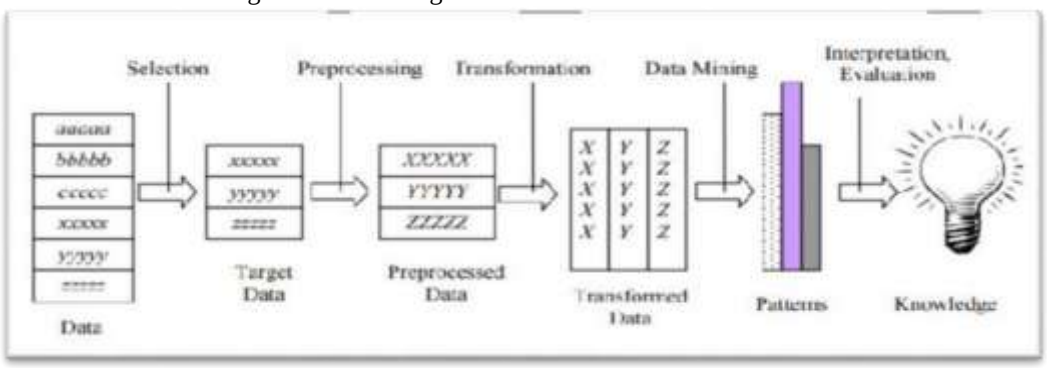

Figure 2: The data mining process

1.4 Data Mining in Education

Educators are liable for designing, planning, building, and maintaining the educational systems, while students use and interact with them. The application of data mining is different for educators as well as students. For students, the aim is to find out activities, resources, and gaining knowledge that enhances their learning, based on their attitude and likings. For educators, the aim is to have more feedback from students for evaluating the structure of the course material, the effectiveness of content on the learning process, to categorize students based on their needs, to discover information to enhance the model and customization of the course, etc.

Although considerableprogress has beenmade in thepastdecade, highereducation in India faces the following common challenges:

1.4.1 Supply and demand gap: India's higher education enrollment rate is very low, only 12% (postgraduate), while China's 26%. The demand for higher education is huge. The Indian government's goal is to achieve a gross enrollment rate of 30% by 2020, which means there will be 40 million university places and an increase of 14 million within 6 years

1.4.2 Poor Quality of teaching and learning: The system faces quality problems in many of its facilities: a shortage of teachers, poor teaching quality, outdated and rigid curriculum, and teaching methods, lack of accountability and quality assurance, and separation of research and teaching.

1.4.3 Limited research Capacity: With a very low level of Ph.D. enrolment (only 1% in the year 2012-13), India does not have enough high-quality researchers; there are few opportunities for interdisciplinary and multidisciplinary working, lack of early-stage research experience; a weak ecosystem for innovation, and low levels of industry engagement.[6]

Data mining technology can fill the knowledge gaps in the education system. The hidden patterns, associations, and anomalies discovered from

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

education data through data mining techniques can improve the decision-making process in the education system. This improvement can bring the following benefits: maximize the efficiency of the education system, reduce the dropout rate, increase the rate of student advancement, increase the retention rate of students, increasing student transition rate, improves the level of education, improves student performance, improves student learning outcomes and reduce system cost.[7] .

1.5 Applications of Data Mining in Educational Data Mining

A few utilizations of information mining in the instruction area are given beneath:

1.5.1 Analysis and Visualization of Data: It is utilized to feature significant data and backing dynamic. In the instructive area, for instance, it very well may be useful for course heads and instructors for examining the use data and understudy's exercises during the course to find out about an understudy's learning. It additionally gives data about provides details regarding month to month or week after week client patterns, utilization rundowns, how much course material understudies will examine, the arrangement wherein they study points, examples of contemplating movement, the fascinating subjects, timing, and sequencing of exercises.

1.5.2 Predicting Student Performance: In understudy execution expectation, we foresee the obscure worth of a variable that characterizes the understudy. In the instructive area, the most anticipated qualities are understudy's presentation, their imprints, information, or score. Various procedures are applied to anticipate understudy's general exhibition. This assessment is advantageous in foreseeing understudy's presentation for example expectation about understudy's accomplishment in a course and forecast about understudy's last grade dependent on information.

1.5.3 Enrolment Management: Enrolment the board is broadly utilized in advanced education to clarify very much arranged procedures and approaches to shape the enrolment of school to meet arranged objectives. It's anything but a hierarchical idea and furthermore a deliberate arrangement of exercises intended to permit instructive foundations to greater affect enrolments. Such practices regularly incorporate maintenance programs, promoting, monetary guide granting, qualification models, and affirmation approaches

1.5.4 Grouping Students: n these case gatherings of understudies are made dependent on their altered highlights, individual attributes, etc. These gatherings of understudies can be utilized by the instructors to assemble a customized learning framework that can accomplish compelling gathering learning.

1.5.5 Predicting Students' Profiling: Data mining can assist the executives with seeing the segment, geographic, and psychographic attributes of understudies dependent on records given by the understudies at the hour of affirmation.

1.5.6 Planning and Scheduling: It is utilized to improve the customary instructive cycle by arranging future courses, course booking, arranging asset distribution which helps in the affirmation and guiding cycles, creating educational program, and so on . Choice trees and Bayesian models have been proposed to assist instructive organizations with investigating the likely impacts of changes in enlistments, confirmations, and courses.

1.5.7 User Modeling: User demonstrating encloses what a student knows, what a student's experience resembles, what a student's conduct and inspiration are, and how they are happy with web based learning. EDM looks for in displaying client information, client conduct, and client experience.

1.5.8 Organization of Syllabus: Presently, the association of the schedule is impacted by numerous variables, for example, partnered, contending projects of colleges, accessibility of instructors, master decisions, and experience.

1.5.9 Detecting Cheating in Online Examination: Now daily's tests are directed online distantly by means of the Internet and on the off chance that misrepresentation happens, one of the essential issues is to know: who is there? Cheating isn't just done by understudies anyway the current embarrassments in business and reporting show that it's anything but a typical practice. Information mining strategies can prescribe models which can helps associations to recognize and forestall cheats in online evaluations. The models created use information containing different understudy's characters, stress circumstances produced by online appraisals, and normal practices utilized by understudies to cheat to get better checks in tests [8]

The rest of the paper is organized into 4 sections. The next section two focuses on the basic concept of EDM followed by the history of EDM

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

with an objective of EDM. Further, phases of EDM and methods of EDM are discussed. Then section three is devoted to the introduction about the Student's performance prediction system. Section four highlights the research trends in student's performance prediction techniques including various author's views on educational outcomes. The final section five concludes the paper with observations based on the survey work.

II. EDUCATIONAL DATA MINING

Educational Data Mining (EDM) is a discipline, concerned with developing methods for exploring data that comes from the educational sector to understand students learning. Educational data is taken from students used for interactive learning, computer-supported learning, or administrative tasks. Issues of time, sequence, and context also play important roles in the study of educational data. In this context, the International Educational Data Mining Society aims to support collaboration and scientific development, through the EDM [9]. Educational Data Mining (EDM) is a sub-domain of Data Mining that deals with data from academic databases which is used to develop various techniques and to recognize unique patterns. The obtained knowledge can then be used to offer suggestions to the academic planners in higher education institutes to enhance their decision-making process, to improve students' academic performance, to decrease failure rates, to understand students' behavior in a better way, to assist instructors, to improve teaching, and to construct regression models and decision trees to predict student performance in terms of their grades or percentage [10]. The EDM can be drawn as thecombination of 3majorregions informatics, education, and statistics. The intersection of these3regionsalsoformsdifferent sub-regions, such as computer-basededucation, machinelearning, and learninganalysis. [11].

2.1 History of Educational Data Mining

EDM is a relatively new discipline that emerged from the application of data mining techniques on educational data. The first international research conference on EDM was in Montreal, Canada in 2008. Journal of Educational Data Mining started publishing in 2009 and the International Educational Data Mining Society was founded in 2011. From that point, EDM continues to grow from different research areas such as data mining and machine learning, pattern recognition, psychometrics and other areas of statistics, artificial intelligence, information visualization, and computational modeling.
[12]

2.2 The Objective of Educational Data Mining

EDM targets to enhance numerous aspects of the education system. The objectives of EDM rely on the end-users (apprentice, educator, administrator, and researcher) and help solve their problems:

2.2.1Student model: this model in the educational domain consists of detailed information such as student characteristics or states such as knowledge, skills, motivation, satisfaction, metacognition, attitudes, experiences, and learning progress, or certain types of issues that they negatively affect learning outcomes. The main aim is here to create or improve a student model from usage information.

2.2.2Predictive modeling: Predict student's overall performance and learning outcomes. The goal is to predict a student's final grades based on data from the course activities.

2.2.3Generation of recommendations: The goal is to recommend to the students the content (or tasks or links) that is most appropriate for them at present.

2.2.4Analyze the learner's behavior: This takes several forms: Apply data mining educational techniques to investigate the behavior of the learner.

2.2.5Maintenance and improvement of the courses: The objective here is to determine how to improve the courses and maintain the system.

2.2.6Students: to provide comments or recommendations adaptable to the students, to respond to the needs of the student, to enhance the performance of learning, etc.

2.2.7Educators: to recognize the learning processes of their students and reflect on their teaching methods, to improve the performance of teaching, to understand social, cognitive, and behavioral aspects, etc.

2.2.8Administrators: evaluate the best way to organize institutional resources (teachers and course material) and their educational system.[13]

2.3 Phases of Educational Data Mining

Educational data mining offersthe interpretationof the latest hidden information from rawfactsgathered from education systems. EDM usuallyincludesthe following phases: Fig. threeshows the Phases of Educational Data Mining which are: The first section of the extraction of educationaldata is to discover the relationships among the data of the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

educationalenvironmentusingtechniques of data extraction (classification, grouping, regression, etc. The secondsection of the extraction of educationaldata is the validation to keep away from uncertainty. The thirdsection is to make predictions for the futurebased on tested relationships in thelearningenvironment. The fourth section is to assist the decision-making process with the help of predictions [14].

Figure 3: Phases of Educational Data Mining

2.4 Methods of Educational Data Mining

EDM methods are including Data Mining and machine learning, psychometrics, and other areas, such as statistics, information visualization, and computational modeling. The following key methods are used for discovering patterns in EDM.

2.4.1Prediction: Ryan S. J. d. Baker has given a detailed explanation of the prediction in his article.

He mentioned that in prediction, the goal is to develop a model that can infer a single aspect of

the data from some combination of other aspects of the data. If we study the prediction exhaustively, we obtain three types of prediction: classification, regression, and density estimation. In any prediction category, the input variables will be categorical or continuous. In the case of classification, categorical or binary variables are used, but continuous input variables are used in the regression. The estimation of the density can be done with the help of several functions of the Kernel.

2.4.2Clustering: In the clustering technique, the data set is divided into various groups, known as clusters. When the data set is already specified, then the clustering is more useful. As per the clustering phenomenon, the data point of one cluster and should be more similar to other data points of the same cluster and more dissimilar to data points of another cluster. There are two ways of initiation of a clustering algorithm. Firstly, start the clustering algorithm with no prior assumption and second is to start the clustering algorithm with a prior postulate

2.4.3Relationship Mining: Relationship mining commonly refers to growing new relationships among variables. It may beperformed in a large data set, having a no. of variables.

Relationship mining is a try tofind out the variable that ismostintentlyassociated withthe specified variable. There are 4kinds ofrelationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining.

2.4.4Discovery with Models: It consists of the design of modelsbased on conceptssuch as prediction, grouping and knowledge engineering, etc. These predictions of newly created models are used to find outa newexpected variable [15].

III. STUDENT'S PERFORMANCE PREDICTION TECHNIQUES

3.1.1 Student performance is an importantcomponent in higher education. This is because anexcellent academicrecord is a key criterion for getting admissions to top universities. There are many definitions of students' performancebut,Usamah et al. (2013) stated that" student's performancecan beobtainedby measuring the learningassessment and co-curriculum". Currently, numerous techniques are being proposed to evaluate students' performance. Data mining is one of themostfamoustechniquesto analyze students' performance. Data mining has been widelyimplemented in the academicarea recently. It is referred to aseducationaldata mining. Educational data mining is a method used to extract usefulrecords from a largeeducational database. The usefuldata and patternsmay beutilized in

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

predicting students' performance. As a result, it'dhelp the educators in presenting an effectiveteaching approach. Besides, educators may alsomonitor their student's progress and students can self-evaluate them. [16]

Student academic performance in higher education (HE) is researched significantly to tackle academic underachievement, increased university dropout rates, graduation delays, among other tenacious challenges. In simple terms, student performance refers to the extent of achieving shortterm and long-term goals in education. However, academicians measure student success from different perspectives, ranging from students' final grades, grade point average (GPA), to future job prospects. To improve student performance in schools and universities data mining and learning analytics techniques are used.

The timely prediction of student performance enables the detection of low-performing students, thus, empowering educators to intervene early during the learning process and implement the required interventions. Fruitful interventions include but are not limited to, student advising, performance progress monitoring, intelligent tutoring systems development, and policymaking. This endeavor is strongly boosted by computational advances in data mining and learning analytics. [17].

3.2 Challenges and Weaknesses of Existing Predictive Models

Several challenges faced by student's performance prediction system is as follows:

3.2.1The prediction of academic performance of student cohorts to help in the automation of course and programlevel outcomes assessment.

3.2.2The use and availability of multiple datasets from various disciplines reinforce the validity of the predictive model. The datasets have to include a huge sample size of students to draw any meaningful conclusions.

3.2.3The inspection of the outcomes of various capabilities at the attainment of scholar consequences to contribute to educational corrective interventions in better training, i.e., the shift from predictive analytics to explanatory analytics. 3.2.4 The use of multiple performance assessment metrics to evaluate the quality of the learning outcomes predictions.

2.4.5The lack of unsupervised learning techniques devised to forecast student attainment of the learning results. [18].

IV. PROBLEM DOMAIN

The motivation of the proposed work is taken from the research article reported in [6]. In this article the author provides a data mining technique to predict the student performance. In addition of they have tried to develop a heat map technique which provides the current performance of students without executing any algorithm. Finally they are able to provide satisfactory accurate prediction outcomes. This article can be extendable for the following two tasks as discussed below.

4.1. Improvement on accuracy for predicting the student's future performance

4.2. Additionally the given technique is extendable to produce recommendations to the students for adopting additional course work for fruitful future.

V. SOLUTION DOMAIN

The proposed data model for student's performance prediction is demonstrated using figure 1. In this diagram the two different kinds of data can be acceptable. First, the historical student records and second is the current student's records. Among them first part of database is used for training of the proposed data model. Additionally the second part of data is used for prediction. The historical data is first used with the k-means clustering algorithm. That algorithm is modified for finding the targeted outcomes. This algorithm results the clusters of students which are low performance, average performance and finally high performance. All the developed clusters are used with the C4.5 decision tree algorithm. The decision tree algorithm is able to organize the available data into "IF then ELSE" rules. These rules are used for predicting the student's future performance as well as recommending the improvements on current study. For prediction of a student's performance, the current student records are used for invoking the developed rules by the c4.5 decision tree algorithm. These rules contain the attribute values and the decisions. The best matched rule based outcome is used as prediction. In addition of for preparing the recommendation for student the tree branches are used. That recommendation helps the students for selecting the options which provide the success in their carrier path.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

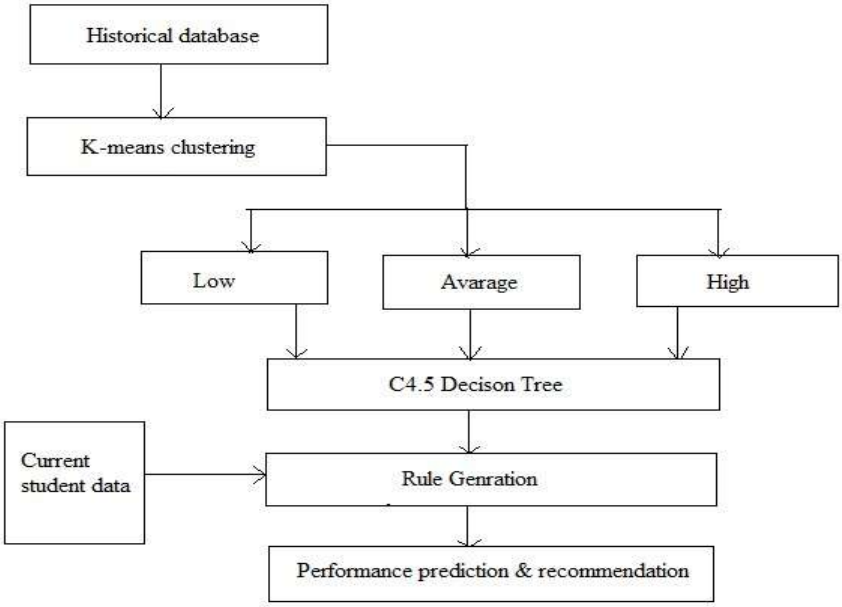

Figure 4: Proposed Data model

VI. LITERATURE SURVEY

This section involves the study of recent contributions and techniques developed for improving the quality of learning and improving the learning ability of students. Thus different methods and techniques are discussed in this section.

In online education systems, for proactive services to students, need to predict student performance on exercising activities. Existing methods exploit the historical records of students, where exercise is represented as manually labeled knowledge, and information contained in the text is still underexplored. Y. Su et al [19] propose an Exercise-Enhanced Recurrent Neural Network (EERNN) by both student exercising records and text of each exercise. For modeling the student exercising process, design a bidirectional LSTM to learn exercise representation from the text description. Then, propose an architecture to trace student states in their sequential process with the combination of exercise. For predictions, design two strategies, with Markov property and Attention mechanism. Experiments demonstrate the effectiveness of the framework. EERNN can deal with the cold start problems from both student and exercise perspectives.

EDM aims to discover knowledge and patterns about student performance. I. A. A. Amra et al [20] propose a student performance prediction model by two algorithms: KNN and Naïve Bayes. The objective is to improve the performance due to the prediction of student performance. Teachers also can take the evaluation to improve student learning. The results show that Naïve Bayes is better than KNN.

For offering proactive services to the students, one of the fundamental tasks is predicting student performance, where it is necessary to track the change of each student's knowledge. Existing approaches can exploit the exercising records and the problem of extracting information in the materials of exercises to achieve precise prediction and interpretable analysis of knowledge acquisition. Q. Liu et al [21] present a study to achieve the goal of performance prediction, first propose an Exercise-Enhanced Recurrent Neural Network (EERNN) by student's exercising records and text content of exercises. EERNN, summarize each student's state into a vector and trace it with RNN, using LSTM to learn the encoding of exercise. For predictions, design two EERNN, with Markov property and Attention mechanism. To track student's knowledge acquisition on multiple concepts, extend EERNN to explainable Exercise-aware Knowledge Tracing, where the student's integrated state vector is extended to a knowledge state matrix. Further, develop a memory network for quantifying how much each exercise can affect the students on multiple concepts. Experiments demonstrate the effectiveness of frameworks.

The student's performance prediction help teachers prevent students from dropping out and identify students need. M. Hussain et al [22] are predicting difficulties that students will encounter. They analyzed the data logged by a technology-enhanced learning system called digital electronics education and design suite (DEEDS) using ML

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

algorithms. The algorithms included ANN, SVMs, logistic regression, Naïve bays, and decision trees. The system allows students to solve exercises with different levels of difficulty. The input was average time, the total number of activities, idle time, number of keystrokes, and total activity for exercise; the output was the student(s) grades. The trained algorithms on the data from the previous session and tested the algorithms on data from the upcoming session. K-fold cross-validation and computed the receiver operating characteristic and RMSE to evaluate the performances. The results show that ANNs and SVMs achieve higher accuracy than others.

MOOC represents an ultimate way to deliver educational content to the students. The research agenda focusing on predicting and explaining the dropout of students and low completion rates has emerged. F. Dalipi et al [23] provide an overview of MOOC student dropout prediction where ML techniques have been utilized. They highlight some solutions used to tackle with dropout problem, provide an analysis of the challenges, and propose valuable insights and recommendations.

C. Verma et al [24] is written to support the real-time prediction of the age group of the students towards four ICT parameters such as development and availability of modern Resources, Attitude of students and mobile technology, Use of ICT and mobile technology in education and benefits. They have collected the data during the year 2017-18 and formulate the idea of feature aggregation to improve the prediction accuracy. They solved a multi-classification problem using 5 supervised ML classifiers i.e. KNN, Random forest, SVM, Bayesian network, and decision tree. The concept of aggregation of features provides maximum accuracy by joint classifiers (RT, SVM, and BN) and the second winner group is (RT, SVM, C5.1). Aggregation of features has increased accuracy in the prediction of a student's age group instead of an individual.

The institutions can utilize EDM to examine the performance of students which can support in recognizing the student's performance. B. K. Francis et al [25] a new prediction algorithm for evaluating student's performance has been developed based on classification and clustering. The result proves that the hybrid algorithm yields are far superior in terms of accuracy in the prediction of the performance of students.

Student performance prediction can benefit, in estimating dropouts, facilitating intervention, and enabling adaptive learning. Existing work targets learning platforms with predefined curriculum and labels. H. Li et al [26] propose an approach using Graph Neural Networks (GNNs) to achieve student performance prediction. Model the relationship between students and questions to construct the student-interaction-question network and present a GNN model, called R2GCN, which works for the heterogeneous networks, to achieve student performance prediction. They evaluate the approach in the problem-solving process. The results show that the approach can achieve higher accuracy.

The measurement of student performance during their progress provides academic leadership with information on each student's likelihood of success. The traditional method used interactions with individual students through class activities and assessments to identify the risk of failure. The modern university, offering online course material, may reduce attendance, making such identification more challenging. Data mining and ML techniques provide accurate predictions of student assessment marks; have focused on large student and wide ranges of data attributes. Many universities comprise small student cohorts, with institutional protocols limiting the student attributes for analysis. E. Wakelam et al [27] describe an experiment, where student data are limited, virtual learning environment accesses, and intermediate assessments. They found potential for predicting individual student interim and final assessment marks with very limited attributes and could be useful to support leaders in identifying students at risk.

Modeling student learning and predicting performance is crucial to personalized education for different students. It remains unclear how to conduct student performance prediction in question pools without wellorganized question orders or tags. H. Wei et al [28] propose to boost student performance prediction by considering student interaction and similarity in questions. Introduce new features based on mouse movement trajectories to problem-solving details. Besides, a heterogeneous information network is applied to students' historical problem-solving on similar questions, enhancing student performance predictions on a new question.
They evaluate the approach using four ML models. The result shows that approach can achieve higher accuracy than the traditional statistical features.

In online education systems, need to predict student performance. Methods exploit the historical records and information. EDM aims to discover knowledge and patterns about student performance. Student performance prediction can benefit, in estimating dropouts, facilitating

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 670-680 www.ijaem.net ISSN: 2395-5252

intervention, and enabling adaptive learning. The performance prediction help teachers prevent students from dropping out and identify students need. The institutions can utilize EDM to examine the performance of students which can support in recognizing the student's performance.

In this context, Y. Su et al [19] propose an Exercise Enhanced Recurrent Neural Network (EERNN) by student exercising records and text of exercise. Design a bidirectional LSTM to learn from the text. An architecture to trace student states with the combination of exercise. For predictions, Markov property, and Attention mechanism. EERNN can deal with the cold start problems. Similarly, Q. Liu et al [21] first proposed an Exercise Enhanced Recurrent Neural Network (EERNN). EERNN, summarize student's state into a vector and trace it with RNN, using LSTM to learn the encoding, with Markov property and Attention mechanism. To track knowledge acquisition, extend it to explainable Exercise-aware Knowledge Tracing.

I. A. A. Amra et al 201] proposes a model by two algorithms: KNN and Naïve Bayes. The objective is to improve the performance. Teachers also evaluate to improve student learning. The results show that Naïve Bayes is better than KNN. M. Hussain et al [22] are predicting difficulties that students will encounter. They analyzed the data log of a technology-enhanced learning

system DEEDS using ML. The algorithms included ANN, SVMs, logistic regression, Naïve bays, and decision trees. The system allows solving exercises. K-fold cross-validation and computed the receiver operating characteristic and RMSE. The results show that ANNs and SVMs achieve higher accuracy.

The research agenda focusing on predicting and explaining the dropout of students. F. Dalipi et al [23] provide a student dropout prediction using ML. They highlight some solutions and analyzed the challenges, propose, and recommendations. C. Verma et al [24] support the real-time prediction of the age group. They have collected the data and formulate the idea of feature aggregation to improve the accuracy. They solved a multiclassification problem using KNN, Random forest, SVM, Bayesian network, and decision tree. Accuracy by joint classifiers (RT, SVM, and BN) and the second winner group is (RT, SVM, C5.1). B. K. Francis et al [25] a new algorithm has been developed to proves that the hybrid algorithm yields are far superior in terms of accuracy. H. Li et al [26] propose an approach using Graph Neural Networks (GNNs) to model the relationship between students and questions to construct the student-interaction question network, called R2GCN, to achieve student performance prediction. The results show that the approach can achieve higher accuracy. Data mining and ML techniques provide accurate predictions of assessment marks. E. Wakelam et al [27] describe student data are limited as a key research issue. H. Wei et al [28] propose to boost prediction by considering student interaction and similar in questions. The features based on mouse movement were used. A heterogeneous information network is applied.

VII. CONCLUSION

In this paper, the main aim is to investigate Educational data mining and its techniques. The EDM is an application of data mining in educational sector data. In this domain, the data mining algorithms were applied to the academic data for different perspectives. Therefore first we studied different EDM models available in the literature. Additionally, the areas of applications in EDM are also reported. Further for understanding the need for student performance prediction a review of existing and recent work has been carried out. Based on the concluded literature and educational data mining technique we proposed a new EDM model in near future. That model is promising for Predicting student's academic performance and recommends the relevant study material. However, we concluded that the EDM methodologies are used to help students and teachers to improve their teaching and learning performance, and reduce the failure ratio of students to improve the quality of learning. In near future, the approach of student performance prediction will be implemented and performance would be measured and reported.

REFERENCES

[1]. Gary M. Weiss, Brian D. Davison," Data Mining", Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons, 2010.

[2]. Kantardzic, Mehmed," Data Mining: Concepts, Models, Methods, and Algorithms". John Wiley & Sons. 2003http://www.unc.edu/~xluan/258/datamining.html.

[3]. Hardeep Kaur,"A Review of Applications of Data Mining in the Field of Education",International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2015.

[4]. MeisamGordan, Zubaidah Ismail, Zainah Ibrahim, and HuzaifaHashim,"Data Mining Technology for Structural Control Systems: Concept, Development, and Comparison",DOI:
[5]. http://dx.doi.org/10.5772/intechopen.88651.
[6]. Brian D. Davison, Gm Weiss"Data Mining",John Wiley and Sons, 2010.
[7]. Jaya Srivastava," Data Mining in Education Sector: A Review",Special Conference Issue: National Conference on Cloud Computing & Big Data. 7.ShikhaDhoriyani,"Data Mining Application and Challenges",IJARIIE-ISSN(O)-2395-4396 Vol-3 Issue-1 2017.
[8]. Hardeep Kaur," A Review of Applications of Data Mining in the Field of Education, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2015, DOI 10.17148/IJARCCE.2015.4492 409.
[9]. AbdulmohsenAlgarni, "Data Mining in Education", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016
[10]. John Jacob, KavyaJha, PaarthKotak, ShubhaPuthran, "Educational Data Mining Techniques and their Application", International Conference on Green Computing and Internet of Things (ICGCloT),2015.
[11]. C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining and Knowledge Discovery, vol. 3, pp.12-27, 2012
[12]. C. Romero, S. Ventura. Educational Data Mining: A Review of the State-of-the-Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 40(**6**), 601-618, 2010.
[13]. R.Jindal, M.D Borah, A Survey on Educational Data Mining and Research trends, International Journal of Database Management System (IJDMS), 5(3), 2013, 53–73.
[14]. N. Upadhyay, V. Katiyar, A Survey on the Classification Techniques in Educational Data Mining, International Journal of Database Management System (IJDMS), 3(11), 2014, 725–728.
[15]. Baker R S J d, Yacef K. The state of educational data mining in 2009: A review and future visions. J EduData Min 2009.
[16]. Amirah Mohamed Shahiria, Wahidah Husain, Nur'aini Abdul Rashida," A Review on Predicting Student's Performance using Data Mining Techniques", 1877-0509 © 2015 The Authors. Published by Elsevier B.V.2015 DOI: 10.1016/j.procs.2015.12.157.
[17]. AbdallahNamoun and Abdullah Alshanqiti,"Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review",https://www.mdpi.com/journal/applsci 2021
[18]. Tsiakmaki, M.; Kostopoulos, G.; Kotsiantis, S.; Ragos, O. Implementing AutoML in educational data mining for prediction tasks. Appl. Sci. 2020, 10, 90.
[19]. Y. Su, Q. Liu, Q. Liu, Z. Huang, Y. Yin, E. Chen, C. Ding, S. Wei, G. Hu, "Exercise-Enhanced Sequential Modeling for Student Performance Prediction", The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)
[20]. I. A. A. Amra, A. Y. A. Maghari, "Students Performance Prediction Using KNN and Naïve Bayesian", 2017 8th International Conference on Information Technology (ICIT), 978-1-5090-6332-1/17/$31.00 ©2017 IEEE
[21]. Q. Liu, Z. Huang, Y. Yin, E. Chen, H. Xiong, Y. Su, G. Hu, "EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction", IEEE Transactions on Knowledge and Data Engineering, arXiv:1906.05658v1 [cs. CY] 7 Jun 2019. 22.M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, S. Ali, "Using machine learning to predict student difficulties from learning session data", © Springer Science+Business Media B.V., part of Springer Nature 2018
[22]. F. Dalipi, A. S. Imran, Z. Kastrati, "MOOC Dropout Prediction Using Machine Learning Techniques: Review and Research Challenges", 2018 IEEE Global Engineering Education Conference (EDUCON).
[23]. C. Verma, Z. Ill´es, V. Stoffova, "Age Group Predictive Models for the Real-Time Prediction of the University Students using Machine Learning: Preliminary Results", 978-1-5386-8158-9/19/$31.00 ©2019 IEEE
[24]. B. K. Francis, S. S. Babu, "Predicting Academic Performance of Students Using a Hybrid Data Mining Approach", Journal of Medical Systems (2019) 43: 162, https://doi.org/10.1007/s10916-019-1295-4
[25]. H. Li, H. Wei, Y. Wang, Y. Song, H. Qu, "Peer-inspired Student Performance Prediction in Interactive Online Question Pools with Graph Neural Network", CIKM

'20, October 19–23, 2020, Virtual Event, Ireland © 2020 Association for Computing Machinery. ACM ISBN 978-1-4503-6859-9/20/10

[26]. E. Wakelam, A. Jefferies, N. Davey, Y. Sun, "The potential for student performance prediction in small cohorts with

[27]. minimal available attributes", British Journal of Educational Technology, doi:10.1111/bjet.12836, Vol 51 No 2 2020 347–370

[28]. H. Wei, H. Li, M. Xia, Y. Wang, H. Qu, "Predicting Student Performance in Interactive Online Question Pools Using Mouse Interaction Features", LAK 2020, March 23 - March 27, 2020, Frankfurt, Germany © 2019 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-6317-4/19/07