Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

A Systematic Survey for Students Performance Prediction with Holistic and Sustainable Education approach using Educational Data Mining

Dheeraj Kumar Singh[1], Narender Kumar[2]

[1,2]*Department of Computer Science, Doon University, Dehradun, Uttarakhand, India*

Abstract:- Information about education is conveyed via text, audio, video, images, and other methods. It is essential to enhance students' learning, perception, and comprehension in the context of education because the majority of public inquiries, work, studies, and research are now conducted online. Some of the algorithms, techniques, and methods employed by Educational Data Mining (EDM) to forecast student performance include the Decision Tree (DT), Outlier Detection, Association Rule, Naive Bayes (NB), K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Neural Network, Relationship Mining, Regression Analysis, Random Forest (RF), and Social Network Analysis (SNA). 126 research articles from the core of EDM and other sources were examined to support the study, out of 4507 research publications that were cited at least once on Google Scholar. This paper will also identify the gaps related to value-based education in study and the occurrence of holistic education cum development through adding Universal Human Values (UHV) to STEM (Science, Technology, Engineering, and Mathematics) (UHV-STEM or UHVSTEM), as well as Time Series Analysis find suitable for student performance prediction during pandemic times. So, finally prediction of holistic and sustainable education of all possible forms can be explore through using Educational Data Mining approaches, tools, and techniques. In this sequence, the fourth sustainable development goal of the United Nations can be satisfied by value-based education, or UHV-STEM education. In India, UHV education already adopted by All India Council for Technical Education, University Grant Commission.

Keywords: *Educational Data Mining, Classification, Clustering, Decision Tree, UHV-STEM, Universal Human Values.*

Highlights:

- Reviewing the various study done on Educational Data Mining tools and techniques used for Students Performance Prediction
- Proposal for exploring Universal Human Values (UHV) through Educational Data Mining
- Proposal for educating UHV with STEM (UHV-STEM) for sustainable education cum development through Educational Data Mining
- Exploring Time Series Analysis for Student's Performance Prediction for forth coming future

1. Introduction

Knowledge Discovery in Databases (Agrawal et al., 1993) is also known as Data Mining and is mainly used for pattern recognition. Educational Data Mining (EDM) refers to the use of data mining techniques for pattern detection, particularly in the field of education. Pattern recognized for educational environment and related stake holders used for future knowledge creation and prediction related to student performance, dropout ratio, course selection, subject selection, and finally programmes selection and its achievements in the form of award, reward, value education, skill development, or enhancement of some or all. The learning and teaching ability of students and teachers can be enhanced by exploration of UHV and better prediction can be possible by using Educational Data Mining tools and techniques, which can respond to the correct or more appropriate prediction of student performance; these learning attributes for students as well teachers may be used for

- Improving students' and teachers' responsibility, accountability, learn-ability, teach-ability, and decision-making ability to understand any subject, course, or programme of education.

- Improving problem-solving abilities and adopting a solution-centric approach to learning rather than a problem-centric approach to learning

- Improvement of the sustainability of a student's life can be recorded and predicted using Educational Data Mining Techniques.

Patterns can be identified using Educational Data Mining Techniques, and these patterns will make all students' performance prediction transformational. Using Educational Data Mining Techniques, Vocational Training Programs can also be investigated, and student job mapping can be predicted. By using the Educational Data Mining Techniques, collaboration and collaborative governance can be maintained and predicted inter and intra organization, and it is possible through demographical details of students such as age, gender, socioeconomic status, and a variety of other parameters that can be established by developing psychological or self-assessment batteries. EDM finally can lead to socially and economically relevant education online and offline through universities, colleges, and other academic institutions.

EDM is most commonly used to predict and improve the process of learning, improving course completion, assisting students in course selection, profiling students, identifying problems that lead to dropping out, students' targeting, curriculum development, predicting student performance, and as a decision-making support at student enrolment.

Through EDM, role of data visualization, social network analysis, feedback for support management, planning and scheduling, student grouping, and detection of undesirable behaviors for better student performance prediction (Romero & Ventura, 2010) is possible.

The study (Kabakchieva, 2012) focuses on the development of data mining models for predicting student performance by employing four data mining algorithms for classification – a Rule Learner, a Decision Tree algorithm, a Neural network, and a K-Nearest Neighbors method. Where, as decision tree algorithm is best suited for predicting student's performance because it is simple to generate prediction rules for more efficient and early prediction (Acharya & Sinha, 2014). The use of data mining techniques in education, prediction of student performance, implications and impact on higher education and e-learning, the importance of aligning with pedagogy, learning theory and design, and research frameworks and ethics policies. Again, the most common techniques used in educational data mining are regression, K-Nearest Neighbor, clustering, neural networks, association rules, decision trees, and classification (Sin &Muthu, 2015).

Educational Data mining, as an interdisciplinary field of study, can use a variety of techniques to predict data for student growth. Students' learning styles are determined by a variety of factors such as time spent on learning tasks, learning in groups, learner behavior in class, classroom decoration, and student motivation to learn (Dutt, 2015). Their performances are compared using accuracy, precision, recall, and specificity performance metrics on a data set comprised of student responses to a real course evaluation questionnaire (Agaoglu, 2016). According to (Algarni, 2016), classification, regression, density estimation, clustering, data distillation for human judgement, model discovery, and relationship mining, which includes Association Rule mining, Correlation mining, sequential pattern mining, and casual data mining, are good approaches for Educational Data mining. For knowledge discovery from large data sets, various data mining techniques such as Classification, Decision Tree, Bayesian Classification, Neural Networks, Clustering, Association Rule Mining, Prediction, Time Series Analysis, Sequential Pattern, Genetic Algorithm, and Nearest Neighbor have been used (Karthikeyan & Kavipriya, 2017). The study (Norm Lien et al., 2020) intends to highlight the usefulness of machine prediction as a teaching guide and contribute to the development and implementation of data mining technology in the field of education by posing research questions, analyzing teacher and student prediction through machine learning, and identifying step-by-step actions in an ill-defined problem-solving approach. This paper is a review of 129 papers, 126 of which were primarily concerned with Educational Data Mining, main goal is to predict students' performance at every stage of life, using binary parameters such as "fail-pass", "success-failure", "good-poor", "below-above", and so on.

This paper is organized as follows: the second section will discuss Methodology, the third section will discuss Educational Data Mining, the fourth section will discuss Literature review on Educational Data Mining, the fifth section will discuss Student Performance Prediction with Various Related Factors, the sixth section will show the Discussion and Open Issues about Educational Data Mining with Gaps Identified, and the last section concluded the paper.

2. Methodology

The primary goal of this study is to review the literature for student performance prediction based on existing papers on Google Scholar, and to identify the challenges, gaps, and research goals in the field of Educational Data Mining. Educational Data Mining techniques are appropriate for this broadening to include all people, including students, teachers, trainers, educators, schools, colleges, institutions, universities, E-Universities, explorer, co-explorer, facilitator, co-facilitator, and all other academicians and stakeholders. This study followed the recommendation given by (Gaur et al., 2010) for determining the purpose, program, potential, of human being as well as the recommendation taxonomy parameter given by Cooper for literature review "Focus, Goal, Perspective, Coverage, Organization, Audience" of Cooper (Cooper, 1988), where the purpose, program, potential, of all human being is the same across the world by (Gaur et al., 2010), described in Table 1.

Table 1: Scope of the paper

Characteristics	Scope
Focus	Research Outcome / EDM Related as per student's performance prediction
Goal	Integration / Criticism
Perspective	Neutral Representation
Coverage	Representative and Pivotal
Organization	Conceptual
Audience	General Scholar / Policy Maker / General Pubic / Specialized Scholar

The Google search engine is used to find research papers using keywords related to the research problem. The methodology for this process is depicted in Figure 1.

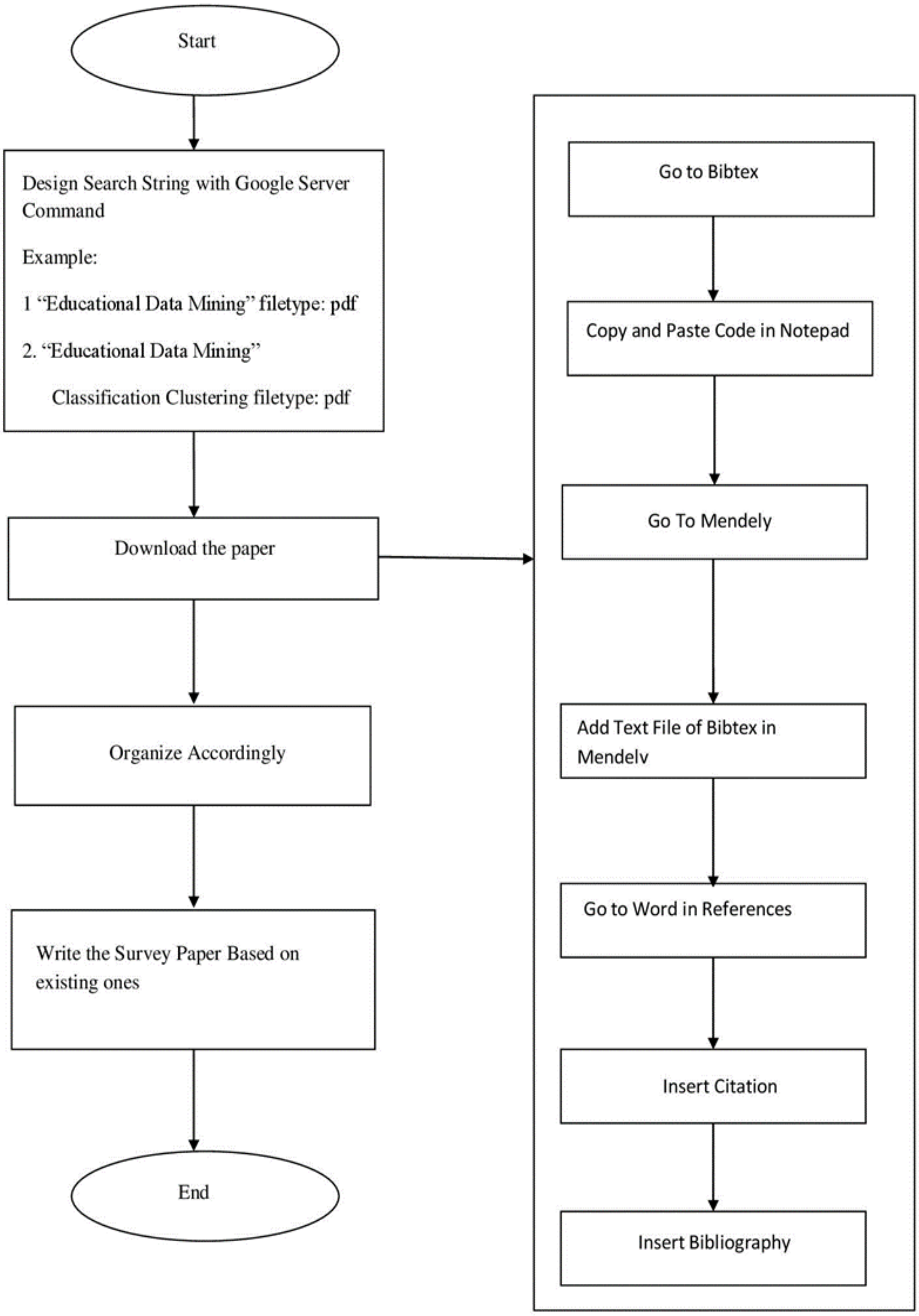


Figure 1: Methodology used for writing the paper

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

3. Educational Data Mining Techniques

Flat files, Relational Databases, Cloud Databases, Data Warehouses, Transactional Databases, Multi Media Databases, Spatial Databases, Time Series Databases, World Wide Web, Unstructured Data, Mobile Data, and Social Networks are all sources of data that can be mined for further knowledge extraction. Data mining is a technique used by various software to discover patterns.

Financial and Economical Analysis, Information Technology Industries, Telematics Industries, Intrusion Detection, Retail Industry, Elementary and Higher Education, Energy and Power Industry, Spatial Data Mining, Biological and Medical Data Analysis are some of the fields where data mining can be used. Educational Data Mining is a specialized field of study in the area of education.

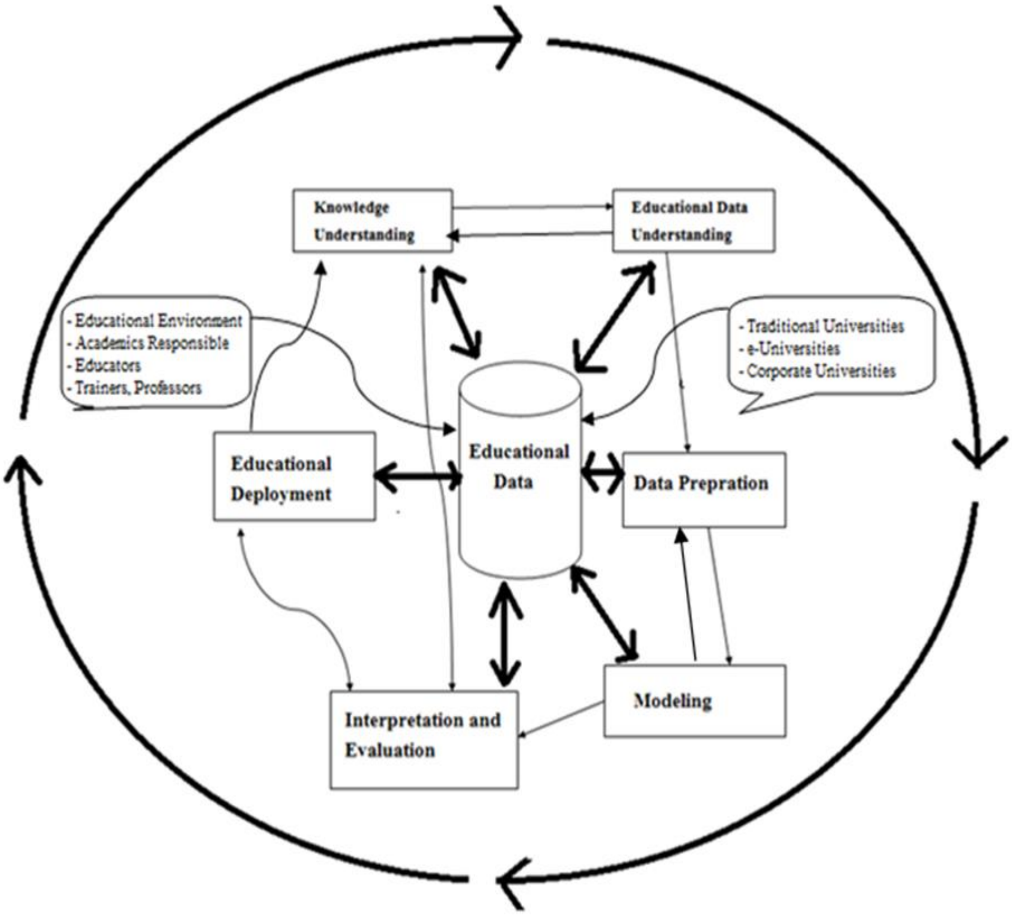
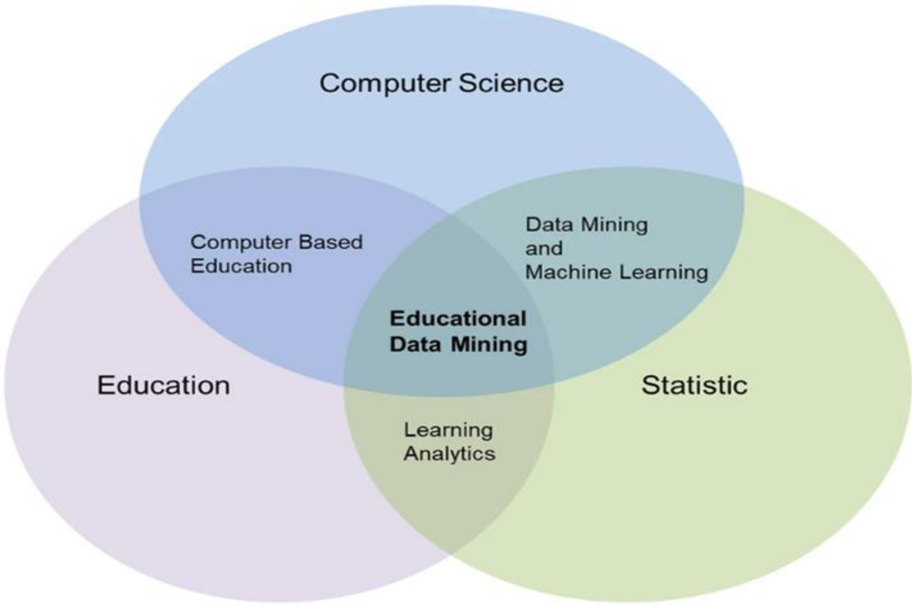


Figure 2: Information flow for Educational Data Mining

This study proposed hypothesis for Educational Data Mining as indicated in Figure 2. Educational Data Mining is the emerging trend in education (Baker & Yacef, 2009), where all the meta data is present at common storage area known to be staging area for current data as well permanent storage for all type and historical educational data, where knowledge understanding to identifying the student's behavior/ motivation is clear in mind, so the further pattern research can be generated and evolved for future decision making. In the diagram above, the educational database or data store is linked to all phases of knowledge understanding, educational data understanding, educational data preparation, modeling (tutor / students / course / outcome / marks / behavior / motivation / feedback etc.), interpretation and evaluation, and educational deployment.

Data transformation happens for better pattern identification, understanding, and modeling of the data for further interpretation and evaluation, and finally onto the level of deployment of this educational data. All the phases are linked in a cyclic order to mutually enrich one another. To better comprehend the progression of student data in various phases, this data will be refined for every instance of time. A more accurate model for student success across all educational dimensions can be predicted by employing the circular, mutually enriching process depicted in the figure 2. (Romero & Ventura, 2013) defined that Educational Data mining is closely related with the computer science, statistics, and education depicted in Figure 3.

From: A bibliometric analysis of Educational Data Mining studies in global perspective



Interdisciplinary structure of educational data mining (Romero & Ventura, 2013)

Figure 3: Information flow for Educational Data Mining

This educational data may be produced by various universities, including traditional universities, e-universities, corporate universities, and by various educational environments, academies, and responsible individuals, such as trainers, researchers, educators, and resource people for education, as well as by web-based education, e-learning, Learning Management Systems, Intelligent Tutoring Systems, Adaptive Educational Systems, Tests, and Questioners.

Students' performance can be possible to predict using various data mining algorithms as well various parameters can also be possible to identify. There may be different Data mining techniques as described below: -

3.1 Text Mining

A method of decomposing a matrix into some consecutive approximation is singular value decomposition (SVD). Singular Value Decomposition (SVD) can process data sets with millions of rows and thousands of attributes. For text mining, this strategy is really useful (Liu et al., 2005). Text mining uses a variety of programmes, apps, and APIs to tag, process, and identify textual data. Text mining employs tools for representing representational relationships and extracting data from text parts of speech, semantic word meaning, and sentence structure (Slater et al., 2017). Finally, text mining can be used to understand syntax and semantic analysis, as in the sentence "Ram eats Stone." In this case, it is discovered that the information provided here follows the common English rule of "subject + verb + object." However, it must be understood that "stone" can never be eaten. So, in the context of text mining, the given information has no significant meaning; thus, semantically, this is completely incorrect information, but syntax is correct. If the sentence is written as "Ram eats food." the text contains correct information as well as syntax which also follow grammar rules. As a result, this is semantic information that must be mined using appropriate text mining tools. So, the main goal of text mining could be to explore both structured and unstructured data.

3.2 Decision Tree

The Decision tree is a non-cyclic, tree-structured flow chart for representing data and predicting outcomes. Decision Tree is a good decision support tool for generating derivation rules. Internal nodes are always non-leaf nodes, and external nodes are always leaf nodes, and these nodes contain prediction properties. Due to the graphical nature of Decision Tree, it has evolved into a technique for visualizing entities in the form of nodes such as root nodes, internal nodes, and leaf nodes (Merceron & Yacef, 2005) (Asif et al., 2017) (Fernandes et al., 2019).

3.3 Classification Analysis

Classification techniques can be applied for doing the better prediction of performance of students of higher studies. The quality of performance by classification techniques depends on the students' attributes taken in particular course. The decision tree may be one of the finest method for classification techniques to apply(Abu Tair & El-Halees, 2012).Basically classification categorizes data for effective and efficient use with respect to time space trade off as well techniques and methods used in it (Durairaj & Vijitha, 2014).Another Naïve Bayes method of classification where, one specific feature present always independent from others, can give better forecast of student's performance (Devasia et al., 2016) (Parhizkar et al.,2023).

3.4 Association Rule Learning

When we take the cases of learning this may be in traditional or may be e-learning selection of multiple courses always taken place with respect to each other by the students. So we can understood the association mining always show some relationship for next chosen courses, and the algebraic and mathematical rules can be developed for such relationship which shows better prediction for student performance and course or material selection and to be understood as association rule mining. Association rule mining can be best for finding the relationship probability based on previous conditions which can be represented in form of "if then else". Association rule mining can be best suited for finding the patterns for e-learning also (Agarwal et al., 2021).

3.5 Anomaly or Outlier Detection

May be possible the student's behavior changes over the time at the time of selection of courses, or performing the examination etc. So due to deviation in events at any point of time, data deviate the behavior in data source, discovering data points is a critical study. Students learning problem can be detected in Educational Data mining through Outlier analysis, by using two approaches such as distance-based approach and density-based approach of outlier detection mining. Outlier detection itself is anomaly detection mining through its derived algorithm versus approaches (Abu Tair & El-Halees, 2012).

3.6 Clustering Analysis

Clustering is unsupervised learning approach for doing Educational data mining, and is useful in scientific data exploration, text mining, spatial data analysis, medical diagnostic, computational biology, web analysis, CRM, Marketing research, and in many more applications (Durairaj & Vijitha, 2014).In this technique of mining data patterns and trends can be explored through available data. A set of educational functionality and realistic pattern can be developed for Educational Data Mining. The implementation and development of statistical method or algorithm with clustering approaches can be developed for better performance of students for different disciplines (Peña-Ayala, 2014).

A typical behavior objects with vide variety of tasks can be grouped altogether by studying the pattern generated by the student's performance of any course in Educational Data Mining. In Clustering, similar pattern objects can be put all together in similar classes as per pattern recognized. Success and failure parameters for students can be generated through these recognized patterns(Asif et al., 2017).For examining information centroid based clustering, graph based clustering, Hierarchical clustering, and partitioned clustering can be used (Ahuja et al., 2019).Clustering can also be one of the fruitful unsupervised learning approaches for identifying pattern of students in blended learning (Nkomo & Nat, 2021).

3.7 Regression Analysis

In statistical approach, calculating one variable in reference to another can be called as dependent versus independent variable, known as Regression analysis. On behalf of pattern recognized better regression model can be formulized. This may lead to better prediction of student's performance. More precisely regression is a numerical evaluation process based on existing data (Jacob et al., 2015).Regression model made can be trained using genetic algorithm to find better results. Multiple linear regression, Decision tree regression, and Support vector regression may be a fit regression model for predicting better performance of students (Mehta et al., 2021).

3.8 Sequential Patterns

To delivering the data in numeric form and in sequence, frequent patterns can be made for such numeric transactions, which can be known as sequential pattern mining. In sequential pattern mining intersession pattern can be find in a very easy way for evaluating the pattern for multidisciplinary purposes (Agrawal & Srikant, 1995)(Romero & Ventura, 2007). Sequential pattern also possible to identify through predictive analysis patterns, and are used for predictions as well for final decisions. The performance of prediction model highly depends on independent variable taken as per the data present and the output observed by dependent variable. Where selection of choice of variables for student's performance prediction must be from reliable data source. In prediction model one has to focus on known results, and have to forecast future outcomes. Also for making better future decisions and predictions the new prediction algorithm can be forecast with the help of machine learning as well as various data mining techniques (Ramaswami & Bhaskaran, 2009) (Anoopkumar & Rahman, 2016) (Asif et al., 2017) (Khasanah & others, 2017) (Aleem & Gore, 2020) (Nahar et al., 2021) (Hantoobi et al., 2021) (Pallathadka et al, 2023).

3.9 Time Series Analysis

In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. The following are some of the most basic components of time series: trend, which describes the movement along the term; Seasonal fluctuations, which are changes in the seasons; cyclical fluctuations, which are periodic but not seasonal; Irregular variations, which are nonrandom origins of series fluctuations. Finally, an observed time series can be decomposed into further components: the trend (long term direction), the seasonal (systematic, calendar related movements) and the irregular (unsystematic, short-term fluctuations), A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency. These fluctuations are usually due to economic conditions, and are often related to the "business cycle". The duration of these fluctuations is usually at least 2 years (Priestley, 1980) (Durbin, 1984). Gaming time series analysis can be applied for gaming of trajectory analysis, and can lead to learning for gaming (Sawyer et al., 2018). Since students all courses are bound in some specific definite time intervals in the case of educational institutions, such as schools, colleges, universities as well as e-university. So, everything regarding a student comes on definite time such as class test, sessional tests, pre-university examinations, and university examination bound with a specific university academic calendar. So, Time Series Analysis possibly may be one of the best approaches for student's performance prediction through Educational Data Mining.

3.10 Pattern Mining

Tracking patterns is a fundamental data mining technique. It involves identifying and monitoring trends or patterns in data to make intelligent inferences about business outcomes. Once an organization identifies a trend in sales data, for example, there's a basis for taking action to capitalize on that insight. Intelligent inferences about business, education, and other related things can be made by using this fundamental tracking pattern mining. Pattern can be recognized in form of data set and in the flow of variable in a period (Zhang et al., 2018).

3.11 Visualization

Data visualization is the main concern for representing the pattern of research, and may be in the form of various type of graphs and tables. This data mining technique will present the data in the more concise way (Prajapati et al., 2012). Understanding trends, patterns, and outlier this technique can be used by means of graphs, charts, and maps which are the visual elements of data representation in Data Visualization, and can also be beneficial for Students performance prediction (Sáiz-Manzanares et al., 2021).

3.12 Neural Network

Human brain operated automatically and follows some specific pattern internally, by putting this all-in mind; some of the related algorithm can be made which can map just like a human being and can predict the student's performance. This fundamental concept can be known as Neural Network and data mining pattern can be generated through such algorithm may be better predict the hidden information related to students. This algorithm may be one of the better algorithms for student's performance prediction, and the best results can be obtained (Godwin-Jones, 2021).

4. Related Work on Educational Data Mining

It is also critical to focus on time-sensitive information during the data mining process, including the quantity and quality of data as well as the data source (Mostow et al., 2005). Because web-based education is becoming more popular in society, it is necessary to devote more time to applying various data extraction techniques such as text mining, association rule mining, and pattern mining, and to integrate such techniques with e-learning systems (Romero & Ventura, 2007). When Educational Data Mining techniques are integrated with fully developed educational systems, and the specified parameters of learning remain the same, major tutor changes will not affect learning efficiency (Cen et al., 2007).

Educational Data Mining also supports pedagogical support, analysis, prediction, and is useful in determining which type of pedagogical support can improve learning. Web mining, text mining, statistics, visualization, and discovering new models can all be used to improve pedagogical support for students (Baker & Yacef, 2009).

Educational data mining models can be more useful in predicting student dropout rates from various courses, and the proper reasons can also be predicted using various classification techniques such as Nave Byes (Kotsiantis, 2009).

Educational Data Mining is a research area for offline face-to-face skill transformation that has moved to the web via e-learning, web mining using a learning management system integrated with an Intelligent Tutoring System (ITS) and an adaptive educational hypermedia system (AEHS). EDM can be used to collect data on student performance, behavior, and curriculum (Romero & Ventura, 2010).

A variety of parameters, including personal, socio-economical, psychological, school location, school type, parental education, marks obtained, and environmental variables, improve and impact student performance in Educational Data Mining. Rules extracted from the Chi-square automatic interaction detection (CHAID) decision Tree may be useful in predicting student performance (Ramaswami & Bhaskaran, 2010). Overfitting occurs when a model has been overfitted to the training data to the point where it expresses even the most unusual special cases and data errors. Because the resulting model is so specialized, it cannot generalize to new data. Overfitting is a significant issue in the educational domain because there are numerous attributes available to design a sophisticated model but only a limited amount of data to effectively learn it. Educational data can be numerical or categorical in nature, with discrete values (Hämäläinen & Vinni, 2011).

For inquiries, representational tools such as the table tool and the graph tool can be used, where the table tool collects data through trials run by students and the graph tool displays two-dimensional relationships between dependent variables and independent variables, which students must understand in order to generate new hypotheses. Finally, an Evidence Centric Model for Students can be predicted to improve students' learning abilities. In the conceptual assessment framework, the evidence model is linked to the task model and the student model. Auto-score and student inquiry performance can be improved using Educational Data Mining techniques (Gobert et al., 2012). The United States Department of Education's National Education Technology Plan (NETP) (U.S. Department of Education 2010a) adopted EDM and Learning Analytics in its education system due to their visualization, prediction, and support capability for students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities (Bienkowski et al., 2012).

Some of the main topics that can be developed through EDM academic benefit include data analysis and visualization, feedback for instructors, recommendations for students, predicting student performance, student modelling, detecting undesirable student behavior, grouping students, social network analysis, developing concept maps, constructing courseware, planning and scheduling of all of the above (Fernández et al., 2014). EDM can forecast learner success, scenario-based instructional sequences that may be more effective for a specific student, students' actions, learning progress, learning environment characteristics, and feedback, as well as visualize and describe raw learning data (Bousbia & Belamri, 2014). Combining Decision Tree and EDM capabilities yields a high level of automation. This allows the models' performance to be evaluated; lift graphics revealed values of 1.74 for Model I and 1.36 for Model II, indicating that the models are capable of prediction (Guruler & Istanbullu, 2014).

It is critical to promote EDM characteristics in order to serve as a technical agent in support of state goals aimed at improving education. EDM techniques for mining educational data can be designed, developed, and deployed as a tool for implementing educational reform (Pena-Ayala & Cárdenas, 2014). Through EDM techniques for checking students' diversity, automated detectors for population can be implemented and can fit in both inter- and intra-culturally (Ocumpaugh et al., 2014).

Methods like Naïve Bayes, Multilayer Perception, and SVM classification algorithms were compared to Students Performance Prediction Network (SPPN) in EDM, and it was determined that SPPN had the best accuracy of the three, with an average accuracy of 77.2 percent. (Guo et al., 2015) describes Deep Learning architecture for forecasting student performance that employs unlabeled student data to learn various levels of representation automatically. Clustering is a pre-processing method used by the EDM. It is an unsupervised data analysis technique used in statistics, machine learning, pattern recognition, data mining, and bioinformatics. It refers to assembling similar objects into a group or cluster. Clustering methods for algorithm implementation can be of the Centroid-based clustering, Graph-based clustering, Grid-based clustering, Density-based clustering, Neural network-based clustering, and cluster mapping can be defined as binary mapping and by degree of belonging (Dutt, 2015). Other attributes, such as student behaviours, skills, and attitudes, play a significant role in prediction from educational data, and suitable Educational Data mining techniques can be used for such prediction (Thakar et al., 2015).

Educational data mining (EDM) techniques can be used to manage information for student analysis and their performance indicators and knowledge can be discovered by incorporating value chain characteristics in design considerations when implementing Data Warehouse for educational data (Moscoso-Zea et al., 2016).

Teachers and instructors are constantly categorizing students based on their knowledge, motivation, and behavior. Using educational data mining techniques, a test using Waikato Environment for Knowledge Analysis (WEKA) tool among four classification methods, the J48, PART, Random Forest, and Bayes Network Classifiers, achieved an accuracy of 99 percent by the Random Forest classification method, as well as minimum errors measured in terms of Mean Absolute Error, Root Mean Square Error, Relative Absolute Error, and Root Relative Squared Error. The Apriori algorithm was also used to discover the association rules and the frequent patterns. All of the results were tested using twenty-four attributes from a data set collected from three colleges in Assam, India (Hussain et al., 2018).

The Multilayer Perceptron algorithm predicted 76.6 percent accuracy, which was higher than the other algorithms used in the research. This was determined by a thorough comparison analysis using Educational Data Mining techniques between Random Forest, Naive Bayes, Multilayer Perceptron, Support Vector Machine, and Decision Tree - J48 (Jalota & Agrawal, 2019).

Greater interpretability, generalizability, transferability, applicability, and clarity of evidence of student performance effectiveness should be objectives in the field of educational data mining. Educational data mining can be applied to collaborative learning, participation and online connections, motivation and engagement, meta-cognition, and self-regulated learning (Baker, 2019).

Using educational data mining techniques, named as Random Forest, Decision Tree, Naive Bayes, Tree Ensemble, and Logistic Regression, the accuracy of students' performance by Logistic Regression has an accuracy of 89.15 percent, which is higher than the other techniques used (Adekitan & Salau, 2019).

Several classification techniques, including K-nearest neighbour (k-NN), random forest (RF), support vector machine (SVM), logistic regression (LR), multi-layer perceptron (MLP), and naive bayes (NB), as well as the Gini index and p-value, were used to categorize the students and predict their performance (Injadat et al., 2020). Gamification techniques combined with EDM can improve learning ability, student engagement, and learning analytics. additionally, an adaptive gamified learning system can be created to help students in improving their grades in any course (Daghestani et al., 2020).

In a blended learning course, learning behavior of the students can be examined using educational data mining techniques. EDM strategies could be used in a blended learning environment to properly predict high-risk learners. The f1-score of the random forest model in this study was 0.83, which was higher than the f1-scores of the decision tree and logistic regression models (Hung et al., 2020).

Educational data mining techniques can support and enhance the students' admission to any or all courses. Early university admission performance can be predicted before admission based on certain pre-admission criteria. According (Mengash, 2020) artificial neural networks have an accuracy rate of over 79 percent, outperforming other classification techniques like decision trees, support vector machines, and Naive Bayes.

The classification models that can most accurately predict student performance include the Nave Bayes classifier (NBC), support vector machine (SVM), and multilayer perceptron (MLP) (Salih & Khalaf, 2021). The challenges of data gathering, data interpretation, database design, and data organization can be reduced by employing EDM approaches. EDM can deliver dependable results to advance comprehension of the data pre-processing stage of student data among academicians and educational policymakers (Feldman- Maggor et al., 2021).

Using a flipped classroom strategy and educational data mining, learning results can be improved. Social media platforms can play vital part in flipped class learning approach using EDM approaches for excellent outcome for learning in critical thinking, communication, compassion, work ethics, and accountability (Su & Lai, 2021). A prediction model for student performance that combines EDM approaches with machine learning can be built, and such models can be used in classical research settings (Hilbert et al., 2021).

It is difficult to identify the variable associated with an outcome when employing EDM in predictive and explanatory models, and these models may contain false negatives and unbalanced binary input and result characteristics, which characterize the difficulties when applying regression approaches (Young & Caballero, 2021).

During a learning video course, a learner who is watching a pedagogical sequence for an instructional video can utilize educational data mining techniques to determine if they pass or fail. K-nearest Neighbours and Multilayer Perceptron algorithms, according to research, may predict learner performance, with the K-NN classifier achieving an average accuracy of 65.07 percent (El Aouifi et al., 2021).

The accuracy of educational data mining techniques depends on the features of the algorithm and the quantity of attributes used. Questionnaires are better for utilizing students' perceptions, which can indicate sustainability in data mining approaches. Compared to Sequential Minimal Optimization (SMO), MLP, R.F., Bagging, and OneR algorithms, RF (Random Forest) and SMO can make superior predictions (Kim et al., 2021). In Self-Regulated Learning (SRL), SMART processes itself including searching, monitoring, assessing, rehearsing, and translating are used to track student behavior. Educational data mining techniques can forecast this behavior (Hutt et al., 2021).

Also, methodical examination of Supervised, Unsupervised, Semi-supervised machine learning (ML) in academic performance highlights the importance of regression, classification and clustering techniques, as well as gaps in population, practical knowledge, and value-based skill development. Popular ML methods include Random Forest, Support Vector Machine, Artificial Neural Network, Logistic and Linear Regression, Decision Tree, Naïve

Bayes, and K-Nearest Neighbor. Demographic and academic characteristics have also been discovered to be significant factors (Issah et al., 2023) (Alalawi et al., 2023).

(Alghamdi & Rahman, 2023) calculated and find Naïve Bayes (NB) as best performing techniques with the accuracy of 99.34 % in students' performance prediction.

(Roslan & Chen, 2023) identify that DT and NB have almost similar highest predictive accuracy (83.9%).

Further (Munshi et al., 2023) proposed new model for classification Elman Neural with Apriori Mining (ENAM) and found better accuracy than compare to SVM, RF, K-NN.

Attribute selection is influenced by genetic algorithms, but student performance is affected in different ways by regression, classification, decision trees, and K-NN algorithms through performance evaluations (Hussain & Khan, 2023).

(Kukkar et al., 2023) proposed a new model based on Clustering, Machine Learning and Deep learning named as Student academic performance predicting (SAPP) system and found 96% accuracy of prediction.

Educational Data Mining versus Learning Analytics and Knowledge

The International Educational Data Mining Society defines EDM as follows: "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." (Baker & Yacef, 2009) (Siemens & Baker, 2012). The Society for Learning Analytics Research defines Learning Analytics (LA) as: "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Baker, 2012).

The best criteria for understanding learning analytics are statistics, visualization, social network analysis, sentiment analysis, influence analytics, discourse analysis, concept analysis, and sense making models as a techniques, semantic web, intelligent curriculum, and systemic interventions as origins. By contrast, categorization, clustering, Bayesian modeling, relationship mining, discovery with models as a technique, educational software, student modeling, and predicting are the best criteria for understanding data science. While EDM places a stronger emphasis on describing and comparing Data Mining approaches, Learning Analytics places more attention on describing data and its findings, where leveraging human judgment is important and automated discovery is the ideal tool (Romero & Ventura, 2013).

EDM may be used to create better and smarter learning technologies and to better inform both students and instructors. Educational data mining and analytics have enormous potential for identifying peoples learn-ability, forecasting learning, and studying real-world learning behavior (Baker, 2014). Implementing EDM and LA in educational settings still confronts several difficulties, such as a lack of a data-driven culture and rapid, thorough, and simple-to-understand technologies that might be incorporated into the most well-liked Learning Management Systems (LMSs) (Calvet Liñán & Juan Pérez, 2015).

EDM and LA can be used to anticipate, identify, diagnose, or monitor students. They can also be used to determine the dropout rate of students enrolled in distant education using a virtual learning environment (ANDRADE et al., 2021). The dropout rate of students in Virtual Learning Environments as well as in all types of Distance Education programmes can be determined using EDM, Learning Analytics, intelligent systems, and collaborative groups (da SILVA et al., 2022).

Learning analytics (LA) and educational data mining (EDM), which place different emphasis on influencing learning, student engagement, teaching tools, and social network analysis and knowledge discovery through educational data sources, techniques, and methods of data analysis, can assess issues and factors of student data and predict outcomes of student online learning based on various environments (Lemay et al., 2021). According to educational statistics, information and communication technology can be used in collaborative learning environments, gamified learning environments, distance education, and other forms of learning environments. A scientific knowledge foundation for learning analytics was created using the study's findings (Fonseca et al., 2021).

5. Student's performance prediction with various related factors

Table 2 provides broad support for the research on the prediction of student performance using Educational Data Mining by identifying the problems and goals of the various authors' studies, along with the algorithm that has been the basis for each author's study.

Table 2: Problem identified and algorithms used by authors

Authors	Problem	Algorithm / Method
(Merceron & Yacef, 2005)	Identifying trends in class data	Classification, Clustering
(Mostow et al., 2005)	Tutor over the internet who is unaware when a student leaves a website.	Relationship mining
(Romero & Ventura, 2007)	Online learning environments	Clustering, Classification, Outlier detection
(Cen et al., 2007)	Intelligent tutoring systems	Learning Factors Analysis (LFA), cognitive model
(Ramaswami & Bhaskaran, 2009)	Model of student performance	Naïve Bayes algorithm, Greedy search method
(Baker & Yacef, 2009)	Put the focus on prediction	Statistics and visualization
(Kotsiantis, 2009)	Anticipate dropout rates and dropout risk in students	Classification, Naive Bayes
(Ahmad, & Shamsuddin, 2010)	Identifying student's learning and behavior through key classifiers such as tree classifier, and classification rule through Weka.	Decision Tree, Classification Rule
(Toscher & Jahrer, 2010)	A student's ability for accurate response, based on previous outcomes	K Nearest Neighbor, Singular value decomposition
(Romero & Ventura, 2010)	Providing input to aid course creators, teachers, and administrators in making decisions is the goal.	Classification, Clustering, Association-rule mining
(Ramaswami & Bhaskaran, 2010)	Create a data mining predictive model for student performance	CHAID prediction model construction, classification tree algorithms,
(Baker & others, 2010)	What degree a pupil might manipulate the system	Prediction, Clustering, Relationship Mining, Density estimation

(Koedinger et al., 2010)	How Cognitive Science and Cognitive neuroscience beneficial in educational Environment. How it can show association with Educational Data Mining.	Classification
(Baradwaj & Pal, 2012)	Prediction of student enrollment in a specific course	Classification
(Scheuer & McLaren, 2012)	Using automated techniques to find patterns in massive volumes of educational data	Text mining, Exploratory analyses
(Hämäläinen & Vinni, 2011)	The ways in which data-driven classification has been used in the field of education	Decision trees, Bayesian classifiers, Neural networks, K-nearest neighbour, Support vector machines, Linear regression
(Kabakchieva, 2012)	Building data mining models to predict student performance	Rule learner, Decision tree classifier, Neural network, Nearest Neighbour classifier
(Siemens & Baker, 2012)	Tools for data mining and analysis to advance the domains of LA and EDM	-
(Gobert et al., 2012)	Data analysis using performance evaluation of inquiry skills and evidence-centered design	Table tool, Graph tool
(Abu Tair & El-Halees, 2012)	To enhance the performance of graduate students	Association, Classification, Clustering, Outlier detection rules
(Antonenko et al., 2012)	To comprehend how students use hyperlinked knowledge sources	Cluster analysis, A hierarchical clustering method (Ward's clustering), Non-hierarchical clustering method (k-Means clustering
(Calders & Pechenizkiy, 2012)	Data-driven decision-making to enhance present educational practice and learning material	Classication, Predictive modeling, Clustering, Biclustering, Frequent pattern mining, Emerging pattern mining,

		Collaborative altering and recommendations, Visual analytics, Process mining
(Chatti et al. , 2013)	Identifying challenges that are partly connected to educational data mining and learning analytics.	Classification, Clustering, Association Rule Mining, Social Network Analysis, Visualization, Learning Analytics
(Romero & Ventura, 2013)	Examining the current situation in the area of educational data mining, including its uses and techniques	Prediction, Clustering, Outlier detecting, Relationship mining, SNA, Process mining, Text mining
(Priya & Kumar, 2013)	Enhancing student performance	ID3 Algorithm, Data classification, Decision tree
(Winne et al., 2013)	Motivation, metacognition, and self-regulated learning problems for educational data mining	Self-Regulated Learning
(Jindal & Borah, 2013)	Focuses on the elements, research trends, related Tools, Techniques, and educational Outcomes of EDM	Classification, Statistics, Clustering, Prediction, Neural Network, Association Rule Mining, Web mining
(Huebner, 2013)	Create models to enhance institutional effectiveness and learning experiences	-
(Acharya & Sinha, 2014)	Early results prediction for students	Classification
(Fernández et al., 2014)	Cloud computing environment's appropriateness for educational data mining	-
(Bhegade, & Shinde, 2016)	Predicting student’s failure, dropout ratio and behavior.	C4.5 Decision Tree

Table 3 provides an understanding of the objectives, problems, and methodologies employed in papers that were published previously. These publications were used to forecast students' success. In these papers', the problem objective and outcome were proposed based on the issue that had been previously noted by many reviewers. However, it has been shown that Educational Data Mining Techniques are a widely employed strategy to address the issue of predicting student performance in educational environments.

Table 3: Problems identified, methods and outcome by authors

Authors	Problem / Objective	Method	Outcome
(Romero et al., 2014)	Difficulties with the pre-processing of academic data	-	PSLC DataShop offers the most data to students.
(Al-Razgan et al., 2014)	The purpose of this study is to provide research-related insights into the current state of EDM	-	The EDM research community focuses mainly on mobile learning, intelligent tutoring, and educational games.
(Papamitsiou & Economides, 2014)	Provides a summary of the empirical data supporting the major goals of the prospective use of LA/EDM in general educational strategy planning	Classification, Clustering, Regression (logistic/multiple)	This study aims to produce actionable advice for students' decision-making in the areas of science, technology, engineering, and mathematics (STEM).
(Ihantola et al., 2015)	How students use EDM and Learning Analytic to tackle programming difficulties	-	The main challenge is to integrate classroom results and practises into ongoing monitoring and improvement of the education.
(Dutt et al., 2017)	The benefits of clustering for education data mining	Clustering	This article's main subject is how students learn, taking into account a variety of factors like how much time is spent on particular learning tasks, how much time is spent studying in groups, how students behave in class, how the classroom is decorated, and how motivated students are to learn.
(Silva & Fonseca, 2017)	Gives a survey of the literature on learning analytics and educational data mining.	Clustering, Prediction, Relationship mining	This article stresses the value of educational data mining.

(Bakhshinategh et al., 2018)	Along with pure research goals, improving and boosting learning quality	Classification, Regression, Clustering, Association rule mining.	predicting student performance and achieving learning objectives
(Rodrigues et al., 2018)	Recognize patterns and keep an eye for potential paths for research, such as behavioural research, teamwork, interaction, and performance in the development of teaching-learning activities.	Clustering, Classication, Regression	Evaluation of educational practises and the traditional classroom in the context of online learning, as well as evaluation of student behaviour, performance, and action
(Aldowah et al., 2019)	Identification and comparison of EDM and LA approaches	Classification, Clusteing, Visual data mining, Outlier detection	The review's primary areas of interest are computer-supported learning analytics (CSLA), computer-supported predictive analytics (CSPA), computer-supported behavioral analytics (CSBA), and computer-supported visualization analytics (CSVA)
(Hernández-Blanco et al., 2019)	This study reviews the research on Deep Learning methods used for EDM from its inception to the present.	-	Deep Learning architectures can be created, put into use, and effectively applied to a variety of supervised and unsupervised activities.
(Romero & Ventura, 2020)	How educational data has been used for mining educational data and learning analytics	-	This paper's focus is on improving the techniques of learning analytics and data mining in education.
(Salloum et al., 2020)	To determine the most recent data mining trends in educational research and to assess the likelihood of	Sequential pattern, Clustering, Prediction, Classification,	EDM implementation is regarded as a future research area.

	applying machine learning in the field of education.	Machine learning models, Association rule analysis	
(Khan & Ghosh, 2021)	Give a thorough analysis of EDM research on learners' performance in the classroom. It emphasises identifying the predictors, the techniques utilised to do so, the timing, and the purpose of the prediction.	Text mining, Social network analysis, Support vector machine, Random forest	With a specific focus on research linked to student performance analysis and prediction, the purpose of this paper is to highlight the gaps or key difficulties.
(Aslam et al., 2021)	The goal of this research analysis is to distinguish between the possible outcomes of evaluating e-learning models using AI techniques like supervised, semi-supervised, and reinforced learning.	Clustering , Decision tree	How machine learning can improve e-learning?
(de Oliveira et al., 2021)	The paper's goal is to apply data mining to identifiable educational data.	Learning Analytics	Learning analytics enhance student engagement, give students a high-quality, personalized experience, and actually prevent students from dropping out of college.
(Baek & Doleck, 2021)	In this article, the two domains of learning analytics and educational data mining are distinguished in different ways.	Learning Analytics, Text Mining	This study focused solely on English-language academic papers and specifically identified the role that learning analytics and educational data mining will play in the development of future educational systems.
(Bachhal et al., 2021)	The goal is to comprehend data mining methods used by researchers in the past and	Classification, Clustering,	must comprehend the EDM's primary objectives, which are to predict students' performance

	contemporary data mining trends in educational studies.	Association, Social network analysis	and to promote education, science, and the development of current models.
(Sobral & Oliveira, 2021)	This article's goal is to examine scientific work that attempts to predict students' performance in introductory programming classes.	Decision tree, Linear regression, Artificial neural network, Support vector machines.	Decision trees, linear regression, artificial neural networks, and support vector machines can all be used to determine the answers to research questions.
(Sathe & Adamuthe, 2021)	Machine learning and data mining can be the two important things for better prediction of studentsperformance in every dimensions.	C5.0 J48, CART, NB, KNN, Random Forest, SVM, Multilayer Perceptron	According to research, C5.0 and Random Forest make more accurate predictions than other algorithms.

Additional examination of the papers' Data Set / Data Source, Tool utilized, and Research Outcome has also been covered in this article to assess the students' work. Table 4 is essentially used to discuss these terminologies.

Table 4: Dataset, and Tools by authors

Authors	**Data Set / Data Source**	**Tool used**	**outcome**
(Merceron & Yacef, 2005)	IDS	Tada-Ed, Excel, Access, SODAS, Logic-ITA	Web-based educational System
(Mostow et al., 2005)	Data Shop team's interviews	LISTEN	Web-based educational System
(Cen et al., 2007)	The study was conducted in a high school of Pittsburgh with 110 students from a total of 6 classes	ANCOVA	Knowledge tracing and cognitive mastery
(Ramaswami & Bhaskaran, 2009)	There were a total of 1969 higher secondary students from various	WEKA	Modeling student performance in relation to feature selection methods

	schools in various districts of the Indian state of Tamil Nadu		
(Baker & Yacef, 2009)	PSLC DataShop	WebCAT	Talk about the research's trends and changes
(Kotsiantis, 2009)	Problem with unbalanced datasets	-	Data mining integration in e-learning systems
(Toscher & Jahrer, 2010)	RMSE results from the combination of the two datasets	-	Data mining for education using cooperative filtering
(Romero & Ventura, 2010)	PSLC DataShop	ITS, LMS	To create studies that are more coherent and cooperative
(Ramaswami & Bhaskaran, 2010)	Students Information was acquired from the college and the Chief Educational Officer's office (CEO) from five distinct schools in three different districts of Tamilnadu, a total of 1000 datasets for the year 2006 were gathered.	CHAID Decision Tree	Building a CHAID prediction model
(Baker , 2010)	National Center for Education Statistics (NCES) and PSLC DataShop data sets	IRT models	Generalization of a prediction model that has been tested
(Baradwaj & Pal, 2012)	MCA (Master of Computer Applications) programme at VBS Purvanchal University in Jaunpur, Uttar Pradesh, provided the data set of 50 students used in this study from the 2007–2010 academic year.	-	Predict the student's performance at the semester's end
(Scheuer & McLaren, 2012)	PSLC Datashop	ITS	Student populations can also use the same student model parameters.
(Hämäläinen & Vinni, 2011)	Data were gathered from the Helsinki University and from multiple universities in Finland,	-	Compared the efficacy of several classification

	possibly over several years, the data sets were relatively extensive (500–20,000 rows, on average 7200 rows)		techniques for common educational data and issues
(Kabakchieva, 2012)	Data regarding students admitted to the university are included in the dataset utilized for the research from University of National and World Economy Sofia, Bulgaria	WEKA	Examined datasets for the predicted target variable in a different format (a nominal variable with five different values: Bad, Average, Good, Very Good, and Excellent).
(Siemens & Baker, 2012)		-	Both LA and EDM share a variety of traits as well as common goals and passions. But different technological, ideological, and methodological philosophies may be employed for the curriculum and instruction in educational institutions as well as in businesses.
(Gobert et al., 2012)		Microworlds Pogram	EDM approaches can record students' performance on inquiries and automatically grade them in a way that takes into account their complexity.
(Abu Tair & El-Halees, 2012)	Student's information gathered from Khanyounis College of Science and Technology	Local Outlier Factor (LOF)	-
(Antonenko et al., 2012)		ANOVA	Cluster analysis is a practical data mining technique for examining how OLEs learn.
(Calders & Pechenizkiy, 2012)	Pittsburgh Science of Learning Center (PSLC) DataShop	ITS	Using a web-based interface, study and visualize the data.
(Romero & Ventura, 2013)	-	ITS	Put emphasis on employing intelligent tutoring systems to predict students' performance.

(Priya& Kumar, 2013)	The data set was contributed by the first 50 students chosen for the experiment from the M.Sc. IT department's 2009 to 2012 class at Hindustan College of Arts and Science in Coimbatore.	ID3, C4.5	For predicting students' performance, ID3 and C4.5 are appropriate Decision Tree algorithms.
(Jindal & Borah, 2013)	PSLC Data shop	EPRules, GISMO, TADAED, O3R, Synergo/ColAT, LISTEN Mining tool, MINEL, LOCO, CIECoF, PDinamet, Meerkat, MMT tool are examples of EDM tools	A significant problem is the mining of heterogeneous data.
(Huebner, 2013)	-	-	There are very few studies that demonstrate enrolment and admission through EDM.
(Acharya & Sinha, 2014)	Students majoring in computer science at a few undergraduate colleges in Kolkata provided the data for this study.	WEKA	Combining multiple classifiers (CMC) may improve predictive performance.
(Fernández et al., 2014)	-	ITS	Improving e-learning prospects by integrating cloud computing with EDM processes

Following an examination of some of the published articles, it became apparent that a number of authors have employed classification, clustering, and other related approaches of various types as either a primary strategy or as a discussion of all these approaches for addressing student performance issues. In educational data mining, these approaches are especially helpful. The problem of forecasting student performance is one that is frequently addressed in the educational environment by educational data mining tools. The challenge of predicting student performance is shown in Table 5 along with several solutions and references.

Table 5: Methods used by different Authors

Methods	References
Classification	(Merceron & Yacef, 2005), (Romero & Ventura, 2007), (Kotsiantis, 2009), (Ramaswami & Bhaskaran, 2010), (Romero & Ventura, 2010), (Hämäläinen & Vinni, 2011), (Kabakchieva, 2012), (Abu Tair & El-Halees, 2012), (Baradwaj & Pal, 2012),(Jindal & Borah, 2013), (Priya& Kumar, 2013), (Romero & Ventura, 2013), (Acharya & Sinha, 2014), (Durairaj & Vijitha, 2014), (Ivančević et al., 2014), (Ocumpaugh et al., 2014), (Papamitsiou & Economides, 2014), (Buniyamin et al., 2015), (Guo et al., 2015), (Sin &Muthu, 2015), (Algarni, 2016), (Devasia et al., 2016), (Bakhshinategh et al., 2018), (Hussain et al., 2018), (Aldowah et al., 2019), (Injadat et al., 2020), (Mengash, 2020), (Prada et al., 2020), (Salloum et al., 2020), (Bachhal et al., 2021), (Salih&Khalaf, 2021), (Issah et al., 2023), (Alalawi et al., 2023), (Batool et al.,2023), (Parhizkar et al.,2023), (Hussain & Khan, 2023)
Clustering	(Mostow et al., 2005), (Merceron & Yacef, 2005), (Romero & Ventura, 2007), (Romero & Ventura, 2010), (Baker, 2010), (Scheuer & McLaren, 2012), (Abu Tair & El-Halees, 2012),(Antonenko et al., 2012), (Calders & Pechenizkiy, 2012), (Antonenko et al., 2012), (Romero & Ventura, 2013), (Jindal & Borah, 2013), (Peña-Ayala, 2014), (Durairaj & Vijitha, 2014), (Ocumpaugh et al., 2014), (Papamitsiou & Economides, 2014), (Sin &Muthu, 2015),(Dutt, 2015), (Algarni, 2016), (Karthikeyan & Kavipriya, 2017), (Asif et al., 2017), (Fernández & Luján-Mora, 2017), (Dutt et al., 2017), (Silva & Fonseca, 2017), (Bakhshinategh et al., 2018), (Rodrigues et al., 2018), (Ahuja et al., 2019), (Prada et al., 2020), (Salloum et al., 2020), (Nkomo& Nat, 2021), (Aslam et al., 2021), (Bachhal et al., 2021), (Araka et al., 2022), (Issah et al., 2023), (Alalawi et al., 2023), (Batool et al.,2023)
Decision Tree	(Merceron & Yacef, 2005), (Ramaswami & Bhaskaran, 2010), (Hämäläinen & Vinni, 2011), (Kabakchieva, 2012), (Abu Tair& El-Halees, 2012), (Priya& Kumar, 2013), (Acharya & Sinha, 2014), (Sin &Muthu, 2015), (Karthikeyan & Kavipriya, 2017), (Asif et al., 2017),(Fernandes et al., 2019), (Adekitan&Salau, 2019), (Hung et al., 2020), (Mengash, 2020), (Mehta et al., 2021), (Aslam et al., 2021), (Sobral& Oliveira, 2021), (Batool et al.,2023), (Parhizkar et al.,2023), (Hussain & Khan, 2023), (Xue & Niu, 2023)
Outlier Detection	(Romero & Ventura, 2007), (Abu Tair& El-Halees, 2012), (Romero & Ventura, 2013), (Aldowah et al., 2019), (Sáiz-Manzanares et al., 2021)
Naïve Bayes	(Devasia et al., 2016), (Salih & Khalaf, 2021), (Batool et al.,2023), (Pallathadka et al, 2023)
Association Rule	(Mostow et al., 2005), (Romero & Ventura, 2007), (Romero & Ventura, 2010), (Abu Tair& El-Halees, 2012), (Jindal & Borah, 2013), (Sin &Muthu, 2015), (Algarni, 2016), (Hussain et al., 2018), (Bakhshinategh et al., 2018), (Norm Lien et al., 2020), (Salloum et al., 2020), (Agarwal et al., 2021), (Bachhal et al., 2021)
K-nearest neighbor (k-NN)	(Hämäläinen&Vinni, 2011), (Kabakchieva, 2012), (Injadat et al., 2020), (El Aouifi et al., 2021), (Issah et al., 2023), (Alalawi et al., 2023), (Batool et al.,2023), (Hussain & Khan, 2023)

Support Vector Machine (SVM)	(Hämäläinen & Vinni, 2011), (Guo et al., 2015), (Jalota& Agrawal, 2019), (Injadat et al., 2020), (Mengash, 2020), (Salih &Khalaf, 2021), (Khan & Ghosh, 2021), (Sobral& Oliveira, 2021), (Mehta et al., 2021), (Issah et al., 2023), (Alalawi et al., 2023), (Batool et al.,2023), (Parhizkar et al.,2023), (Pallathadka et al, 2023)
Neural Network	(Hämäläinen&Vinni, 2011), (Kabakchieva, 2012), (Jindal & Borah, 2013), (Dutt, 2015), (Karthikeyan & Kavipriya, 2017), (Fernandes et al., 2019), (Mengash, 2020), (Godwin-Jones, 2021), (Sobral& Oliveira, 2021)
Relationship Mining	(Mostow et al., 2005), (Baker & others, 2010), (Romero & Ventura, 2013), (Algarni, 2016), (Silva & Fonseca, 2017)
Regression Analysis	(Hämäläinen & Vinni, 2011), (Ocumpaugh et al., 2014), (Ivančević et al., 2014), (Papamitsiou & Economides, 2014), (Jacob et al., 2015), (Sin &Muthu, 2015), (Algarni, 2016), (Bakhshinategh et al., 2018), (Rodrigues et al., 2018), (Adekitan & Salau, 2019), (Injadat et al., 2020), (Hung et al., 2020), (Mehta et al., 2021), (Young & Caballero, 2021), (Sobral& Oliveira, 2021), (Hussain & Khan, 2023)
Random Forest	(Zaffar et al., 2017), (Hussain et al., 2018), (Baker, 2019), (Adekitan & Salau, 2019), (Injadat et al., 2020), (Hung et al., 2020), (Kim et al., 2021), (Khan & Ghosh, 2021), (Issah et al., 2023), (Alalawi et al., 2023), (Batool et al.,2023), (Parhizkar et al.,2023), (Xue & Niu, 2023)
Social Network Analysis	(Romero & Ventura, 2010), (Fernández et al., 2014), (Lemay et al., 2021), (Khan & Ghosh, 2021), (Bachhal et al., 2021)

As per the above Table 5, the figure 4 shows, have been drawn for the visualization as non-categorical data on count plot, for better understanding of the EDM techniques used by various authors. Classification, and Clustering has been identified the most usable techniques of EDM as per study done.

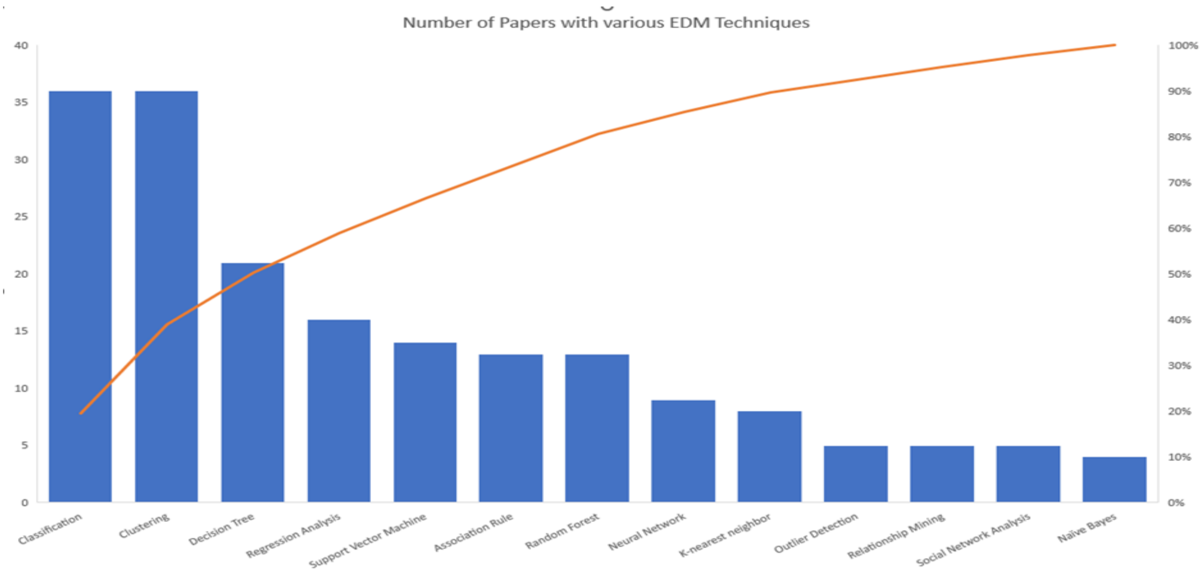
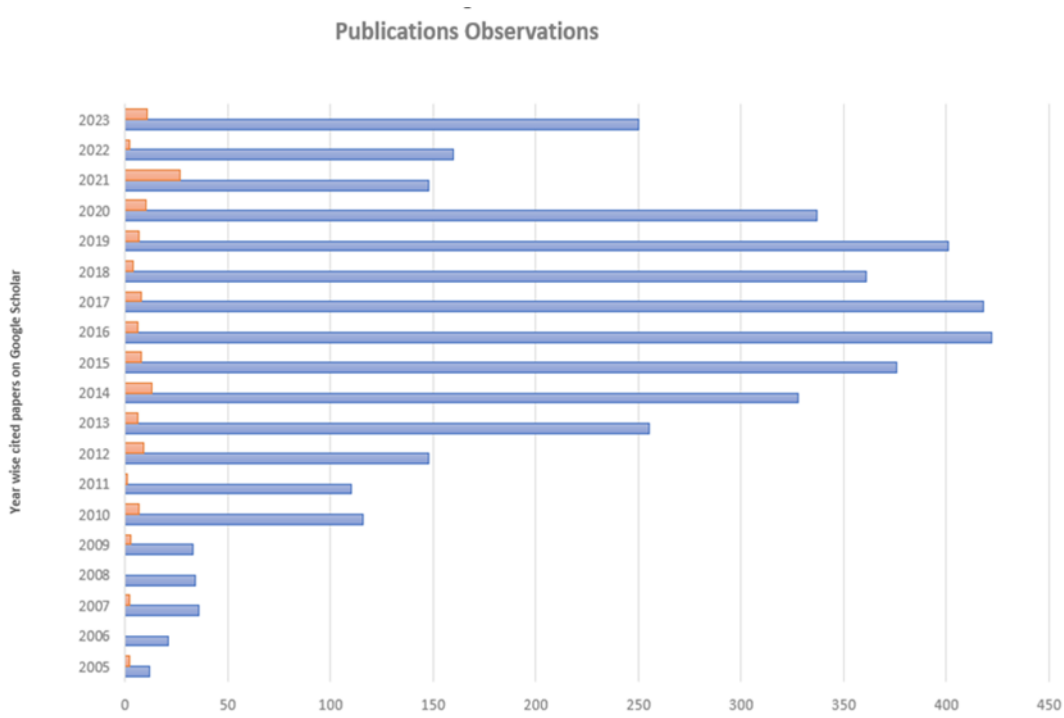


Figure 4: Study of papers with various EDM Techniques : Count Plot

Since, the majority of authors in this study used classification and clustering as well as linked related techniques and methodologies as their primary approaches for predicting the performance of the students. A total of 129 papers were included in this analysis, 10 of which supported data mining and other ideas discussed in the paper. Finally, total 126 papers were directly or indirectly related to educational data mining tools, related methodologies, or making predictions on their behalf. Some findings from the research on publications are presented in Figure 5's bar chart, which compares the 126 year-wise investigated articles with the 4507 total published articles on Google Scholar over the same period that have received at least one citation.



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Referenced Papers	2	0	2	0	3	7	1	9	6	13	8	6	8	4	7	10	27	2	11
Published Papers cited at least once	12	21	36	34	33	116	110	148	255	328	376	422	418	361	401	337	148	160	250

Figure 5: Annual Publication Bar chart for number of studies versus total articles

The created bar chart clearly shows that the range of educational data mining for pedagogical practices is continuously expanding over time. Future predictions of students' performance in a variety of subjects, including those pertaining to their grades, marks, percentages, job mapping, dropout rates, course and subject choices, behavioral projections, and programme selection, are best made using educational data mining approaches.

6. Aim of Prediction – Identified Research gap

Ultimate aim of prediction is to identify the Students Performance corresponding to the grades and behavior, and also identifying that the prediction of education should be based towards skill or should be based towards values, while the good approach may be to skill acquired must be guided by values. This is the gap present across all the research done till now, that’s the reason, agitation among students is increasing across the world. In prediction, education approach plays a very important role.

In this section, ongoing research and prediction earlier the skill-based education has been identified and prioritized more, due to that sustainability has become crucial in the organizations. Before prediction of Education through Educational Data Mining tools and techniques, it is necessary to discuss here the definition of “Education” and “Sanskar”. As per study research question generated today:

RQ: Is Role of Education for Holistic development and transformation to Human Consciousness?

The role of “Education-Sanskar” is to enable this transformation by way of ensuring the development of the competence to live with human consciousness and definite human conduct

For this, it has to ensure

1. Right understanding in every child
2. The capacity to live in relationship with the other human being
3. The capacity to identify the need of physical facility,

the skills and practice for sustainable production of more than what is required – leading to the feeling of prosperity (Gaur et al., 2010)

As per oxford definition “School” is “a place where children go to be educated” and “Education” is “a process of teaching, training and learning, especially in schools, colleges or universities, to improve knowledge and develop skills”

So, as per the oxford dictionary “Education” definition leads only towards the skill development and while we take the reference of Gaur et al., 2010 the “role of education” leads towards the holistic development of students as well as all Human being. So as per study oxford definition for education leads only towards the body centric development, while at another end Gaur et al., 2010 definition for education leads towards the self-centric development, so lead towards the harmony in Human being. Educational Data Mining can be useful for exploring and predicting such sustainable and holistic education based on (Gaur et al., 2010) and for prediction of behavior of students. Here values centric approach for sustainable education cum development also discussed.

The two approaches of education have been discussed here as:

6.1 Skill Based Education

Still, the education system whatever being followed in the world is related to maximum skill-based education, due to this system of education students only predict the courses according to seeing the research questions such as:

RQ1: How economical the course, subject, and program are?

RQ2: How much economics they will get after completion of course, subjects and Program?

RQ3: How quick relocation is possible in that field of learning for economic growth?

Even this skill-based learning sometimes leads towards the failure to sustain the students or learner (employee), even while this student or learner knows that moving him can lead towards the product failure, sometimes the organization failures. So, prediction of course, subject, and program by student is economic, skill centric nowadays. Definitely marks prediction can be possible in the system by applying the classification, clustering, and other related approaches of learning.

6.2 Values Based Education

(Gaur et al., 2010) said that all the universal human values parameters are universal, time invariant, rational, verifiable, and leading to harmony. Change in geographical locations has no impact on parameters and learning of Universal Human Values. So, if the students have to be educated in such a way, when core of learning in students is based on Universal Human Values than such type of students will work with more stability in the organizations, behave well, and will generate the harmonious environment for the society. Since Values directly related to the behavior of the students, so by applying or generating different psychological assessment batteries, the behavior of students also can be predicted. Since psychological assessment batteries will generate the parameters which will be categorical or numeric in nature so classification, clustering, and related techniques also can be applied for identifying the behavior of students, and right understanding can be ensured.

All the prediction of data of students today is in relation to skill acquired, while the aim of prediction also should be value based or more appropriately it should be skill guided by values.

7. Discussion and Open Issues

Following the Corona period, many educational institutions adopted online and software-based learning approaches, and online education has also developed into a great model towards continuing the education in multi-disciplinary fields of skill and value development among the students. This is evident from the implementation possibilities of Educational Data Mining for predicting students' performance around the world. The methods of classification, clustering, and aggregation are evolving and improving performance prediction for students. According to studies, a lot of research has been done on the contemporary educational system, but less has been done on the Indian educational system, which has been intertwined with value education throughout the world for many centuries.

The belief that the body is "Nashwar," or “Mortal”, the self is “Shashwat” or "Immortal" through reincarnation after the death of a human body, and that space is "Vyapak," or “All-pervasive”, has persisted throughout ancient Indian culture (Gaur et al., 2010).

Through the use of educational data mining methods and techniques, it is possible to investigate sustainable or holistic education and predict student performance in terms of both skill and value centric development. In the near future, educational institutions like schools, colleges, universities, and academies will be able to design programmes, courses, and subjects in this regard, and student performance will be predicted using EDM tools and techniques, where classification and clustering techniques may be useful for the prediction. This type of educational system can also be explored across the globe using educational data mining. Techniques for educational data mining can be useful for teaching not only the skill obtained by body, but also the value earned by Self.

According to research, Educational Data Mining can be used to map skill development and value centric development to helping students find jobs through campus placement, walk-in interviews, or off-campus placement. By matching up students with their skills, pre-placement processes on campus can produce positive outcomes that can be utilized to forecast the future of students' careers. Predicting student performance was a big challenge in both school and higher education during the Corona period for everyone.

The Data Mining technique called Time Series Analysis, find, most suited for predicting student performance over such a time period, and can also be used in such a challenging time. The performance of students in general always based on their attendance during semester, class test, sessional examinations, pre-university examination, by using a specific educational calendar followed by schools, Institutes, Academies, Educational Environments, Universities, as well as E-learning environments.

As per studies Social Cognitive Theory (SCT) and Social Cognitive Career Theory (SCCT) (Lent, R. W., 1994) can be utilized always on every form of learning. Observing that the range of research in the fields of science, technology, engineering, and mathematics (STEM) is expanding globally since it is a prerequisite in education for all nations' governments worldwide in schools, colleges, and universities. According to research, any STEM-related progress will only be sustained if it is based on universal human values (UHV). Therefore, educating students or society about UHV-STEM (Universal Human Values with STEM) is more advantageous, and this job may be accomplished relatively quickly using tools from Educational Data Mining.

Nationality may be one of the key criteria for understanding the human values internationally relating to families, civil, and spiritual. Somehow the role of religion in establishing Universal Human values can exist. So, values can be imparted to the pupils as per the countries, according to its cultural and traditional requirements (Kostina et al., 2015).

The priority among right understanding, relationship, and physical facility can be decided and can be teach to every human being, and by knowing this thing in right method the “existence is co-existence” can be comprehend by every human, and can produce a rational change to the whole humanity. Learning about universal human values and professional ethics can make it possible (Gaur et al., 2010).

Influences from gender, religion, caste, sects, culture, family background, parents and teachers, and the educational environment might be some of the factors that affect how well students succeed. Decide on the aforementioned fundamental issues, and then learn that Educational Data Mining techniques are appropriate for generating miracle answers to the mentioned difficulties and predictions.

Educational Data Mining tools and techniques are beneficial for implementing National Education Policy of India 2020 (NEP 2020), as e-University formation process in India is on trends nowadays, and also fulfill the Education4.0 objectives to the world, where automation is very easy through it. Educational Data Mining can be used to attract, maintain, predict the students to achieve the profitability of educational institutes, universities, schools, Olympiads. Educational Data mining techniques can assist and help organizations, trainers, resource persons, educators, and students to manage training and learning events, self-paced courses, body centric courses, self-centric courses, and blended learning programs / environments.

8. Conclusion and potential future work

For knowledge discovery, understanding, and determining which practice is beneficial, it is essential to analyze the student data. This paper is basically set up so that it demonstrates the approaches to Data mining and Educational Data mining, as well as the applicability of various algorithms and methods, tools, datasets, and data sources used by authors from around the world for a better understanding of educational data mining. Through the Classification, Clustering, and related methodology of Educational Data Mining, student’s performance can be predicted where Classification follows supervised learning approaches and clustering follow unsupervised learning approaches.

Even in the global epidemic, educational data mining can be used to identify and predict student performance, dropout rates, as well as teacher performance. It can assist teachers and students in monitoring academic progress in order to raise everyone's level of competency and improve the effectiveness and efficiency of the teaching and learning process. It can also assist students in choosing courses and in the management of their learning and behaviour. Combining machine learning techniques with educational data mining methods can result in models for students and instructors that are accurate and supported by evidence, as well as help make sense of data for future predictions of educational trends.

Prediction of EDM for various developments of Human being / students is explained in Table 6, more predictions may be made. A better resource recommendation system can be projected for the student's better performance in their studies and in their choice of job through the use of EDM tools, techniques, and modelling methodologies. Students' self-evaluation, self-awareness, self-reflection, self-learning, and self-exploration will undoubtedly increase as a result of the pattern developed through EDM approaches, which will ultimately result in better performance prediction. This performance will also raise the level of happiness and prosperity among students as well as in tutors, trainers, or educators.

This paper finally comes to the conclusion that sustainable or holistic education can be investigated through educational data mining techniques, and can lead to universal human values in the self, family, society, nature, and existence, and with this harmonious level of living of human being "existence is co-existence" can be better understood.

Table 6: Prediction of EDM for various developments of human being

1.	Are EDM methodologies helpful for investigating both value and skill-based education? **Observation:** Today's focus is primarily on using EDM tools and techniques to explore skill-based courses, which is essentially the goal of the present global education system. **Recommendation:** It is strongly advised to use the EDM tools and techniques for exploring not only skill-based courses across the world but also the value-based courses, since EDM provides the quickest

	way to learn about this education system globally. This is because the world has seen significant behavioral changes in human nature as a result of the current educational system.
2	Are EDM techniques helpful for exploring courses on the growth of Human Consciousness, or are they simply helpful for the development of body-centric development? **Observation:** Since this is a body-centric approach to study and its development, the emphasis is mostly on improving the courses related to the establishment of additional physical facilities and their use by everyone. **Recommendation:** Since self and body are often studied separately in Indian cultural education and since the self is immortal and value-based education is a significant educational system that has existed here since ancient times, it is advised that by using EDM tools and techniques, human consciousness development courses can be explored across the globe quickly, along with current skill-based and body-centric courses, subjects, and programmes for its development.
3	Are self-centric courses prediction can be achieved through EDM, or only it is approach for predicting body centric courses? **Observation:** The contemporary educational systems around the world, where people of various sects and religions have various preconditions, place more emphasis on gathering physical facilities for obtaining happiness, prosperity and sustainable development. The majority of the attention in the course selection is therefore on a body-centric approach. **Recommendation:** Since many people and societies around the world now adhere to the idea of Universal Human Values, self-centric courses are being prioritized around the globe. These courses place a strong emphasis on value-based education systems in addition to skill development. Therefore, EDM techniques may be used to forecast the self- and body-centric courses both.
4.	Is sustainable education can be predicted through EDM? Observation: In current education system physical facility, and body-based education / skill development is the prime focus, which can never be sustainable without having right understanding in human being as well human – human relation, its development, and without human value and ethics in students. So, this not at all leading to humanity as well as no sustainable development. Recommendation: Since Right understanding in human being, and relationship between human being having more priority for continuous happiness and prosperity, and it can lead to sustainable development also such education system can be explored and can predicted through EDM and it may be model of exploring Universal Human value in self and in education system. By using the UHV based system new educational system, or course, subjects, program, can be predicted.
5.	Is mapping of employability of students can be predicted through their performance in any courses through EDM techniques? Observation: Still this work is not done through EDM in the world as per study.

	Recommendation: Since student's performance prediction can be possible through EDM tools and techniques, this shows that we can also be done mapping of skills versus their job or employability in the future for further growth of students.

Finally, after addressing the entire study, mention that there is still room for future work in the area of educational data mining, including: 1. Mapping of student performance, value, and skill acquisition with career/job mapping. 2. Investigating education approach and identification of skill guided by values 3. Investigating UHV-STEM as a subject, course, and programme from the perspective of holistic education and development. 4. Using time series analysis as a data mining tool to forecast student performance during challenging periods, such as pandemics. 5. Teaching, recognizing, and forecasting students' academic and behavioral outcomes for the development of their life skills in areas like Universal Human Values online and offline. Techniques from educational data mining, such as classification and clustering, can be used to investigate all of this.

Conflict of Interest

The researcher declares that they have no conflict of interest.

Refrences

[1] Abu Tair, M. M., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. International Journal of Information, 2(2). http://hdl.handle.net/20.500.12358/25066 Corpus ID: 661468

[2] Acharya, A., & Sinha, D. (2014). Early Prediction of Students Performance using Machine Learning Techniques. International Journal of Computer Applications, 107(1), 37–43. https://doi.org/10.5120/18717-9939

[3] Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon, 5(2), e01250. https://doi.org/10.1016/j.heliyon.2019.e01250

[4] Agaoglu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. IEEE Access, 4, 2379–2387. https://doi.org/10.1109/ACCESS.2016.2568756

[5] Agarwal, A., Hasan, R., Naidu, V. R., Saqib, M., Srinivas, S., &Jesrani, K. (2021, January). Educational association mining on the use of media platforms for e-learning. In 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 309-314). IEEE. https://doi.org/10.1109/ICCAKM50778.2021.9357727

[6] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (pp. 207-216).

[7] Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. Proceedings of the Eleventh International Conference on Data Engineering, 3–14. https://doi.org/10.1109/icde.1995.380415

[8] Ahmad, N. B. H., & Shamsuddin, S. M. (2010). A comparative analysis of mining techniques for automatic detection of student's learning style. In 2010 10th International Conference on Intelligent Systems Design and Applications (pp. 877-882). IEEE. https://doi:10.1109/isda.2010.5687150

[9] Ahuja, R., Jha, A., Maurya, R., & Srivastava, R. (2019). Analysis of educational data mining. In Harmony Search and Nature Inspired Optimization Algorithms (pp. 897–907). Springer.https://doi.org/10.1007/978-981-13-0761-4_85

[10] Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. Engineering Reports, e12699.

[11] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

[12] Aleem, A., & Gore, M. M. (2020). Educational data mining methods: A survey. 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), 182–188.https://doi.org/10.1109/csnt48778.2020.9115734

[13] Algarni, A. (2016). Data mining in education. International Journal of Advanced Computer Science and Applications, 7(6).

[14] Alghamdi, A. S., & Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. Education Sciences, 13(3), 293.

[15] Al-Razgan, M., Al-Khalifa, A. S., & Al-Khalifa, H. S. (2014). Educational data mining: A systematic review of the published literature 2006-2013. Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), 711–719. https://doi.org/10.1007/978-981-4585-18-7_80

[16] ANDRADE, T. L. de, Rigo, S. J., & Barbosa, J. L. V. (2021). Active Methodology, Educational Data Mining and Learning Analytics: A Systematic Mapping Study. Informatics in Education, 20(2). https://doi.org/10.15388/infedu.2021.09

[17] Anoopkumar, M., & Rahman, A. M. J. M. Z. (2016). A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration. 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 122–133.https://doi.org/10.1109/sapience.2016.7684113

[18] Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 60(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

[19] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers \& Education, 113, 177–194.https://doi.org/10.1016/j.compedu.2017.05.007

[20] Aslam, S. M., Jilani, A. K., Sultana, J., & Almutairi, L. (2021). Feature evaluation of emerging e-learning systems using machine learning: An extensive survey. IEEE Access, 9, 69573–69587.https://doi.org/10.1109/access.2021.3077663

[21] Bachhal, P., Ahuja, S., & Gargrish, S. (2021). Educational Data Mining: A Review. Journal of Physics: Conference Series, 1950(1), 12022.https://doi.org/10.1088/1742-6596/1950/1/012022

[22] Baek, C., & Doleck, T. (2021). Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019. Interactive Learning Environments, 1–23.https://doi.org/10.1080/10494820.2021.1943689

[23] Baker, R. S. (2014). Educational data mining: An advance for intelligent systems in education. IEEE Intelligent Systems, 29(3), 78–82.https://doi.org/10.1109/mis.2014.42

[24] Baker, R. S., &Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of educational data mining, 1(1), 3-17. https://doi.org/10.5281/zenodo.3554657

[25] Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. Journal of Educational Data Mining, 11(1), 1-7. https://doi.org/10.5281/zenodo.3554745

[26] Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies, 23(1), 537–553.https://doi.org/10.1007/s10639-017-9616-z

[27] Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies, 28(1), 905-971.

[28] Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. ArXiv Preprint ArXiv:1201.3417. https://doi.org/10.14569/ijacsa.2011.020609

[29] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. Office of Educational Technology, US Department of Education. https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

[30] Bhegade, K. B., & Shinde, S. V. (2016). Student performance prediction system with educational data mining. International Journal of Computer Applications, 146(5), 32-35. https://doi.org/10.5120/ijca2016910704

[31] Bousbia, N., & Belamri, I. (2014). Which contribution does EDM provide to computer-based learning environments? In Educational data mining (pp. 3–28). Springer.https://doi.org/10.1007/978-3-319-02738-8_1

[32] Buniyamin, N., bin Mat, U., & Arshad, P. M. (2015, November). Educational data mining for prediction and classification of engineering students achievement. In 2015 IEEE 7th International Conference on Engineering Education (ICEED) (pp. 49-53). IEEE.

[33] Calders, T., & Pechenizkiy, M. (2012). Introduction to the special section on educational data mining. AcmSigkdd Explorations Newsletter, 13(2), 3–6. https://doi.org/10.1145/2207243.2207245

[34] CalvetLiñán, L., & Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. International Journal of Educational Technology in Higher Education, 12(3), 98–112. https://doi.org/10.7238/rusc.v12i3.2515

[35] Cen, H., Koedinger, K. R., & Junker, B. (2007). Is Over Practice Necessary?-Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining. Frontiers in Artificial Intelligence and Applications, 158, 511. https://dl.acm.org/doi/abs/10.5555/1563601.1563681

[36] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2013). A reference model for learning analytics. International Journal of Technology Enhanced Learning, 4(5-6), 318-331.https://doi.org/10.1504/ijtel.2012.051815

[37] Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104-126.https://doi.org/10.1007/bf03177550

[38] da SILVA, L. M., DIAS, L. P. S., BARBOSA, J. L. V, RIGO, S. J., dos ANJOS, J., GEYER, C. F. R., & LEITHARDT, V. R. Q. (2022). Learning Analytics and Collaborative Groups of Learners in Distance Education: A Systematic Mapping Study. Informatics in Education, 21(1).https://doi.org/10.15388/infedu.2022.05

[39] Daghestani, L. F., Ibrahim, L. F., Al-Towirgi, R. S., & Salman, H. A. (2020). Adapting gamified learning systems using educational data mining techniques. Computer Applications in Engineering Education, 28(3), 568–589.https://doi.org/10.1002/cae.22227

[40] de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How Does Learning Analytics Contribute to Prevent StudentsDropout in Higher Education: A Systematic Literature Review. Big Data and Cognitive Computing, 5(4), 64. https://doi.org/10.3390/bdcc5040064

[41] Devasia, T., Vinushree, T. P., & Hegde, V. (2016). Prediction of students performance using Educational Data Mining. 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 91–95.https://doi.org/10.1109/sapience.2016.7684167

[42] Durairaj, M., &Vijitha, C. (2014). Educational data mining for prediction of student performance using clustering algorithms. International Journal of Computer Science and Information Technologies, 5(4), 5987–5991. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.567.8824&rep=rep1&type=pdf Corpus ID: 17278167

[43] Durbin, J. (1984). Present Position and Potential Developments: Some Personal Views Time Series Analysis. Journal of the Royal Statistical Society: Series A (General), 147(2), 161-172.https://doi.org/10.2307/2981673

[44] Dutt, A. (2015). Clustering Algorithms Applied in Educational Data Mining. International Journal of Information and Electronics Engineering. https://doi.org/10.7763/ijiee.2015.v5.513

[45] Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. Ieee Access, 5, 15991–16005.https://doi.org/10.1109/access.2017.2654247

[46] El Aouifi, H., El Hajji, M., Es-Saady, Y., & Douzi, H. (2021). Predicting learners performance through video sequences viewing behavior analysis using educational data-mining. Education and Information Technologies, 26(5), 5799–5814.https://doi.org/10.1007/s10639-021-10512-4

[47] Feldman-Maggor, Y., Barhoom, S., Blonder, R., & Tuvi-Arad, I. (2021). Behind the scenes of educational data mining. Education and Information Technologies, 26(2), 1455–1470. https://doi.org/10.1007/s10639-020-10309-x

[48] Fernández, A., Peralta, D., Ben\'\itez, J. M., & Herrera, F. (2014). E-learning and educational data mining in cloud computing: an overview. International Journal of Learning Technology, 9(1), 25–52. https://doi.org/10.1504/ijlt.2014.062447

[49] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research, 94, 335–343.https://doi.org/10.1016/j.jbusres.2018.02.012

[50] Fonseca, D., Garc\'\ia-Peñalvo, F. J., & Camba, J. D. (2021). New methods and technologies for enhancing usability and accessibility of educational data. In Universal access in the information society (Vol. 20, Issue 3, pp. 421–427). Springer.https://doi.org/10.1007/s10209-020-00765-0

[51] Gaur, R.R.,Sangal, R., & Bagaria, G.P.(2010). A Foundation course in Human Values and Professional Ethics. 2010, Excel Books Private Limited.

[52] Gobert, J. D., Sao Pedro, M. A., Baker, R. Sj., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. Journal of Educational Data Mining, 4(1), 104–143. https://doi.org/10.5281/zenodo.3554646

[53] Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. https://www.semanticscholar.org/paper/Big-data-and-language-learning%3A-Opportunities-and-Godwin-Jones/8bf4ae96185d8ad84a420e30710af5dad89074b6 Corpus ID: 232108946

[54] Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting students performance in educational data mining. 2015 International Symposium on Educational Technology (ISET), 125–128.https://doi.org/10.1109/iset.2015.33

[55] Guruler, H., &Istanbullu, A. (2014). Modeling student performance in higher education using data mining. In Educational Data Mining (pp. 105–124). Springer.https://doi.org/10.1007/978-3-319-02738-8_4

[56] Hämäläinen, W., & Vinni, M. (2011). Classifiers for educational data mining. Handbook of Educational Data Mining, Chapman \& Hall/CRC Data Mining and Knowledge Discovery Series, 57–71.https://doi.org/10.1201/b10274-7

[57] Hantoobi, S., Wahdan, A., Al-Emran, M., & Shaalan, K. (2021). A review of learning analytics studies. Recent Advances in Technology Acceptance Models and Theories, 119–134.https://doi.org/10.1007/978-3-030-64987-6_8

[58] Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. Complexity, 2019.https://doi.org/10.1155/2019/1306039

[59] Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., &Stachl, C. (2021). Machine learning for the educational sciences. Review of Education, 9(3), e3310.https://doi.org/10.31234/osf.io/3hnr6

[60] Huebner, R. A. (2013). A Survey of Educational Data-Mining Research. Research in Higher Education Journal, 19. https://eric.ed.gov/?id=EJ1064637

[61] Hung, H.-C., Liu, I.-F., Liang, C.-T., & Su, Y.-S. (2020). Applying educational data mining to explore students learning patterns in the flipped learning approach for coding education. Symmetry, 12(2), 213.https://doi.org/10.3390/sym12020213

[62] Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students academic performance using WEKA. Indonesian Journal of Electrical Engineering and Computer Science, 9(2), 447–459.https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

[63] Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science, 10(3), 637-655.

[64] Hutt, S., Ocumpaugh, J., Ma, J., Andres, A. L., Bosch, N., Paquette, L., Biswas, G., & Baker, R. S. (2021). Investigating SMART Models of Self-Regulation and Their Impact on Learning. International Educational Data Mining Society. https://eric.ed.gov/?id=ED615568

[65] Ihantola, P., Vihavainen, A., Ahadi, A., Butler, M., Börstler, J., Edwards, S. H., Isohanni, E., Korhonen, A., Petersen, A., Rivers, K., & others. (2015). Educational data mining and learning analytics in programming: Literature review and case studies. Proceedings of the 2015 ITiCSE on Working Group Reports, 41–63. https://doi.org/10.1145/2858796.2858798

[66] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992. https://doi.org/10.1016/j.knosys.2020.105992

[67] Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decision Analytics Journal, 100204.

[68] Ivančević, V., Knežević, M., & Luković, I. (2014). A course exam scheduling approach based on data mining. Smart Digital Futures 2014, 262, 132.

[69] Jacob, J., Jha, K., Kotak, P., & Puthran, S. (2015). Educational data mining techniques and their applications. 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 1344–1348.https://doi.org/10.1109/icgciot.2015.7380675

[70] Jalota, C., & Agrawal, R. (2019). Analysis of educational data mining using classification. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 243–247. https://doi.org/10.1109/comitcon.2019.8862214

[71] Jindal, R., & Borah, M. D. (2013). A survey on educational data mining and research trends. International Journal of Database Management Systems, 5(3), 53.

[72] Kabakchieva, D. (2012). Student performance prediction by using data mining classification algorithms. International Journal of Computer Science and Management Research, 1(4), 686–690.

[73] Karthikeyan, K., & Kavipriya, P. (2017). On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 7(5), 935–941. https://doi.org/10.23956/ijarcsse/sv7i5/0348

[74] Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and Information Technologies, 26(1), 205–240. https://doi.org/10.1007/s10639-020-10230-3

[75] Khasanah, A. U. (2017, June). A comparative study to predict student's performance using educational data mining techniques. In IOP Conference Series: Materials Science and Engineering (Vol. 215, No. 1, p. 012036). IOP Publishing. https://doi.org/10.1088/1757-899x/215/1/012036

[76] Kim, K., Kim, H.-S., Shim, J., & Park, J. S. (2021). A study in the early prediction of ICT literacy ratings using sustainability in data mining techniques. Sustainability, 13(4), 2141. https://doi.org/10.3390/su13042141

[77] Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2010). The knowledge-learning-instruction (KLI) framework: Toward bridging the science-practice chasm to enhance robust student learning. Cognitive https://doi.org/10.1111/j.1551-6709.2012.01245.x

[78] Kostina, E., Kretova, L., Teleshova, R., Tsepkova, A., & Vezirov, T. (2015). Universal human values: Cross-cultural comparative analysis. Procedia-Social and Behavioral Sciences, 214, 1019–1028.https://doi.org/10.1016/j.sbspro.2015.11.696

[79] Kotsiantis, S. (2009). Educational data mining: a case study for predicting dropout-prone students. International Journal of Knowledge Engineering and Soft Data Paradigms, 1(2), 101–111.https://doi.org/10.1504/ijkesdp.2009.022718

[80] Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. Education and Information Technologies, 1-30.

[81] Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. Computers and Education: Artificial Intelligence, 2, 100016.https://doi.org/10.1016/j.caeai.2021.100016

[82] Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. Journal of vocational behavior, 45(1), 79-122.https://doi.org/10.1006/jvbe.1994.1027

[83] Liu, N., Zhang, B., Yan, J., Chen, Z., Liu, W., Bai, F., & Chien, L. (2005). Text representation: From vector to tensor. Fifth IEEE International Conference on Data Mining (ICDM'05), 4--pp.https://doi.org/10.1109/icdm.2005.144

[84] Mehta, M. H., Chauhan, N. C., & Gokhale, A. (2021). Predicting Institute Graduation Rate with Genetic Algorithm Assisted Regression for Education Data Mining. ICTACT Journal on Soft Computing, 11(2), 2266–2278. DOI : 10.21917/ijsc.2021.0324

[85] Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. IEEE Access, 8, 55462–55470.https://doi.org/10.1109/access.2020.2981905

[86] Merceron, A., & Yacef, K. (2005). Educational Data Mining: a Case Study. AIED, 467–474. https://www.semanticscholar.org/paper/Educational-Data-Mining%3A-a-Case-Study-Merceron-Yacef/9eb1b8cd6f7e379e505ac4f1ea5a95d16b8e259fCorpus ID: 10949661

[87] Moscoso-Zea, O., Luján-Mora, S., & others. (2016). Datawarehouse design for educational data mining. 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET), 1–6. https://doi.org/10.1109/ithet.2016.7760754

[88] Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., & Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell. Proceedings of the Workshop on Educational Data Mining, National Conference on Artificial Intelligence, 15–22. https://www.aaai.org/Library/Workshops/2005/ws05-02-003.php

[89] Munshi, M., Shrimali, T., & Gaur, S. (2023). An intelligent graph mining algorithm to analyze student performance in online learning. Education and Information Technologies, 28(6), 6667-6693.

[90] Nahar, K., Shova, B. I., Ria, T., Rashid, H. B., & Islam, A. H. M. (2021). Mining educational data to predict students performance. Education and Information Technologies, 26(5), 6051–6067. https://doi.org/10.1007/s10639-021-10575-3

[91] Nkomo, L. M., & Nat, M. (2021). Student Engagement Patterns in a Blended Learning Environment: an Educational Data Mining Approach. TechTrends, 65(5), 808–817.https://doi.org/10.1007/s11528-021-00638-0

[92] Norm Lien, Y. C., Wu, W. J., & Lu, Y. L. (2020). How Well Do Teachers Predict Students' Actions in Solving an Ill-Defined Problem in STEM Education: A Solution Using Sequential Pattern Mining. IEEE Access, 8, 134976–134986. https://doi.org/10.1109/ACCESS.2020.3010168

[93] Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. British Journal of Educational Technology, 45(3), 487-501.

[94] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings, 80, 3782-3785.

[95] Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology \& Society, 17(4), 49–64. ISSN : 1436-4522 (Electronic); 1176-3647 (Print) https://www.jstor.org/stable/jeductechsoci.17.4.49?seq=1

[96] Parhizkar, A., Tejeddin, G., & Khatibi, T. (2023). Student performance prediction using datamining classification algorithms: Evaluating generalizability of models from geographical aspect. Education and Information Technologies, 1-19.

[97] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications, 41(4), 1432–1462. https://doi.org/10.1016/j.eswa.2013.08.042

[98] Peña-Ayala, A., & Cárdenas, L. (2014). How educational data mining empowers state policies to reform education: The Mexican case study. In Educational Data Mining (pp. 65–101). Springer. https://doi.org/10.1007/978-3-319-02738-8_3

[99] Prada, M. A., Dominguez, M., Vicario, J. L., Alves, P. A. V., Barbu, M., Podpora, M., ... & Vilanova, R. (2020). Educational data mining for tutoring support in higher education: a web-based tool case study in engineering degrees. IEEE Access, 8, 212818-212836.

[100] Prajapati, J., Bhatt, M., & Prajapati, D. J. (2012). Detection and summarization of genuine review using visual data mining. Int. J. Comput. Appl, 975, 8887.https://doi.org/10.5120/6148-8522

[101] Priestley, M. B. (1980). State-dependent models: A general approach to non-linear time series analysis. Journal of Time Series Analysis, 1(1), 47–71. https://doi.org/10.1111/j.1467-9892.1980.tb00300.x

[102] Priya, K. S., & Kumar, A. V. S. (2013). Improving the student's performance using educational data mining. International Journal of Advanced Networking and Applications, 4(4), 1806. https://www.semanticscholar.org/paper/Improving-the-Student%27s-Performance-Using-Data-Priya-Kumar/47edbb034f69000abc8431406a24251621a9218dCorpus ID: 16834716

[103] Ramaswami, M., & Bhaskaran, R. (2009). A study on feature selection techniques in educational data mining. ArXiv Preprint ArXiv:0912.3924.https://doi.org/10.48550/arXiv.0912.3924

[104] Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. ArXiv Preprint ArXiv:1002.1144.https://doi.org/10.48550/arXiv.1002.1144

[105] Rodrigues, M. W., Isotani, S., & Zarate, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. Telematics and Informatics, 35(6), 1701–1717. https://doi.org/10.1016/j.tele.2018.04.015

[106] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

[107] Romero, C., & Ventura, S. (2010). 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.

[108] Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12–27.https://doi.org/10.1002/widm.1075

[109] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

[110] Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. In Educational data mining (pp. 29–64). Springer. https://doi.org/10.1007/978-3-319-02738-8_2

[111] Roslan, M. H. B., & Chen, C. J. (2023). Predicting students' performance in English and Mathematics using data mining techniques. Education and Information Technologies, 28(2), 1427-1453.

[112] Sáiz-Manzanares, M. C., Rodr\'\iguez-D\'\iez, J. J., D\'\iez-Pastor, J. F., Rodr\'\iguez-Arribas, S., Marticorena-Sánchez, R., & Ji, Y. P. (2021). Monitoring of student learning in learning management systems: An application of educational data mining techniques. Applied Sciences, 11(6), 2677. https://doi.org/10.3390/app11062677

[113] Salih, N. Z., & Khalaf, W. (2021). Prediction of studentâ€™ s performance through educational data mining techniques. Indonesian Journal of Electrical Engineering and Computer Science, 22(3), 1708-1715. https://doi.org/10.11591/ijeecs.v22.i3.pp1708-1715

[114] Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Mining in educational data: review and future directions. The International Conference on Artificial Intelligence and Computer Vision, 92–102.https://doi.org/10.1007/978-3-030-44289-7_9

[115] Sathe, M. T., & Adamuthe, A. C. (2021). Comparative Study of Supervised Algorithms for Prediction of Students' Performance. International Journal of Modern Education & Computer Science, 13(1). https://doi.org/10.5815/ijmecs.2021.01.01

[116] Sawyer, R., Rowe, J., Azevedo, R., & Lester, J. (2018). Filtered Time Series Analyses of Student Problem-Solving Behaviors in Game-Based Learning. International Educational Data Mining Society. https://eric.ed.gov/?id=ED593118 Corpus ID: 49188485

[117] Scheuer, O., & McLaren, B. M. (2012). Educational data mining. Encyclopedia of the Sciences of Learning, 13, 10-1007.

[118] Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: towards communication and collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 252–254.https://doi.org/10.1145/2330601.2330661

[119] Silva, C., & Fonseca, J. (2017). Educational data mining: a literature review. Europe and MENA Cooperation Advances in Information and Communication Technologies, 87–94.https://doi.org/10.1007/978-3-319-46568-5_9

[120] Sin, K., & Muthu, L. (2015). "Application of Big Data in Education Data Mining and Learning Analytics – a Literature Review ". ICTACT Journal on Soft Computing, 05(04), 1035–1049. https://doi.org/10.21917/ijsc.2015.0145

[121] Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. Journal of Educational and Behavioral Statistics, 42(1), 85–106. https://doi.org/10.3102/1076998616666808

[122] Sobral, S. R., & Oliveira, C. F. de. (2021). Predicting students performance in introductory programming courses: a literature review.https://doi.org/10.21125/inted.2021.1485

[123] Su, Y.-S., & Lai, C.-F. (2021). Applying educational data mining to explore viewing behaviors and performance with flipped classrooms on the social media platform Facebook. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.653018

[124] Thakar, P., Mehta, A., & others. (2015). Performance analysis and prediction in educational data mining: A research travelogue. ArXiv Preprint ArXiv:1509.05176. https://doi.org/10.48550/arXiv.1509.05176

[125] Toscher, A., & Jahrer, M. (2010). Collaborative filtering applied to educational data mining. KDD Cup. https://www.semanticscholar.org/paper/Collaborative-Filtering-Applied-to-Educational-Data-Jahrer/e4d8a9bab5c68f5405c44764dcd2c2eba99af670Corpus ID: 9248221

[126] Winne, P. H., Baker, R. Sj., & others. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. Journal of Educational Data Mining, 5(1), 1–8.https://doi.org/10.5281/zenodo.3554620

[127] Xue, H., & Niu, Y. (2023). Multi-Output Based Hybrid Integrated Models for Student Performance Prediction. Applied Sciences, 13(9), 5384. https://doi.org/10.3390/app13095384

[128] Young, N. T., & Caballero, M. D. (2021). Predictive and explanatory models might miss informative features in educational data. ArXiv Preprint ArXiv:2103.14513.https://doi.org/10.48550/arXiv.2103.14513

[129] Zhang, D., Lee, K., & Lee, I. (2018). Hierarchical trajectory clustering for spatio-temporal periodic pattern mining. Expert Systems with Applications, 92, 1–11.https://doi.org/10.1016/j.eswa.2017.09.040