IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Educational Data Mining Techniques for Personalized Learning and Student Performance Prediction

Ram Jaysing Salunke[1] and Dr. Rajeev Kumar[2]
[1]Research Scholar, Department of Mathematics
[2]Associate Professor, Department of Mathematics
Vikrant University, Gwalior, M.P

Abstract: *Educational Data Mining (EDM) has emerged as an important research area for transforming large volumes of educational data into meaningful information for improving teaching and learning processes. The increasing use of learning management systems, online examinations, digital classrooms, student information systems, and e-learning platforms generates extensive data related to students' academic activities, learning behaviour, engagement, attendance, assessment results, and interaction patterns. EDM techniques can analyse these data to identify hidden patterns and develop predictive models capable of estimating student performance and supporting personalized learning.*
This research paper examines the major Educational Data Mining techniques used for personalized learning and student performance prediction, with particular emphasis on classification, clustering, regression, association-rule mining, decision trees, random forests, support vector machines, k-nearest neighbours, artificial neural networks, and ensemble learning approaches. The study discusses how these techniques can be applied to identify students with different learning needs, predict academic outcomes, recommend suitable learning resources, and provide early interventions for academically at-risk students. A comparative analysis is presented to examine the advantages, limitations, and educational applications of different EDM techniques. The paper further proposes an integrated framework in which educational data are collected, pre-processed, analysed, and transformed into personalized learning recommendations.

Keywords: Educational Data Mining, Personalized Learning, Predictive Analytics, Learning Analytics, Classification, Clustering, Academic Achievement

I. INTRODUCTION

The rapid development of digital technologies has fundamentally changed the way educational institutions collect, store, and utilize information about learners. Modern educational environments generate large quantities of data through student registration systems, attendance records, examination results, learning management systems, online quizzes, digital assignments, discussion forums, and educational applications. Although these data contain valuable information about student behaviour and academic progress, conventional methods of educational analysis are often unable to identify complex relationships within large datasets. Educational Data Mining provides a systematic approach for discovering useful patterns and knowledge from educational data.

Educational Data Mining can be understood as the application of data mining, statistics, machine learning, and computational techniques to educational datasets. Its objective is not simply to collect information but to transform educational data into actionable knowledge. Such knowledge can assist teachers in understanding learner behaviour, enable institutions to identify students who may require additional support, and help learners receive educational resources that correspond to their individual needs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Personalized learning represents one of the most important applications of EDM. Traditional education frequently follows a relatively uniform instructional approach in which students receive similar content, activities, and assessments. However, learners differ considerably in terms of prior knowledge, learning speed, interests, motivation, academic ability, and preferred methods of learning. EDM techniques can analyse these differences and support the development of individualized learning pathways. For example, clustering techniques can group students according to similar learning behaviours, while classification algorithms can predict whether a student is likely to achieve a particular performance level.

Student performance prediction is another significant application. Predictive models can use historical academic records, attendance, assessment scores, engagement indicators, and interaction patterns to estimate future academic outcomes. Early prediction can allow teachers and institutions to intervene before academic difficulties become severe. Instead of waiting until the end of a semester to identify poor performance, educators can use predictive information to provide additional learning materials, mentoring, remedial instruction, or counselling.

The effectiveness of EDM depends on the quality and relevance of educational data. Data must generally be cleaned, transformed, and prepared before applying mining algorithms. Missing values, inconsistent records, duplicated observations, and irrelevant attributes can negatively affect predictive performance. Feature selection is therefore an important stage in the EDM process because it helps identify variables that have meaningful relationships with student outcomes.

EDUCATIONAL DATA MINING TECHNIQUES FOR PERSONALIZED LEARNING

Different EDM techniques provide different forms of educational insight. Classification is particularly useful when the objective is to assign students to predefined categories such as high, medium, or low performance. Decision trees, random forests, support vector machines, naïve Bayes, k-nearest neighbours, and neural networks can be employed for this purpose.

Clustering is an unsupervised technique that identifies naturally occurring groups within educational data. Unlike classification, it does not require predefined categories. Students may be grouped according to learning behaviour, attendance, assessment patterns, participation, or interaction with online resources. These groups can subsequently be used to develop differentiated instructional strategies.

Regression techniques are useful for predicting continuous outcomes such as examination scores, grade-point averages, or expected achievement levels. Association-rule mining identifies relationships between different educational activities. For instance, patterns may reveal that students who regularly access particular learning resources and complete practice exercises tend to achieve better assessment results.

Table 1. Major Educational Data Mining Techniques

EDM Technique	Main Purpose	Educational Application	Major Advantage	Limitation
Decision Tree	Classification	Predicting performance categories	Easy to interpret	Can overfit
Random Forest	Classification/Prediction	Identifying at-risk students	Strong predictive capability	Less interpretable
Support Vector Machine	Classification	Student performance classification	Effective with complex data	Requires parameter selection
K-Nearest Neighbour	Classification	Learner similarity analysis	Simple implementation	Sensitive to feature scaling
Neural Network	Prediction	Academic performance prediction	Captures complex relationships	Requires more data and computation

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 5, July 2026 Impact Factor: 8.2

Clustering	Group formation	Creating learner profiles	Discovers hidden groups	Results may be difficult to interpret
Regression	Numerical prediction	Predicting grades or scores	Suitable for continuous outcomes	Assumptions may affect results
Association Rules	Relationship discovery	Finding learning behaviour patterns	Reveals useful associations	May produce many irrelevant rules

Personalized learning can be developed by combining several EDM techniques. First, clustering can identify groups of learners with similar characteristics. Second, classification or prediction models can estimate the likely performance of individual learners. Third, recommendation mechanisms can use these results to suggest appropriate learning resources. Such an approach transforms EDM from a purely analytical activity into a learner-support mechanism.

For example, students demonstrating low engagement and declining assessment performance may be classified as academically at risk. The system can subsequently recommend additional practice exercises, instructional videos, revision materials, or teacher support. High-performing students may instead receive advanced learning activities designed to extend their knowledge.

STUDENT PERFORMANCE PREDICTION USING EDM

Student performance prediction is a major area of EDM research because early identification of academic difficulties can support timely intervention. Prediction models generally use historical student information to estimate future academic outcomes. Variables may include previous examination marks, attendance, assignment completion, participation, learning-resource usage, demographic characteristics, and online activity.

The general EDM prediction process involves data collection, preprocessing, feature selection, model development, model training, evaluation, and interpretation. During data collection, relevant educational information is obtained from institutional databases and learning platforms. Preprocessing removes errors and prepares the dataset for analysis. Feature selection then identifies the most informative variables.

Classification models can predict categories such as pass/fail or high/medium/low achievement. Regression models can predict numerical scores. Ensemble techniques combine several predictive models and can improve robustness in complex datasets.

Table 2. Example Variables for Student Performance Prediction

Variable Category	Example Variables	Expected Relevance
Academic	Previous grades, test scores, GPA	Very High
Attendance	Attendance percentage, absence frequency	High
Engagement	Login frequency, activity completion	High
Assessment	Assignment marks, quiz scores	Very High
Learning Behaviour	Resource access, study time	High
Participation	Discussion activity, classroom participation	Moderate
Support	Tutoring, mentoring, remedial activities	Moderate
Demographic	Age, educational background	Variable

Model evaluation is essential because a highly accurate model on one dataset may not perform equally well on another. Common evaluation measures include accuracy, precision, recall, F1-score, mean absolute error, and root mean square error, depending on whether the problem is classification or regression.

Early-warning systems represent an important practical application of performance prediction. Such systems continuously analyse student information and generate alerts when performance indicators suggest possible academic difficulty. Teachers can then investigate the underlying reasons and provide appropriate support.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

However, prediction should not be interpreted as an unquestionable judgment about a learner's ability. Educational outcomes are influenced by many factors that may not be represented in the available dataset. Predictive models should therefore support teacher decision-making rather than replace professional educational judgment.

COMPARATIVE ANALYSIS OF EDM MODELS

The selection of an EDM technique depends on the nature of the educational problem, characteristics of the dataset, required interpretability, computational resources, and desired prediction accuracy. Simple algorithms such as decision trees have an important advantage because their predictions can be explained relatively easily to teachers and administrators. More complex models, such as neural networks and ensemble methods, may capture sophisticated relationships but can be more difficult to interpret.

Table 3. Comparative Evaluation of EDM Approaches

Model	Interpretability	Prediction Capability	Data Requirement	Computational Complexity	Suitable Application
Decision Tree	High	Good	Moderate	Low	Performance classification
Random Forest	Moderate	Very Good	Moderate/High	Moderate	Risk prediction
SVM	Moderate/Low	Very Good	Moderate	Moderate/High	Complex classification
KNN	Moderate	Good	Moderate	Moderate	Similar learner identification
Neural Network	Low	Excellent for complex patterns	High	High	Advanced prediction
Clustering	Moderate	Exploratory	Moderate	Moderate	Learner profiling
Regression	High	Good for numerical outcomes	Moderate	Low/Moderate	Grade prediction
Association Rules	High	Pattern-based	Moderate/High	Moderate	Behaviour analysis

The comparative analysis indicates that no single EDM algorithm is universally optimal. Decision trees are valuable when interpretability is a priority. Random forests are useful where predictive robustness is more important. Neural networks can be effective for complex nonlinear relationships, while clustering is particularly appropriate when learner groups are not known beforehand.

A hybrid approach can provide stronger educational support by combining multiple methods. For example, clustering may first identify learner profiles, followed by a classification model for predicting performance. A recommendation component can then provide individualized learning resources. This multi-stage approach reflects the complexity of real educational environments more effectively than reliance on a single algorithm.

PROPOSED FRAMEWORK AND DISCUSSION

A personalized learning framework based on EDM can be organized into five major stages: data acquisition, data preprocessing, predictive analysis, learner profiling, and personalized intervention. Educational data may be collected from student information systems, learning management platforms, examinations, assignments, attendance records, and digital learning environments.

In the preprocessing stage, incomplete and inconsistent information is handled, duplicate records are removed, and relevant variables are transformed into an appropriate analytical format. Feature selection can reduce unnecessary variables and improve model efficiency.

The next stage involves applying EDM algorithms. Classification models can predict performance levels, regression models can estimate academic scores, and clustering methods can identify learner groups. The resulting information can be incorporated into a personalized recommendation mechanism.

Table 4. Proposed EDM-Based Personalized Learning Framework

Stage	Process	Output
1	Educational Data Collection	Student dataset
2	Data Cleaning and Preprocessing	Quality-controlled dataset
3	Feature Selection	Relevant educational attributes
4	EDM Modelling	Prediction and learner profiles
5	Performance Analysis	Risk and achievement indicators
6	Personalization	Individual learning pathway
7	Intervention	Targeted educational support
8	Feedback	Updated learner information

The framework should operate continuously rather than as a one-time analytical exercise. Student behaviour changes over time, and therefore predictions should be updated when new assessment and engagement data become available. Continuous feedback also enables the system to evaluate whether recommended interventions are actually improving learning outcomes.

Despite its potential, several challenges must be considered. Educational datasets may contain missing or inaccurate information. Privacy is another major concern because student data can contain sensitive academic and behavioural information. Institutions must establish appropriate data-governance practices and restrict access to authorized users. Algorithmic bias must also be monitored because predictive systems can unintentionally reproduce inequalities present in historical data.

Interpretability is equally important. Teachers are more likely to trust predictive recommendations when they understand why a system has identified a student as potentially at risk. Therefore, explainable prediction mechanisms should accompany automated recommendations wherever possible.

II. CONCLUSION

Educational Data Mining offers significant opportunities for improving educational decision-making, personalized learning, and student performance prediction. By transforming educational data into meaningful patterns and predictions, EDM can help educators understand learner behaviour, identify academic difficulties at an early stage, and provide individualized learning support. Classification, clustering, regression, association-rule mining, decision trees, random forests, support vector machines, and neural networks each provide valuable capabilities for different educational applications.

The comparative discussion demonstrates that algorithm selection should depend on the specific educational objective rather than simply selecting the most complex model. Interpretable techniques remain important for teacher-oriented applications, whereas advanced ensemble and neural approaches may provide stronger predictive performance when sufficient data are available. Combining multiple techniques can provide a more comprehensive framework for personalized education.

Future research can focus on developing real-time EDM systems that integrate data from multiple educational platforms. Adaptive learning environments can increasingly use continuous student interaction data to modify learning content according to individual progress. Explainable artificial intelligence can further improve the transparency of predictions and recommendations. Future studies should also investigate privacy-preserving EDM, fairness-aware algorithms, longitudinal student modelling, multimodal learning data, and hybrid prediction systems.

Overall, Educational Data Mining should be regarded as a decision-support technology rather than a replacement for teachers. Its greatest value lies in combining computational intelligence with educational expertise. When implemented responsibly, EDM can contribute to more responsive, evidence-based, and learner-centred educational environments while supporting improved student performance and reducing the risk of academic failure.

REFERENCES

1. Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135–146.
2. 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.
3. Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery, 3*(1), 12–27.
4. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education, 113,* 177–194.
5. Khasanah, A. U., & Harwati. (2019). Educational data mining techniques approach to predict student's performance. *International Journal of Information and Education Technology, 9*(2), 115–118.
6. Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education, 143*, 103676.
7. Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings, 47*(15), 5260–5267.
8. Trakunphutthirak, R., & Lee, V. C. S. (2022). Application of educational data mining approach for student academic performance prediction using progressive temporal data. *Journal of Educational Computing Research, 60*(3), 742–776.
9. Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*, Article 11.
10. 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.
11. Kord, A., Aboelfetouh, A., & Shohieb, S. M. (2026). Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques. *Journal of Computing in Higher Education, 38*, 38–76

DOI: 10.48175/568