

The Impact of Machine Learning-Based Data Mining Techniques to Enhance Higher Educational Outcomes

Shilpa K

Research Scholar and Associate Professor
Institute of Computer Science and Information Science
Srinivas University, Mangalore
Nrupathunga University, NT road Bengaluru
shilpakr@gmail.com

Abstract – Purpose: Student performance is a key factor in addressing educational issues and evaluating learning outcomes. Educational Data Mining (EDM) has emerged to leverage data knowledge for enhancing educational systems. **Objectives:** This study aims to improve the predictive accuracy of academic success using eight popular machine learning techniques. It focuses on developing methods for analyzing data from learning environments to enhance student experiences and provide precise insights. **Methodology/Design/Approach:** The study utilizes EDM to analyze data from platforms like Learning Management Systems (LMS), Massive Open Online Courses (MOOC), and Intelligent Tutoring Systems (ITS). Eight machine learning techniques, including k-nearest neighbors (KNN), support vector machine (SVM), and random forest (RF), are employed. Performance indicators such as recall, accuracy, precision, and F-measure evaluate the prediction model's effectiveness. **Findings/Result:** EDM aids in automated grading, recommendations, and adaptive systems, significantly improving student performance. Predicting students' grades helps identify at-risk individuals, allowing timely interventions that benefit parents, educators, and students. **Originality/Value:** This study underscores the significance of EDM in extracting meaningful insights from academic data and improving educational outcomes through advanced predictive models.

Keywords – Student performance, Educational Data Mining (EDM), Learning Management Systems (LMS), Massive Open Online Courses (MOOC), Intelligent Tutoring Systems (ITS), Machine Learning, Predictive Accuracy, Educational Outcomes.

I. INTRODUCTION

The combination of "big data" with technological advancements, particularly in the form of expanded instructional media, has the potential to revolutionize learning processes both within and outside of traditional education. The need to use data mining technologies to assist students in data analysis and the production of insightful insights that can enhance their learning is what is driving this development. The vast amounts of data generated by teaching and learning activities have been analyzed, making Educational Data Mining (EDM) a vital tool (Nguyen, et al. (2020) [1]). Teachers, students, and educational administrators can gain a comprehensive understanding of students' learning progress and make fact-based decisions by analyzing these large datasets. Recommendations and interventions tailored to each child can be created based on these decisions. The field of educational data mining (EDM) describes itself as one that is concerned with developing methods for analyzing data from educational environments. These methods are used to improve students' understanding of themselves and their learning environments. The major objective of EDM is to develop and improve methods for analyzing educational data, which often has multiple layers of substantial structure, in order to obtain new insights into how students learn. This has significantly advanced the subject of learning sciences by making it easier for scholars to explore learning theories (Baker, et al. (2010) [2]).

Owing to its unique characteristics, educational data is sometimes unsuitable for conventional data mining (DM) methods. As a result, EDM requires certain DM strategies

meant to handle educational concerns. Numerous disciplines, including computer modeling, machine learning, psychometrics, and visualization (such as infographics), are the origins of these techniques. Despite the fact that association analysis, classification, and clustering are common DM techniques used in EDM, educational systems need unique approaches to address specific problems. Prediction, which forecasts future events based on historical data, is a critical component of EDM (Romero, et al. (2020) [3]). One area where it is particularly useful to teachers is in identifying students who are likely to succeed, need remediation, or are at a higher risk of failing and dropping out. One common prediction method in EDM is linear regression, which takes one or more categorical or continuous input variables and uses them to estimate continuous values (Alshehri, et al. (2020) [4]).

The process of putting students in groups based on similar performance patterns is called clustering. It can be applied to several forms of educational data sources, including class discussions, student assessments, interview data, and standardized test results. The main purpose of clustering is to identify student groups who perform or behave similarly, such as those who are very likely to drop out of school or do exceptionally well on impending tests (Romero, et al. (2013) [5]). One of the many possible advantages of combining big data with EDM in education is the improvement of learning outcomes. By using advanced data mining technologies, educators can gain deeper insights into student behaviors and learning processes, enabling them to make well-informed decisions that will improve students' accomplishment. Because educational data is unique, the EDM community has developed and refined

specific DM algorithms (Prakash et al. (2021) [7]). These techniques aid in the development of a more efficient and individualized learning environment by forecasting student performance and spotting patterns and connections in educational data.

A few applications for EDM include determining the learning requirements of certain student groups, identifying at-risk pupils to support early intervention, and implementing strategies to increase graduation rates. It also provides assessments of how well educational institutions are doing, makes it easier to manage campus resources efficiently, and promotes continuous course development.

EDM Process

The preprocessing, data mining, and post-processing stages of the EDM process are the same as those of the DM process. An essential aspect of data mining (DM) is the process of comprehending and interpreting raw data. The data mining process encompasses several distinct procedures for extracting knowledge: Data selection refers to the act of choosing a preprocessed dataset for further examination. Data pre-processing involves the identification and removal of missing values and irrelevant information in order to enhance the reliability of the data (Prakash et al. (2021) [12]). Data transformation involves the utilization of clustering and classification techniques to categorize data in preparation for data mining. Data mining is a process that involves using techniques such as clustering, regression, and classification to extract valuable patterns from data (Papadogiannis, et al. (2020) [14]). Pattern evaluation is the process of identifying and evaluating specific patterns to achieve certain objectives. Knowledge representation is the visual display of the acquired knowledge, which helps users interpret the findings and gain a comprehensive understanding. Figure 1 depicts the sequential steps involved in the process of data mining with the purpose of extracting valuable knowledge.

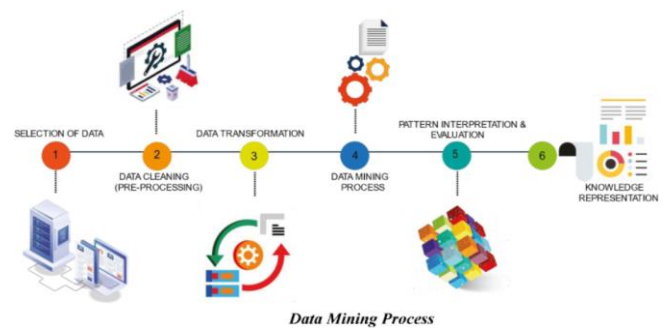


Figure 1. Steps of the data mining process for extracting knowledge (Yaw et al. 2023).

II. RELATED WORK

The rapid advancement of machine learning and data mining tools in recent years has opened up new possibilities for enhancing academic outcomes in higher education. This study examines the potential impact of these novel tools on student accomplishment, retention, and overall academic success. Machine learning algorithms leverage vast quantities of educational data to uncover hidden patterns and insights that are often overlooked by traditional methods. The objective of this study is to determine the applicability of these techniques in detecting students who are vulnerable, forecasting student achievement, and customizing educational experiences. The findings of the research have the potential to impact educational practices and policies, hence enhancing instructional strategies and academic outcomes [18]. Understanding the role of machine learning in education can help promote the use of a data-driven approach to education, ensuring that students have the necessary support and resources to thrive in their academic pursuits.

S. No.	Methods	Focus	Reference
1	Machine Learning in Educational Data Mining	Application of common machine learning algorithms for student assessment and prediction	Ahmad et al. (2023) [6]
2	Deep Learning in MOOCs	Forecasting dropout rates, feature processing, and model building	Sun et al. (2021) [8]
3	Deep Learning Models in Education	Overview of traditional Deep Learning models (SAE, DBN, DNN) and methods for various kinds of educational data	Zhang et al. (2018) [9]
4	EDM Tactics and Approaches	Knowledge inquiry, judgment, proposal generation, behavior screening, skill computation, course recommendations, intelligent suggestions, data visualization, risk identification, functionality forecasting, and student grouping	Buenaño et al. (2019) [10]
5	Data Mining in Education	Early detection of student behavior and achievement for benefiting businesses and learners	Khan et al. (2018) [11]
6	Classification Methods	Use of Decision Trees, Naive Bayes, and KNN to classify student data and forecast academic progress	Asif et al. (2017) [13]
7	Data Mining in Higher Education	Challenges in utilizing data mining to improve learning environments and decision-making, importance of SAP prognosis	Kong et al. (2019) [30]
8	Predictive Models in Education	Use of various data mining approaches to forecast academic development, employing methods such as J48, decision	Kumar et al. (2017) [15]

		trees, SVM, ANN, KNN, SMO, Linear Regression, and Naive Bayes	
9	Supervised Learning in Education	Classification of university students based on academic results using supervised learning	Kabakchieva et al. (2013) [16]
10	Data Mining for Undergraduate Performance	Evaluating academic achievements and analyzing student growth to offer targeted support and interventions	Yağcı et al. (2022) [17]

III. PROBLEM STATEMENT

Given the limitations of traditional methods in meeting the diverse needs of students and addressing institutional challenges, it is imperative to employ innovative approaches to enhance educational outcomes in response to the rapid expansion of higher education. The research topic "The Impact of Machine Learning-Based Data Mining Techniques to Enhance Higher Educational Outcomes" investigates the potential of advanced data mining (DM) and machine learning (ML) algorithms to optimize student performance, retention rates, and institutional efficiency in higher education. While machine learning and deep learning can provide valuable insights from vast educational datasets, its integration into educational systems is yet incomplete. This research seeks to assess the efficacy of these technologies in predicting academic achievement, identifying at-risk students, tailoring learning experiences, and informing policy decisions [19]. The project aims to demonstrate the efficacy and benefits of various machine learning-based data mining strategies in enhancing higher education outcomes through systematic evaluation. Ultimately, it will provide a structure for educational institutions to use data-driven approaches for enhancing student achievement and ongoing enhancement.

Objectives

- With the improvement of student models—which characterize and group a student's knowledge, motivation, metacognition, and attitudes—EDM seeks to forecast learners' actions.
- In order to improve learning systems, EDM aims to discover the essential, influential learning components by gathering empirical data to construct and support pedagogical theories, frameworks, and educational phenomena.

IV. PROPOSED METHODOLOGIES

The suggested model in this study consists of four components: data preprocessing, hyperparameter tweaking, recommender model, and model evaluation. However, additional elements are integrated into these main constituents. The model's general design is depicted in Figure 1. The initial stage of the dataset collection process involves acquiring information from institutes of higher education. Subsequently, we implemented three data preprocessing procedures. Data preprocessing involves several procedures, including data reduction, cleaning, and classification, which are necessary to prepare the dataset for training data mining algorithms. Subsequently, to

determine the most informative features, we employed feature extraction. Afterwards, hyperparameter optimization was utilized to enhance the approach. Hyperparameter optimization is achieved by fine-tuning a model's hyperparameters automatically. Hyperparameters refer to the model parameters that remain fixed during the learning process of an algorithm. They are employed to configure the gradient descent algorithm in a manner that minimizes the cost function of the learning rate. We apply this to the features provided to the algorithm. In this work, hyperparameter tweaking is used just to enhance the model learning process and determine the hyperparameter values that result in the minimum error on the validation set. Therefore, it is essential to allocate a loss and reserve a validation set. In this study, we employed various prediction models and utilized a clustering technique to anticipate a student's ultimate result. The factors considered for prediction were gender, locality, entrance result, number of previous attempts, studied credits, and impairment. Model building involves the application of prediction techniques to create several models.

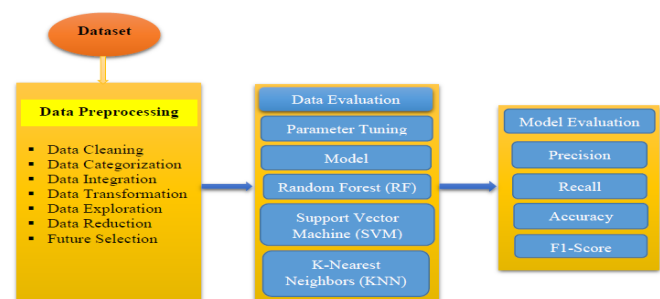


Fig. 2. Research steps for the student performance prediction model

Data Understanding

The dataset utilized in this study incorporates academic data obtained from higher education institutions, specifically generated from learning management systems. This state-of-the-art technology enables the simultaneous delivery of educational information to several internet-connected devices. The Experience API is a platform designed to monitor and record student actions for the purpose of data collection. Providers can utilize the Experience API to define the components of the learning experience, including products, activities, and students.

Table 1. Student features and their description.

Feature Category	Feature	Description
Demographical Features	Gender	The gender of the student (female or male)
	Address	Student Place
	Relation	Student's Guardian
Academic Background Features	StageID	Stage student belongs such as (1 st , 2 nd , 3 rd and 4 th)
	GradeID	Grade student belongs such as (A, B, C and F)
	SectionID	Section student belongs such as (A, B, C).
	Topic	Course topic such as (Engg, Science, MBA etc)
	Student Absence Days	Student absence days (Above-5, Under-5)
	Semester	College year semester

The collection contains a total of 480 instances, with each instance having 16 associated attributes. These traits can be categorized into three groups: The user's text can be rewritten as follows: (1) Demographic factors, such as gender and nationality; (2) Academic background factors, including education, socioeconomic status, and grade level; (3) Social factors, such as participation in class discussions, access to scientific learning resources, willingness to respond to parent surveys, and parental satisfaction with the school. The dataset is distributed in the following manner: Table 1 shows that there were 142 instances for the high class, 211 instances for the medium class, and 127 instances for the low class.

Modeling

At this stage, a meticulous choice is made regarding the modeling techniques, protocols, or their combined effects. Subsequently, the optimal parameter values for the selected algorithms are determined. In general, when faced with a specific situation, there are multiple potential modeling techniques to take into account. Specific methodologies can be adjusted to suit specific formats or constraints related to the quality of data. Therefore, if the models do not meet the requisite quality standards, this strategy often involves repetitive iterations.

Random forest

The DT (Decision Tree) and random forest methodologies are intimately interconnected. DTs are well-known for their low bias and high variability. The RF ensemble reduces the variance component of the model by aggregating many DTs. The practice of averaging projections in an RF can be used to generate forecasts for unknown samples [20].

$$X = \frac{1}{M} \sum_{M=1}^M f(Y) \quad (1)$$

The estimate given by the i-th decision tree (DT) is denoted as prediction i, and M represents the total number of individual decision trees in the ensemble. The RF method utilizes a set of Decision Trees (DT) to do data analysis and find the most effective plan of action by combining forecasts from each tree. Furthermore, Random Forest (RF) has the capability to handle datasets that contain missing values due to its foundation in ensemble learning and utilization of the bagging technique [21].

Support vector machines

Support vector machines select the hyperplane from the training dataset that maximizes the separation between classes for the purpose of data classification. The mathematical definition of this hyperplane is as follows:

$$f(y) = b^K y + d \quad (2)$$

b represents the dimensional coefficient.
d represents the offset.

One of the primary advantages of SVM is its ability to choose from a variety of kernels, allowing it to effectively handle complex and structured datasets by utilizing many kernel options. Support Vector Machines (SVM) also exhibited reduced susceptibility to overfitting. The crucial importance of selecting the suitable kernel makes the task challenging. Conversely, Support Vector Machines (SVM) may require significant computational time when dealing with larger datasets [22].

K-nearest neighbor

As part of the KNN procedure, the dataset is evaluated to find the K instances that are closest to the observed data point. Subsequently, the computer utilizes the obtained results to generate predictions for the variable y of the observed instance. The Euclidean distance is the metric employed to calculate the distance between two observations. The formula for Euclidean distance is as follows:

$$f(k_j, l_j) = \sqrt{(k_{j,0} - l_{j,0})^2 + \dots + (k_{j,u} - l_{j,u})^2} \quad (3)$$

The KNN technique is renowned for its low computational demands because it learns directly from the dataset during prediction generation without the need for an initial training step. The simplicity of its implementation arises from the fact that it just necessitates the K value and the distance function [23]. Nevertheless, the process of managing extensive datasets poses challenges, and employing the method on data with many dimensions diminishes its effectiveness.

V. RESULTS AND DISCUSSIONS

The research was conducted on a desktop computer equipped with four 2.30GHz Intel CPUs and eight

gigabytes of RAM. The project was developed using Python Notebook version 3.9.13. The suggested classification models were evaluated and compared using toolkits such as Pandas, NumPy, Matplotlib, and Scikit-Learn. Furthermore, a 10-fold cross-validation technique was employed to produce training and testing subsets of the dataset.

After conducting a comprehensive analysis of several machine learning algorithms, we performed extensive simulations. As a result, prediction models were used to establish groups of pupils based on their performance levels: low, medium, and high. A variety of performance metrics were calculated and assessed, including as F1-score, accuracy, precision, sensitivity, specificity, and the area under the receiver operating characteristic (ROC-AUC) curve. Below are the mathematical formulations for each measure: true positive, true negative, false positive, and false negative.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1 - Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Feature Selection

In order to obtain the most accurate prediction outcome, we utilized four separate classifiers in our analysis: Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). We systematically evaluated each classifier in order to determine the optimal result. The Random Forest classifier achieved the highest performance in the initial iteration. The Random Forest classifier consistently outperformed the other classifiers in every metric in subsequent rounds. Table 2 presents the rankings of the classifiers based on their accuracy levels.

Table 2. Features selection using classification techniques results.

	F1	Accuracy	Precision	Recall	RoC_AUC
RF	0.798	0.768	0.799	0.788	0.85
SVM	0.636	0.636	0.634	0.659	0.77
KNN	0.624	0.624	0.628	0.683	0.74

Based on our analysis, it has been determined that the RF classifier exhibited superior performance, as evidenced by the data presented in Table 2. Our main focus is on evaluating the F1 and ROC AUC scores because of the binary nature of the classification problem and the presence of imbalanced data. The ROC AUC score and the F1 score, which represent the balanced combination of precision and recall, are essential measures. It is crucial to differentiate these categories from each other because if marketing strategies are handled improperly, users can suffer significant financial consequences.

Confusion Matrix

The ML models were trained with an emphasis on reducing bias and correcting overfitting concerns following the implementation of our recommended hyperparameter tuning technique. Subsequently, cross-validation testing was conducted on the models to ensure the accuracy of the results and reduce variation, with the primary goal of preventing data leakage. Table 3 displays the confusion matrices for each classifier.

Table 3. Results of the algorithms utilized in the study's confusion matrix.

		Predicted		
		L	M	H
Confusion Matrix (RF)				
Actual	L	0.18	0.0038	0
	M	0.029	0.24	0.069
	H	0	0.036	0.32
Confusion Matrix (SVM)				
Actual	L	0.23	0.016	0
	M	0.012	0.023	0.046
	H	0	0.07	0.021
Confusion Matrix (KNN)				
Actual	L	0.20	0.019	0
	M	0.046	0.25	0.07
	H	0	0.11	0.16

VI. DISCUSSIONS

Education researchers are increasingly utilizing data mining (DM) methods to analyze the behaviors and academic achievements of their students. A diverse range of data mining techniques, such as decision trees, association rules, nearest neighbors, neural networks, evolutionary algorithms, exploratory factor analysis, and regression, are used to process substantial volumes of educational data [24]. These strategies enhance the quality of instruction by helping teachers identify pupils in need of additional guidance or academic counseling.

Predicting Student Performance

- Regression: Regression is a metric prediction technique that compares one or more independent variables with a dependent variable. Regression is employed in educational data mining (EDM) to predict the future performance and behavior of students.
- Classification Techniques: Support vector machines, backpropagation, and k-nearest neighbor classifiers are among the techniques employed to forecast student progress and identify pupils in need of additional assistance.
- This study assessed the efficacy of various data mining approaches, such as OneR, C4.5 (J48), Multi-Layer Perceptron, and IB1, in evaluating student performance. The IB1 classifier is a highly effective tool for improving traditional classification techniques, as it has shown the highest level of accuracy (82.00%) and the lowest rate of error (0.20%). Student performance was significantly influenced by characteristics such as high school grade (HSG),

mother's qualification (Mqual), family annual income (FAIn), and senior secondary school grade (SSG) [25].

- The study on Student Dropout: The primary objective of this study was to forecast the rate at which undergraduate Systems Engineering students will discontinue their studies. The researchers identified reliable indicators of student dropout by employing decision trees, logistic regression, Naive Bayes, and random forest algorithms. This study demonstrates that these simple algorithms may effectively forecast student attrition [26].

Grouping and Categorizing Learners

- Clustering: Clustering algorithms can be used to identify learning styles and student behavior by grouping individuals together based on common characteristics. These insights empower educators to create adaptable content and foster effective community education through the use of personalized learning frameworks.
- The study indicates that clustering and classification analysis greatly enhance the effectiveness of decision-making and prediction in e-learning. A study found that performing clustering prior to classification reduces the root mean squared error (RMSE) by employing K-means clustering and comparing the mean squared error (MSE) between data that has been clustered and data that has not been clustered. Linear Regression outperformed Support Vector Machine (SVM) in predicting students' final GPA [27].
- This study employed test results and data from the Moodle database to categorize students who exhibited comparable characteristics. Contrary to previous forecasts, researchers discovered that pupils who engaged in extracurricular activities and had a plenty of free time performed poorly on tests. The study highlighted the influence of a short dataset on the number of clusters and variables, as well as the importance of grouping data before analysis [28].

Decision Making by Educators

- Educational Decision Making Framework: Determining problems, establishing objectives, reviewing pertinent information, and carrying out selected actions are all parts of decision-making in the educational field. As a reflection of the basic variations in natural laws, distinct domains could contain distinct content at every step.
- A framework for educational decision-making that takes into account the rules and features of student growth was put forth by Yahya et al. (2019) [29]. Through a case study, they illustrated the framework's efficacy by utilizing Weka software to evaluate graduate student data using the pruned C4.5 method. This method demonstrated the usefulness of decision trees in forecasting educational outcomes by offering tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

Various strategies are utilized in Educational Data Mining (EDM) to enhance comprehension and administration of student learning processes. EDM offers insightful

information that supports educators in providing individualized, high-quality instruction, from forecasting student performance to organizing students into groups and assisting with decision-making [31]. These strategies' ability to greatly improve educational outcomes is demonstrated by how well they work in practical settings.

VII. CONCLUSION

The reason the EDM method of student performance prediction works so well is that it can identify students who are at risk of failing because of their subpar performance. Early detection can assist in taking the appropriate steps, thereby preventing students from failing or being dropped. Performance prediction is important, as demonstrated by the large quantity of research being done in this area. These approaches primarily entail distinct phases of the data mining procedure, employ diverse classification algorithms, and showcase their conclusions by implementing their methodology on diverse datasets. Parents and students can both gain from such techniques, which can be very valuable for educational institutions [32][33]. But it's also critical that the performance prediction method be dependable and have undergone thorough examination. Inaccurate conclusions and misleading outcomes might arise from process flaws. In this study, we carefully curated an improved dataset and thoroughly investigated feature selection strategies. The ensemble learning approach yielded the highest overall accuracy of 86% among the numerous prediction models we used. In addition, our study highlighted the notable improvements obtained by including hyperparameter adjustment and cross-validation into the suggested methods.

REFERENCES

1. Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education: an integrated view. *J Inf Syst Educ.*, 31(1), 61–71. Google Scholar□□CrossRef/DOI□
2. Baker, R. (2010). Data mining for education. *Int Encycl Educ.*, 7(3), 112–118. Google Scholar□□CrossRef/DOI□
3. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *Wiley Interdiscip Rev: Data Min Knowl Discov.*, 10(3), e1355. Google Scholar□□CrossRef/DOI□
4. Alshehri, E., Alhakami, H., Baz, A., & Alsubait, T. (2020). A comparison of EDM tools and techniques. *Int J Adv Comput Sci Appl.*, 11(12), 824–831. Google Scholar□□CrossRef/DOI□
5. Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdiscip Rev Data Min Knowl Discov.*, 3(1), 12–27. Google Scholar□□CrossRef/DOI□
6. Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Al-Fuqaha, A. (2023). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on*

- Learning Technologies*, 99(1), 1-20. Google Scholar□□CrossRef/DOI□
7. Prakash, K. & Prasad, J. P. (2021). Mobile and Location Based Service using Augmented Reality: A Review. *European Journal of Electrical Engineering and Computer Science*, 5(2), 13–18. Google Scholar□□CrossRef/DOI□
8. Sun, Z., Harit, A., Yu, J., Cristea, A.I., & Shi, L. (2021). A brief survey of deep learning approaches for learning analytics on moocs, in: Intelligent Tutoring Systems: 17th International Conference, ITS 2021, Virtual Event, June 7–11, 2021, Proceedings 17, Springer, pp. 28–37. Google Scholar□□CrossRef/DOI□
9. Zhang, Q., Yang, L.T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42(1), 146–157. Google Scholar□□CrossRef/DOI□
10. Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability*, 11(10), 28-33. Google Scholar□□CrossRef/DOI□
11. Khan, A., & Ghosh, S. K. (2018). Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, 23(1), 1677-1697. Google Scholar□□CrossRef/DOI□
12. Prakash, K., Prasad, J. P. & Sunay, K. (2021). IOT based Smart Healthcare Wheelchair for Independent Elderly. *European Journal of Electrical Engineering and Computer Science*, 5(5), 4-9. Google Scholar□□CrossRef/DOI□
13. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & education*, 113, 177-194. Google Scholar□□CrossRef/DOI□
14. Papadogiannis, I., Pouloupoulos, V., & Wallace, M. (2020). A Critical Review of Data Mining for Education: What has been done, what has been learnt and what remains to be seen. *International Journal of Educational Research Review*, 5(4), 353-372. Google Scholar□□CrossRef/DOI□
15. Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on student's performance prediction in education using data mining techniques. *International Journal of Education and Management Engineering*, 7(6), 40-49. Google Scholar□□CrossRef/DOI□
16. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, 13(1), 61-72. Google Scholar□□CrossRef/DOI□
17. Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. Google Scholar□□CrossRef/DOI□
18. Yaw, B. A. (2023). Handling Big Data in Education: A Review of Educational Data Mining Techniques for Specific Educational Problems. *AI, Computer Science and Robotics Technology* 2023(2), 1–16. Google Scholar□□CrossRef/DOI□
19. Ahmad, Z., & Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. *Bulletin of Education and Research*, 40(3), 157–164. Google Scholar□□CrossRef/DOI□
20. Alshantiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access*, 8(1), 203827–203844. Google Scholar□□CrossRef/DOI□
21. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113(1), 177–194. Google Scholar□□CrossRef/DOI□
22. Babić, I. D. (2017). Machine learning methods in predicting the student academic motivation. *Croatian Operational Research Review*, 8(2), 443–461. Google Scholar□□CrossRef/DOI□
23. Burgos, C., Campanario, M. L., De, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, 66(2018), 541–556. Google Scholar□□CrossRef/DOI□
24. Casquero, O., Ovelar, R., Romo, J., Benito, M., & Alberdi, M. (2016). Students' personal networks in virtual and personal learning environments: A case study in higher education using learning analytics approach. *Interactive Learning Environments*, 24(1), 49–67. Google Scholar□□CrossRef/DOI□
25. Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using Learning Analytics to improve teamwork assessment. *Computers in Human Behavior*, 47(1), 149–156. Google Scholar□□CrossRef/DOI□
26. García-González, J. D., & Skrita, A. (2019). Predicting academic performance based on students' family environment: Evidence for Colombia using classification trees. *Psychology, Society and Education*, 11(3), 299–311. Google Scholar□□CrossRef/DOI□
27. Kardaş, K., & Güvenir, A. (2020). Analysis of the effects of Quizzes, homeworks and projects on final exam with different machine learning techniques. *EMO Journal of Scientific*, 10(1), 22–29. Google Scholar□□CrossRef/DOI□
28. Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. *Higher Education*, 80(5), 875–894. Google Scholar□□CrossRef/DOI□
29. Yahya, A. A. (2019). Swarm intelligence-based approach for educational data classification. *Journal of King Saud University-Computer and Information Sciences*, 31(1), 35–51. Google Scholar□□CrossRef/DOI□
30. Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining

and applications. *Journal of Network and Computer Applications*, 132(1), 86–103. Google Scholar□□CrossRef/DOI□

31. Lu, X., Zhu, Y., Xu, Y., & Yu, J. (2021). Learning from multiple dynamic graphs of student and course interactions for student grade predictions. *Neurocomputing*, 431, 23–33. Google Scholar□□CrossRef/DOI□
32. Mai, T. T., Bezbradica, M., & Crane, M. (2022). Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. *Future Generation Computer Systems*, 127, 42–55. Google Scholar□□CrossRef/DOI□
33. Huang, T. C. K., Huang, C. H., & Chuang, Y. T. (2016). Change discovery of learning performance in dynamic educational environments. *Telematics and Informatics*, 33(3), 773–792. Google Scholar□□CrossRef/DOI□