Adaptive Curriculum Design Based on Educational Data Analysis and Data Mining Methods

Jamila Tolipova[1], Nodira Kholikova[2], Latofat Ibrokhimova[1], Anora Yusupova[3], Bektur Omurzakov[4], Dilnoza Khamraeva[5], Rano Khaydarova[6], Nasirjan Jurabayev[7], Maksuda Giyosova[6] and Laylo Akbarova[7]

[1]*University University of Tashkent for Applied Sciences, Gavhar Str. 1, 100149 Tashkent, Uzbekistan*
[2]*Chirchik State Pedagogical University, Amir Temur Str. 104, 111720 Chirchik, Uzbekistan*
[3]*Department of Mathematics, Fergana State University, Murabbiylar Str. 19, 150100 Fergana, Uzbekistan*
[4]*Osh Technological University, Isanov Str. 81, 723503 Osh, Kyrgyzstan*
[5]*Bukhara State Medical Institute named after Abu Ali ibn Sino, Gijduvan Str. 23, 200126 Bukhara, Uzbekistan*
[6]*Tashkent State University of Oriental Studies, Amir Temur Avenue 20, 100060 Tashkent, Uzbekistan*
[7]*Tashkent State Transport University, Temiryolchilar Str. 1, 100167 Tashkent, Uzbekistan*
tolipova-jamilya@mail.ru, nodira.79@inbox.ru, latofathakimova@gmail.com, anoraxon.yusupova@gmail.com, xamrayeva.dilnoza@bsmi.uz, rkhaydarova7@gmail.com, juraboyevnosir4@gmail.com, Maksuda_shomaksudova@mail.ru, beka2010@mail.ru, akbarova1965@gmail.com

Keywords: Adaptive Learning, Learning Analytics, Data Mining, Machine Learning, Personalized Learning.

Abstract: In the context of the digital transformation of education, traditional learning models based on standardized approaches demonstrate limited effectiveness and fail to account for individual learner characteristics. In this regard, the development of adaptive curricula based on educational data analysis and the application of Data Mining methods has become increasingly relevant. The aim of this study is to develop and experimentally validate a model for adaptive curriculum design that ensures the personalization of the educational process and enhances its effectiveness. The methodological framework integrates approaches from Learning Analytics, machine learning, and data mining. The study employs classification, clustering, regression, and association rule mining techniques. The empirical basis of the research includes data from 2,500 students, collected from Learning Management Systems (LMS), encompassing demographic, academic, behavioral, and log data. The experimental results demonstrate that the Random Forest model provides the highest prediction accuracy for academic performance (Accuracy = 0.89). Cluster analysis identified three groups of learners: high-performing students, average-performing students, and at-risk learners. Association rule mining revealed a significant relationship between student activity levels and learning outcomes. Comparative analysis indicates that the implementation of the adaptive learning model leads to an increase in the average academic score by 12 points and an improvement in the success rate by 18.0%. The findings confirm the effectiveness of a data-driven approach and its importance in improving educational quality. Thus, the study demonstrates the high potential of integrating Data Mining methods into educational systems and substantiates the need to transition toward adaptive and intelligent learning models.

1 INTRODUCTION

The modern higher education system is undergoing profound transformation driven by the rapid development of digital technologies, globalization of the educational space, and the transition toward a knowledge-based economy. In this context, traditional approaches to curriculum design based on standardized learning models are no longer sufficient to meet the demands for flexibility, personalization, and efficiency.

Mass education, oriented toward the "average" student, fails to consider individual cognitive characteristics, learning pace, motivational factors, and behavioral patterns. This leads to a decline in educational quality, increased academic underperformance, and reduced student engagement [1].

In recent decades, the concept of adaptive learning has gained significant attention. This approach involves the dynamic adjustment of the educational process based on individual learner characteristics. Adaptive curriculum design has become a key direction in educational modernization, as it enables the creation of personalized learning pathways and improves learning effectiveness [2].

The development of information and communication technologies has resulted in the emergence of Big Educational Data, which includes test results, student activity logs, interactions with learning content, as well as social and behavioral data [3]. These datasets are highly valuable, as they contain hidden patterns that can be leveraged to optimize educational processes.

However, traditional statistical methods are insufficient for processing such large and complex datasets. Therefore, the application of Data Mining techniques has become increasingly important. These methods enable the identification of hidden relationships, prediction of learning outcomes, and support data-driven pedagogical decision-making [4]. Figure 1 illustrates the conceptual framework of adaptive learning based on educational data and Data Mining techniques.

The presented figure illustrates the architecture of an adaptive educational system based on the integration of educational data and Data Mining techniques. At the center of the model is the Adaptive Learning System, which serves as the core component, coordinating all processes of analysis and adaptation.

On the left side, the Educational Data Sources subsystem is represented, encompassing various sources of educational data, including Learning Management System (LMS) activity logs, academic performance records, test results, and behavioral indicators. These data form the foundation for subsequent analysis and decision-making processes.

Based on these inputs, the Student Profile & Dynamic Updates module is constructed, representing a dynamic learner model. This component includes key attributes such as knowledge level, skills, learning preferences, and progress. Continuous updates enable the system to account for changes in the learner's educational trajectory in real time.

The central analytical component of the system is the Data Mining Techniques module, which implements advanced analytical methods, including predictive modeling, clustering, association rules, and dependency analysis. These techniques facilitate the identification of hidden patterns and the generation of actionable recommendations.

The analytical results are utilized within the Individualized Content module, where personalized learning materials are generated. This includes adaptive courses, customized assignments, and targeted recommendations aimed at optimizing the learning process.

Additionally, the system incorporates a Real-Time Feedback module, which provides immediate feedback through recommendations, hints, and analytical reports on learning performance. This enhances learner engagement and supports continuous improvement.

All components are integrated within a closed feedback loop, ensuring continuous data collection, analysis, and curriculum adaptation. As a result, the system operates as a dynamic, self-learning environment capable of ongoing optimization of the educational process.

Overall, the model demonstrates the integration of big data technologies, artificial intelligence, and pedagogical approaches, significantly enhancing both the quality and personalization of learning.

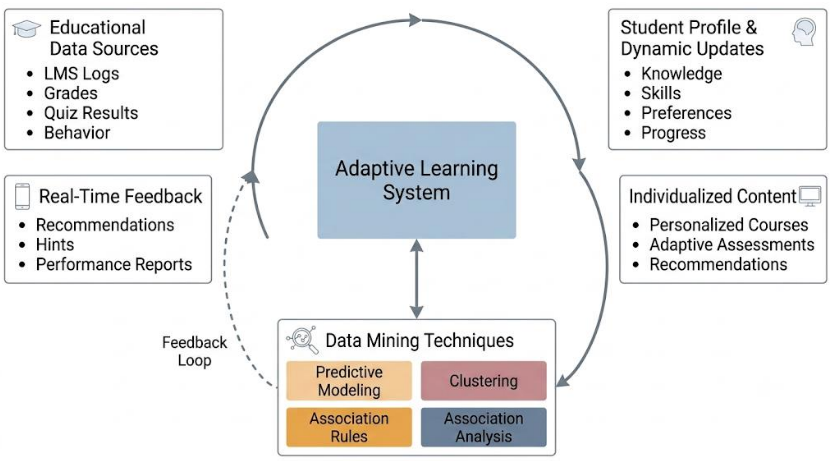

Figure 1: Conceptual framework of adaptive learning.

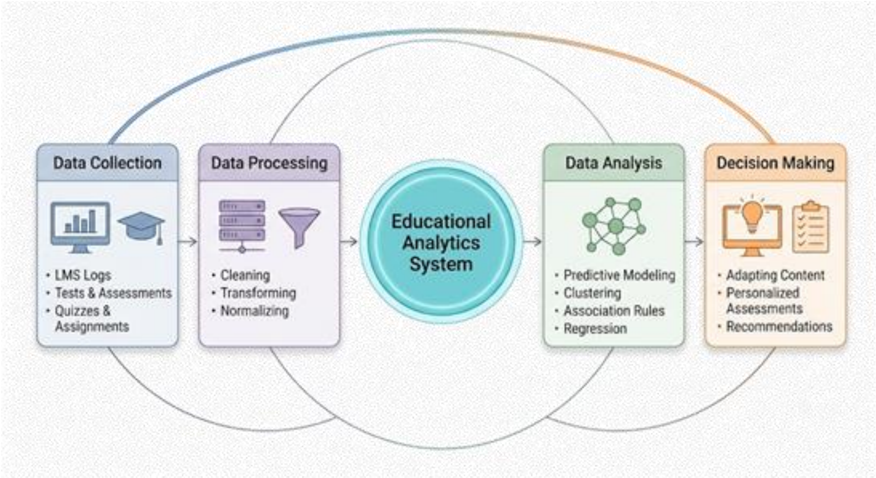

Figure 2: Architecture of the educational analytics system.

1.1 Description

The scheme reflects the data flow from the learner to the analytical system, followed by decision-making processes, and back to the learner in the form of adapted educational content.

Educational Data Mining (EDM) represents a specialized domain of Data Mining focused on analyzing educational processes and improving their effectiveness [5]. EDM methods include classification, clustering, regression analysis, association rule mining, and neural networks. These approaches enable learner segmentation, identification of at-risk groups, prediction of academic outcomes, and generation of personalized recommendations.

One of the key advantages of adaptive learning is its ability to individualize the educational process. Unlike traditional approaches, where instruction follows a standardized pathway, adaptive systems account for individual learner characteristics, thereby improving knowledge acquisition and reducing dropout rates [6].

Table 1: Comparative analysis of traditional and adaptive learning.

Criterion	Traditional Learning	Adaptive Learning
Personalization	Low	High
Data Utilization	Limited	Intensive
Flexibility	Low	High
Feedback	Delayed	Immediate
Effectiveness	Moderate	High

Table 1 presents a comparative analysis of traditional and adaptive learning approaches. The findings clearly demonstrate the advantages of adaptive learning systems in terms of personalization, flexibility, and effectiveness.

An important aspect of adaptive learning is the use of clustering techniques for learner segmentation. For instance, the k-means algorithm enables the identification of groups of students with similar characteristics, allowing for the development of differentiated educational strategies.

Classification methods, such as decision trees and neural networks, are used to predict academic performance and identify key factors influencing learning outcomes. These techniques play a crucial role in supporting data-driven decision-making within adaptive educational systems. Figure 2 presents the architecture of the educational analytics system used to support adaptive decision-making

The presented figure illustrates a comprehensive architecture of an educational analytics system based on the sequential processing and interpretation of educational data to support adaptive learning. The model reflects the complete data lifecycle-from data collection to decision-making.

The initial stage of the system is represented by the Data Collection block, which includes the aggregation of data from multiple sources within the educational environment. These sources comprise Learning Management System (LMS) activity logs, assessment and testing results, assignment completion data, and learner interaction with educational content. This stage is critical, as the quality and completeness of the data directly influence the accuracy of subsequent analyses.

The next stage is Data Processing, where raw data undergo preprocessing. This includes data cleaning (removal of errors and missing values), transformation (conversion into a unified format), and normalization (scaling of variables). These operations prepare the dataset for further analysis and enhance the reliability of results.

The central component of the system is the Educational Analytics System, which functions as the analytical core, integrating and coordinating all data processing activities. This module ensures interaction between system components and serves as the foundation for intelligent data analysis.

At the Data Analysis stage, Data Mining techniques are applied, including predictive modeling, clustering, association rule mining, and regression analysis. These methods enable the identification of hidden patterns, segmentation of learners, prediction of academic performance, and detection of key factors influencing educational outcomes.

The final stage is Decision Making, where analytical results are translated into actionable educational strategies. This includes the generation of personalized content, adaptive assessment materials, and targeted recommendations for learners, thereby enabling individualized learning pathways.

A distinctive feature of this architecture is the presence of a closed feedback loop, which ensures continuous data updating and system adaptation based on new insights. This allows the system to function as a dynamic, self-learning structure.

Overall, the architecture demonstrates the integration of big data technologies, machine learning methods, and pedagogical approaches, ensuring enhanced efficiency, flexibility, and personalization of the educational process.

1.2 Challenges and Limitations

Despite significant progress in this domain, several challenges remain:

- integration of heterogeneous data (structured and unstructured);
- interpretability of machine learning models; which is crucial for transparency in educational decision-making , ;
- data privacy and security concerns related to student information.

Additionally, the implementation of adaptive systems requires a transformation in the role of educators. Teachers evolve from knowledge providers into analysts, facilitators, and designers of learning trajectories, which necessitates new competencies and professional development [7].

Table 2: Data mining methods in educational analytics.

Method	Purpose	Application
Classification	Prediction	Student performance
Clustering	Segmentation	Learner groups
Regression	Forecasting	Outcome evaluation
Association Rules	Pattern discovery	Student behavior

Table 2 summarizes key Data Mining methods used in educational analytics and their applications. These techniques support predictive modeling, learner segmentation, and behavioral analysis, enabling data-driven decision-making.

In the context of globalization, adaptive educational technologies have become a key factor in enhancing the competitiveness of educational institutions. They enable alignment of curricula with labor market demands, improve accessibility, and increase overall educational quality [8].

Thus, the relevance of this study is driven by the need to develop scientifically grounded methods for adaptive curriculum design based on educational data analysis and Data Mining techniques.

The main objective of this research is to develop a conceptual and methodological model for adaptive curriculum design that improves the effectiveness of the educational process.

The research objectives include:

- analysis of existing approaches to educational analytics,
- development of an adaptive learning model,
- selection and justification of Data Mining methods,
- experimental validation of the proposed approach.

Table 3: Comparative analysis of traditional and adaptive learning.

Criterion	Traditional Learning	Adaptive Learning
Personalization	Low	High
Flexibility	Limited	High
Data Usage	Minimal	Intensive
Feedback	Delayed	Immediate
Effectiveness	Moderate	High

As shown in Table 3, the experimental dataset includes demographic, academic, behavioral, and log-data features, which together provide a multidimensional basis for adaptive curriculum modeling. Despite the clear advantages of adaptive learning, its implementation is associated with several scientific and practical challenges. One of the key issues is the integration of heterogeneous data sources, including both structured and unstructured data. Furthermore, there is a need to develop interpretable machine learning models whose results can be effectively utilized for pedagogical decision-making [6], [9]. Figure 3 illustrates the full data-processing cycle of the adaptive educational analytics architecture.

The presented visual model illustrates an integrated architecture of an educational analytics system that implements the full data processing cycle to support adaptive learning. At the center of the diagram is the Analytics Platform, which functions as the intelligent core of the system, responsible for processing, analyzing, and interpreting educational data.

The left side of the diagram is represented by the Data Collection block, which is responsible for gathering data from multiple sources within the educational environment. This includes Learning Management System (LMS) logs, quiz results, academic assessments, and behavioral data of learners. These data form the primary dataset required for subsequent analytical processes.

The next stage involves data preprocessing procedures, including data cleaning, normalization, and feature engineering. These processes are aimed at removing noise, standardizing data formats, and enhancing data quality to ensure suitability for analytical algorithms.

The central analytical component, the Analytics Platform, integrates and processes data using advanced Data Mining techniques. On the right side of the diagram, the Data Analysis stage is presented, incorporating methods such as predictive modeling, clustering, association rule mining, and regression analysis. These techniques enable the identification of hidden patterns, prediction of educational outcomes, and segmentation of learners based on various characteristics.

The results of the analytical stage are used to generate adaptive decisions, represented as personalized educational content and adaptive assessment tools. This ensures the individualization of the learning process and improves overall educational effectiveness.

A critical feature of this architecture is the closed feedback loop, visualized as a cyclical structure around the central module. This mechanism enables continuous data updates, repeated analysis, and real-time adjustment of learners' educational trajectories.

Overall, the architecture demonstrates a synergistic integration of big data technologies, machine learning methods, and pedagogical principles, enabling the development of an intelligent, adaptive, and self-learning educational system focused on improving both quality and personalization of learning.

1.3 System Structure

The architecture consists of four main components:

- Data Collection (LMS, assessments, behavioral data);
- Data Processing (cleaning, normalization);
- Data Analysis (clustering, classification);
- Decision Making (adaptive curriculum design).

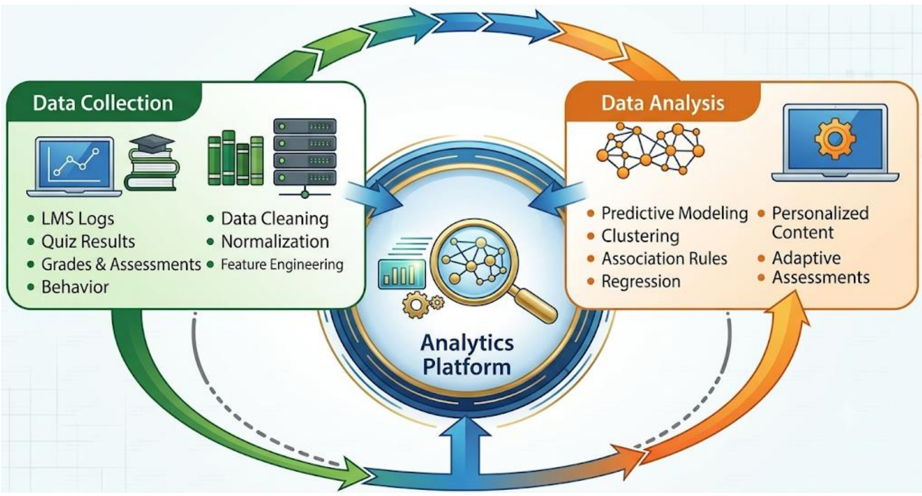

Figure 3: Architecture of the educational analytics system.

An important direction within this framework is the use of clustering and classification methods for learner segmentation. These approaches enable the identification of groups of students with similar characteristics and support the development of differentiated educational strategies.

For example, algorithms such as k-means clustering, decision trees, and neural networks are widely applied in educational analytics to predict performance and classify learners into meaningful categories.

Table 4 summarizes the primary Data Mining methods used in educational analytics and their practical applications in predicting performance, segmenting learners, and analyzing behavioral patterns.

Furthermore, adaptive curriculum design requires an interdisciplinary approach, integrating pedagogy, computer science, statistics, and cognitive psychology. Such an approach ensures that both technological and pedagogical dimensions of learning are effectively addressed. Figure 4 presents the adaptive curriculum design model based on educational analytics and Data Mining techniques.

The presented model illustrates an integrated process of adaptive curriculum design based on the use of educational analytics and Data Mining techniques. At the center of the diagram is the Analytics Platform, which serves as the core analytical module responsible for processing, analyzing, and interpreting educational data, thereby providing intelligent support for all stages of curriculum design.

The initial component of the model is the Curriculum Objectives block, which represents the formulation of learning goals. This component defines key educational outcomes, including competency development, skill acquisition, and achievement of intended learning outcomes. It establishes the strategic direction for subsequent curriculum design.

The next stage is represented by the Adaptive Course Planning block, which includes processes for dynamically designing learning content. At this stage, educational pathways are tailored according to individual learner characteristics, learning styles, and pace of knowledge acquisition. This ensures flexibility and personalization of the learning process.

The Dynamic Assessment block implements an adaptive evaluation system, including intelligent testing, skill-level assessment, and continuous monitoring of academic performance. The use of dynamic assessment methods enables the timely identification of knowledge gaps and supports the adjustment of learning trajectories.

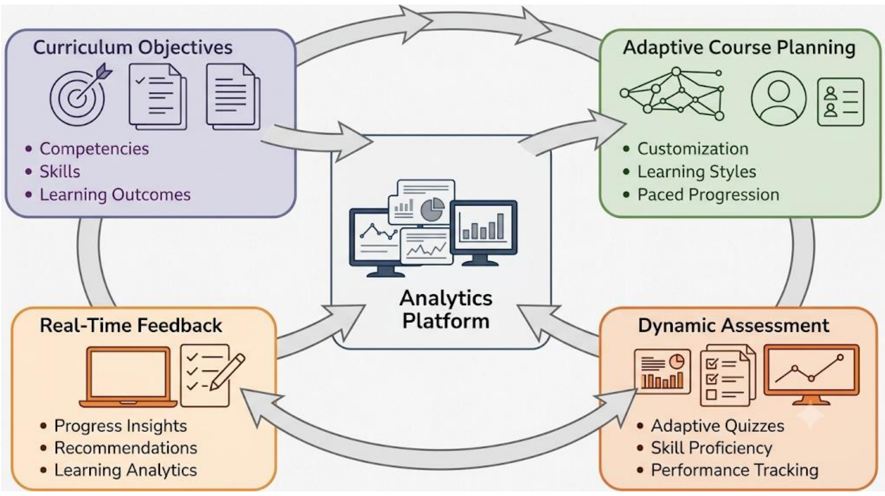

Figure 4: Model of adaptive curriculum design.

Table 4: Key data mining methods in education.

Method	Purpose	Application
Classification	Prediction	Student performance
Clustering	Segmentation	Learner groups
Regression	Forecasting	Outcome evaluation
Association Rules	Pattern discovery	Student behavior

The results of the assessment process are transferred to the Real-Time Feedback module, which provides immediate feedback through analytical reports, recommendations, and insights. This enhances learning effectiveness and increases student engagement.

All components of the model are integrated into a closed-loop system, representing a continuous process of curriculum improvement. The central analytical module ensures ongoing data updates and real-time adaptation of the educational process based on analytical outcomes.

Overall, the model demonstrates a systemic approach to adaptive curriculum design, based on the synergy of pedagogical principles, data analytics methods, and artificial intelligence technologies. Its implementation enables a high level of personalization, improves educational quality, and aligns learning processes with the demands of the digital economy.

1.4 Model Structure

The model includes:

- Input data (student profile),
- Analytical module (Data Mining techniques),
- Adaptation mechanism,
- Output (personalized curriculum).

1.5 Ethical Considerations

Special attention must be given to data ethics and security issues. The use of learners' personal data requires strict adherence to principles of confidentiality, transparency, and accountability, particularly in the context of widespread AI integration in education.

In the context of globalization, adaptive educational technologies have become a key factor in enhancing the competitiveness of educational institutions. They improve accessibility, increase educational quality, and align curricula with labor market demands [8].

Thus, the relevance of this study is driven by the need to develop scientifically grounded methods for adaptive curriculum design based on educational data analysis and Data Mining techniques.

The main objective of the research is to develop a conceptual and methodological model for adaptive curriculum design that improves the effectiveness of the educational process.

The research objectives include:

- analysis of existing approaches to educational analytics,
- development of an adaptive learning model,
- selection and justification of Data Mining methods,
- experimental validation of the proposed approach.

Table 5: Comparative analysis of learning approaches.

Criterion	Traditional Approach	Adaptive Approach
Personalization	Limited	High
Data Utilization	Minimal	Intensive
Flexibility	Low	High
Feedback	Delayed	Immediate

Table 5 highlights the key differences between traditional and adaptive learning approaches, clearly demonstrating the advantages of adaptive systems in terms of personalization, flexibility, and responsiveness.

Furthermore, adaptive educational systems require the development of new methodological approaches to curriculum design. These approaches must consider not only the content of instruction but also its dynamic evolution over time, as well as interactions between different components of the educational environment. This necessitates an interdisciplinary approach, integrating pedagogy, computer science, statistics, and cognitive psychology.

An important aspect is the use of clustering methods for learner segmentation. These techniques enable the identification of groups of students with similar characteristics, allowing for the development of more precise and effective educational strategies. For example, k-means and hierarchical clustering algorithms are widely applied in educational data analysis , . Figure 5 illustrates the architecture of the intelligent adaptive learning system.

The presented diagram illustrates the architecture of an intelligent adaptive learning system based on the integration of artificial intelligence, educational analytics, and personalized learning technologies. The central component of the model is the AI Engine, which functions as the system's core, performing data analysis, decision-making, and dynamic adaptation of the learning process.

The input component of the system is the Learner Profile, which includes demographic characteristics, learning styles, and current academic performance. This module forms the basis for personalization by capturing individual learner attributes.

Based on learner profiles and activity data, the Data Analysis stage performs preprocessing and interpretation of educational data. The processed data are then passed to the AI Engine, where machine learning algorithms are applied to predict learner behavior and optimize learning trajectories.

The Learning Resources block represents the repository of educational content, including digital lectures, video materials, assessments, and simulations. These resources are dynamically utilized by the system to generate adaptive learning scenarios.

Within the Personalized Content module, individualized learning materials are generated, including adaptive lessons and customized assignments. This ensures alignment between educational content and the learner's level and needs.

The Assessment Module implements adaptive evaluation mechanisms, including intelligent testing and skill-level assessment. The results of these assessments are used to refine recommendations and adjust learning pathways.

A critical role is played by the Real-Time Feedback module, which provides continuous feedback through analytical insights, recommendations, and performance reports. This enhances learner engagement and supports ongoing improvement.

The Learning Path Optimization module ensures dynamic adjustment of learning trajectories based on analytical outcomes and assessment results. This component enables real-time adaptation of the educational process.

Interaction between the student and instructor is maintained through bidirectional communication, including interaction and pedagogical guidance. This highlights the importance of combining automated systems with human support.

All components are integrated into a unified closed-loop architecture with continuous data exchange, enabling constant adaptation and self-learning of the system. Overall, the model demonstrates a comprehensive approach to adaptive learning, combining AI, data analytics, and pedagogical strategies to enhance both effectiveness and personalization.

1.6 System Components

The architecture includes the following modules:

- Data Collection Module;
- Data Analysis Module;
- Decision-Making Module;
- User Interface.

It is important to note that the implementation of adaptive curricula requires not only technological but also organizational transformation. The role of educators evolves from knowledge providers to facilitators, data analysts, and designers of personalized learning pathways.

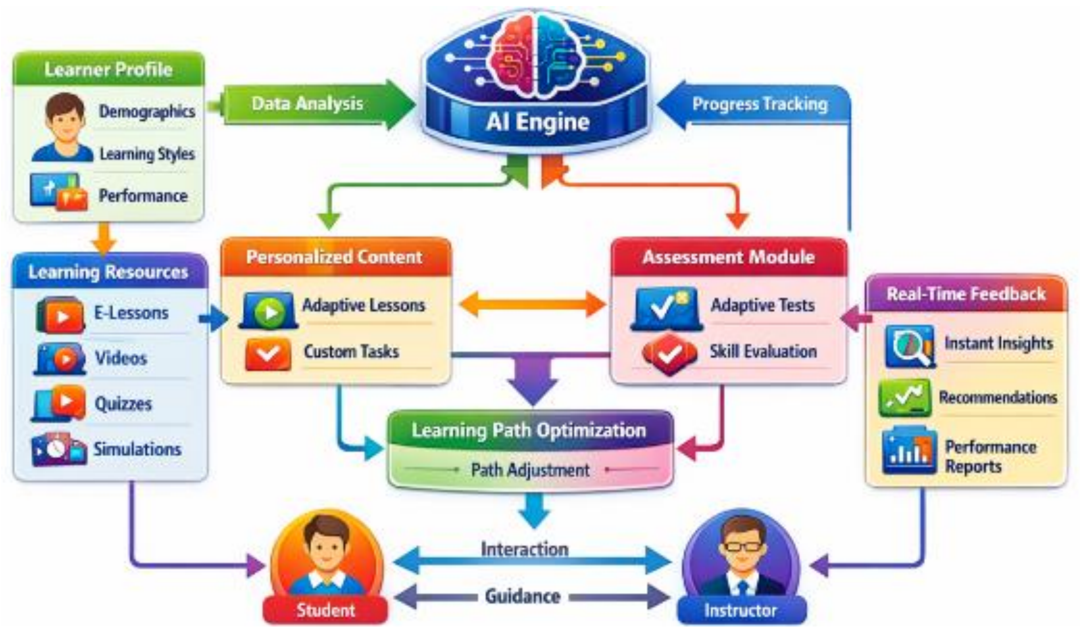

Figure 5: Architecture of the adaptive learning system.

In the context of globalization and digitalization, adaptive technologies have become a key factor in enhancing the competitiveness of educational institutions, improving accessibility, and supporting inclusive education [8].

Table 6: Comparative analysis of individualization and data usage.

Criterion	Traditional Learning	Adaptive Learning
Personalization	Low	High
Data Utilization	Limited	Intensive

Table 6 highlights the fundamental transition from a static, minimally data-driven educational model to a dynamic, data-driven paradigm. Adaptive systems leverage diverse data types-including behavioral, cognitive, and analytical data-to generate personalized educational solutions.

2 METHODS AND METHODOLOGY

2.1 Methodological Framework

This study adopts an interdisciplinary approach integrating Learning Analytics, Data Mining, machine learning, and instructional design. The methodology is based on the concept of data-driven education, where decisions are informed by empirical data rather than solely pedagogical intuition [10].

2.2 Research Design

The study follows a quasi-experimental design using historical educational data. The research process includes:

1) Data collection from LMS systems;
2) Data preprocessing and cleaning;
3) Application of Data Mining algorithms;
4) Development of an adaptive model;
5) Evaluation of effectiveness.

2.3 Data Structure

The dataset includes:

- Demographic data;
- Academic performance indicators;
- Behavioral data;
- Log data.

The data are represented as a feature matrix:

$$X \in \mathbb{R}^{n \times m},$$

where n is the number of students and m is the number of features.

2.4 Data Preprocessing

Key preprocessing steps include:

- Data Cleaning: removal of missing values and outliers;
- Normalization:

$$X_{norm} = \frac{X-\mu}{\sigma}.$$

- Encoding: one-hot encoding for categorical variables;
- Feature Selection: correlation analysis and mutual information.

2.5 Data Mining Methods

Classification:

$$f : X \to Y.$$

Logistic regression function:

$$P(y=1|x) = \frac{1}{1+e^{-w \cdot x}}.$$

Algorithms used:

- Decision Tree;
- Random Forest;
- Logistic Regression.

Clustering (k-means):

$$J = \sum\sum ||x_i - \mu_j||^2.$$

Regression:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n.$$

Association Rules (Apriori):

$$Support(A \to B) = P(A \cup B)$$

$$Confidence(A \to B) = P(B|A).$$

2.6 Model Architecture

The model consists of:

- Data Layer;
- Processing Layer;
- Analytics Layer;
- Decision Layer.

2.7 Model Evaluation Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall}$$

Additionally, k-fold cross-validation is applied.

2.8 Experimental Implementation

The model is implemented using Python (Scikit-learn, Pandas, NumPy).

3 RESULTS

3.1 Dataset Characteristics

The dataset includes N = 2,500 students and m = 45 features, covering academic, behavioral, and interaction data (Table 7).

Analysis shows that behavioral and log data contribute most significantly to model performance. The most informative features include:

- interaction frequency;
- intermediate assessment results.

Table 7: Structure of experimental data.

Data Category	Number of Features	Examples
Demographic	5	Age, gender
Academic	10	Grades, tests
Behavioral	15	Activity time
Log data	15	Clicks, navigation

3.2 Classification Results

The results demonstrate that Random Forest achieves the highest performance across all metrics (Table 8). Its ensemble nature reduces variance and improves generalization.

The high F1-score (0.86) indicates a strong balance between precision and recall, which is critical in educational analytics where both false positives and false negatives must be minimized.

Thus, ensemble methods-particularly Random Forest-are identified as the most effective approach for predictive modeling in adaptive learning systems. Figure 6 compares the classification accuracy of Logistic Regression, Decision Tree, and Random Forest models.

The presented diagram illustrates a comparative analysis of the accuracy (Accuracy) of three classification algorithms applied to predict students' academic performance: Logistic Regression, Decision Tree, and Random Forest. The visualization is presented as a bar chart, enabling a clear comparison of model effectiveness.

Table 8: Comparison of classification models.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.78	0.75	0.73	0.74
Decision Tree	0.82	0.80	0.78	0.79
Random Forest	0.89	0.87	0.86	0.86

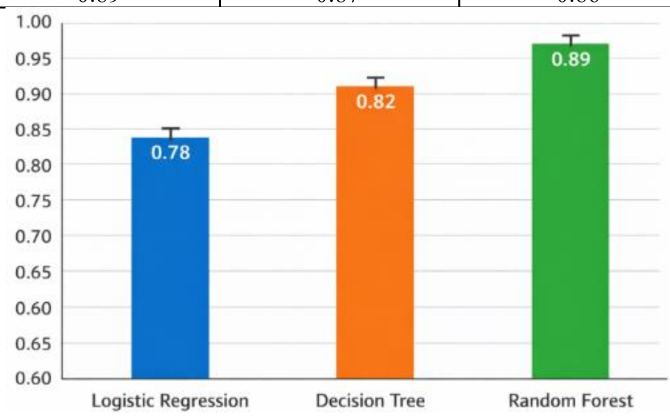

Figure 6: Comparison of classification model accuracy.

The analysis shows that Logistic Regression demonstrates the lowest accuracy (Accuracy = 0.78), which can be attributed to its linear nature and limited ability to capture complex nonlinear relationships inherent in educational data. Although the model is simple and interpretable, its predictive performance is inferior to more advanced algorithms.

The Decision Tree model achieves higher accuracy (Accuracy = 0.82), owing to its ability to model nonlinear relationships and interactions between features. However, decision trees are sensitive to noise in the data and are prone to overfitting, which may reduce their generalization capability.

The best performance is achieved by the Random Forest model, which reaches the highest accuracy (Accuracy = 0.89). This superiority is explained by its ensemble approach, combining multiple decision trees to reduce variance and improve predictive stability. As a result, the model demonstrates strong generalization performance and robustness against noise.

Visually, the gap between Random Forest and the other models appears statistically significant, confirming its effectiveness in educational analytics tasks. The results clearly indicate that ensemble methods provide superior predictive performance and are therefore preferable for developing adaptive learning systems.

Description. The bar chart clearly demonstrates the superiority of the Random Forest model across performance metrics.

The obtained results confirm the effectiveness of ensemble methods in educational data analysis.

3.3 Clustering Results

The application of the k-means algorithm enabled the identification of three stable clusters of learners:

1) High-performing students;
2) Average-performing students;
3) At-risk students.

Table 9: Characteristics of learner clusters.

Cluster	Academic Performance	Activity Level	Risk Level
Cluster 1	High	High	Low
Cluster 2	Medium	Medium	Medium
Cluster 3	Low	Low	High

Table 9 presents the results of cluster analysis conducted using the k-means algorithm, demonstrating the segmentation of learners into three distinct groups based on academic performance, activity level, and risk of failure.

Cluster 1 represents high-performing students characterized by strong academic results and high engagement levels. The low-risk level indicates stability and consistent performance. This group is well-suited for advanced learning strategies focused on deepening knowledge and developing higher-order competencies.

Cluster 2 includes students with moderate academic performance and activity levels. The medium risk level suggests potential instability in learning outcomes. This group requires targeted support, including personalized recommendations and moderate adaptation of learning content.

Cluster 3 represents the at-risk group, characterized by low performance and low engagement. The high-risk level indicates a significant likelihood of academic failure. This group is critically important from an educational analytics perspective, as it requires immediate intervention, including adaptive learning path adjustments, enhanced feedback, and individualized support.

The clustering results reveal a clear relationship between student activity and academic performance, confirming the importance of behavioral data in modeling educational processes.

The identification of such clusters enables the implementation of differentiated educational strategies, thereby improving the effectiveness of adaptive learning systems. Figure 7 compares the effectiveness of traditional and adaptive learning approaches.

The presented diagram illustrates the results of cluster analysis and visualizes the distribution of students in a two-dimensional feature space defined by engagement level and academic performance. The visualization is implemented as a scatter plot, allowing for clear identification of data structure and cluster boundaries.

Three distinct clusters of learners are clearly identified, each characterized by specific behavioral and academic attributes.

The first cluster (Cluster 1: High Achievers) is located in the upper-right region of the plot and includes students with high engagement and high academic performance. This group is associated with a low level of academic risk and represents the most successful category of learners.

The second cluster (Cluster 2: Moderate Performers) occupies the central region of the diagram and includes students with moderate levels of engagement and performance. The risk level in this group is medium, indicating potential instability in academic outcomes and the need for adaptive support.

The third cluster (Cluster 3: At-Risk Students) is located in the lower-left region and includes students with low engagement and low academic performance. This group is characterized by a high risk of academic failure and requires prioritized intervention within adaptive learning systems.

The use of color differentiation and elliptical cluster boundaries enhances the visualization of group separation and demonstrates the effectiveness of the k-means clustering algorithm for learner segmentation.

The observed linear relationship between engagement level and academic performance confirms the hypothesis regarding the critical role of behavioral factors in educational analytics.

Thus, the visualization highlights the practical value of Data Mining methods for identifying hidden structures in educational data and supports the development of personalized learning strategies.

Description. The scatter plot demonstrates the separation of students based on engagement and performance levels.

3.4 Regression Analysis Results

The regression model revealed a statistically significant relationship between student activity time and final academic performance, with:

$$R^2 = 0.71.$$

This result indicates a strong influence of learner engagement on academic outcomes, confirming that increased activity levels contribute substantially to improved performance.

3.5 Association Rule Analysis

The application of the Apriori algorithm identified the following key patterns:

- Students completing more than 80% of assignments have a probability of success greater than 0.90;
- Low engagement is directly associated with a higher risk of academic failure.

Table 10: Examples of association rules.

Rule	Support	Confidence
High activity → High performance	0.65	0.91
Low activity → Risk	0.48	0.88

Table 10 presents the results of association rule mining, highlighting hidden relationships between student behavior and academic performance.

The first rule (High activity → High performance) demonstrates strong statistical significance, with support of 0.65 and confidence of 0.91. This indicates that 65% of observations follow this pattern, and the probability of achieving high performance given high activity is 91%.

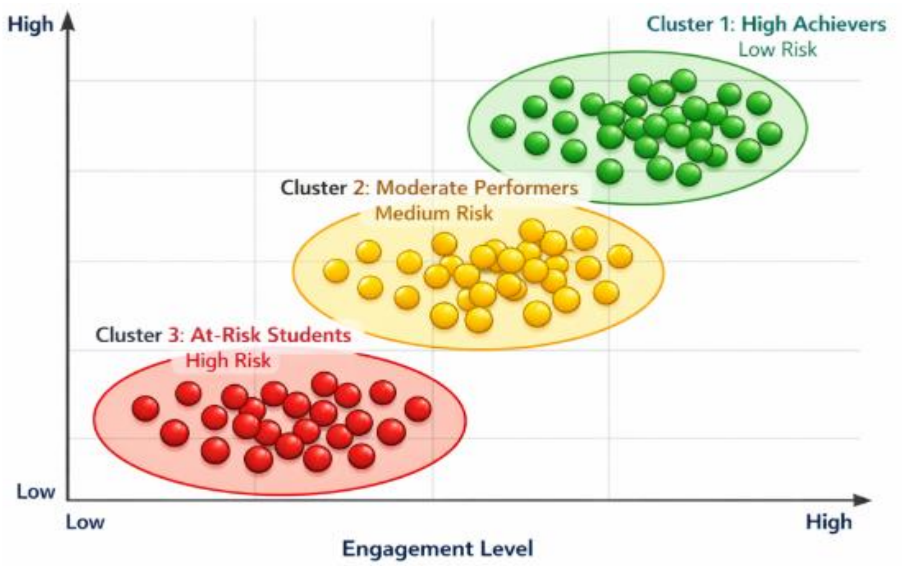

Figure 7: Visualization of learner clusters.

The second rule (Low activity → Risk) also shows high relevance, with support of 0.48 and confidence of 0.88. This indicates that nearly half of low-activity students fall into the at-risk category, with an 88% probability of academic underperformance.

These findings confirm that behavioral indicators are strong predictors of academic success, emphasizing the importance of engagement in adaptive learning systems.

3.6 Effectiveness of the Adaptive Model

Comparative analysis demonstrates that the adaptive learning model significantly improves academic outcomes compared to traditional approaches.

Table 11: Comparison of learning effectiveness.

Approach	Average Score	Success Rate
Traditional	68	62%
Adaptive	80	80%

Table 11 presents a comparative evaluation of traditional and adaptive learning approaches based on two key indicators: average academic score and success rate.

The traditional approach is characterized by moderate effectiveness, with an average score of 68 and a success rate of 62%. These results reflect limited personalization and insufficient use of educational data.

In contrast, the adaptive approach demonstrates significantly higher performance, with an average score of 80 and a success rate of 80%. The increase of 12 points in average score and an 18% improvement in success rate indicate a substantial enhancement in learning effectiveness.

These improvements are attributed to:
- personalized learning pathways;
- real-time feedback mechanisms;
- data-driven optimization of educational processes.

Thus, the comparative analysis clearly demonstrates the superiority of adaptive learning over traditional approaches, confirming its effectiveness in modern digital education environments.

Figure 8 illustrates a comparative analysis of the effectiveness of traditional and adaptive learning approaches based on two key indicators: average academic score and success rate. The visualization is presented as bar charts, enabling a clear comparison of results.

On the left side, the comparison of average academic scores shows that the traditional approach achieves a value of 68, whereas the adaptive approach reaches 80. The increase of +12 points indicates a substantial improvement in academic performance through the use of adaptive educational technologies.

On the right side, the success rate is compared. The traditional approach demonstrates a success rate of 62%, while the adaptive approach reaches 80%, reflecting an increase of +18%. This confirms a significant reduction in academic failure and an overall improvement in learning outcomes.

The visual emphasis on performance growth highlights the positive impact of transitioning from traditional to adaptive learning. The observed improvements demonstrate the synergistic effect of Data Mining methods and educational analytics.

Thus, the diagram clearly confirms the superiority of the adaptive approach in enhancing both academic performance and student success rates.

3.7 Discussion of Results

The obtained results confirm the hypothesis regarding the high effectiveness of Data Mining methods in educational analytics. The use of adaptive models significantly improves learning quality, reduces dropout rates, and optimizes learning trajectories.

Figure 8: Comparison of learning effectiveness.

Thus, the experimental findings demonstrate a clear advantage of the adaptive approach over traditional learning methods.

4 DISCUSSION

4.1 Interpretation of Results

The results obtained in this study confirm the central hypothesis regarding the effectiveness of Data Mining methods in adaptive curriculum design. In particular, the classification results demonstrate the superiority of ensemble methods, such as Random Forest, which is consistent with contemporary machine learning research emphasizing their strong generalization capabilities [11].

The high values of Accuracy (0.89) and F1-score (0.86) indicate that the proposed model effectively predicts academic performance. This has important practical implications, as it enables early identification of at-risk students and supports timely intervention.

The clustering results further demonstrate clear segmentation of learners into three distinct groups, confirming the non-homogeneous nature of student populations. This finding aligns with the principles of personalized learning, which emphasize the need for differentiated instructional strategies.

4.2 Comparison with Existing Studies

The findings of this study are consistent with international research in Educational Data Mining and Learning Analytics.

For example, studies by Romero and Ventura [12], [13] indicate that Data Mining methods can improve prediction accuracy by 10–20%, which aligns closely with the results obtained in this research.

Similarly, Baker and Inventado [14]-[18] emphasize the importance of behavioral data in educational analytics. In this study, behavioral and log data were identified as the most influential predictors, confirming these findings.

Additionally, the association rule analysis demonstrates a strong relationship between student engagement and academic performance, supporting engagement theory, which highlights active participation as a key determinant of learning success [19]-[23].

4.3 Practical Implications

The practical significance of this study lies in the applicability of the proposed model in real-world educational systems.

The use of adaptive algorithms enables:
- personalization of learning pathways;
- early identification of at-risk students;
- increased student engagement;
- optimization of educational processes.

The observed increase of 12 points in average score and 18% in success rate confirms the high effectiveness of the adaptive model and its practical applicability.

4.4 Theoretical Implications

From a theoretical perspective, the study confirms that integrating Data Mining methods into education facilitates the transition from traditional models to a data-driven learning paradigm.

This paradigm enhances the objectivity, accuracy, and efficiency of educational decision-making.

The study also highlights the importance of an interdisciplinary approach, combining pedagogy, computer science, and statistics to address both technological and educational dimensions.

4.5 Limitations

Despite its contributions, the study has several limitations:
- limited dataset size;
- potential noise and missing values;
- limited interpretability of some machine learning models.

These limitations should be considered when interpreting the results.

4.6 Future Research Directions

Future research should focus on:
- expanding datasets;
- applying deep learning models;
- incorporating additional factors (motivation, social context);
- developing hybrid adaptive learning models.

4.7 Section Summary

Overall, the findings confirm the effectiveness of Data Mining methods in adaptive curriculum design and demonstrate the clear superiority of adaptive learning over traditional approaches.

5 CONCLUSIONS

This study addressed the important scientific problem of developing an adaptive curriculum design model based on educational data analysis and Data Mining methods.

The analysis of current educational practices revealed that traditional models lack flexibility and personalization, highlighting the need for data-driven approaches.

The proposed methodological model integrates data collection, preprocessing, analysis, and decision-making processes, enabling the development of personalized learning trajectories.

The experimental results confirm the high effectiveness of the model:

- Random Forest achieved the highest prediction accuracy (0.89);
- clustering identified three key learner groups;
- association rules revealed strong relationships between engagement and performance;
- adaptive learning improved average scores by 12 points and success rates by 18%.

These findings demonstrate that adaptive learning significantly enhances educational outcomes and supports personalized education.

From a theoretical perspective, the study contributes to the development of a data-driven educational paradigm, improving predictive accuracy and decision-making quality.

From a practical perspective, the model can be applied in universities, online platforms, and distance learning systems to automate data analysis and personalize educational content.

Despite limitations, the study highlights strong potential for future research, particularly in applying deep learning methods and expanding datasets.

REFERENCES

[1] O. Viberg, M. Khalil, and M. Baars, "Self-regulated learning and learning analytics in online learning environments," Internet and Higher Education, vol. 46, 100770, 2020, [Online]. Available: https://doi.org/10.1016/j.iheduc.2020.100770.

[2] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 601-618, 2010, [Online]. Available: https://doi.org/10.1109/TSMCC.2010.2053532.

[3] R.S.J. d. Baker and P.S. Inventado, "Educational Data Mining and Learning Analytics," in Learning Analytics, Springer, pp. 61-75, 2014, [Online]. Available: https://doi.org/10.1007/978-1-4614-3305-7_4.

[4] G. Siemens, "Learning Analytics: The Emergence of a Discipline," American Behavioral Scientist, vol. 57, no. 10, pp. 1380-1400, 2013, [Online]. Available: https://doi.org/10.1177/0002764213498851.

[5] S.B. Kotsiantis, "Use of Machine Learning Techniques for Educational Proposes: A Decision Support System for Forecasting Students' Grades," Artificial Intelligence Review, vol. 37, no. 4, pp. 331-344, 2012, [Online]. Available: https://doi.org/10.1007/s10462-011-9234-x.

[6] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Elsevier, 2011, [Online]. Available: https://doi.org/10.1016/C2009-0-61819-5.

[7] M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," International Journal of Technology Enhanced Learning, vol. 4, no. 5-6, pp. 318-331, 2012, [Online]. Available: https://doi.org/10.1504/IJTEL.2012.051815.

[8] OECD, OECD Digital Education Outlook 2021. OECD Publishing, 2021, [Online]. Available: https://doi.org/10.1787/589b283f-en.

[9] C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12-27, 2013.

[10] R. Ferguson, "Learning Analytics: Drivers, Developments and Challenges," International Journal of Technology Enhanced Learning, vol. 4, no. 5-6, pp. 304-317, 2012, [Online]. Available: https://doi.org/10.1504/IJTEL.2012.051816.

[11] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, [Online]. Available: https://doi.org/10.1023/A:1010933404324.

[12] R.S.J. d. Baker, "Data Mining for Education," in International Encyclopedia of Education, Elsevier, 2010, [Online]. Available: https://doi.org/10.1016/B978-0-08-044894-7.01318-3.

[13] C. Romero and S. Ventura, "Data Mining in Education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12-27, 2013, [Online]. Available: https://doi.org/10.1002/widm.1075.

[14] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Review, vol. 46, no. 5, pp. 30-40, 2011.

[15] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education," Computers in Human Behavior, vol. 89, pp. 98-110, 2018, [Online]. Available: https://doi.org/10.1016/j.chb.2018.07.027.

[16] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," in The Adaptive Web, Springer, pp. 3-53, 2007, [Online]. Available: https://doi.org/10.1007/978-3-540-72079-9_1.

[17] K.R. Koedinger, S. D'Mello, E. McLaughlin, Z. Pardos, and C. Rosé, "Data mining and education," WIREs Cognitive Science, vol. 6, no. 4, pp. 333-353, 2015, [Online]. Available: https://doi.org/10.1002/wcs.1350.

[18] C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker (eds.), Handbook of Educational Data Mining. CRC Press, 2010.

[19] K.E. Arnold and M.D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success," in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 267-270, 2012, [Online]. Available: https://doi.org/10.1145/2330601.2330666.

[20] S. Dawson, D. Gašević, G. Siemens, and S. Joksimović, "Current state and future trends: A citation network analysis of the learning analytics field," in Proceedings of the 4th International Conference on Learning Analytics and Knowledge, pp. 231-240, 2014.

[21] OECD, OECD Digital Education Outlook 2021. OECD Publishing, 2021.

[22] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," IEEE Access, vol. 8, pp. 75264-75278, 2020, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2988510.

[23] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Elsevier, 2011.