Review

Available online at
www.heca-analitika.com/jeml

Journal of Educational Management and Learning

Vol. 3, No. 1, 2025

Techniques and Tools in Learning Analytics and Educational Data Mining: A Review

Teuku Rizky Noviandy [1], Ghazi Mauer Idroes [2], Maria Paristiowati [3] and Rinaldi Idroes [4,*]

[1] Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; rizky_si@abulyatama.ac.id (T.R.N.)
[2] Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I.)
[3] Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, Jakarta 13220, Indonesia; maria.paristiowati@unj.ac.id (M.P.)
[4] School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 9 March 2025
Revised 16 May 2025
Accepted 23 May 2025
Available Online 31 May 2025

Keywords:
Predictive modeling
Student engagement
Personalized learning
Dropout prediction
Curriculum analytics

Abstract

Learning analytics and educational data mining are rapidly evolving fields that leverage data-driven methods to enhance teaching, learning, and institutional decision-making. This review provides a comprehensive overview of the key analytical techniques and tools employed in learning analytics and educational data mining, including classification, clustering, regression, association rule mining, and data visualization. It also highlights the integration of advanced methods such as deep learning and adaptive systems for personalized education. The paper examines various platforms and technologies, including learning management systems, open-source tools, and AI/ML libraries, to evaluate their capabilities, scalability, and practical adoption. Key application areas, such as dropout prediction, engagement analysis, personalized learning, and curriculum design, are examined through selected case studies spanning K–12 and higher education. The review emphasizes the growing importance of ethical considerations, interpretability, and usability in the application of educational analytics. By synthesizing current practices and trends, this work aims to inform educators, researchers, and developers seeking to harness educational data for improved learning outcomes and strategic planning.

Copyright: © 2025 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License. (https://creativecommons.org/licenses/by-nc/4.0/)

1. Introduction

Learning analytics and educational data mining are interdisciplinary fields that focus on the collection, measurement, analysis, and interpretation of data about learners and their contexts [1–3]. Learning analytics is primarily concerned with understanding and optimizing learning and the environments in which it occurs, often involving stakeholders like educators, administrators, and students [4–6]. Educational Data Mining, on the other hand, emphasizes the development and application of computational and statistical methods to explore data generated from educational settings [7–9]. Although they share common goals, educational data mining tends to focus more on algorithmic and methodological rigor, while learning analytics often prioritizes actionable insights for decision-making in educational practice.

In the current era of digital education, data-driven approaches have become increasingly essential for

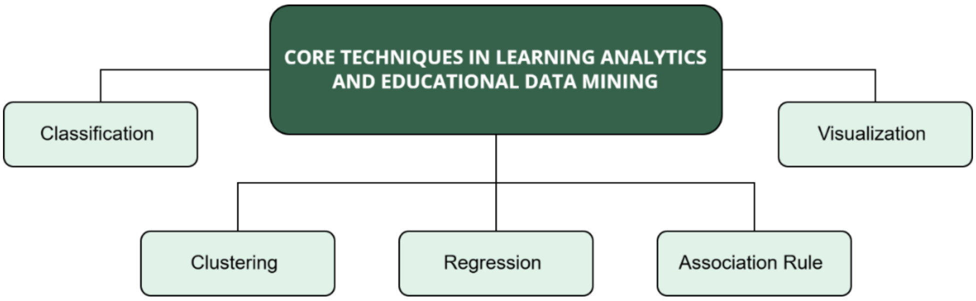

Figure 1. Core techniques in learning analytics and educational data mining.

enhancing learning outcomes, personalizing instruction, and informing policy decisions [10–12]. The proliferation of learning management systems (LMS) [13], intelligent tutoring systems [14], Massive Open Online Courses (MOOCs) [15], and other digital platforms has led to an unprecedented amount of educational data. This wealth of information enables researchers and practitioners to uncover patterns in student behavior, predict academic performance, and design interventions tailored to individual learning needs [16, 17]. Such approaches contribute significantly to improving the quality, accessibility, and efficiency of education.

As educational settings become more varied and complex, it's increasingly important to find effective ways to make sense of the data they generate. Techniques such as machine learning [18, 19], natural language processing [20, 21], and network analysis [22] provide valuable tools for understanding how students learn. Still, their success depends on addressing concerns such as data privacy, ethical use, and the ease with which results can be interpreted. With more institutions transitioning to online and hybrid learning, there is a growing need for flexible tools that can handle various types of data, such as click logs, written responses, and video content, and provide timely feedback. Choosing the right methods and tools depends on matching them to the specific goals of each learning environment.

The purpose of this review is to provide an overview of the key techniques and tools used in learning analytics and educational data mining, highlighting their applications across different educational settings. By examining the capabilities, limitations, and use cases of these approaches, this review aims to offer insights for educators, researchers, and developers seeking to leverage data to enhance teaching and learning processes.

2. Core Techniques in Learning Analytics and Educational Data Mining

The core techniques used in learning analytics and educational data mining encompass a broad range of statistical, machine learning, and visualization methods aimed at interpreting educational data and supporting evidence-based decision-making. These include classification, clustering, regression, association rule mining, and visualization, each serving a distinct purpose in analyzing and enhancing learning environments. These methods help uncover patterns in student behavior, predict academic outcomes, and inform the design of targeted interventions to support student success. Figure 1 illustrates these core techniques, providing a visual overview of the analytical methods that underpin much of the work in learning analytics and educational data mining.

2.1. Classification Techniques

Classification involves predicting a discrete label or category for a given input based on learned patterns from labeled data. Common classification tasks in education include predicting whether a student will pass or fail a course, identifying students at risk of academic difficulties, or detecting instances of cheating behavior.

- Decision Trees operate by recursively splitting the dataset into subsets based on feature values that maximize information gain (e.g., using entropy or Gini impurity) [23]. They are easy to interpret because they mimic human decision-making processes. For example, a tree might split on whether a student submitted more than 80% of assignments before predicting their final grade category.
- Support Vector Machines (SVMs) construct a hyperplane in a high-dimensional space to separate classes with the largest possible margin [24]. SVMs are particularly effective when there are many features and a limited number of instances, such as predicting dropout based on hundreds of behavioral metrics logged in an LMS.
- Logistic Regression, though technically a regression model, is widely used for binary classification (e.g., pass/fail). It estimates the probability of a binary outcome using a sigmoid function applied to a linear combination of input features [25].

2.2. Clustering Techniques

Clustering is an unsupervised learning method used to group similar data points based on their features, without predefined labels [26, 27]. This technique is particularly useful in segmenting students into behavioral profiles or identifying patterns in course engagement.

- K-Means Clustering partitions the data into k groups by minimizing the sum of squared distances between data points and their cluster centroid [26]. For example, students can be grouped based on time spent on tasks, forum activity, and quiz attempts to identify "highly engaged," "moderately engaged," and "disengaged" learners.
- Hierarchical Clustering builds nested clusters either bottom-up (agglomerative) or top-down (divisive), and is often visualized with a dendrogram [28]. It's useful when the number of clusters is unknown or when the structure among groups is of interest.

These methods help educators uncover hidden structures in student data, which can inform targeted interventions or curriculum adjustments.

2.3. Regression Analysis

Regression techniques model the relationship between one or more independent variables and a continuous or categorical outcome. They are often used to predict academic performance metrics, such as final exam scores, GPA, or time-on-task [29].

- Linear Regression estimates the relationship between input variables (e.g., attendance rate, assignment scores) and a continuous target variable (e.g., final grade) by fitting a line that minimizes the squared error. It assumes a linear relationship and is easy to interpret [30].
- Multivariate and Regularized Regressions (e.g., Lasso and Ridge regression) are employed when datasets have many correlated features. These methods help prevent overfitting by penalizing large coefficients, making them useful for high-dimensional student data [31].

In educational settings, regression analysis is often used to quantify the contribution of specific factors, such as participation, frequency of resource access, or time management, to academic outcomes.

2.4. Association Rule Mining

Association rule mining is used to identify frequent patterns, correlations, or associations among variables in large datasets [32]. It is particularly effective for understanding co-occurrence relationships in learning behaviors.

- Apriori Algorithm identifies frequent itemsets by iteratively expanding them and pruning those that do not meet minimum support thresholds [33]. For instance, it might reveal that students who frequently access lecture slides and complete quizzes early are also more likely to earn high grades.
- FP-Growth Algorithm improves on Apriori by using a prefix-tree structure to reduce computational complexity, making it suitable for larger educational datasets [34].

The output of these algorithms is a set of rules in the form *if A, then B*, typically evaluated by support (frequency), confidence (reliability), and lift (interestingness). These rules can inform curriculum design, recommend resources, or detect common pathways to success or failure.

2.5. Visualization Techniques

Visualization plays a crucial role in making complex data understandable and actionable for educators and learners [35]. Effective visual analytics enables users to identify patterns, outliers, and trends that may not be apparent in raw data.

- Dashboards aggregate student metrics such as attendance, grades, and participation in visual summaries [36]. These are commonly embedded in LMS platforms to support real-time monitoring.
- Heatmaps and Time-Series Plots visualize engagement over time, such as login frequencies or resource access peaks [37]. These tools can help identify periods of low activity that might warrant instructional intervention.
- Network Graphs are used to visualize social interactions in collaborative learning environments, such as discussion forums or peer feedback systems [38, 39]. Nodes represent users and edges represent interactions, helping to highlight isolated learners or overly dominant participants.

Visualization techniques support quick interpretation and are essential for feedback loops in learning environments. However, their effectiveness depends on thoughtful design and alignment with user needs.

2.6. Comparison of Traditional vs. Emerging Methods

Traditional data analysis methods, such as decision trees, linear regression, and k-means clustering, are valued for their interpretability, low computational cost, and ease of

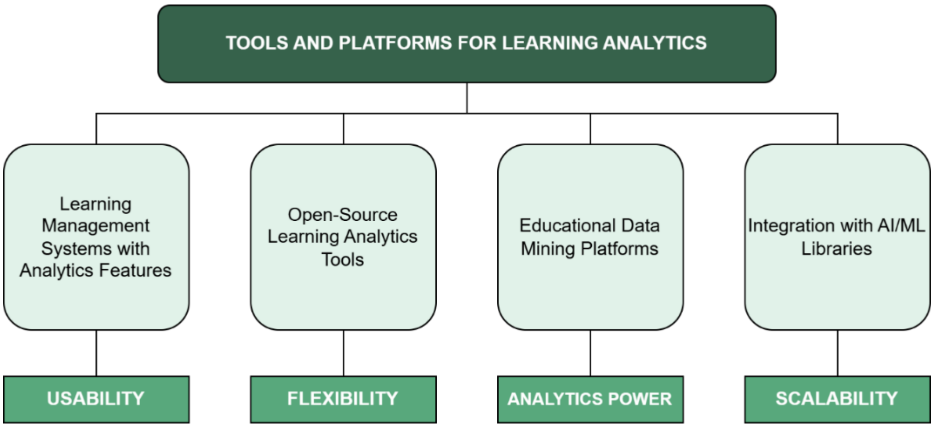

Figure 2. Tools and platforms for learning analytics are categorized by usability, flexibility, analytics power, and scalability.

implementation. They are especially useful in educational settings where transparency is essential for gaining stakeholder trust.

However, emerging techniques such as:

- Deep Neural Networks (DNNs), which consist of multiple layers of interconnected neurons, can model highly non-linear relationships and handle unstructured data like text and video, but are often considered "black boxes" due to their low interpretability [40].
- Ensemble Methods such as Random Forests and Gradient Boosting (e.g., XGBoost, CATBoost) combine multiple weak learners (like decision trees) to improve accuracy and robustness [41]. These are frequently used in predictive modeling tasks such as dropout prediction due to their strong performance on tabular data.

While emerging methods often outperform traditional ones in terms of accuracy and flexibility, they also present challenges, including interpretability, higher resource demands, and ethical concerns (e.g., bias in training data). Selecting the right approach requires balancing technical capabilities with the needs and constraints of the educational setting.

3. Tools and Platforms for Learning Analytics

The application of learning analytics and educational data mining relies heavily on a variety of tools and platforms designed to collect, process, analyze, and visualize educational data. These tools vary in terms of functionality, ease of use, scalability, and suitability for different educational settings. This section provides an overview of widely used tools and platforms in both research and practice, highlighting their features and implications for adoption.

As illustrated in Figure 2, tools for learning analytics can be grouped into four main categories based on their primary strengths: Learning Management Systems with analytics features, open-source learning analytics tools, educational data mining platforms, and integration with AI/ML libraries. These categories reflect the trade-offs between usability, flexibility, analytical power, and scalability that institutions must consider when selecting appropriate solutions.

3.1. Learning Management Systems with Analytics Features

LMS platforms such as Moodle, Canvas, and Blackboard have built-in analytics tools that offer real-time dashboards, progress tracking, and performance reporting [42]. For example, Canvas Analytics provides instructors with visual insights into student participation, grades, and submission timelines. These tools are widely adopted due to their integration within existing instructional environments, making them user-friendly and accessible to educators with limited technical expertise. However, analytical capabilities are often limited to descriptive statistics and predefined reports, which may not be sufficient for in-depth learning analytics investigations.

3.2. Open-Source Learning Analytics Tools

Open-source solutions offer flexibility and customization, making them popular among researchers and institutions aiming to tailor analytics systems to specific needs [43]. Tools built with R Shiny or Python Dash enable the creation of interactive, web-based dashboards that visualize trends in student data, such as attendance, engagement, and outcomes. These tools require programming skills but offer powerful capabilities for exploratory and diagnostic analytics, especially in research-intensive settings.

3.3. Educational Data Mining Platforms

General-purpose data mining tools, such as RapidMiner, WEKA, and Orange, are frequently used in educational data mining research [44]. These platforms support a wide range of machine learning algorithms and data preprocessing techniques through intuitive graphical user interfaces, enabling seamless integration and efficient data analysis. RapidMiner is known for its user-friendly workflow design and extensibility, while WEKA is widely adopted in academia due to its simplicity and robust algorithm library. Orange offers visual programming for interactive data analysis and has been increasingly used for educational data exploration. These tools are ideal for prototyping and testing machine learning models without requiring deep coding knowledge.

3.4. Integration with AI/ML Libraries

For advanced and customized analytics, integration with programming libraries such as Python, Scikit-learn, TensorFlow, and PyTorch is essential [45]. These libraries enable the development of sophisticated models for prediction, classification, natural language processing, and deep learning. Python, in particular, has become the dominant language in learning analytics and educational data mining due to its vast ecosystem and community support [46]. This approach is highly scalable and adaptable to diverse educational data types, but typically requires more technical expertise and computational resources.

3.5. Usability, Scalability, and Adoption Considerations

The choice of tool often depends on the user's technical background, institutional infrastructure, and the scope of analysis. LMS-based tools offer ease of use and are readily adopted in teaching-focused environments but may lack advanced analytical power [47]. Open-source and education-specific data mining platforms offer greater depth and flexibility, but require technical investment [48]. AI/ML library integration ensures high scalability and sophistication, ideal for large-scale learning or big data research [49].

4. Key Application Areas

Learning analytics and educational data mining are being increasingly applied across various educational settings to enhance learning outcomes, improve instructional strategies, and support institutional decision-making. The following are some of the most impactful application areas, supported by practical use cases in both K–12 and higher education settings.

4.1. Dropout Prediction and Early Intervention

One of the most common applications of learning analytics and educational data mining is predicting student dropout or academic failure. Machine learning models, such as decision trees, logistic regression, and support vector machines, are trained on historical student data, including grades, attendance records, and engagement metrics, to identify students at risk [50]. These predictive models enable early interventions, such as personalized academic support, counseling, or tailored communication strategies, helping institutions improve retention and student success.

4.2. Student Engagement and Behavior Analysis

Understanding how students interact with digital learning environments is crucial for enhancing motivation and achieving better outcomes. Engagement analysis uses behavioral data (e.g., clickstream logs, discussion forum activity, quiz attempts) to identify participation patterns [51, 52]. Clustering and time-series analysis can segment students based on behavior types, such as passive viewers versus active participants. These insights support the design of differentiated instruction and more engaging learning environments.

4.3. Adaptive Learning and Personalization

Adaptive learning systems aim to provide personalized pathways by adjusting content, difficulty, and pacing based on individual learner performance [53]. Techniques such as reinforcement learning, Bayesian modeling, and rule-based engines are employed to optimize content sequencing. These systems are particularly beneficial in large-scale online learning platforms and intelligent tutoring systems, where personalization improves both learner satisfaction and achievement.

4.4. Curriculum Design and Resource Planning

Data analytics also play a key role in curriculum refinement and institutional planning. By analyzing assessment results, resource usage, and course completion rates, educators can identify learning bottlenecks and areas requiring redesign [54]. Visualization tools and item analysis help instructors pinpoint ineffective materials or poorly performing test items. At the administrative level, aggregated data informs strategic decisions about course offerings, faculty allocation, and infrastructure investments.

4.5. Summary of Selected Studies

Table 1 highlights selected studies that demonstrate the practical application of learning analytics and educational

Table 1. Selected studies demonstrating learning analytics and educational data mining applications.

No.	Authors	Educational Level	Method(s)	Tool(s)	Key Findings
1	Yağcı [25]	Higher Education	Random Forest, SVM, Logistic Regression, Naïve Bayes, k-NN	Academic Records (Midterm, Department, Faculty Data)	ML models predicted final exam grades with 70–75% accuracy, supporting the early identification of at-risk students with limited input data, which is valuable for learning analytics in higher education.
2	Alruwais and Zakariah [51]	Higher Education	CATBoost, XGBoost, Random Forest, MLP	VLE Data	CATBoost outperformed other models in detecting student engagement with ~92.23% accuracy, 100% recall, and an AUC of 0.9624; it surpassed the AISAR model in all metrics.
3	Demartini et al. [53]	Primary & Secondary Education	Learning Analytics, AI-based Dashboard	Adaptive Learning Tools	Developed an AI-based dashboard to enhance adaptive learning and support educators; aimed to reduce dropout, improve collaboration, writing, and computational thinking; highlighted resistance in K-12 to analytics adoption.
4	Goren et al. [55]	Higher Education	Neural Networks, XGBoost	LMS & Administrative Data	Developed a 'studentship' feature combining cognitive and social elements, enhancing early dropout prediction accuracy. Link
5	Johar et al. [56]	Higher Education	Systematic Literature Review	Learning Analytics	Identified behavioral engagement as the most analyzed dimension; emphasized the need for multifaceted engagement analysis. Link
6	Peng et al. [57]	Higher Education	Smart Learning Environments	Adaptive Learning Systems	Proposed a framework for personalized adaptive learning, focusing on individual characteristics and performance. Link
7	Park et al. [58]	Higher Education	Decision Tree, Random Forest (RF), SVM, Deep Neural Networks (DNN)	LMS (Cyber University) & Learner Statistics	Using 7 years of LMS data, RF demonstrated the highest prediction accuracy for early dropout; DNN was also found to be suitable for integration into LMS systems.
8	Zhang et al. [59]	Higher Education	Data Mining (Classification, Association Rules)	University IMS	Applied data mining to improve university information systems; identified course enrollment patterns and enhanced administrative data analysis and decision-making capabilities.
9	Koc and Akın [60]	Secondary Education	Beta Regression, Support Vector Regression, Random Forest, Decision Tree	Socioeconomic & Educational Statistics	Beta regression and random forest models effectively predicted high school entrance exam success rates; key predictors included the divorce rate, GDP, and illiteracy rates. Random forest showed slightly better overall performance.
10	Aldhafeeri and Alotaibi [61]	Secondary Education	Experimental Study (DES Model)	Engagement Checklist Survey	The Digital Education Shifting (DES) model improved both internal and observable engagement in online classes, demonstrating that digital learning, when enhanced with teacher-student interaction, can rival traditional education in effectiveness.

data mining techniques. These cases demonstrate how institutions have leveraged data to predict dropout risks, enhance student engagement, personalize instruction, and refine curriculum design.

The studies summarized in Table 1 reflect the diverse and impactful ways in which Learning Analytics and Educational Data Mining are being applied across educational levels. A common trend among these cases is the use of machine learning models, such as Random Forest, XGBoost, and Neural Networks, for tasks like dropout prediction and engagement analysis, demonstrating their effectiveness in producing actionable insights.

Several studies focus on higher education, where large-scale data from LMS platforms support personalized learning, early warning systems, and curriculum optimization. Notably, adaptive learning frameworks and AI dashboards show promising outcomes in improving learner satisfaction and performance. At the same time, applications in K–12 education highlight both the potential and the challenges, such as resistance to technology adoption and integrating analytics in more structured and traditional environments.

5. Conclusion and Future Directions

The integration of advanced techniques and tools in learning analytics and educational data mining has significantly expanded the ability of educators, researchers, and institutions to enhance educational outcomes. By combining statistical models, machine learning algorithms, and interactive visualization tools, learning analytics and educational data mining provide actionable insights into learner behaviors, instructional effectiveness, and institutional processes. The synergy between these analytical techniques and technological platforms enables data-informed decisions that are more targeted, timely, and scalable.

Despite the progress in the field, several challenges remain. Data quality remains a significant concern, as inconsistent, incomplete, or biased data can compromise the accuracy and fairness of analytics outcomes. Tool accessibility and scalability are also critical issues, particularly in under-resourced educational settings where institutions may lack the infrastructure or technical expertise to adopt advanced analytics systems. Furthermore, educator training is essential to ensure that insights derived from learning analytics and educational data mining are effectively interpreted and applied. Without adequate professional development, the full potential of these technologies may go unrealized.

Looking ahead, there are numerous opportunities to advance the impact of learning analytics and educational data mining. The development of real-time analytics systems can provide learners and instructors with immediate feedback, fostering more dynamic and responsive educational environments. Additionally, the incorporation of ethical AI principles, such as transparency, privacy protection, and bias mitigation, will be crucial to building trust and ensuring that analytics systems support equitable learning experiences. As the volume and complexity of educational data continue to grow, so does the potential for intelligent, personalized, and ethical learning systems.

Finally, realizing the full promise of learning analytics and educational data mining requires interdisciplinary collaboration. Expertise from computer science, education, psychology, data ethics, and domain-specific pedagogies must converge to design analytics systems that are not only technically sound but also pedagogically meaningful and socially responsible. As the field evolves, continued collaboration among educators, researchers, developers, and policymakers will be essential to shaping the future of data-informed education.

Author Contributions: Conceptualization, T.R.N. and G.M.I.; methodology, T.R.N., M.P. and R.I.; software, T.R.N. and M.P.; validation, T.R.N., M.P. and R.I.; formal analysis, T.R.N. and G.M.I.; investigation, T.R.N. and G.M.I.; resources, M.P. and R.I.; data curation, G.M.I. and M.P.; writing—original draft preparation, T.R.N. and G.M.I.; writing—review and editing, M.P. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Data Availability Statement: All data discussed are available within the article itself.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Baek, C., and Doleck, T. (2023). Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019, *Interactive Learning Environments*, Vol. 31, No. 6, 3828–3850. doi:10.1080/10494820.2021.1943689.
2. Rabelo, A., Rodrigues, M. W., Nobre, C., Isotani, S., and Zárate, L. (2024). Educational Data Mining and Learning Analytics: A Review of Educational Management in e-Learning, *Information Discovery and Delivery*, Vol. 52, No. 2, 149–163. doi:10.1108/IDD-10-2022-0099.
3. Romero, C., and Ventura, S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey, *WIREs Data Mining and Knowledge Discovery*, Vol. 10, No. 3. doi:10.1002/widm.1355.
4. Kew, S. N., and Tasir, Z. (2022). Learning Analytics in Online Learning Environment: A Systematic Review on the Focuses and the Types of Student-Related Analytics Data, *Technology, Knowledge and Learning*, Vol. 27, No. 2, 405–427. doi:10.1007/s10758-021-09541-2.

5. Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., and Yamada, M. (2021). Putting Learning Back into Learning Analytics: Actions for Policy Makers, Researchers, and Practitioners, *Educational Technology Research and Development*, Vol. 69, No. 4, 2131–2150. doi:10.1007/s11423-020-09909-8.

6. Gray, G., Schalk, A. E., Cooke, G., Murnion, P., Rooney, P., and O'Rourke, K. C. (2022). Stakeholders' Insights on Learning Analytics: Perspectives of Students and Staff, *Computers & Education*, Vol. 187, 104550. doi:10.1016/j.compedu.2022.104550.

7. Shu, X., and Ye, Y. (2023). Knowledge Discovery: Methods from Data Mining and Machine Learning, *Social Science Research*, Vol. 110, 102817. doi:10.1016/j.ssresearch.2022.102817.

8. Alam, A. (2023). Improving Learning Outcomes through Predictive Analytics: Enhancing Teaching and Learning with Educational Data Mining, *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 249–257. doi:10.1109/ICICCS56967.2023.10142392.

9. Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., and Damasevicius, R. (2021). Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review, *Applied Artificial Intelligence*, Vol. 35, No. 13, 983–1021. doi:10.1080/08839514.2021.1922847.

10. Idroes, R., Subianto, M., Zahriah, Z., Afidh, R. P. F., Irvanizam, I., Noviandy, T. R., Sugara, D. R., Mursyida, W., Zhilalmuhana, T., and Idroes, G. M. (2023). Digital Transformations in Vocational High School: A Case Study of Management Information System Implementation in Banda Aceh, Indonesia, *Journal of Educational Management and Learning*, Vol. 1, No. 2, 48–54. doi:10.60084/jeml.v1i2.128.

11. Lin, L., Zhou, D., Wang, J., and Wang, Y. (2024). A Systematic Review of Big Data Driven Education Evaluation, *Sage Open*, Vol. 14, No. 2. doi:10.1177/21582440241242180.

12. Gao, P., Li, J., and Liu, S. (2021). An Introduction to Key Technology in Artificial Intelligence and Big Data Driven E-Learning and e-Education, *Mobile Networks and Applications*, Vol. 26, No. 5, 2123–2126. doi:10.1007/s11036-021-01777-7.

13. Alturki, U., and Aldraiweesh, A. (2021). Application of Learning Management System (LMS) during the COVID-19 Pandemic: A Sustainable Acceptance Model of the Expansion Technology Approach, *Sustainability*, Vol. 13, No. 19, 10991. doi:10.3390/su131910991.

14. Lin, C.-C., Huang, A. Y. Q., and Lu, O. H. T. (2023). Artificial Intelligence in Intelligent Tutoring Systems toward Sustainable Education: A Systematic Review, *Smart Learning Environments*, Vol. 10, No. 1, 41. doi:10.1186/s40561-023-00260-y.

15. Stracke, C. M., and Trisolini, G. (2021). A Systematic Literature Review on the Quality of MOOCs, *Sustainability*, Vol. 13, No. 11, 5817. doi:10.3390/su13115817.

16. Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., and Gates, K. M. (2022). Predicting Student Outcomes Using Digital Logs of Learning Behaviors: Review, Current Standards, and Suggestions for Future Work, *Behavior Research Methods*, Vol. 55, No. 6, 3026–3054. doi:10.3758/s13428-022-01939-9.

17. Dwivedi, D. N., Mahanty, G., and Dwivedi, V. nath. (2024). The Role of Predictive Analytics in Personalizing Education, 44–59. doi:10.4018/979-8-3693-2169-0.ch003.

18. Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions, *Computers*, Vol. 12, No. 5, 91. doi:10.3390/computers12050091.

19. Guleria, P., and Sood, M. (2023). Explainable AI and Machine Learning: Performance Evaluation and Explainability of Classifiers on Educational Data Mining Inspired Career Counseling, *Education and Information Technologies*, Vol. 28, No. 1, 1081–1116. doi:10.1007/s10639-022-11221-2.

20. Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., and Galligan, L. (2022). A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis, *IEEE Access*, Vol. 10, 56720–56739. doi:10.1109/ACCESS.2022.3177752.

21. Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., and Albekairy, A. M. (2023). The Emergent Role of Artificial Intelligence, Natural Learning Processing, and Large Language Models in Higher Education and Research, *Research in Social and Administrative Pharmacy*, Vol. 19, No. 8, 1236–1242. doi:10.1016/j.sapharm.2023.05.016.

22. Wulff, P. (2023). Network Analysis of Terms in the Natural Sciences Insights from Wikipedia through Natural Language Processing and Network Analysis, *Education and Information Technologies*, Vol. 28, No. 11, 14325–14346. doi:10.1007/s10639-022-11531-5.

23. Mienye, I. D., and Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications, *IEEE Access*, Vol. 12, 86716–86727. doi:10.1109/ACCESS.2024.3416838.

24. Li, H. (2024). Support Vector Machine, *Machine Learning Methods*, Springer Nature Singapore, Singapore, 127–177. doi:10.1007/978-981-99-3917-6_7.

25. Yağcı, M. (2022). Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms, *Smart Learning Environments*, Vol. 9, No. 1, 11. doi:10.1186/s40561-022-00192-z.

26. Noviandy, T. R., Hardi, I., Zahriah, Z., Sofyan, R., Sasmita, N. R., Hilal, I. S., and Idroes, G. M. (2024). Environmental and Economic Clustering of Indonesian Provinces: Insights from K-Means Analysis, *Leuser Journal of Environmental Studies*, Vol. 2, No. 1, 41–51. doi:10.60084/ljes.v2i1.181.

27. Sasmita, N. R., Khairul, M., Sofyan, H., Kruba, R., Mardalena, S., Dahlawy, A., Apriliansyah, F., Muliadi, M., Saputra, D. C. E., Noviandy, T. R., and Watsiq Maula, A. (2023). Statistical Clustering Approach: Mapping Population Indicators Through Probabilistic Analysis in Aceh Province, Indonesia, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 63–72. doi:10.60084/ijds.v1i2.130.

28. Ran, X., Xi, Y., Lu, Y., Wang, X., and Lu, Z. (2023). Comprehensive Survey on Hierarchical Clustering Algorithms and the Recent Developments, *Artificial Intelligence Review*, Vol. 56, No. 8, 8219–8264. doi:10.1007/s10462-022-10366-3.

29. Hahs-Vaughn, D. L. (2021). Multivariate Statistics, *Applied Multivariate Statistical Concepts*. doi:10.4324/9781315816685-6.

30. Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., and Gopakumar, G. (2021). Educational Data Mining for Predicting Students' Academic Performance Using Machine Learning Algorithms, *Materials Today: Proceedings*, Vol. 47, 5260–5267. doi:10.1016/j.matpr.2021.05.646.

31. Kour, S., Kumar, R., and Gupta, M. (2021). Analysis of Student Performance Using Machine Learning Algorithms, *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, 1395–1403. doi:10.1109/ICIRCA51532.2021.9544935.

32. Dol, S. M., and Jawandhiya, P. M. (2023). Classification Technique and Its Combination with Clustering and Association Rule Mining in Educational Data Mining — A Survey, *Engineering Applications of Artificial Intelligence*, Vol. 122, 106071. doi:10.1016/j.engappai.2023.106071.

33. Chen, S., Xue, Y., and Cui, X. (2024). Information Literacy of College Students from Library Education in Smart Classrooms: Based on Big Data Exploring Data Mining Patterns Using Apriori Algorithm, *Soft Computing*, Vol. 28, No. 4, 3571–3589. doi:10.1007/s00500-023-09621-8.

34. Yu, H. (2024). Research on the Application of Association Rule Algorithm Based on FP-Growth in College English Information Teaching, *Proceedings of the International Conference on*

Modeling, Natural Language Processing and Machine Learning, ACM, New York, NY, USA, 23–27. doi:10.1145/3677779.3677783.

35. Feng, G., and Fan, M. (2024). Research on Learning Behavior Patterns from the Perspective of Educational Data Mining: Evaluation, Prediction and Visualization, *Expert Systems with Applications*, Vol. 237, 121555. doi:10.1016/j.eswa.2023.121555.

36. Ramaswami, G., Susnjak, T., Mathrani, A., and Umer, R. (2023). Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies, *Technology, Knowledge and Learning*, Vol. 28, No. 3, 959–980. doi:10.1007/s10758-022-09613-x.

37. DOBASHI, K., HO, C. P., FULFORD, C. P., and Lin, M.-F. G. (2019). A Heat Map Generation to Visualize Engagement in Classes Using Moodle Learning Logs, *2019 4th International Conference on Information Technology (InCIT)*, IEEE, 138–143. doi:10.1109/INCIT.2019.8912068.

38. Becheru, A., Calota, A., and Popescu, E. (2018). Analyzing Students' Collaboration Patterns in a Social Learning Environment Using StudentViz Platform, *Smart Learning Environments*, Vol. 5, No. 1, 18. doi:10.1186/s40561-018-0063-0.

39. Rabbany, R., Elatia, S., Takaffoli, M., and Zaïane, O. R. (2014). Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective, 441–466. doi:10.1007/978-3-319-02738-8_16.

40. Huang Yi, Sun Shiyu, Duan Xiusheng, and Chen Zhigang. (2016). A Study on Deep Neural Networks Framework, *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, IEEE, 1519–1522. doi:10.1109/IMCEC.2016.7867471.

41. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

42. Suryadevara, G., and Pachipulusu, P. (2025). Integrating Real-Time Data Streams, 67–90. doi:10.4018/979-8-3373-3952-8.ch003.

43. Ahammad, N., Saiful Bahri, F. D., and Husaini, H. (2024). Exploring the Open-Source Impact on Bangladesh Academic Library Service Sustainability, *Journal of Information, Communication and Ethics in Society*, Vol. 22, No. 4, 478–493. doi:10.1108/JICES-06-2024-0080.

44. Thilagaraj, T., and Sengottaiyan, N. (2017). A Review of Educational Data Mining in Higher Education System, 349–358. doi:10.15439/2017R87.

45. Raschka, S., Patterson, J., and Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence, *Information*, Vol. 11, No. 4, 193. doi:10.3390/info11040193.

46. Ihantola, P., Vihavainen, A., Ahadi, A., Butler, M., Börstler, J., Edwards, S. H., Isohanni, E., Korhonen, A., Petersen, A., Rivers, K., Rubio, M. Á., Sheard, J., Skupas, B., Spacco, J., Szabo, C., and Toll, D. (2015). Educational Data Mining and Learning Analytics in Programming, *Proceedings of the 2015 ITiCSE on Working Group Reports*, ACM, New York, NY, USA, 41–63. doi:10.1145/2858796.2858798.

47. Munna, M. S. H., Hossain, M. R., and Saylo, K. R. (2024). Digital Education Revolution: Evaluating LMS-Based Learning and Traditional Approaches, *Journal of Innovative Technology Convergence*, Vol. 6, No. 2, 21–40. doi:10.69478/JITC2024v6n002a03.

48. Sáiz-Manzanares, M. C., Rodríguez-Díez, J. J., Díez-Pastor, J. F., Rodríguez-Arribas, S., Marticorena-Sánchez, R., and Ji, Y. P. (2021). Monitoring of Student Learning in Learning Management Systems: An Application of Educational Data Mining Techniques, *Applied Sciences*, Vol. 11, No. 6, 2677. doi:10.3390/app11062677.

49. Althati, C., Tomar, M., and Shanmugam, L. (2024). Enhancing Data Integration and Management: The Role of AI and Machine Learning in Modern Data Platforms, *Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023*, Vol. 2, No. 1, 220–232. doi:10.60087/jaigs.v2i1.154.

50. Noviandy, T. R., Zahriah, Z., Yandri, E., Jalil, Z., Yusuf, M., Yusof, N. I. S. M., Lala, A., and Idroes, R. (2024). Machine Learning for Early Detection of Dropout Risks and Academic Excellence: A Stacked Classifier Approach, *Journal of Educational Management and Learning*, Vol. 2, No. 1, 28–34. doi:10.60084/jeml.v2i1.191.

51. Alruwais, N., and Zakariah, M. (2023). Student-Engagement Detection in Classroom Using Machine Learning Algorithm, *Electronics*, Vol. 12, No. 3, 731. doi:10.3390/electronics12030731.

52. Maulana, A., Idroes, G. M., Kemala, P., Maulydia, N. B., Sasmita, N. R., Tallei, T. E., Sofyan, H., and Rusyana, A. (2023). Leveraging Artificial Intelligence to Predict Student Performance: A Comparative Machine Learning Approach, *Journal of Educational Management and Learning*, Vol. 1, No. 2, 64–70. doi:10.60084/jeml.v1i2.132.

53. Demartini, C. G., Sciascia, L., Bosso, A., and Manuri, F. (2024). Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study, *Sustainability*, Vol. 16, No. 3, 1347. doi:10.3390/su16031347.

54. Bennett, L., and Abusalem, A. (2024). Artificial Intelligence (AI) and Its Potential Impact on the Future of Higher Education, *Athens Journal of Education*, Vol. 11, No. 3, 195–212. doi:10.30958/aje.11-3-2.

55. Goren, O., Cohen, L., and Rubinstein, A. (2024). Early Prediction of Student Dropout in Higher Education Using Machine Learning Models, *Proceedings of the 17th International Conference on Educational Data Mining*, 349–359.

56. Johar, N. A., Kew, S. N., Tasir, Z., and Koh, E. (2023). Learning Analytics on Student Engagement to Enhance Students' Learning Performance: A Systematic Review, *Sustainability*, Vol. 15, No. 10, 7849. doi:10.3390/su15107849.

57. Peng, H., Ma, S., and Spector, J. M. (2019). Personalized Adaptive Learning: An Emerging Pedagogical Approach Enabled by a Smart Learning Environment, *Smart Learning Environments*, Vol. 6, No. 1, 9. doi:10.1186/s40561-019-0089-y.

58. Park, H. S., and Yoo, S. J. (2021). Early Dropout Prediction in Online Learning of University Using Machine Learning, *JOIV : International Journal on Informatics Visualization*, Vol. 5, No. 4, 347. doi:10.30630/joiv.5.4.732.

59. Zhang, M., Fan, J., Sharma, A., and Kukkar, A. (2022). Data Mining Applications in University Information Management System Development, *Journal of Intelligent Systems*, Vol. 31, No. 1, 207–220. doi:10.1515/jisys-2022-0006.

60. Koc, T., and Akin, P. (2022). Estimation of High School Entrance Examination Success Rates Using Machine Learning and Beta Regression Models, *Journal of Intelligent Systems: Theory and Applications*, Vol. 5, No. 1, 9–15. doi:10.38016/jista.922663.

61. Aldhafeeri, F. M., and Alotaibi, A. A. (2022). Effectiveness of Digital Education Shifting Model on High School Students' Engagement, Education and Information Technologies, Vol. 27, No. 5, 6869–6891. doi:10.1007/s10639-021-10879-4.