UNSUPERVISED AND SEMI-SUPERVISED LEARNING TECHNIQUES IN CONTEMPORARY EDUCATIONAL APPLICATION

Veljko Aleksić*
[0000-0003-2337-1288]

University of Kragujevac,
Faculty of Technical Sciences in Čačak,
Čačak, Serbia

Abstract:

The paper examines the educational application of unsupervised and semi-supervised learning techniques. The comprehensive analysis evaluated diverse approaches, methods, and algorithms. Findings indicate that k-means clustering effectively differentiated student performance groups, while dimensionality reduction techniques offered valuable visualization capabilities for complex educational data. The semi-supervised learning paradigm demonstrated particular utility in environments characterized by abundant unlabelled data. The effectiveness of the presented analytical approaches significantly depends on data quality, appropriate algorithm selection, and domain expertise. As educational datasets grow increasingly complex, various computational methods will become essential in developing personalized learning, adaptive educational interventions, and innovative evidence-based teaching practices. The research contributes to the ever-evolving field of educational technology by systematically evaluating the strengths and limitations of various machine learning and artificial intelligence approaches, providing a foundation for future research.

Keywords:

Unsupervised Learning, Semi-Supervised Learning, Educational Technology, Algorithms.

INTRODUCTION

Artificial intelligence and machine learning are closely related terms often used interchangeably or as synonyms. However, they represent distinct concepts within the field of computer science. Artificial intelligence refers to the development of computational systems capable of performing tasks that typically require human intelligence, such as audio-visual perception, speech recognition, language translation, and even decision-making. The ultimate objective of artificial intelligence technology is the creation of virtual machines capable of thinking and acting in ways that simulate or surpass human cognitive abilities [1]. The domain of artificial intelligence encompasses a broad spectrum of approaches and techniques, including rule-based systems, expert systems, neural networks, and evolutionary algorithms [2]. These approaches are aimed at enabling computers to reason, learn, adapt, and solve problems in ways that emulate human intelligence [3]. It is noteworthy that artificial intelligence has, since its inception in the mid-20th century, progressed through cycles of increased and diminished enthusiasm and research investment [4].

Correspondence:
Veljko Aleksić

e-mail:
veljko.aleksic@ftn.kg.ac.rs

DOI: 10.15308/Sinteza-2025-259-266

Despite these fluctuations, significant advancements have been achieved in recent decades, driven by developments in computer hardware and software, networking capabilities, the creation and accessibility of very large datasets (e.g., Big Data), and innovative algorithms. Machine learning should be perceived as a subfield of artificial intelligence focused on developing algorithms and statistical models that enable computers to learn and improve performance on specific tasks without explicit programming [5]. The basic idea is that computer software can learn from data, identify patterns, and make decisions or predictions with minimal human intervention [6]. Machine learning algorithms are generally categorized into three primary types: Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL). The field of machine learning has advanced significantly in recent years, particularly with the emergence of the Deep Learning (DL) model. DL should be perceived as a subdivision of machine learning that employs Artificial Neural Networks (ANNs) with multiple layers for learning and processing hierarchical data representations [7]. DL has achieved remarkable success in tasks such as image classification, speech recognition, and natural language processing. UL is a branch of machine learning directed toward uncovering hidden patterns, structures, and relationships within unlabelled data [8]. Unlike supervised learning, where algorithms learn from labeled input-output pairs, UL algorithms learn from the data with no explicit labels or target outputs to identify inherent structures, similarities, or clusters. UL algorithms possess significant potential in educational data mining and learning analytics, enabling the discovery of educational data hidden patterns, structures, and relationships. These algorithms can provide valuable insights into student behaviors, learning processes, and knowledge acquisition, thereby facilitating the design of personalized learning experiences, adaptive interventions, and curriculum enhancements [9] [10].

2. UNSUPERVISED LEARNING

Educational datasets often contain substantial quantities of unrelated data, including selection/click data, textual information, or multimedia content that is challenging to analyze and interpret without the assistance of UL algorithms. In the educational context, these algorithms are employed to address clustering challenges; grouping students based on their educational behaviors, preferences, or academic achievement levels; creating recommendations for elective subjects and course selection; and organizing diverse educational resources [11]. UL algorithms can generally be classified into three categories:

- *Grouping (clustering) algorithms* partition input instances into unrelated discrete or overlapping groups, wherein instances within each group have greater similarity to one another compared to the instances in other groups. Prominent clustering algorithms are *k-means, Hierarchical Clustering, and Gaussian Mixture Models* (GMM) [12]. The k-means algorithm was applied to group students based on their academic performance and engagement levels in a blended learning environment [13], identifying four distinct student groups: high-performing, average, low-performing, and disengaged students, thereby assisting teachers in recognizing students at greater risk of academic failure and providing timely support and feedback;
- *Dimensionality reduction algorithms* transform high-dimensional input space into lower-dimensional representations while preserving the essential structure and information within the data [14]. This technique facilitates the visualization of complex datasets such as student interaction networks or conceptual maps, which can reduce computational complexity and enhance learning task performance. By reducing data dimensionality, educators can identify the most significant characteristics or variables that contribute to student learning outcomes and uncover hidden patterns or relationships among them. For instance, dimensionality reduction algorithms have demonstrated considerable usefulness in visualizing student learning trajectories within Massive Open Online Courses (MOOCs). Students who followed structured and linear pathways through MOOCs exhibited superior completion rates and higher overall results compared to those pursuing exploratory and non-linear trajectories [15]. Prominent dimensionality reduction algorithms are *Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding* (t-SNE), and autoencoders [16].
- *Representation learning algorithms* derive new representations of input instances that detect fundamental variation factors or explanatory factors within the data. These learned representations can serve as input characteristics for supervised learning (classification or regression. both)

or unsupervised learning tasks (clustering or dimensionality reduction). In educational contexts, representation learning algorithms are used to extract meaningful and interpretable characteristics from educational data (e.g., student essays, forum posts, or lecture videos). The features identified and learned through these methods enhance automated essay scoring, knowledge tracking, and personalized recommendations or student feedback generation. Prominent representation learning algorithms are *Deep Belief Networks* (DBNs), *Restricted Boltzmann Machines* (RBMs), and variational autoencoders.

The previously mentioned algorithms enable researchers and educational practitioners to distill complex, multifaceted educational data into interpretable or structured representations that reveal fundamental learning patterns while maintaining statistical fidelity to the original information structure. Such approaches are essential as educational datasets grow both in size and complexity, necessitating sophisticated analytical techniques that can extract meaningful insights without overwhelming computational resources or analytical capacity. By identifying latent features and structures within educational data, UL algorithms provide educators with deeper insights into student learning patterns, conceptual understanding, and engagement behaviors that may not be immediately apparent through traditional analysis methods, thus significantly enhancing the potential for adaptive learning systems capable for dynamic respond to individual student needs and learning trajectories.

2.1. *k*-MEANS CLUSTERING METHOD

Clustering is one of the most frequently utilized UL techniques in educational data mining and learning analytics [17]. Clustering algorithms partition datasets into groups or clusters, ensuring that the data within a specific cluster exhibit greater similarity to each other than the data in other clusters. Similarity between the data is typically measured using metric distances, such as Euclidean distance or cosine similarity [18]. K-means clustering is a prominent partitional clustering algorithm that iteratively assigns data points to k clusters based on their similarity to cluster centroids. The algorithm starts with random initialization of k cluster centroids, followed by the two steps alternating execution: (1) assigning each data point to its nearest cluster center, and (2) updating centroids based on the average value of data points assigned to each cluster. This process repeats until convergence is achieved, i.e., when assignments can no longer be changed [19]. In educational environments, the k-means clustering method is applied for grouping students based on their performance, learning behaviors, or engagement patterns [20]. For instance, this algorithm was employed to identify distinct student groups in MOOCs based on their interaction patterns with learning materials, quizzes, and discussion forums [21], revealing four clusters: active learners, passive learners, assessment-oriented learners, and dropout students. Clustering algorithms can also be utilized to group educational resources and evaluate their similarity in personalized education recommendations and curriculum design [22]. K-means algorithms are used for grouping educational videos based on content characteristics and usage patterns, thus enhancing personalized recommendations. These insights can inform teachers and other educators to adapt instructional design and support strategies for various student groups.

Hierarchical clustering can be performed by merging smaller clusters into larger ones (i.e., agglomerative clustering) or dividing larger clusters into smaller ones (i.e., divisive clustering). Agglomerative hierarchical clustering begins with each data point as a separate cluster and iteratively merges the closest clusters until a singular cluster is formed. Conversely, the divisive hierarchical clustering starts with all data points in one cluster and recursively divides clusters until each data point becomes a cluster [23]. In the educational context, hierarchical clustering is used to explore educational resources, curricula, and knowledge domain structures [24].

2.2. PRINCIPAL COMPONENT ANALYSIS

Dimensionality reduction represents another critical UL domain that aims to transform high-dimensional data into lower-dimensional representations while preserving the original essential structure and information. These techniques are particularly valuable for complex educational data visualization and exploration, such as student performance across multiple sequential assessments or the relationships between varying skill levels and knowledge components [25]. Principal Component Analysis (PCA) represents a frequently utilized linear dimensionality reduction technique that projects high-dimensional data onto a lower-dimensional subspace based on principal components that represent orthogonal directions capturing maximum variance within the data [26]. The first principal component corresponds to the direction of the highest variance, with subsequent components progressively capturing smaller variance portions while maintaining orthogonality to preceding components. The retention of only the highest-ranked principal components can effectively reduce data dimensionality while preserving critical information. This technique is often used in educational data mining for student achievement data analysis and visualization, identification of the key factors influencing learning outcomes, and student behavioral patterns detection [27]. PCA can also be used for exploring relationships between student motivation, cognitive engagement, and academic performance in e-learning environments [28]. Numerous educational datasets contain a large number of characteristics and redundant or irrelevant information, impeding analytical performance and interpretability. PCA can serve as an effective mechanism for identifying the most informative characteristics and data dimensionality reduction, thus enhancing the efficiency and effectiveness of data mining and machine learning algorithms.

The widespread application of PCA in education is the interpretation of principal components because they are linear combinations of the original characteristics and may not have clear pedagogical or other specific meanings. In addition, PCA is sensitive to scaling or normalization of the original characteristics, as various pre-processing methods can lead to different results and interpretations. Educational datasets often contain a mixture of continuous, categorical, and binary characteristics, along with missing values, necessitating meticulous cleaning and transformation of the data before the analysis. As PCA assumes linear relationships between original characteristics it may fail to identify nonlinear structures or interactions in educational data. Nevertheless, PCA techniques continue to advance and are expected to play a significant role in data-driven decision-making, personalized learning, and evidence-based educational interventions.

2.3. *t*-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING

t-Distributed Stochastic Neighbor Embedding (t-SNE) represents a nonlinear dimensionality reduction technique that is particularly efficient for visualizing 2D or 3D high-dimensional data. This methodology preserves local data structure by modeling similarities between data points in high-dimensional space and embedding them in low-dimensional space, thereby minimizing the divergence between probability distributions representing similarity pairs in both high- and low-dimensional spaces [29]. In comparison to the PCA, this technique is more suitable for recognizing nonlinear relationships and preserving local data structure. However, it requires more processing time and is not the most efficient solution if the global data structure should be preserved [30]. By focusing on preserving local similarities between samples, this technique can reveal complex patterns and clusters in the data that may be "hidden" in linear projections. t-SNE requires very complex computing, exhibits sensitivity to hyperparameters selection, and encounters difficulties in preserving global data structure. In educational contexts, t-SNE is often used for visualizing and exploring complex student data, such as learning paths, skill mastery patterns, and, notably, misconception analysis. t-SNE is also utilized for visualizing and comparing student subgroups in educational data, thereby detecting various clustered groups of successful and unsuccessful students based on the identified key factors. When t-SNE is applied, the appropriate hyperparameters should be selected (e.g., controlling the balance between preserving local and global data structure and learning rate). As hyperparameters can significantly influence the quality and interpretability of the resulting visualization, different values are appropriate for various datasets and tasks [31]. Working with large educational datasets creates the problem of scalability, as the computational complexity of this technique quadratically increases with the number of instances. In addition, proper interpretation and evaluation require domain expertise, as although t-SNE can reveal interesting data patterns and clusters, the actual meaning or relevance of these patterns may not always be obvious in the educational context.

2.4. ASSOCIATION RULE MINING

Association Rule Mining (ARM) is a popular UL technique often used for discovering interesting relationships and patterns within large transactional or categorical datasets. This method aims to identify item sets that frequently co-occur and to generate associative rules describing appearance patterns and their interdependencies. ARM is implemented in two primary steps: (1) generating frequent item sets (e.g., the identification of sets that satisfy established constraints), and (2) generating rules from frequent item sets that meet specified criteria. The Apriori algorithm is usually combined with ARM as it uses a search strategy for generating frequent item sets. The algorithm is initiated by generating all frequent item sets containing only one item by dataset scanning and counting. It then iteratively generates k-candidates from the frequent (k-1) item sets, eliminates certain candidates based on the Apriori principle that every subset of a frequent item set must also be frequent, and recounts the remaining candidates by scanning the dataset. The process continues until all frequent item sets are generated. Finally, the algorithm generates "strong rules" from frequent item sets by calculating their confidence and verifying constraints [32]. ARM can also be performed via an FP-growth algorithm, that utilizes in-depth search and compressed tree structure for generating frequent item sets without generating the candidates. The algorithm first scans the dataset to identify all frequent one-item sets and sorts them in descending order of support. Then it rescans the dataset and constructs a tree by inserting each transaction as a tree path, while transaction items are arranged according to the frequency of one-item order. This is a compact dataset representation that enables efficient mining of frequent item sets. FP-growth algorithm recursively explores the tree by generating conditional pattern databases and constructs conditional trees for each frequent item. Frequent item sets are generated by combining frequent items in conditional trees with their corresponding prefix paths [33]. In educational contexts, ARM is used to discover patterns in student behavior, academic performance and success, and course selection structure. Having in mind that educational datasets often contain a large number of categorical variables (e.g., demographic characteristics, activity logs, learning activities, grades, etc.), they are suitable for this technique, allowing teachers and researchers to gain insights into the relationships between different variables and generate recommendations for teaching and learning improvement [34].

3. SEMI-SUPERVISED LEARNING

Semi-supervised learning (SSL) represents a machine learning paradigm that utilizes both labeled and unlabelled data for predictive model performance enhancement. Contrary to the unlabelled, labeled data are usually scarce or expensive to obtain. SSL aims to exploit the hidden structure and information within unlabelled datasets to improve the learning process and reduce the reliance on labeled sets [35]. The SSL model assumes that the distribution of unlabelled data carries valuable information about the basic data structure and can assist in constraining and directing predictive model learning. The smoothness assumption states that data points situated in proximity to one another in the input space are likely to have similar output values or belong to the same class. Leveraging this assumption, SSL algorithms can propagate information from labeled examples to unlabelled instances and enhance the generalization performance of learned models [36]. SSL has several major approaches:

- *Self-training* constitutes an iterative process wherein a classifier is initially trained on labeled data and subsequently used to predict labels for unlabelled instances. The most confident predictions are incorporated into the labeled dataset, and the classifier is retrained on the expanded dataset. The process repeats until a stopping criterion is achieved or all unlabelled instances are labeled [37]. Self-training is applied in text content classification, image recognition, and speech recognition;
- *Co-training* assumes that the input characteristics can be partitioned into two or more independent and sufficient views, each capable of predicting the target variable. The algorithm trains separate classifiers of each view using labeled data and then uses the most confident predictions from each classifier to expand the labeled set for other views. The classifiers are iteratively retrained on expanded datasets until a stopping criterion is met [38]. This approach is used for web page classification, natural language processing, and bioinformatics;
- *Tri-training* extends co-training without requiring multiple views. Instead, three classifiers are trained on different labeled data samples obtained via bootstrap techniques, and the predictions of two classifiers are used to label the unlabelled instances for the third classifier. The process repeats for each classifier until a stopping criterion is achieved [39]. Tri-training is efficient for text categorization, image classification, and software defect prediction;
- *Graph-based method* represents data as a graph where nodes correspond to labeled and unlabelled instances, and edges reflect similarity between instances. The main idea is to propagate the label information from the labeled nodes to unlabelled ones based on the graph structure, assuming that the connected nodes have similar labels. This method is used in computer vision, natural language processing, and recommendation systems [40];

- *Generative models* analyze the joint distribution of input characteristics and target variables using both labeled and unlabelled data. The primary concept involves training a generative model to produce realistic samples from the data distribution and to employ it to infer the missing labels of unlabelled instances. Generative models are based on various frameworks, including Gaussian Mixture Models, Hidden Markov Models, and Variational Autoencoders [41], and are applied in text classification, speech recognition, and image generation.

SSL has numerous educational applications. It can be used for predicting student grades in distance learning courses utilizing both labeled and unlabelled data to enhance prediction accuracy [42], or creating personalized recommendations within e-courses employing available grades and implicit feedback from student interactions with the learning management system [43]. SSL performance primarily depends on the quality and representativeness of unlabelled data, as well as the validity of assumptions regarding data distribution. If unlabelled data sets are biased or contain instances from different domains or distributions, SSL will yield poor performance [44].

4. CONCLUSION

The paper presented the multifaceted landscape of unsupervised and semi-supervised learning techniques in educational applications. The investigation of clustering algorithms, dimensionality reduction methods, representation learning approaches, and association rule mining has revealed their significant potential for extracting meaningful patterns from complex educational datasets. *K-means* clustering has demonstrated substantial utility in identifying distinct student groups based on performance and engagement metrics, enabling educators to implement targeted interventions. PCA and t-SNE have proven valuable for visualizing high-dimensional educational data, though each presents distinct advantages regarding linear versus nonlinear relationship detection. ARM techniques have facilitated the discovery of meaningful correlations between student behaviors and academic outcomes, providing actionable insights for instructional design. SSL approaches are particularly promising in educational contexts in case labeled data are scarce and unlabelled data are abundant. Self-training, co-training, tri-training, graph-based methods, and generative models each offer unique capabilities for propagating label information across datasets, potentially enhancing the accuracy of predictive models while reducing reliance on extensively labeled examples. Nevertheless, the efficacy of these techniques critically depends on data quality, representativeness, and the validity of underlying assumptions regarding data distribution. Careful algorithm selection, hyperparameter tuning, and domain expertise remain essential for ensuring optimal results in educational applications.

As educational datasets continue to grow in both size and complexity, the presented sophisticated analytical approaches will play increasingly vital roles in personalized learning, adaptive learning, and evidence-based educational design. Future research will focus on enhancing the interpretability of numbered techniques, particularly for educational stakeholders without extensive data science expertise, thereby maximizing their practical usefulness in improving the educational process.

5. ACKNOWLEDGEMENTS

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-136/2025-03/200132 with University of Kragujevac, Faculty of Technical Sciences Čačak.

REFERENCES

[1] H. Pedersen, “Existentialism and Artificial Intelligence in the 21st Century,” in *The Routledge Handbook of Contemporary Existentialism*, Routledge, 2024, pp. 36-45.

[2] N. Vlassis, A concise introduction to multiagent systems and distributed artificial intelligence, Morgan & Claypool Publishers, 2007.

[3] R. Verganti, L. Vendraminelli and M. Iansiti, "Innovation and design in the age of artificial intelligence," *Journal of product innovation management*, vol. 37, no. 3, pp. 212-227, 2020.

[4] L. Floridi, "AI and its new winter: From myths to realities," *Philosophy & Technology*, vol. 33, pp. 1-3, 2020.

[5] Z. H. Zhou, *Machine learning,* Springer Nature, 2021.

[6] S. Vieira, W. H. L. Pinaya and A. Mechelli, "Introduction to machine learning," in *Machine learning*, Academic Press, 2020, pp. 1-20.

[7] T. J. Sejnowski, The deep learning evolution, MIT Press, 2018.

[8] D. Valkenborg, A. J. Rousseau, M. Geubbelmans and T. Burzykowski, "Unsupervised learning," *American Journal of Orthodontics and Dentofacial Orthopedics,* vol. 163, no. 6, pp. 877-882, 2023.

[9] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, vol. 10, no. 3, p. e1355, 2020.

[10] A. Dutt, M. A. Ismail and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991-16005, 2017.

[11] E. Cam and M. E. Ozdag, "Discovery of Course Success Using Unsupervised Machine Learning Algorithms," *Malaysian Online Journal of Educational Technology*, vol. 9, no. 1, pp. 26-47, 2021.

[12] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651-666, 2010.

[13] A. Dutt, S. Aghabozrgi, M. A. B. Ismail and H. Mahroeian, "Clustering algorithms applied in educational data mining," *International Journal of Information and Electronics Engineering*, vol. 5, no. 2, p. 112, 2015.

[14] A. K. Rastogi, S. Taterh and B. S. Kumar, "Dimensionality Reduction Algorithms in Machine Learning: A Theoretical and Experimental Comparison," *Engineering Proceedings,* vol. 59, no. 1, p. 82, 2023.

[15] Z. Alj, A. Bouayad and M. O. Cherkaoui Malki, "Dimensionality reduction for predicting students dropout in MOOC," in *International Conference on Digital Technologies and Applications*, Cham: Springer Nature Switzerland, 2023.

[16] Y. Bengio, A. Courville and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013.

[17] M. Z. Rodriguez, C. H. Comin, D. Casanova, O. M. Bruno, D. R. Amancio, L. D. F. Costa and F. A. Rodrigues, "Clustering algorithms: A comparative approach," *PloS one*, vol. 14, no. 1, p. e0210236, 2019.

[18] G. Gan, C. Ma and J. Wu, Data clustering: theory, algorithms, and applications, Society for Industrial and Applied Mathematics, 2020.

[19] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access,* vol. 8, pp. 80716-80727, 2020.

[20] A. Moubayed, M. Injadat, A. Shami and H. Lutfiyya, "Student engagement level in an e-learning environment: Clustering using k-means," *American Journal of Distance Education,* vol. 34, no. 2, pp. 137-156, 2020.

[21] M. Khalil and M. Ebner, "Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories," *Journal of computing in higher education*, vol. 29, pp. 114-132, 2017.

[22] V. Balakrishnan and E. Lloyd-Yemoh, "Stemming and lemmatization: A comparison of retrieval performances," *Lecture notes on software engineering,* vol. 2, no. 3, p. 262, 2014.

[23] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn and C. Mathieu, "Hierarchical clustering: Objective functions and algorithms," *Journal of the ACM (JACM),* vol. 66, no. 4, pp. 1-42, 2019.

[24] A. Triayudi and I. Fitri, "A new agglomerative hierarchical clustering to model student activity in online learning," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1226-1235, 2019.

[25] B. K. Tripathy, S. Anveshrithaa and S. Ghela, "Comparative Analysis of Dimensionality Reduction Techniques," in *Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization*, Routledge, 2021, p. 137–149.

[26] M. Greenacre, P. J. Groenen, . T. Hastie, A. I. d'Enza, A. Markos and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers,* vol. 2, no. 1, p. 100, 2022.

[27] G. Feng, M. Fan and Y. Chen, "Analysis and prediction of students' academic performance based on educational data mining," *IEEE Access,* vol. 10, pp. 19558-19571, 2022.

[28] H. Hershcovits, D. Vilenchik and K. Gal, "Modeling engagement in self-directed learning systems using principal component analysis," *IEEE Transactions on Learning Technologies*, vol. 13, no. 1, pp. 164-171, 2019.

[29] M. Balamurali, "t-Distributed Stochastic Neighbor Embedding," in *Encyclopedia of Mathematical Geosciences*, Cham: Springer International Publishing, 2021, pp. 1-9.

[30] M. Wattenberg, F. Viégas and I. Johnson, "How to use t-SNE effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.

[31] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *The journal of machine learning research,* vol. 15, no. 1, pp. 3221-3245, 2014.

[32] D. M. D. Angeline, "Association rule generation for student performance analysis using apriori algorithm," *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, vol. 1, no. 1, pp. 12-16, 2013.

[33] L. Zhiyuan, "Research on Association Rule Mining Algorithm Based on FP-tree," in *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, IEEE, 2024.

[34] L. Zhou and S. Yau, "Association rule and quantitative association rule mining among infrequent items," in *Proceedings of the 8th international workshop on Multimedia data mining:(associated with the ACM SIGKDD 2007)*, ACM, 2027.

[35] Z. Chen and B. Liu, "Lifelong Semi-supervised Learning for Information Extraction," in *Lifelong Machine Learning,* Cham: Springer International Publishing, 2017, pp. 81-88.

[36] Z. Miao, Y. Zhang, X. Piao, Y. Chu and B. Yin, "Region feature smoothness assumption for weakly semi-supervised crowd counting," *Computer Animation and Virtual Worlds,* vol. 34, no. 3-4, p. e2173, 2023.

[37] J. Jeong, S. Lee and N. Kwak, "Self-Training using Selection Network for Semi-supervised Learning," in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods,* 2020.

[38] C. Yabi and Z. Yongzhao, "Co-training semi-supervised active learning algorithm based on noise filter," in *2009 WRI Global Congress on Intelligent Systems* (Vol. 3), IEEE, 2009.

[39] Y. Zhang and S. Yan, "Semi-supervised active learning image classification method based on tri-training algorithm," in *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*, IEEE, 2020.

[40] K. Avrachenkov and M. Dreveton, "Graph-based semi-supervised learning," in *Statistical Analysis of Networks, Now Publishers*, 2022, p. 5.

[41] H. Yoon and H. Kim, "Label-noise robust deep generative model for semi-supervised learning," *Technometrics*, vol. 65, no. 1, pp. 83-95, 2023.

[42] G. Kostopoulos, S. Kotsiantis, N. Fazakis, G. Koutsonikos and C. Pierrakeas, "A semi-supervised regression algorithm for grade prediction of students in distance learning courses," *International Journal on Artificial Intelligence Tools,* vol. 28, no. 04, p. 1940001, 2019.

[43] A. Elbadrawy and G. Karypis, "Domain-aware grade prediction and top-n course recommendation," in *Proceedings of the 10th ACM conference on recommender systems,* ACM, 2016.

[44] G. Abdelrahman, Q. Wang and B. Nunes, "Knowledge tracing: A survey," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1-37, 2023.