Journal of Intelligent Computing

DLINE JOURNALS

Print ISSN: 0976-9005
Online ISSN: 0976-9013

JIC 2026: 17 (1)
https://doi.org/10.6025/jic/2026/17/1/1-13

Data Mining Algorithms and Models for Distance Education Management

Ye Qing
Universiti Sains Malaysia
USM, 11800. Gelugor
Penang. Malaysia
Ewrwerw45@yandex.com

ABSTRACT

This paper explores the application of data mining techniques to enhance the management and evaluation of distance education. It begins by contextualizing China's educational reforms, emphasizing decentralization and institutional autonomy, which have increased the need for data driven decision making. The study highlights the challenges of online learning remarkably low supervision, poor instructional quality, and high dropout rates and proposes Educational Data Mining (EDM) as a solution. EDM leverages algorithms like decision trees (especially C4.5), K-means, Apriori, SVM, KNN, and Naive Bayes to analyze student behavior, predict performance, and support timely interventions. Among these, the decision tree algorithm is selected for its interpretability, accuracy, and efficiency in handling diverse data types. The paper details the algorithm's computational framework, including information entropy and gain calculations, and presents an intelligent model for distance education management. Experimental results demonstrate that the optimized C4.5-based model improves both accuracy and processing speed compared to traditional methods. Association rule mining reveals significant behavioral patterns linked to student success, such as homework scores and forum participation. The study concludes that integrating data mining into distance education enables proactive, personalized support and more effective administrative oversight. While traditional assessment persists, algorithmic approaches offer a scalable, equitable pathway to enhance teaching, learning, and institutional management in the era of big data. Further refinement of these models is recommended for broader and more robust application.

Keywords: Data Mining Algorithm, Distance Education Management, Intelligent Mode

Received: 11 October 2025, Revised 23 December 2025, Accepted 29 December 2025

1. Introduction

During the transformation of China's educational management framework, we notice that its fundamental developmental trajectory is evident in the following areas [1]. At the onset of educational reform, the management system shifts focus from a centralized authority to a more localized approach. This is achieved by redistributing management powers from the central government and its educational administration to regional governments and local educational institutions [2]. By modifying this structure, our educational management system can better align with actual teaching practices. Following these reforms, the transfer of knowledge authority does not significantly affect the overarching educational framework and remains consistent with our initial system [3]. The previously centralized educational management has now been delegated to local governments and their respective educational administration departments.

As a result of this gradual decentralization, the motivation for teaching among local authorities and educational institutions has seen a significant boost, enabling these institutions to swiftly adapt their educational approaches to the evolving societal demands. The efficiency of information gathering has also improved, thereby enhancing the effectiveness of educational management indirectly [4]. After the redistribution of powers, schools gain autonomy in student admissions, which plays a crucial role in advancing the development of higher education. By tailoring subjects and specialties to the characteristics of the schools, a distinctive operational model has been developed, allowing educational institutions in our country to showcase a diverse and multifaceted approach to education.

2. Related Studies

In the process of externalizing the previous system, it is essential to establish a teaching model that incorporates the historical governance of the past, integrating the government's role in education with societal needs [6]. This approach fosters synergy between school instruction and societal demands, ensuring that the curriculum is better aligned with societal expectations, thereby making graduates more welcome in the community. The educational functions previously held by the government, as a vital social endeavor, must resonate with societal needs [7]. Traditional educational practices in China require reform. The quality of education advocated in our nation is shaped by global educational trends while being enhanced by unique Chinese cultural elements.

Following the implementation of a series of edcational reforms, the primary objective is now to deliver higher-quality education [8]. To elevate the standards of teaching resources and improve the overall educational quality of our nation, it is imperative to develop a sound and robust educational framework, bolster the growth of academic institutions, and create a management model that reflects a distinctive socialist ethos [9]. Throughout this endeavor, we must consistently prioritize a people centered educational approach [10]. The fundamental purpose of education is to cultivate talent, foster students' comprehensive development, and improve academic quality.

3. Advancing Distance Education Evaluation through Educational Data Mining and Algorithmic Approaches

In recent years, distance education has undergone a profound transformation driven by the rapid advancement of educational informatisation. Online learning platforms have become a dominant mode of instruction, offering learners unprecedented flexibility by dissolving traditional temporal and spatial barriers. Students can now access a rich array of courses at any time and from virtually any location, significantly broadening educational opportunities. However, this increased accessibility is accompanied by notable challenges. The absence of robust supervision mechanisms in many digital learning environments has led to concerns about low instructional quality and alarmingly high student dropout rates (11-15). These issues underscore the urgent need for more effective evaluation and intervention strategies particularly those grounded in data driven insights.

To address these gaps, researchers and educators have increasingly turned to Educational Data Mining (EDM) a rapidly evolving interdisciplinary field that applies computational techniques to uncover meaningful patterns, correlations, and predictive rules within the vast datasets generated during digital learning processes (16,17). EDM draws from psychometrics, learning analytics, and statistical modeling to transform raw institutional data into actionable intelligence. Its applications are wide ranging: predicting student performance, diagnosing systemic challenges within learning environments, identifying root causes of academic underachievement, and supporting timely, evidence-based interventions. By comprehensively analyzing diverse student data ranging from login frequency and assignment submissions to forum interactions EDM strengthens the reliability of academic assessments and enhances the efficacy of process oriented evaluation models (Bilici & Özdemir, 2021; Bonde & Kirange, 2018; Tekin & Öztekin, 2018).

Central to EDM's success are various machine learning algorithms, each offering unique strengths in analyzing educational data. Among these, decision trees have emerged as one of the most widely adopted methods in academic research. Valued for their high accuracy, computational efficiency, interpretability, and ability to handle heterogeneous data types, decision trees are particularly well suited for evaluating learning outcomes and identifying key performance predictors (21-23). For example, Hamoud et al. (24) employed the C4.5 decision tree algorithm to predict students' outcomes on English graduation exams, achieving a classification accuracy of 81.62%. Similarly, Di and Xu tested an enhanced variant of the decision tree model and reported an accuracy peak of 81%, demonstrating the algorithm's robustness in academic forecasting (25). Despite these promising results, researchers acknowledge that performance still falls short of ideal benchmarks, suggesting room for further refinement and hybridization with other techniques (26).

Beyond prediction, decision trees have proven instrumental in uncovering the behavioral and academic factors that most strongly influence student success. Weinberg and Last integrated an interpretable decision-tree induction method within a big-data parallel computing framework. They identified mastery of foundational knowledge as the single strongest predictor of academic achievement (27). In another study, Wu (281) applied an improved C4.5-based model to analyze learner behaviors in Massive Open Online Courses (MOOCs). The findings revealed that homework scores, participation in discussion forums, and frequency of communication with instructors or peers were significantly correlated with course completion and failure rates. These insights not only inform pedagogical adjustments but also enable early identification of at-risk students, allowing for timely academic support.

Further reinforcing the utility of decision trees, Yaacob et al. developed a supervised data mining framework that employs complex decision tree classifiers to forecast students' academic trajectories (29). Meanwhile, Fiarni et al. designed an academic decision support system built on decision tree logic to assist students in selecting submajors within information systems programs, demonstrating how algorithmic tools can also guide strategic academic planning (30).

While decision trees offer strong interpretability and performance, they are part of a broader algorithmic ecosystem used in EDM. Tuncay Sevindik (31) conducted a comparative analysis of several prominent algorithms including K-means clustering, Apriori association rules, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Naive Bayes to evaluate their effectiveness in assessing distance education platforms across three critical dimensions: user experience, institutional management, and teaching support. The study concluded that different algorithms excel in distinct contexts: K-means is effective for student segmentation, Apriori uncovers hidden behavioral patterns, SVM and KNN provide high predictive accuracy for outcome classification, and Naive Bayes offers simplicity and speed for real-time applications. This diversity underscores the need for a multifaceted, context-sensitive approach to educational data analysis.

As student activity increasingly migrates to digital platforms, the volume and complexity of educational data continue to grow. In response, specialized EDM methodologies are becoming indispensable for preventing academic failure and promoting student success (32). Rather than relying solely on traditional summative assessments which often arrive too late to influence learning trajectories modern evaluation systems leverage real-time analytics to foster responsive, adaptive, and learner-centred education.

In conclusion, while traditional assessment methods remain in use, the integration of algorithmic and data-driven strategies particularly through EDM and decision tree models offers a robust and scalable pathway toward more effective, equitable, and proactive distance education systems. By harnessing the predictive and diagnostic capabilities of machine learning, educators and institutions can move beyond reactive measures and instead cultivate learning environments that anticipate challenges, personalize support, and ultimately empower every student to succeed.

Despite the continued use of traditional assessment methods in distance education, evolving approaches—particularly algorithmic ones are emerging to enhance evaluation across user, management, and teaching dimensions. This study analyzes and compares algorithms like K-means, Apriori, C4.5, SVM, KNN, and Naive Bayes for evaluating distance education platforms. Findings indicate which algorithms most effectively analyse student behaviour, support management decisions, and aid teachers in making impactful instructional choices. (31)

Online learning has become a popular mode of learning due to the rapid development of educational informatisation. The rise of online learning platforms has significantly enhanced students' learning experience by removing temporal and spatial constraints, enabling them to study a variety of courses at any time, from any location. However, due to the lack of an effective supervision mechanism, online learning is undermined by low quality tutoring and a substantial dropout rate among students. (11-15)

The decision tree is one of the most widely used algorithms in data mining research, valued for its high accuracy, computational efficiency, rapid processing, and ability to handle diverse data types [21-23]. Current applications in online education primarily focus on evaluating learning outcomes and identifying key factors that influence student performance.

4. Methodology

4.1 Pre-Preparation and Functional Classification of Data Mining Algorithms

The computational capabilities of data mining algorithms differ across various computing domains. In general, these algorithms can perform calculations across most areas. To date, we have employed data mining algorithms for classification. Data mining algorithms can detect patterns within data and categorize data types according to established rules. This represents the most basic function of data mining and is one of the most frequently utilized. Another computational approach merges data mining algorithms with association rules to facilitate research. This method is commonly used to evaluate the relationships between events; one event may trigger another, thereby improving forecasts of future occurrences. There is also a clustering algorithm in data mining that searches for patterns in the data and classifies the data. The data mining algorithm possesses a predictive function: by examining event patterns, it can forecast future trends. Data mining inherently involves relationships. Data mining algorithms are computer processes that weave together these connections and enhance their impact. The algorithm's computations are performed by iteratively applying the function. Additionally, the unique computational tasks of data mining algorithms lead to variations in computational forms and procedural steps. A comprehensive analysis will follow at the end.

Overall, the primary function of data mining algorithms is to identify patterns and rules within data. The calculations performed by data mining algorithms serve to illustrate relationships in the initial computation. Through the calculations and investigations of data mining algorithms, connections among data are recognized and categorized. This method of calculation eventually progressed into predictive analysis. The evaluation of predictive features is grounded in the study of relationships among variables; the resulting dataset serves as the foundation for subsequent computations. Currently, all data mining algorithms are computed through these two foundational algorithms. The data mining algorithms discussed in this paper are no different; we integrate prediction and description across two types of computational content. This calculation model is more intricate. We must combine both the computational steps and forms, and scrutinize the standard formula of the data mining algorithm.

Research into pattern discovery is categorized into six primary directions. The first is association rules, which illustrate relationships between data. The second is sequential patterns, which allow research on association rule models to assess both feasibility and potential timing. The third is the classification model, which necessitates developing a mathematical framework that enables reasonable data partitioning. The fourth is the clustering model, which divides data into distinct categories. The fifth is the regression model; conventional methods for processing regression models often complicate the handling of discrete data. Finally, the last model focuses on deviation analysis. This data mining algorithm aims to identify and scrutinize erroneous data within the dataset. This form of calculation is widely applied in our everyday analyses. The computations are also explored further, utilizing Figure 1 to dissect the computational steps of the data mining algorithm employed.

4.2 Formula Based on Data Mining Algorithm and Intelligent Model Construction in Distance Education Management

Data mining techniques take various computational forms; however, the most frequently employed algorithms are the decision tree, K-nearest neighbour, and binary classification algorithms three distinct types. Before delving into the computational formulas of data mining algorithms, it is essential to assess the computational capabilities of these three algorithms and select the appropriate data mining algorithm that aligns with this paper's objectives and optimization needs. The subsequent Table 1 is utilized for analysis and comparison to evaluate the computational power of the three algorithms.

From the preceding table, it is evident that the decision tree classification algorithm excels in training-set learning speed, classification speed, and in handling discrete, binary, and continuous data. Yet it falls short in accuracy, handling of noisy data, and incremental learning. The K nearest neighbor algorithm demonstrates strong performance in training set learning speed, overfitting, and incremental learning; however, its classification speed is sluggish, and accuracy requires enhancement. The binary algorithm demonstrates strong accuracy, classification speed, and dependency property measurement, but it exhibits slow convergence during training on a training set. It can be concluded that each of the three classification algorithms has its strengths and weaknesses. In practical applications, the choice of classification algorithms should be guided by the available data and specific requirements. Based on our investigation presented in this paper, we determine that the decision tree algorithm is the most suitable for overall calculations. This algorithm aligns more closely with the demands of this paper, leading us to employ a decision tree algorithm for data mining. Further analysis reveals that the C4.5 algorithm is particularly well-suited to our computations. The following section analyses the C4.5 algorithm. The C4.5 algorithm is also utilized to create the decision tree separator and retains the benefits of the ID3 algorithm. To manage continuous descriptive attributes, these attributes must be discretized.

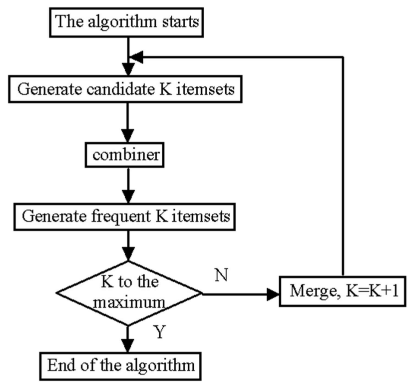

Figure 1. The calculation steps of the data mining algorithm used in this article

project	Computer algorithm		
	Decision tree	K nearest neighbor	Binary
accuracy	excellent	good	good
Classification speed	good	medium	good
Leakage measurement	excellent	excellent	good
Noise data processing	good	medium	medium
Incremental learning	good	good	excellent
Model parameter processing	excellent	medium	good
Measurement of unrelated attributes	medium	excellent	medium
Measurement of redundant attributes	good	medium	excellent

Table 1. Comparing the Computational Capabilities of the Three Algorithms

The minimum value of the continuous attribute is denoted as *MIN*, while the maximum value is indicated as *MAX*.

(1) *N* is divided into [*M I N* and *MAX*] intervals, and each breakpoint *B* is obtained. The expression of the breakpoint is:

$$B = MIN + \frac{MAX - MIN}{N} * i \tag{1}$$

The average amount of information is used to assess uncertainty, that is, information entropy.

$$S = -\sum_{I} \left(P_i * \lg(P_i)\right) \tag{2}$$

The amount of information used to predict the category of a record in S is calculated according to the following formula:

$$Info(S) = Info(S_P, S_n) = -\left(\frac{x}{x+y} * \lg\frac{x}{x+y} + \frac{y}{x+y} * \lg\frac{y}{x+y}\right) \tag{3}$$

The amount of information in the classification of the class in the subclass is:

$$Info(D, S) = \sum_{i}^{k} \frac{x_1 + y}{x + y} * Info(S_p, S_n) \tag{4}$$

The information increment of nodes is greater than the information increment of all nodes. The information increment of variable D is:

$$Gain(D) = Info(S) - Info(A, S) \tag{5}$$

The following is the general definition of information gain.

$$Gain(D) = \sum_{i}^{n} \left(\frac{|S_1|}{S_1}\right) * Info(S_1) \tag{6}$$

Thus, the computational formula for the employed data mining algorithm is complete. We have reviewed the computation steps for data mining algorithms. This computational method can facilitate quicker analysis and calculations pertinent to this paper. Using the computational model of the data mining algorithm, an appropriate calculation formula and an intelligent model for distance education management can be established. The computations for this model primarily rely on the data mining algorithm, leading to a data processing approach grounded in it, which can be examined and studied through the intelligent model data processing situation diagram in remote education management, as depicted in Figure 2 below.

5. Analysis and Discussion

In this work, the methodology and steps involved in the data mining process are examined. Furthermore, the data-mining formula is evaluated and implemented, resulting in a computer model. However, to ensure the precision of the enhanced data mining algorithm and its practical applicability, a thorough analysis and testing must be conducted before deployment. A comparative study is utilised in this paper to assess and analyse data mining. Initially, we investigate the data mining process, followed by a comparison between the traditional

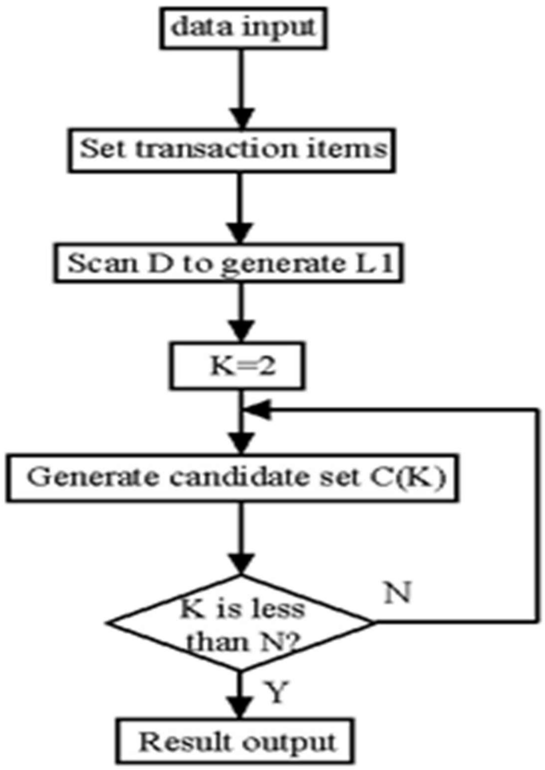

Figure 2. Data processing situation of the intelligent model in distance education management

algorithm and the optimized version. Based on the data mining steps, a specific set of project sets is established for computation and analysis, and the results are evaluated using the computational model described in the previous section. The findings from this collection are presented in Table 2 below.

Candidate Set										
Iterms	**A2,Z3**	**A2,E1**	**A2,E2**	**A3,Z3**	**A3,E1**	**A3,E2**	**A4,Z3**	**A4,E1**	**A4,E2**	**Z3,E2**
Support%	1	3	2	10	2	9	4	3	3	2.5
Confidence%	10	20	20	30.8	16	60	53.3	9.2	60	12.5

Large Project Set					
Iterms	**A3Z3**	**A3Z2**	**A4E3**	**A4E2**	**Z3E2**
Support%	20	9	4	3	10
Confidence%	30.8	60	53.3	60	33.3

Table 2. Association Rules Data Mining Algorithm Secondary Calculation Results

Through the calculation in our previous table, we analyze according to our specific data analysis method, and the preliminary calculation results in our table are correct and can be easily proven. The calculation results meet our requirements and the accuracy of our calculation. But this is not enough. We need to analyse the data further. Our final result is shown in Table 3.

Candidate set of items			**Candidate set of items**	
Iterms	**A3 Z3 E2**	**A4 Z3 E2**	**Iterms**	**A3Z3E2**
Support	8%	45.7%	Support	8%
Confidence	2%	40%	Confidence	45.7%

Table 3. Association Rules Data Mining Algorithm Final Calculation Results

The last three entries present the outcomes of our refined data mining algorithm. As indicated in the table, the optimized algorithm not only generates the required results but also computes multiple sets of results simultaneously, with each set being accurate and meeting our criteria for utilizing the computed outcomes. The confidence level and support for the three experimental sets calculated are correct; the confidence for the A_4, Z_3, E_2 group stands at 40%, while the A_3, Z_3, E_2 group achieves a confidence of 45.7%, both of which fulfill our usage requirements. The aforementioned tests demonstrate that the enhanced data mining algorithm can accurately yield the required results; however, we still lack a precise understanding of its computational efficiency. This necessitates further testing and analysis. Two experimental groups have been created to compare the traditional and optimized data mining algorithms. By assessing the computation time, we can validate the efficiency of our optimized algorithm. The dimensions ranging from 8 to 24 were segmented into five experimental groups for subsequent calculations and analyses. Figure 3 compares our experimental outcomes.

The computation time for both algorithms is evaluated and analyzed. It is observed that the computation time

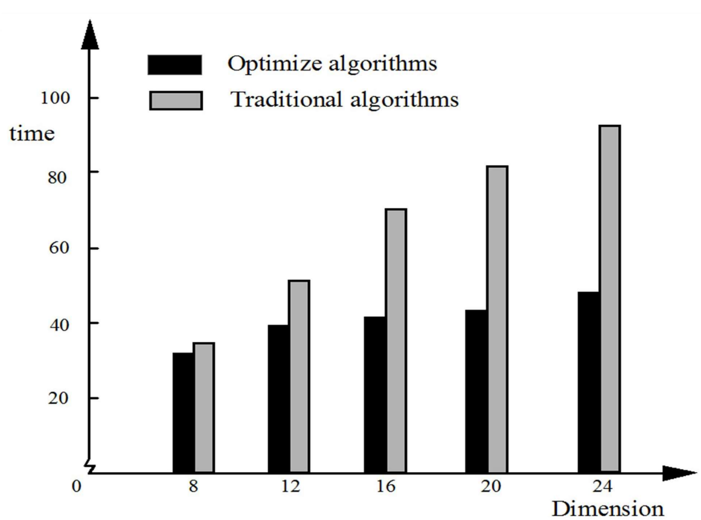

Figure 3. Contrast test time comparison table

for both algorithms increases with the computational dimensions, but the rate of increase varies significantly.The traditional algorithm's computation time escalates more rapidly, while our optimized algorithm's computation time remains comparatively lower. The computation time for both algorithms rises with the increase in dimensions, extending from an initial 5 seconds to a final 40 seconds.This confirms that the computational efficiency of our optimized algorithm surpasses that of the traditional algorithm. Finally, regarding task data ranging from 100 to 400, the response time for the data mining algorithm is tested, with the results displayed as follows:

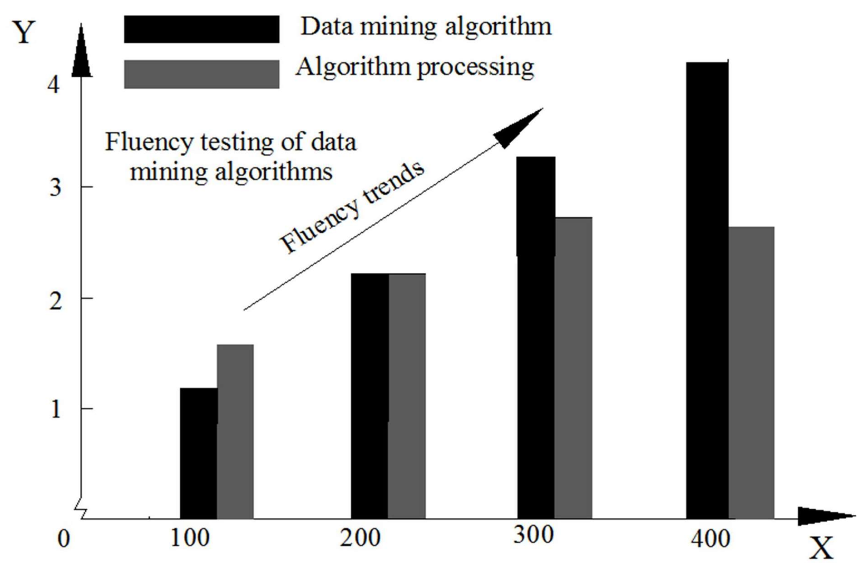

Figure 4. Optimization of processing speed for data mining algorithms

As illustrated in the preceding image, the overall processing duration increased with the volume of information. Without incorporating a data algorithm, the entire processing function becomes constrained. The data processing capabilities of the whole platform experience significant delays, while the server operates smoothly when handling fewer complex data processing tasks. However, once the algorithm is implemented, overall data-processing performance improves notably.

6. Conclusion

With the surge in social information, we have entered the big data era. In response to the challenges of this new phase, physical education (PE) instruction and its reforms are being examined to understand this evolving trend. This paper employed a data mining algorithm to conduct an in depth analysis of trends in the management of physical education in higher education institutions. Utilizing data mining algorithms provides us with substantial computational power to address the challenges we face. [33] Furthermore, our comprehensive evaluation of the refined data-mining algorithm demonstrates its ability to compute the required outcomes accurately. Multiple results can be obtained simultaneously. Additionally, we assessed the computational efficiency of the algorithm, and the results indicate that it performs efficiently. As the dimensionality of computation increases, the processing time for both algorithms exceeds 40 seconds at 24 dimensions. While we have achieved significant progress in optimizing the algorithm, further investigation and the development of even more advanced algorithms remain necessary for broader application.

References

[1] Mohanty, A. K., Senapati, M. R., Lenka, S. K. (2016). Retraction Note to: An improved data mining technique for classification and detection of breast cancer from mammograms. *Neural Computing & Applications* 27 (1), 249-249.

[2] Robson, B., Boray, S. (2016). Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple-choice medical licensing examinations. *Computers in Biology Medicine* 73 (73), 71.

[3] Hong, H., Pourghasemi, H. R., Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 259, 105-118.

[4] Rathod, R. R., Garg, R. D. (2016). Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. *International Journal of Electrical Power & Energy Systems,* 78, 368-374.

[5] Bandaru, S. N. g., A. H. C., Deb, K. (2017). Data mining methods for knowledge discovery in multi-objective optimization: Part B - New developments and applications. Expert Systems with Applications, 70, 119-138.

[6] Giabbanelli, P. J., Jean, A. (2016). Identifying small groups of foods that can predict achievement of key dietary recommendations: data mining of the UK National Diet and Nutrition Survey, 2008 12. *Public Health Nutrition* 19 (9), 1.

[7] Wu, W., Peng, M. (2017). A Data Mining Approach Combining K -Means Clustering With Bagging Neural Network for Short-Term Wind Power Forecasting. *IEEE Internet of Things Journal,* 4 (4), 979-986.

[8] Tayefi, M., Tajfard, M., Saffar, S., et al. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer Methods Programs in Biomedicine,* 141, 105-109.

[9] Chen, W., Pourghasemi, H. R., Naghibi. S. A. (2017). Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. *Bulletin of Engineering Geology the Environment*, 1-19.

[10] Alborzi, M., Khanbabaei, M. (2016). Using data mining and neural networks techniques to propose a new hybrid customer behaviour analysis and credit scoring model in banking services based on a developed RFM analysis method. *International Journal of Business Information Systems*, 23 (1), 1.

[11] Wang, C., Chang, L., Liu, T. (2022). Predicting Student Performance in Online Learning Using a Highly Efficient Gradient Boosting Decision Tree. In: Shi, Z., Zucker, JD., An, B. (eds) Intelligent Information Processing XI. IIP 2022. *IFIP Advances in Information and Communication Technology,* vol 643. *Springer, Cham.*

[12] Breiman, L. (2001). Random forests. Mach. Learn. 45(1), 5–32

[13] Karalar, H., Kapucu, C., Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *Int. J. Educ. Technol. Higher Educ.* 18(1), 1–18.

[14] Namoun, A., Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: a systematic literature review. *Appl. Sci.* 11(1), 237.

[15] Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., Lim, H.: Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Comput.* 24 (24), 18777–18793 (2020)

[16] Hotaman, D. (2020). Öðrenci baþarýsýnýn deðerlendirilmesinde eðitsel veri madenciliðinin kullanýlmasý. Ulakbilge Dergisi, 8(48), 577–587. https://doi.org/10.7816/ulakbilge-08-48-08

[17] Tan, S. S., Gökta, Y., Koçak, Ö. (2018, December 12-13). Veri madenciliði ile ÖSYM verileri kullanýlarak akademik baþarý tahmini [Conference presentation abstract] 2. Uluslararasý Uzaktan Eðitim ve Yenilikçi Eðitim Teknolojileri Konferansý, Amasya.

[18] Bilici, Z., Özdemýr, D. (2021). Data mining studies in education: Literature review for the years 2014- 2020. Bayburt Eðitim Fakültesi Dergisi. 17 (33), 342 - 376. https://doi.org/10.35675/befdergi.849973.

[19] Bonde, S. N., Kirange, D. K. (2018). Educational data mining survey for predicting student's academic performance. A. P. Pandian, T. Senjyu, S. M. S.

[20] Tekýn, A., Öztekýn, Z. (2018). Eðitsel veri madenciliði ile ilgili 2006-2016 yýllarý arasýnda yapýlan çalýþmalarýn incelenmesi. Eðitim Teknolojisi Kuram ve Uygulama, 8(2), 108-124. https://doi.org/10.17943/etku.351473.

[21] Ramya K., Teekaraman Y., and Ramesh Kumar K. A., Fuzzy-based energy management system with decision tree algorithm for power security system, *International Journal of Computational Intelligence Systems*. (2019) 12, no. 2, 1173–1178, https://doi.org/10.2991/ijcis.d.191016.001.

[22] Raut, A. B., Nichat M. A. A. (2017). Students performance prediction using decision tree, International *Journal of Computational Intelligence Research*. 13, no. 7, 1735–1741.

[23] Rani, R. M., Pushpalatha, M. (2019). Generation of Frequent sensor epochs using efficient Parallel Distributed mining algorithm in large IOT, Computer Communications. 148, no. 1, 107–114, https://doi.org/10.1016/j.comcom.2019.09.006, 2-s2.0-85072538303.

[24] Hamoud, A. K., Hashim, A. S., Awadh, W. A., (2018). Predicting student performance in higher education institutions using decision tree analysis, *International Journal of Interactive Multimedia and Artificial Intelligence*. 5, no. 2, 26–31, https://doi.org/10.9781/ijimai.2018.02.004.

[25] Di, J. Xu Y. (2019). Decision tree improvement algorithm and its application, International Core *Journal of Engineering*. 5, no. 9, 151–158.

[26] Jin M., Wang H., Zhang Q. et al. (2018). Financial management and decision based on decision tree algorithm, *Wireless Personal Communications*. 7, no. 2, 118–147.

[27] Weinberg, A. I., Last, M. (2018). Interpretable decision-tree induction in a big data parallel framework, *International Journal of Applied Mathematics Computer Science*. 27, no. 4, 737–748.

[28] Wu, Q. (2019). MOOC learning behavior analysis and teaching intelligent decision support method based on improved decision tree C4.5 algorithm, International Journal of Emerging Technologies in Learning (IJET). 14, no. 12, https://doi.org/10.3991/ijet.v14i12.10810, 2-s2.0-85069762620.

[29] Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., Sobri, N. M. (2019). Supervised data mining approach for predicting student performance, *Indonesian Journal of Electrical Engineering and Computer Science*. 16, no. 3, 1584–1592, https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592, 2-s2.0-85073562534.

[30] Fiarni, C., Sipayung, E. M., Tumundo, P. B. T. (2019). Academic decision support system for choosing information systems sub majors programs using decision tree algorithm, *Journal of Information Systems Engineering and Business Intelligence*. 5, no. 1, 57–66, https://doi.org/10.20473/jisebi.5.1.57-66.

[31] Sevindik, Tuncay., Cömert, Zafer. (2010). Using algorithms for evaluation in web based distance education, *Procedia Social and Behavioral Sciences*, Volume 9, Pages 1777-1780).

[32] Oliveira, de., C. F., Sobral, S. R., Ferreira, M. J., Moreira, F. (2021). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing,* 5(64), 1-33. https:// doi.org/10.3390/bdcc5040064.

[33] Islam, H. Wang (Eds). (2018). In Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2018): Vol.31. (p. 293-302). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-24643-3