



International Journal of Data Mining and Bioinformatics

ISSN online: 1748-5681 - ISSN print: 1748-5673

<https://www.inderscience.com/ijdmb>

Integrating data mining techniques for analysing implicit user behaviours in online courses

Guizhi Li, Dake Jiang, Dawei Kong

DOI: [10.1504/IJDMB.2026.10077083](https://doi.org/10.1504/IJDMB.2026.10077083)

Article History:

Received:	28 October 2025
Last revised:	07 December 2025
Accepted:	15 January 2026
Published online:	29 May 2026

Integrating data mining techniques for analysing implicit user behaviours in online courses

Guizhi Li*

School of Economics and Management,
Yingkou Institute of Technology,
Yingkou, 115004, China
Email: lizhigreat@126.com
*Corresponding author

Dake Jiang

Public Security General Teaching Department,
Criminal Investigation Police University of China,
Shenyang, 110854, China
Email: jiangdake@dlut.edu.cn

Dawei Kong

School of Economics and Management,
Yingkou Institute of Technology,
Yingkou, 115004, China
Email: wei2white@163.com

Abstract: The rapid expansion of online education has made the analysis of users' implicit behaviours – viewed through the lens of nonlinear and complex data – a crucial avenue for enhancing educational effectiveness. To address this, we introduce a random forest–fuzzy comprehensive evaluation (RF-FCE) method embedded within a clustering framework. Leveraging multiple clustering techniques, we first identify distinct category-specific influence patterns across different courses. Subsequently, we integrate fuzzy comprehensive evaluation with machine learning to analyse implicit behavioural data, examining both the intrinsic factors that affect course outcomes and the complex interactions between these factors and course quality. Our findings reveal significant variations in user engagement and learning outcomes across courses of differing quality, with these variations exerting a substantial influence on learning behaviours. In summary, this study offers a structured and robust analytical approach for examining implicit user behaviours in online education, demonstrating both methodological innovation and practical utility for improving course design and delivery.

Keywords: online course; data mining technology; implicit behaviour; cluster analysis; online education; course quality evaluation; behaviour analysis.

Reference to this paper should be made as follows: Li, G., Jiang, D. and Kong, D. (2026) 'Integrating data mining techniques for analysing implicit user behaviours in online courses', *Int. J. Data Mining and Bioinformatics*, Vol. 30, No. 6, pp.21–48.

Biographical notes: Guizhi Li is a PhD from Dongbei University of Finance and Economics, is currently an Associate Professor in the Big Data Management and Application program at Yingkou Institute of Technology. Her research interests include big data analysis and mining, information fusion, intelligent algorithms, and their applications.

Dake Jiang he is a Doctor of Business Administration School of Management, Jilin University and currently a Lecturer at the Public Security General Teaching Department, Criminal Investigation Police University of China. In recent years, his research interests have focused on entrepreneurship, police science, police management, social research methods and public relations.

Dawei Kong is a Master of Engineering in Software Engineering from Shandong University, is currently a Lecturer in the Big Data Management and Application program at Yingkou Institute of Technology. His research interests include big data technology, natural language processing, image processing, and related fields.

1 Introduction

With the rapid advancement of information technology, the demand for the sharing of educational resources has significantly increased. Massive open online courses (MOOCs) have quickly become widespread in this context, enabling more users to access educational resources through online platforms. However, due to the autonomy and unstructured nature of MOOCs, the analysis of online users' implicit behaviours has become increasingly important. More and more users are engaging in online learning through open educational resources and online course platforms on the internet, making online education a significant component of the educational field. Online education breaks the constraints of time and space associated with traditional education, offering flexible learning opportunities to users worldwide. By accessing course content at any time and from any place via online platforms, students greatly enhance the flexibility and convenience of their learning (Alam, 2023). Research indicates that online education does not show a significant difference in learning outcomes compared to traditional face-to-face education but has obvious advantages in flexibility and accessibility (Okewu et al., 2020). As technology continues to advance and educational demands become more diverse, the application of online courses in the education sector will continue to expand and deepen, becoming an important part of future educational models. Currently well-known online learning platforms include Coursera from Stanford University, edX co-founded by Harvard University and the Massachusetts Institute of Technology, and FutureLearn created by the Open University in the UK. These platforms use advanced educational technology to provide rich teaching tools and resources, significantly improving the quality and efficiency of online learning, having a profound impact on promoting educational equity and driving innovation in educational models.

Overall, the rise and popularity of online education have not only changed traditional education models but also provided users with more learning choices and pathways. Online education shows great potential in promoting educational equity, enhancing the utilisation of educational resources, and driving educational innovation. However, despite the unparalleled advantages of online education platforms over traditional education, the

abundance of online course resources with varying quality presents increasing challenges. For instance, how to balance the diverse needs of global users in course design to ensure every student gains the best learning experience; how to stimulate students' learning motivation and increase persistence to avoid dropouts; and how to ensure fairness in course evaluation to increase user engagement, and how to provide effective technical support (Zhang and Chen, 2022). Especially given the growing and varied quality of online course resources, continuous innovation in course design and implementation is necessary, as well as more technical support and policy guarantees to ensure the sustainable development of high-quality online courses.

To address these challenges, this study aims to fill a critical gap in the existing literature: while prior research has explored various aspects of online learning, there is a lack of integrated approaches that leverage data mining techniques to comprehensively analyse implicit user behaviours in MOOCs. For example, while studies like Luo et al. (2024) have applied knowledge mapping algorithms to enhance learner performance, they often overlook the subtle implicit behaviours that precede learning outcomes. Similarly, although advanced techniques such as transfer learning have been utilised in other domains (Shu, 2025), their application to implicit behaviour analysis in online education remains underexplored. This gap limits the ability to derive actionable insights for personalised learning support and course improvement. The importance of this research is driven by several key factors: the high dropout rates in MOOCs due to insufficient engagement tracking, the need for equitable and adaptive assessment methods in diverse learner populations, and the potential of data-driven decision-making to optimise educational resource allocation. These drivers are exacerbated by the scalability of online platforms, which generate vast amounts of behavioural data yet lack systematic analysis frameworks to translate this data into meaningful pedagogical interventions (e.g., Rizvi and Khalid, 2025). By integrating data mining techniques, this study seeks to provide a novel perspective on understanding and enhancing user experiences in online courses. Therefore, research on the implicit behaviour patterns of online course users is particularly important to help identify and promote high-quality educational resources.

2 Related research

In recent years, research on online courses has rapidly expanded, covering multiple topics. Many studies focus on utilising user behaviour to analyse different aspects of online courses, including how to assess online course quality, the characteristics of high-quality courses, and the factors influencing learning outcomes. From the perspective of data mining, course quality and platform design are crucial for ensuring user learning outcomes. According to the study by Zhang and Li (2020), the quality of online courses is typically evaluated across several dimensions, including learning outcomes, student engagement, course content, course design and structure, interaction and feedback, and technology platform support.

Some studies have constructed online course quality assessment systems using big data analysis technology to better understand the relationship between user behaviour and learning outcomes. For example, Mai et al. (2022) emphasised the importance of data mining techniques in behavioural data analysis in educational research. Meanwhile, other research focuses on quality standards and evaluation indicators for online courses.

Zimmerman et al. (2020) discussed the implementation of online course quality standards in their research and pointed out that clear quality standards help improve course design consistency and learning outcomes. Soffer et al. (2017), through electronic assessment based on student activity and perception, highlighted the importance of identifying areas for improvement in online courses. Research on the characteristics of high-quality online courses is also an important area. The application of data mining technology in online course research has also yielded fruitful results. Nilashi et al. (2022) explored the factors influencing student satisfaction in MOOCs using data mining techniques, including sentiment analysis and clustering analysis. They emphasised the importance of personalised learning experiences and interactive elements in online education. The study revealed that students' comments, feedback, interaction records, and individual differences all had varying degrees of influence on satisfaction. Alhothali et al. (2022) examined how various machine learning methods can be used to analyse students' learning behaviours and characteristics in order to predict their academic performance and outcomes. This series of machine learning methods covers logistic regression (LR), support vector machine (SVM), random forest (RF), and deep learning, among others. The research found that data mining technologies such as machine learning have great potential in online education research, particularly in enhancing student engagement, learning outcomes, and predicting learning outcomes (2020). In the field of cluster analysis research, Romero and Ventura (2020) emphasised that cluster analysis, as an unsupervised learning method, is widely applied in educational data mining, particularly for analysing student learning behaviours and performance patterns, which are of significant importance. Similarly, Rakic et al. (2020) used a hybrid approach to cluster students' learning behaviours, revealing the complex relationships between these behaviours and academic performance. Nalli et al. (2021) further explored the practical applications of clustering algorithms in online education platforms. They conducted clustering experiments on students' behavioural data from online courses, effectively identifying distinct groups of students based on differences in learning styles, knowledge levels, and engagement. Mohamed Nafuri et al. (2022) demonstrated that clustering techniques can effectively distinguish academic performance among higher education students, providing data-driven insights to identify at-risk students and optimise resource allocation. Feng et al. (2022) further highlighted the practicality of clustering in predicting academic performance. Their research argued that combining clustering with predictive models can significantly improve prediction accuracy, allowing for earlier identification of students who require intervention. Overall, these studies collectively highlight the value of cluster analysis in enhancing educational practices through data-driven insights, particularly in understanding student performance and engagement. In addition, Shang and Du (2023) analysed 50 factors influencing online course quality through data mining, revealing the hierarchical and dependent relationships among these factors. In terms of user behaviour analysis, Kizilcec et al. (2017) studied how self-regulated learning strategies affect learning behaviour and outcomes in MOOCs and suggested providing personalised support based on user characteristics. Personalised learning support is considered an important means to improve the quality of online courses. By analysing data from learning management systems, customised learning paths can be provided for different users to enhance learning experience and achievement. Scholars also suggest incorporating more social learning elements in course design, such as group discussions and collaborative learning, to enhance user interaction and support (Wright et al., 2023).

To improve the quality and learning outcomes of online courses, scholars believe that course design should emphasise interactivity and feedback, ensuring that users can actively participate and receive timely guidance. Okewu et al. (2021) emphasised the important role of artificial neural networks in the field of educational data mining. These methods have yielded promising predictive outcomes in areas such as student performance prediction, learning behaviour analysis, and educational evaluation. However, despite the strong predictive power of artificial intelligence technologies, such as artificial neural networks, in online education research (Towfek et al., 2024), significant challenges persist regarding data quality and model interpretability. At the same time, reliable technical support and user-friendly learning platforms are also crucial for improving learning outcomes. Online course learning plays an increasingly important role in higher education, and evaluating and improving course quality is key to ensuring learning outcomes. Through systematic quality standards, innovative course design, effective technical support, and personalised learning paths, online courses can provide users with a better educational experience. The literature on online courses emphasises the importance of factors such as student perception, teacher presence, engagement, and course design in enhancing online learning outcomes. These studies provide valuable recommendations for optimising the delivery of online courses and improving student success rates. Despite the richness of the literature on online courses, research has also revealed some deficiencies. Firstly, behavioural data studies primarily reflect surface engagement, failing to deeply understand students' cognitive processes or emotional experiences (Aldowah et al., 2019). At the same time, current research rarely considers the potential influence of external factors (such as technical issues or personal time arrangements), which can lead to data bias and affect the accuracy of analysis results. Secondly, there is also a lack of in-depth research on implicit behaviour patterns revealed by behavioural data, especially in the context of complex and diverse user data. Similarly, researchers may encounter problems such as high model complexity or insufficient generalisation ability during analysis, requiring more suitable methods to better fit the complex relationships between learning behaviour and performance. Additionally, data quality issues are also worth noting. Online data is often incomplete or inaccurate, which can affect the results of data analysis (Zhou et al., 2020). For example, during data mining, a large amount of irrelevant or noisy data may need to be processed to ensure the accuracy of analysis results (Abdollahzadeh et al., 2024).

Research finds that data analysis and mining have important advantages in studying online course behavioural data. Researchers need to comprehensively use various data mining technologies and consider data completeness and student engagement behaviour for analysis to improve the effectiveness and accuracy of the research (El-Kenawy et al., 2024). However, some data mining algorithms used in research rely on large amounts of data (Luo et al., 2024), which is less effective for emerging courses or courses with fewer students. Especially for the lack of analysis of implicit behaviour data, the universality of behavioural analysis methods is also not strong. From the perspective of data mining technology, research still faces challenges such as data dependence, model portability, and computational resource requirements. Designing overly complex models can lead to overfitting issues, limiting the application of these methods in real educational environments. To avoid the aforementioned issues, this paper designs a data mining method suitable for collecting process behaviour data patterns, adopting a multi-model clustering method and fuzzy comprehensive evaluation method, and incorporating principal component analysis (PCA) technology and RF technology (Wu, 2024) to

achieve dimensionality reduction visualisation and implicit behaviour analysis of users. Under multiple category labels, it conducts category analysis and evaluation of online course quality. Through various clustering analysis methods, in-depth mining and analysis of online course behaviour data are carried out to reveal the behavioural patterns of different users and their impact on course quality. Firstly, online course behaviour data is characterised by being massive, diverse, and complex, making traditional analysis methods insufficient to fully reveal the latent behavioural patterns in the data (Cai et al., 2022). This study innovatively adopts various clustering methods, including but not limited to K-means clustering, hierarchical clustering, and density clustering, to cluster and compare the behavioural data of online learning users. Secondly, based on behavioural analysis, this paper further uses the fuzzy comprehensive evaluation method to evaluate the course quality of different categories of users. This method comprehensively considers users' multidimensional behavioural characteristics, avoiding the limitations of a single evaluation indicator, and provides a more comprehensive and objective course quality assessment. The combination of multiple methods in data mining technology not only accurately identifies the heterogeneity of user behaviour but also effectively divides different types of user groups. In addition, this study further explores the complex relationship between user behavioural characteristics and course quality, providing scientific evidence for online education platforms to optimise course design and improve teaching quality.

3 Methodology

Clustering is one of the mature data mining techniques that has received extensive attention and research. Clustering is the process of mining potential information from existing data and ultimately dividing the data into multiple clusters. During the clustering process, the key issue is how to achieve better clustering effects by combining practical scenarios, distinguishing the differences and commonalities between data. When dealing with high-dimensional data, due to the different focus and sensitivity of various algorithms in extracting information features from behavioural data, there are significant differences in the applicability of various clustering methods. Through the comparison of multiple clustering algorithms and the realisation of high-dimensional data visualisation combined with PCA, it is possible to ensure that the extracted low-dimensional features better retain the information and structure of the original data while making the data differences more apparent, achieving better clustering effects.

This paper conducts clustering on the behaviour patterns of online course users, experimenting by setting different clustering numbers and dividing category samples into different levels to provide users with more suitable learning resources and services. By observing the distinction between various user groups, different implicit behaviour patterns of users are mined. The clustering methods adopted in this paper include the KCA, HCA, and DBSCAN algorithm for comparison.

3.1 *K-means clustering algorithm*

The K-means clustering algorithm (KCA) is an iterative technique used to find the local optimum of a target function. Given a sample set $D = \{x_1, x_2, \dots, x_n\}$, where $x_i = \{x_{i1}, x_{i2},$

$\dots, x_{im}\} \in R^m$, the algorithm seeks to minimise the squared error E for the clusters $C = \{C_1, C_2, \dots, C_K\}$, expressed as:

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|_2^2 \quad (1)$$

where $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ is the mean vector of cluster C_i . Intuitively, equation (1) to some extent characterises the compactness of the samples within the cluster around the mean vector. A smaller E value indicates higher similarity among samples within the cluster. The k-means algorithm uses a greedy strategy to approximate solving the minimisation of equation (1) through iterative optimisation.

3.2 Hierarchical clustering algorithm

The hierarchical clustering algorithm (HCA) attempts to partition the data at different levels to form a tree-like clustering structure. The partitioning of data collection is divided into a ‘bottom-up’ aggregation strategy and a ‘top-down’ splitting strategy. The HCA with the bottom-up aggregation strategy is adopted in this paper. Firstly, each sample in the dataset is regarded as an initial clustering cluster. Then, in each step of the algorithm, the two closest clusters are found and merged continuously until the preset number of clusters is reached. The distance between clustering clusters is calculated using equation (2). Suppose clusters C_i and C_j , the average distance expression used is:

$$d_{avg}(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{z \in C_j} dist(x, z) \quad (2)$$

The core difference between clustering methods lies in the construction of the clustering loss function, which is used to measure the relationship between clustering results and input data. It is generally related to the clusters and centroids divided from the samples. K-means clustering is a distance-based clustering algorithm that can quickly and effectively classify large-scale datasets, and the clustering results may be more inclined to discover more obvious behavioural patterns in the data. This algorithm is more sensitive to noise and outliers in data, so in some complex behavioural datasets, it may result in inaccurate classification. However, the advantage of K-means is that it can reveal the core characteristics of different user groups in the dataset, which is of great significance for quickly identifying implicit behaviour patterns of users. In contrast, the HCA is a method that constructs a hierarchical structure and gradually merges or splits data points to finally form a tree-like clustering result. This method has advantages in processing high-dimensional data and identifying potential hierarchical relationships in data. Therefore, the category values of the HCA can provide more detailed hierarchical information of behaviour patterns, which helps to deeply understand the complexity and diversity of user behaviour. The results of hierarchical analysis are particularly effective for analysing data with hierarchical structures or multi-level relationships. Observing the clustering tree structure can help understand the layer-by-layer evolution process of user behaviour, which provides a richer information basis for course designers to optimise teaching strategies.

3.3 Density-based spatial clustering of application with noise

Density-based spatial clustering of applications with noise (DBSCAN) is a density-driven clustering algorithm. In contrast to partitioning and hierarchical methods, DBSCAN defines clusters as the largest set of density-connected points. It excels at segmenting high-density regions into clusters and detecting arbitrarily shaped clusters in noisy, multidimensional datasets.

The algorithm defines the density of a point in a dataset as the count of points within a specified radius centred on that point. Two parameters must be configured for this process: the neighbourhood radius (Eps) and the minimum number of points ($MinPts$). Let the number of proximate points be denoted as $NumPts$. When proximate points $\in NEps(P)$, where $NEps(P) = \{q|q \in D, \text{distance}(p, q) \leq Eps\}$, the following applies:

- If $NumPts > MinPts$: the proximate point is labelled as a core point, and all points within its neighbourhood are directly density-reachable from the core point.
- Else: the proximate point is classified as a border point, provided it is within the neighbourhood radius of a core point. Otherwise, it will be labelled as noise.

When multiple points in the same direction are directly density-reachable and connected, they are referred to as density-reachable. In the dataset D , if an object O is density-reachable to both P and Q , then P and Q are considered density-connected. The DBSCAN algorithm aims to identify the largest set of density-connected points, treating this set as a cluster. Its strength lies in its density-based approach, which allows for the detection of clusters with arbitrary shapes and sizes. However, the algorithm is less effective for datasets with significant density variation and is particularly challenging for high-dimensional data where defining density is complex.

In summary, the k-means algorithm is computationally efficient for small to medium-sized datasets, with a complexity of $O(n*k*I)$, where n is the number of data points, k is the number of clusters, and I is the number of iterations. However, its efficiency decreases with increasing dataset size or feature dimensionality. Hierarchical clustering has a complexity of $O(n^2 \log n)$ and requires calculating distances between all pairs of points, making it time-consuming for large datasets and unsuitable for real-time applications. DBSCAN, in contrast, performs well with large datasets and clusters of complex shapes, with an approximate complexity of $O(n^2)$, though this depends on the use of spatial indexing. In conclusion, KCA is suitable for datasets with regular cluster shapes and moderate sizes, HCA is best for small datasets or those requiring specific structures, and DBSCAN is most effective for large datasets with irregular cluster shapes.

3.4 Evaluation metrics

Clustering evaluation involves various metrics, such as the silhouette coefficient, CHI, NMI, rand index (ARI), and cluster stability (CS). The suitability of each metric varies depending on its application, and the choice of metric should align with the specific use case. For the online course data analysed in this study, it was observed that the sample sizes across clusters were relatively balanced, and the problem fell under the unsupervised learning category. To ensure high intra-cluster cohesion while minimising computational complexity, the CHI, known for its low computational cost, was selected

for evaluation. For assessing the stability of clustering results, the CS metric was employed.

3.4.1 Calinski-Harabasz index

The Calinski-Harabasz index (CHI), also known as the variance ratio criterion, measures clustering effectiveness by calculating the ratio of between-cluster variance to within-cluster variance. Generally, a higher CHI indicates better clustering performance.

The index is computed using the total covariance matrix T and the within-cluster covariance matrix W_k as follows:

$$\text{Calinski - Harabasz} = \frac{\text{Tr}(B_k)(k-1)}{\text{Tr}(W_k)(n-k)} \quad (3)$$

where $B_k = T - W_k$, $\text{Tr}(B_k)$ is the trace of the between-cluster covariance matrix, $\text{Tr}(W_k)$ is the trace of the within-cluster covariance matrix, k denotes the number of clusters, and n represents the total number of samples in the dataset.

3.4.2 Category stability

Category stability (CS) evaluates the stability effect after clustering. For each $b = 1, 2, \dots, n$, apply the clustering algorithm A to dataset S_b twice, resulting in two clustering outcomes, C_b and C_b' . For the clustering results C_b and C_b' , calculate the distance $d(C_b, C_b')$ between the different clustering results. Sum all the inter-cluster distances and then divide by the square of the number of datasets to represent the instability of the clustering. The calculation formula is as follows:

$$\text{Instability} = \frac{1}{n^2} \sum_{b=1}^n d(C_b, C_b') \quad (4)$$

when the normalised mutual information (NMI) is closer to 1, the consistency of the clustering result with the true label is more evident, reflecting higher clustering accuracy. When stability is closer to 1, it indicates that the evaluated algorithm is less sensitive to data perturbations, suggesting better consistency and reliability of the clustering results.

3.5 Fuzzy comprehensive evaluation incorporating random forest

RF is an ensemble learning technique based on decision trees that enhances model accuracy and stability by combining the predictions of multiple decision trees. Each tree is independently trained using distinct bootstrap samples and randomly selected subsets of features. This design allows the model to capture diverse characteristics and patterns in the data. The inherent randomness of the RF model strengthens its generalisation capabilities, reduces the likelihood of overfitting, and makes it highly effective for analysing high-dimensional and noisy datasets. RFs are well-suited for capturing complex nonlinear relationships between features and offer significant advantages in calculating feature importance based on the contribution of split nodes. The fuzzy comprehensive evaluation method, grounded in fuzzy mathematics, utilises membership functions to provide structured assessments. In this study, the integration of RF into the fuzzy comprehensive evaluation method is implemented through the following steps:

Assume a dataset $D = \{x_{ij}\}$, where $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$.

Step1 Combine the clustering-derived class labels $C_1, C_2, \dots, C_k$ with the dataset D to construct a labelled dataset $Dset = \{x_{ij}, Cr\}$, $r = 1, 2, \dots, k$.

Step2 Define the set of evaluation factors F , calculate their membership vectors, and construct the membership matrix R .

The set of factors F , representing the evaluation objects, is defined as $F = \{f_j\}$, $j = 1, 2, \dots, m$. After normalising the sample set D , the membership matrix $R = \{r_{ij}\}$ is derived, where $i = 1, 2, \dots, n$, and $\sum_{j=1}^m r_{ij} = 1$.

Step3 Employ the RF model to derive the weight vector W_F .

Bootstrap sampling is used to repeatedly extract data with replacement, enabling the construction of multiple decision trees. Each decision tree randomly selects p features to create a new data subset $Dset_{np} \subset Dset$. The CART algorithm is applied to train the model, with C_r serving as the class label. The feature importance W_F obtained from the training process serves as the weight vector for the evaluation factors.

Step 4 According to the calculation rule (5), the designed model integrates W_F and W_E' into the total score method to compute the comprehensive membership vector S and the comprehensive evaluation value (the overall score) μ .

$$S = W_F R, \mu = W_E' S^T \quad (5)$$

3.6 Research design

This paper identifies implicit behaviour patterns of learning users by analysing the behaviour data of online courses and explores the relationship between these behaviours and course quality. First, detailed behaviour data were collected from a large online learning platform for multi-dimensional description of users, using various clustering analysis methods to cluster the behaviour data of online course users to identify different user behaviour patterns. The use of these clustering methods can more comprehensively reveal the heterogeneity of user behaviour and provide descriptions of core characteristics of different groups. On this basis, the paper further uses the fuzzy comprehensive evaluation method based on RFs to evaluate the course experience of different categories of users based on their high-dimensional behavioural characteristics. The research design framework is shown in Figure 1.

4 User behaviour analysis based on multi-model clustering

4.1 Dataset

The dataset utilised in this study comprises behavioural data from 289 online courses offered by Stanford University. The dataset features include diverse information such as the types of course content, the number of participants, student engagement, and interaction levels. The data was obtained through the platform's API, ensuring both its

completeness and accuracy. The data collection spans all open course data available since the platform's launch, and the data is stored in CSV format for convenient subsequent processing and analysis. The experimental platform was developed using PyCharm 3.9 along with the Anaconda integrated library. The characteristics of user behaviour data from the online courses are described in Table 1.

Figure 1 Research design framework (see online version for colours)

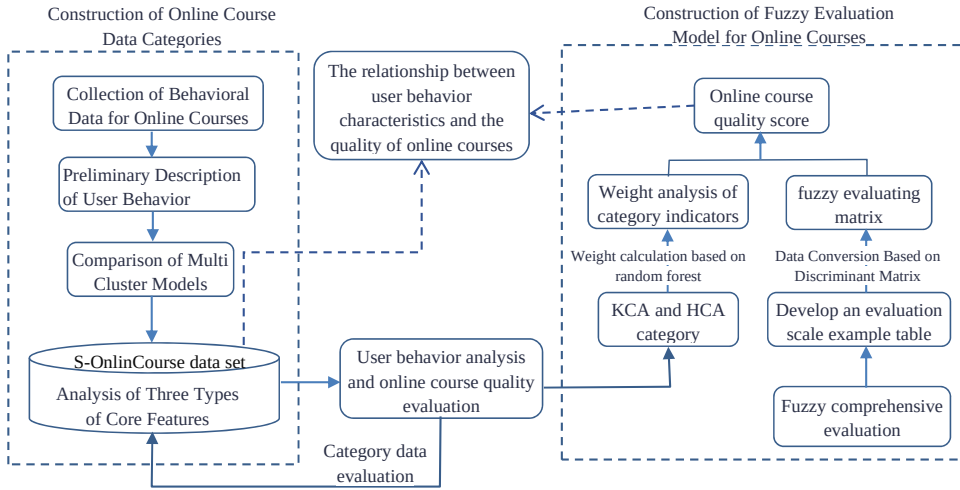


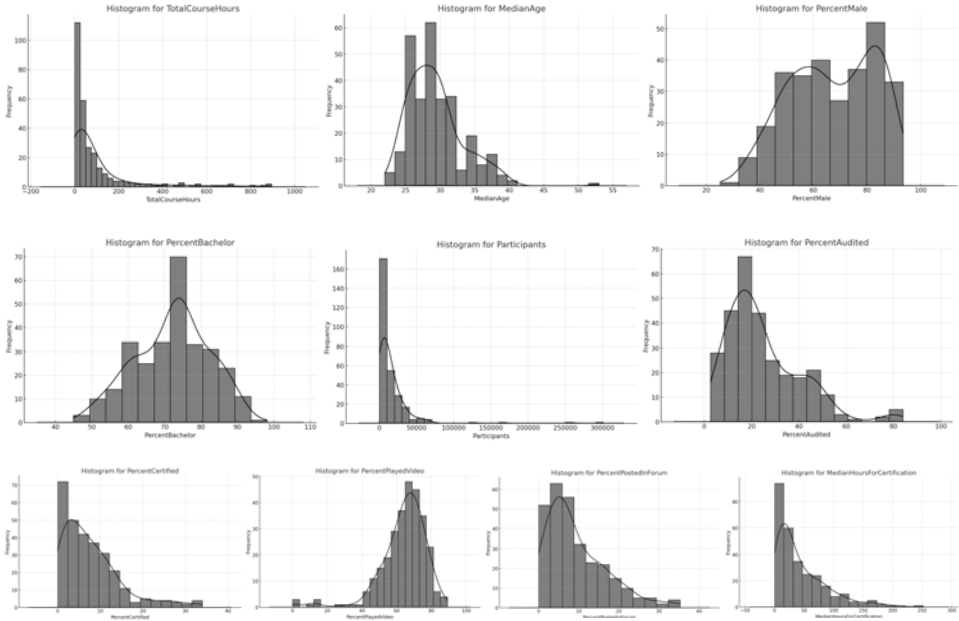
Table 1 Dataset properties description

Constituent elements	Specific indicators	Name
Course information	Course duration	Year
	Certification availability	HonorCodeCertificates
	Institution share of course	Institution
	Total course duration in hours	TotalCourseHours
	Course name	CourseTitle
	Course topic	CourseSubject
User information	Median user age	MedianAge
	Percentage of male users	PercentMale
	Percentage of users with Bachelor's degree or higher	PercentBachelor
Course learning information	Number of enrolled students	Participants
	Percentage of users completing 50% of the course	PercentAudited
	Certification rate	PercentCertified
	Video play rate	PercentPlayedVideo
	Forum participation rate	PercentPostedInForum
	Median hours to certification	MedianHoursForCertification

To handle missing data, this study introduces a k-means clustering-based similar mean imputation method, which performs imputation within clusters after categorisation. This

method effectively leverages the intrinsic structure of the data, mitigating the loss of variability often associated with direct mean imputation and significantly enhancing imputation accuracy. Furthermore, categorical features are handled using approaches such as one-hot encoding and data transformation to ensure minimal impact of data quality on the analysis results.

Figure 2 Distribution trend and description of data properties (see online version for colours)



4.2 Preliminary description and analysis of user behaviour

This study provides a descriptive analysis of specific indicators from Stanford University’s online course data. Histograms and trend curves were used to illustrate the distribution of behavioural characteristics, providing an initial understanding of central tendencies, dispersion, and distribution patterns in the behaviour data (see Figure 2).

Figure 2 shows that MedianAge, PercentMale, and PercentBachelor describe the basic information of users. The distribution of MedianAge indicates that most online course users are young, mainly between 25 and 32 years old, representing a demographic of young adults who pursue further education after college graduation. The distribution of PercentMale reveals a significantly higher number of male users compared to female users, reflecting a substantial gender disparity in online learning participation. The distribution of PercentBachelor shows that a high percentage (60% to 90%) of users possess a bachelor’s degree, indicating that the online learning platform attracts a highly educated user base.

The distribution of PercentAudited and PercentCertified illustrates user engagement in courses and certification outcomes. The distribution of PercentAudited indicates that most users tend to audit courses rather than pursue certifications, with most

PercentAudited values falling between 5% and 25%, suggesting similar learning behaviours among users.

The distribution of PercentPlayedVideo and PercentPostedInForum shows user activity and performance in the courses. The histogram indicates that while most users view video content, a relatively lower proportion of users post in forums, possibly indicating a preference for passive learning over active interaction. TotalCourseHours and MedianHoursForCertification are features related to study time, with their distributions showing the variation in total course duration and the average study time needed for certification. The distribution of TotalCourseHours is relatively dispersed, reflecting considerable differences in total duration among courses. In contrast, MedianHoursForCertification is concentrated around lower hours, suggesting that most courses have low certification time requirements, which may help increase the certification rate.

4.3 Multi-model clustering of user behaviour

Prior to model analysis, a correlation test was conducted on the high-dimensional indicators, and Pearson correlation coefficients were calculated. It was found that the two indicators PercentCertified50 and PercentGradeHigherThanZero were highly correlated, thus, these two indicators were removed.

4.3.1 Comparison of clustering methods

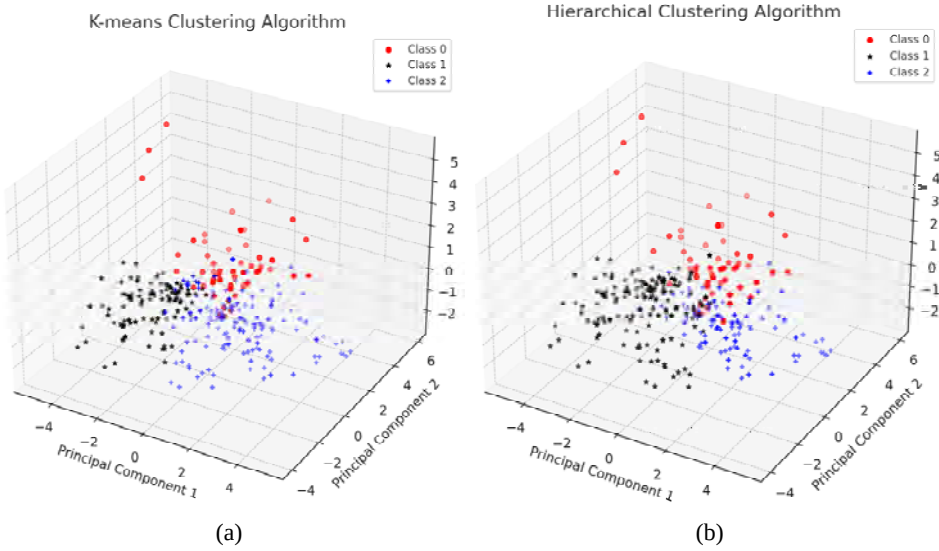
This study employed three methods for clustering analysis: KCA, HCA, and DBSCAN. In selecting clustering methods, this study comprehensively considered data distribution characteristics, algorithmic effectiveness, computational efficiency, and interpretability to ensure the chosen methods effectively reveal the intrinsic structure of the data and meet application requirements. The comparison process is outlined below:

From the perspective of data distribution, online courses data typically include user behaviour records, course interactions, and engagement metrics, characterised by high dimensionality and complex hierarchical structures. As a centroid-based clustering method, KCA assumes spherical data distribution and employs Euclidean distance for similarity measurement. Therefore, this method exhibits high computational efficiency and is suitable for scenarios with clear category boundaries and uniform data distribution. To address KCA's limitations in handling non-convex shapes and noise, hierarchical clustering and density-based DBSCAN were also considered. The hierarchical structure provided by hierarchical clustering enables administrators of educational platforms to analyse clustering results at varying granularities. DBSCAN excels at identifying arbitrarily shaped clusters and handling noise, but its sensitivity to parameters (e.g., neighbourhood radius ϵ and minimum points minPts) and unstable performance in high-dimensional settings due to density variations pose challenges.

To quantitatively evaluate the effectiveness of these clustering methods, the study employed two key metrics: the CH and CS. Experimental results showed that KCA and HCA achieved high CH index values (24.87 and 26.12, respectively), indicating reasonable cluster partitioning. In contrast, DBSCAN's lower CH index (12.50) suggested suboptimal performance in certain scenarios. Additionally, CS measures the stability of clustering outcomes, representing the consistency of cluster assignments under different data sampling or perturbation conditions. Experimental results

demonstrate that KCA and HCA maintain stable cluster partitioning across varying data samples, with CS values of 0.79 and 0.81, respectively. In contrast, DBSCAN achieves a lower CS value (0.52), indicating its heightened sensitivity to data variations and weaker robustness in dynamic data environments.

Figure 3 Visualisation effect of, (a) KCA and (b) HCA after dimensionality reduction (see online version for colours)



Additionally, practical considerations such as computational efficiency and interpretability also influenced the selection. While hierarchical clustering offers strong interpretability, its high computational complexity limits its suitability for large-scale datasets in online education platforms. Conversely, KCA’s efficiency makes it a viable benchmark for big data analysis. After balancing clustering performance, computational efficiency and interpretability, the study selected KCA and HCA as the primary clustering methods for subsequent analysis of online education data, ensuring a robust framework for educational quality assessment and personalised learning recommendations.

To better visualise the clustering results, dimensionality reduction was performed using PCA to project the data into a three-dimensional space. The reduction process was constrained by the requirement of retaining at least 80% of the cumulative contribution rate, ensuring the preservation of data interpretability and enhancing the discriminability of the clustering patterns. The visualisation of the clustering patterns is presented in Figure 3.

Figure 3 categorises online courses into three types. The next step involves combining the clustering analysis results to further identify the quality of each course and uncover the implicit relationships between features and course quality.

4.3.2 Weight analysis

Based on the classification results obtained in the previous section, this section applies the RF algorithm to compute the weights of various features, as shown in Table 2. The

indicator *PercentPlayedVideo* was selected for its direct reflection of student engagement with core course materials. Similarly, features such as *TotalCourseHours*, *PercentPostedInForum*, and *PercentCertified* were included, as they capture aspects of sustained learning, active participation in course discussions, and achievement of certifications – key areas identified by research for enhancing learning experiences, improving knowledge retention, and fostering social connections. Prior studies have demonstrated a strong correlation between these indicators and learning outcomes, positioning them as essential metrics for user engagement and course quality. To substantiate these choices, feature importance scores derived from the RF model were analysed (Table 2). The findings confirmed that these features are pivotal in assessing course quality, aligning with and validating the initial selections. In the feature importance evaluation after using KCA and HCA for classification, four features – *PercentMale* (16.52%), *PercentFemale* (12.01%), *TotalCourseHours* (13.07%), and *PercentCertified* (11.24%) – showed high levels of distinction, indicating these features' ability to differentiate learning outcomes and gender preference groups, leading to significant separation among different groups. *TotalCourseHours* and *PercentCertified* reflect participants' learning investment and outcomes, naturally dividing groups with different learning habits and results. *PercentPlayedVideo* (8.37%) and *PercentPostedInForum* (3.26%) suggest that user participation behaviours partly influence group classification. *MedianHoursForCertification* (7.76%) and *PercentBachelor* (7.10%) may be related to users' backgrounds and behaviour patterns. Particularly in hierarchical clustering, learning behaviour characteristics such as *MedianHoursForCertification*, *PercentPlayedVideo* (7.87%), and *PercentPostedInForum* (6.15%) were also significant, reflecting user engagement and interaction levels.

Table 2 Comparison of feature weights between KCA and HCA

<i>Feature</i>	<i>KCA</i>	<i>HCA</i>	<i>Feature</i>	<i>KCA</i>	<i>HCA</i>
<i>PercentMale</i>	0.1652	0.1532	<i>PercentBachelor</i>	0.0710	0
<i>TotalCourseHours</i>	0.1307	0.1436	<i>Participants</i>	0.0608	0.0654
<i>PercentFemale</i>	0.1201	0.1094	<i>PercentAudited</i>	0.0343	0.0405
<i>PercentCertified</i>	0.1124	0.1546	<i>PercentPostedInForum</i>	0.0326	0.0615
<i>PercentPlayedVideo</i>	0.0837	0.0787	<i>Year</i>	0.0196	0.0295
<i>MedianAge</i>	0.0833	0.0320	<i>HonorCodeCertificates</i>	0.0088	0.0584
<i>MedianHoursForCertification</i>	0.0776	0.0497			

5 Evaluation of online course quality

5.1 Construction steps

5.1.1 Determination of indicator

Weights based on the clustering analysis results from the previous section, where the clustered category labels were used as dependent variables, this section employs the RF method to assess the importance of each feature (see Table 3 for the results). The indicator weight vector is denoted as $W = (w_1, w_2, \dots, w_n)$.

5.1.2 Development of evaluation scale table and score conversion

To eliminate the effects of different data scales, this section discretises the feature data. Based on expert suggestions and the feature values described in the data, an equal-width binning method is used to determine the partition intervals. Scores ranging from 1 to 5 are then assigned based on each interval range, creating an evaluation scale table (see Table 2 for details). Finally, the discretised evaluation value r_{ij} for behaviour data is obtained.

5.1.3 Construction of the fuzzy evaluation matrix

Based on the discretised evaluation values r_{ij} , the membership degree of each online course for each indicator is calculated to construct the fuzzy evaluation matrix R . In this matrix, R_{ij} represents the fuzzy membership degree of the i^{th} online course concerning the j^{th} indicator.

5.1.4 Performing the fuzzy comprehensive evaluation

A comprehensive evaluation is conducted using fuzzy matrix multiplication. According to equation (6), this process yields the score vector V for the online courses.

$$V = W \cdot R = (w_1, w_2, \dots, w_n) \cdot \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \quad (6)$$

5.2 Comparison of multiple models

Existing online education classification models, such as SVM, LR, and deep learning (Ali and Simmou, 2025) (long short-term memory network, LSTM (Li et al., 2022)), typically depend on historical evaluation labels. This dependence limits their effectiveness when applied to scenarios where prior knowledge or historical evaluation indicators are unavailable. To address this challenge, this paper introduces and develops a random forest-fuzzy comprehensive evaluation (abbreviated as RF-FCE) method within a clustering framework. The aim is to explore how to overcome the limitations of traditional evaluation methods in complex tasks, such as education quality assessment, especially in the absence of prior knowledge and historical evaluation labels. The proposed method seeks to improve the model's prediction accuracy, interpretability, and computational efficiency.

5.2.1 Statistical validation

To strengthen the reliability and interpretability of the results, we first conducted statistical validation experiments on the four baseline models: RF, LR, SVM and LSTM. As LR is a generalised linear model, it allows direct computation of p-values and confidence intervals. However, RF, SVM and LSTM are non-parametric machine learning methods, meaning they do not inherently support direct calculation of p-values or confidence intervals. First, this research employs the permutation test method by

randomly shuffling the target variable. Through multiple iterations of model retraining and computation of performance distributions, we compare the model performance before and after permutation to obtain the p-value, thereby testing whether the model performance is statistically superior to random guessing. To address this limitation, we employed permutation tests and bootstrap confidence intervals to assess the statistical significance of these models. Additionally, we employed the bootstrap confidence interval method, which estimates confidence intervals for statistical metrics based on repeated resampling. This approach involves repeatedly resampling the dataset to train the model and computing confidence intervals for key performance metrics such as classification accuracy and the area under the curve (AUC), thereby assessing the stability of model performance. The experiments were implemented using Python.

The main experimental procedure includes:

- 1 For LR model, p-values and 95% confidence intervals were calculated for each variable to assess the statistical significance of feature effects on the response variable.
- 2 For RF, permutation tests and bootstrap confidence intervals were applied to calculate the confidence intervals for feature importance, identifying which variables made significant contributions to the model.
- 3 For SVM and LSTM, bootstrap confidence intervals were used to assess classification accuracy and prediction error stability, while permutation tests verified whether their predictive power surpassed random models.

Key results

- 1 LR: critical features exhibited $p = 0.0032$ (< 0.05), with confidence intervals excluding zero, confirming significance.
- 2 RF: 12 key features (Table 2) showed importance p-values < 0.5 and narrow confidence intervals, indicating stable and significant contributions.
- 3 SVM and LSTM: classification accuracy 95% confidence intervals were (0.814, 0.837) and (0.051, 0.068), respectively, with permutation test p-values of 0.004 and 0.009, confirming statistical superiority over random guessing.

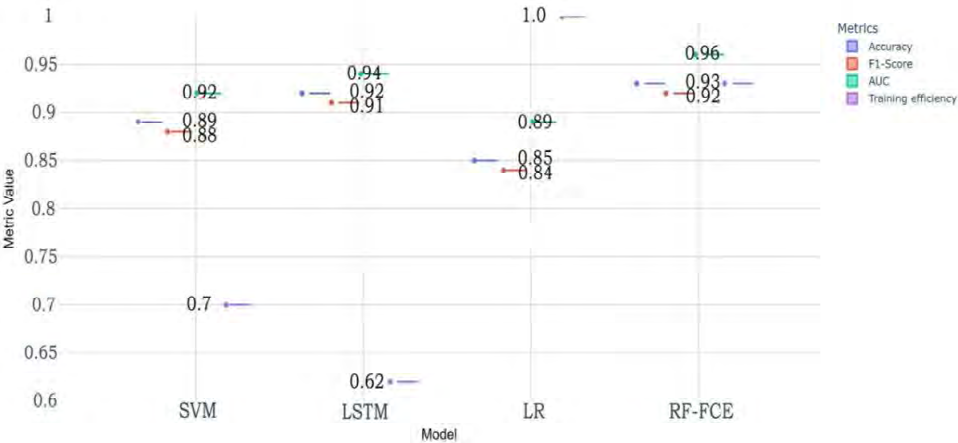
The experimental results demonstrate that the statistical tests of LR confirm the rationality of the feature selection, with variables exhibiting significant impacts on the model. Furthermore, the importance analysis of RF revealed that specific variables make substantial contributions, which can be effectively utilised for feature screening. Additionally, the outcomes from SVM and LSTM models indicated stable predictive performance, both significantly outperforming random baseline models. These statistically validated findings enhance the credibility of the experimental outcomes, demonstrating that the baseline models achieve statistically significant classification and prediction performance. Such statistical verification effectively mitigates the potential impact of data randomness on model results.

5.2.2 Model performance comparison

In selecting machine learning models, the proposed model was compared with three traditional benchmark models: SVM, LR, and LSTM. The models were evaluated across

four dimensions: prediction accuracy, F1-score, AUC, and training efficiency. The results of the comparison are presented in the form of a violin plot chart as shown in Figure 4. (sorted by total values in descending order).

Figure 4 Violin chart comparing performance values of multiple models (see online version for colours)



The experimental results clearly show that the RF-FCE model excels in accuracy, achieving a score of 0.93, which significantly outperforms SVM (0.89), LR (0.85), and LSTM (0.92). This finding validates the superiority of RF-FCE in solving multi-dimensional evaluation problems, particularly in scenarios where historical evaluation labels are not used. RF-FCE effectively classifies samples and offers superior accuracy without relying on past evaluation data. The F1-score, which balances precision and recall, is a key metric for assessing classification performance. RF-FCE's F1-score of 0.92 is notably higher than those of SVM (0.88), LR (0.84), and LSTM (0.91). This result demonstrates that RF-FCE not only ensures classification accuracy but also effectively avoids false negatives and false positives, particularly in multi-dimensional complex data, where it can better capture genuine evaluation features. Moreover, the RF-FCE model exhibits outstanding performance in terms of AUC value, achieving 0.96, which significantly surpasses other models (SVM: 0.92, LR: 0.89, LSTM: 0.94). The closer the AUC value is to 1, the stronger the model's discriminative ability. The high AUC value of RF-FCE indicates its superior capability to distinguish between different categories when dealing with complex datasets, thereby enhancing the model's robustness and generalisation ability. Finally, this paper compares the training efficiency of the models (i.e., taking the reciprocal of the model training time and normalising it; the shorter the training time, the higher the training efficiency). The training efficiency of RF-FCE is 0.93, higher than that of SVM (0.70) and LSTM (0.62), demonstrating a significant computational efficiency advantage. LSTM typically requires substantial computational resources and longer training times, whereas RF-FCE can complete training and prediction in a shorter time without the need for complex hardware support, showcasing its superiority in practical application scenarios with limited computational resources.

Table 3 Evaluation scale example table

Feature	1 points	2 points	3 points	4 points	5 points
PercentMale	25.172–38.88	38.88–52.52	52.52–66.16	66.16–79.8	79.8–93.44
TotalCourseHours	0.00–179.875	179.09–358.07	358.07–537.05	537.05–716.03	716.03–895.01
PercentCertified	0.00–6.83	6.796–13.592	13.592–20.388	20.388–27.184	27.184–33.98
PercentPlayedVideo	0.00–17.9171	17.828–35.656	35.656–53.484	53.484–71.312	71.312–89.14
MedianAge	21.969–28.2	28.2–34.4	34.4–40.6	46.8–53.0	53.0–64.21
MedianHoursForCertification	0.00–50.616	50.364–100.728	100.728–151.092	151.092–201.456	201.456–251.82
PercentBachelor	44.897–55.582	55.582–66.214	66.214–76.846	76.846–87.478	87.478–98.11
Participants	21.24–60474.0	60.474.0–120.626.0	120.626.0–180.778.0	240.930.0–301.082.0	30.1082.0–328.405.15
PercentAudited	2.549–18.896	18.896–35.162	35.162–51.428	51.428–67.694	67.694–83.96
PercentPostedInForum	0.00–7.0913	7.056–14.112	14.112–21.168	21.168–28.224	28.224–35.28

5.2.3 Interpretability analysis

In the field of education, the interpretability of models is particularly crucial. While deep learning models such as LSTM can deliver high predictive accuracy, their ‘gray-box’ nature makes the results difficult to comprehend and interpret. This poses significant limitations, especially in addressing the widespread fuzzy semantics and dynamic decision-making needs in educational platform data, which are critical for course design and improvement. In contrast, the RF-FCE framework establishes an interpretable model that integrates ‘white-box’ transparency with scenario adaptability, demonstrating superior interpretability. This can be attributed to the following reasons:

- The RF-FCE method provides a hierarchical interpretation mechanism based on membership semantics. This model maps prediction results into multi-level probability distributions (e.g., (high-quality course: 0.62, medium-quality course: 0.28, low-quality course: 0.10)) using membership functions. The output directly aligns with the grading standards in educational evaluations, avoiding the semantic discontinuity issues inherent in gray-box models. This interpretability allows educators and decision-makers to clearly understand evaluation results and optimise course content and teaching strategies accordingly.
- The RF-FCE method employs verifiable fuzzy rule-driven decision logic. Through explicit weight judgments (e.g., when the proportions of PercentCertified, TotalCourseHours, and PercentMale increase, the membership degree for high course quality tends to rise), the FCE ensures transparent decision-making. Educational experts can directly participate in rule definition and weight calibration, guaranteeing consistency between the model and domain knowledge. In contrast, SVM relies on kernel functions for high-dimensional mapping, making its decision boundaries difficult to translate into understandable domain rules (e.g., SVM cannot explain the semantic relationship between ‘support vectors’ and educational indicators). While LR offers interpretable linear coefficients, it fails to capture nonlinear synergistic effects among features (e.g., interactions between course levels and data characteristics). These results demonstrate that RF-FCE, through its white-box architecture and editable fuzzy logic, overcomes the limitations of traditional models in interpretability depth and scenario adaptability. This approach provides a novel perspective for online course quality assessment, balancing theoretical rigor and practical value. Consequently, RF-FCE exhibits enhanced interpretability and holds significant application potential in domains such as online course evaluation.

5.2.4 Experimental summary

5.2.4.1 Comparative analysis of base models

This study systematically compares multiple machine learning base models – including SVM, LR, and LSTM – to validate the superior performance of RF-FCE in online education quality assessment. Statistical verification demonstrates that the feature importance metrics in RF-FCE’s RF model exhibit p-values consistently below 0.5 and narrow confidence intervals, indicating stable variable selection and significant contributions to classification tasks. In contrast, while SVM and LSTM achieve satisfactory classification accuracy, their lack of interpretability hinders direct

quantification of feature contributions. RF-FCE outperforms these models in feature selection, model stability, and classification capability, effectively mitigating misleading conclusions caused by data randomness. In terms of performance metrics, RF-FCE surpasses SVM, LR, and LSTM across key indicators: accuracy (0.93), F1-score (0.92), and AUC (0.96), demonstrating robust classification and generalisation capabilities in complex data environments. Furthermore, RF-FCE achieves a training efficiency of 0.93, significantly exceeding LSTM (0.62), thereby avoiding the computational resource constraints inherent to deep learning models and enhancing its suitability for real-time analysis in online education platforms.

5.2.4.2 Adaptability and interpretability of RF-FCE

A core strength of RF-FCE lies in its adaptability to the unique characteristics of online education data and its interpretability. Online education datasets often exhibit high dimensionality, fuzziness, and dynamic variability – such as student learning behaviours, course feedback, and instructor evaluations – which are challenging to model accurately with traditional linear approaches. Unlike SVM's opaque decision boundaries and LSTM's high-dimensional embedded features, RF-FCE employs a fuzzy comprehensive evaluation method to decompose complex course quality tasks into interpretable fuzzy rules. For instance, the model maps quality assessments into discrete categories (e.g., high, medium, and low-quality courses) via membership functions, enabling education administrators to transparently trace decision logic and optimise course content. Additionally, RF-FCE's rule-driven architecture allows experts to directly adjust rule parameters, enhancing model adaptability and credibility – a capability absent in SVM and LSTM, which lack mechanisms for human intervention or visual tuning.

5.2.4.3 Integrated advantages and practical value

In summary, RF-FCE not only outperforms SVM, LR, and LSTM in prediction accuracy and computational efficiency but also uniquely integrates RF's feature selection strengths with the interpretability of fuzzy comprehensive evaluation. This integration enables RF-FCE to address the complex demands of online education quality assessment more effectively. The model's transparent decision-making process and flexible adjustment mechanisms provide significant practical value in this domain, offering education managers a precise, interpretable analytical tool to support data-driven decision-making.

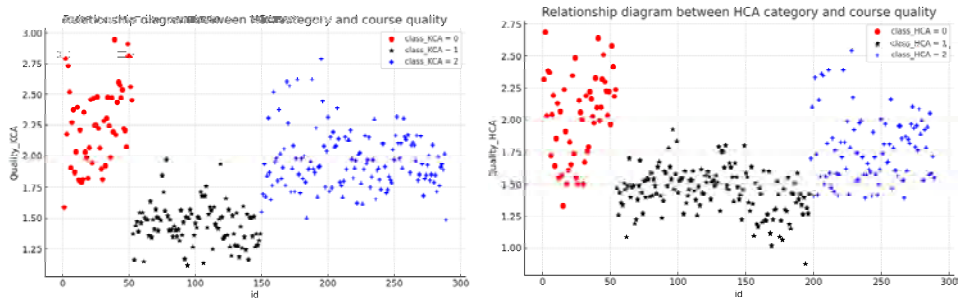
5.3 Online course quality evaluation

This study first obtained group category labels using two clustering algorithms and then performed a comprehensive calculation based on the value evaluation metrics of the transformed high-dimensional user data, resulting in quality scores for online courses. This section further analyses and identifies user behaviour patterns and their impact on course quality, exploring the relationship between course quality and user characteristics across different clustering categories (as shown in Figure 5).

Firstly, we assessed the quality of online courses according to their categories. In the KCA category analysis, the average quality scores for online courses in class 0, class 1, and class 2 were 2.25, 1.45, and 1.97, respectively. In the HCA category analysis, the average quality scores for online courses in class 0, class 1, and class 2 were 2.04, 1.44,

and 1.77, respectively. Both analyses revealed that class 0 consistently received higher quality scores for online courses, while class 1 had lower scores. There are notable differences in course quality among the categories: class 0 has the highest quality but with significant variability, class 1 has relatively low quality, and class 2 has high quality with relative stability. Applying these findings could be beneficial for course improvement and quality enhancement.

Figure 5 Relationship diagram between category and course quality (see online version for colours)



5.4 Analysis of the relationship

To further analyse the relationship between user characteristics and course quality, this section visualises the course quality scores obtained earlier, based on different clustering categories (KCA and HCA). The relationships between the classified course quality results and various behavioural characteristics are analysed. These comparative characteristics include PercentAudited, PercentCertified, PercentPlayedVideo, PercentPostedInForum, TotalCourseHours, MedianHoursForCertification, MedianAge, PercentMale, and PercentBachelor. Each user characteristic is represented using different symbols, as shown in Figure 6, which provides a visual representation of the relationships between these characteristics and course quality.

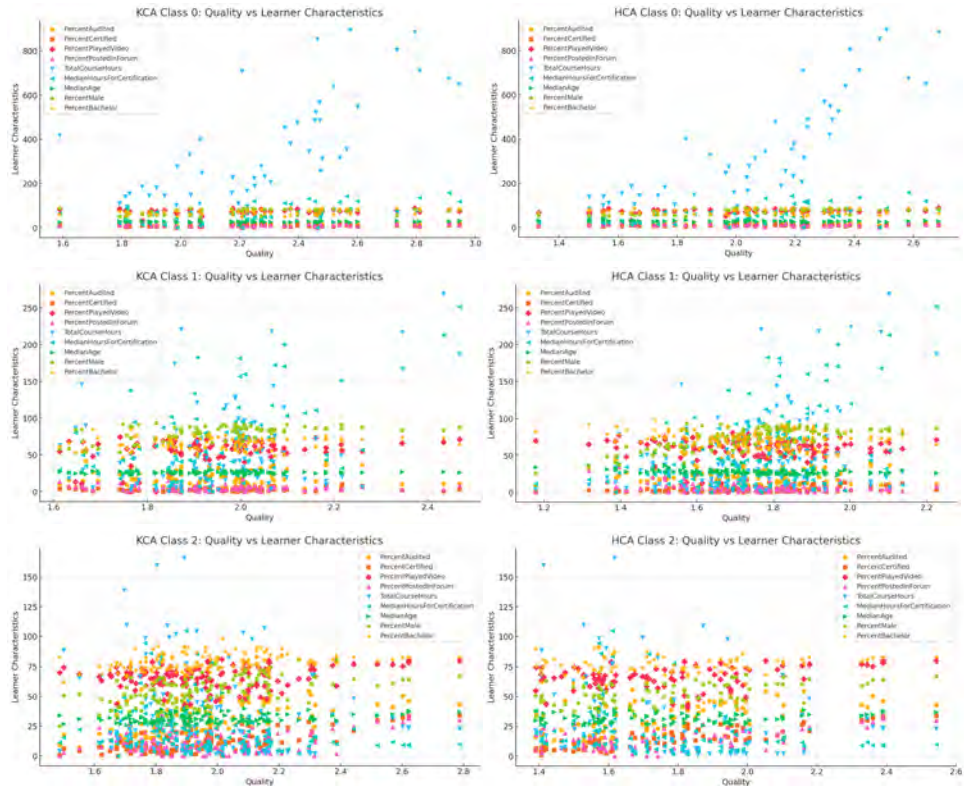
Based on the results from each graph in Figure 6, we divided course quality into seven levels and calculated the average values for each characteristic within these levels, examining the correlation between course quality and different user behavioural characteristics.

A comparative analysis using category labels obtained from KCA and HCA revealed that in different categories, features such as the number of participants (participants), percentage of video views (PercentPlayedVideo), percentage of forum posts (PercentPostedInForum), and total course hours (TotalCourseHours) are closely related to course quality.

From the KCA label results, it can be observed that categories with higher course quality tend to have a greater total course duration (TotalCourseHours) and a higher number of participants (participants). Additionally, these categories usually show higher percentages of video views (PercentPlayedVideo) and forum posts (PercentPostedInForum). Data analysis of overall course behaviour indicates that the total course duration is generally higher for online courses offered by prestigious universities, suggesting that learners are more concerned with whether a course can provide comprehensive learning content and rich resources. A high value for Participants

indicates that online courses are more likely to attract more participants due to effective course content or successful promotion and marketing strategies. Similarly, the ongoing appeal of a course is a core indicator of its quality; the video view rate (PercentPlayedVideo) and forum interaction percentage (PercentPostedInForum) reflect whether an online course maintains consistent appeal.

Figure 6 The relationship between behavioural characteristics and curriculum quality (see online version for colours)



The HCA label results show that the relationship between course quality and various characteristics has some similarities to the KCA results but also reveals some unique findings. Categories with higher median ages (MedianAge) generally have higher quality scores, indicating that high-quality courses are more likely to attract mature learners. Moreover, a higher proportion of certificate attainment is found in high-quality courses, suggesting that mature learners typically have clear learning objectives and pathways, placing greater importance on course quality and certification outcomes. Similar to the KCA category results, high-quality courses have higher rates of video views and forum participation, once again emphasising the significance of student interaction and engagement in learning outcomes.

5.5 *Results and discussion*

By integrating diverse data mining techniques such as cluster analysis, PCA, RF, and fuzzy comprehensive evaluation, this study performs an in-depth analysis of the implicit behaviours of users in Stanford University's online courses. The findings reveal crucial connections between user behaviour patterns and course quality, providing actionable insights to guide course design for other online education platforms.

5.5.1 *Differentiated analysis of learning user behaviour*

The study reveals significant differences in user behaviours across various clusters, particularly regarding gender, learning investment, learning behaviours, and background characteristics. Gender differences were found to be a critical factor in both clustering algorithms, indicating their substantial impact on user behaviour and learning outcomes. Key metrics such as total course hours (TotalCourseHours) and certification rate (PercentCertified) reflect user time commitment and learning achievements, serving as primary indicators for segmenting user groups. Behavioural characteristics, including video completion rates (PercentPlayedVideo) and forum participation rates (PercentPostedInForum), further illustrate variations in engagement and interaction styles. High-quality courses often see increased forum participation, emphasising the positive role of user interaction in enhancing course quality. To improve user engagement and overall course quality, online education platforms should prioritise fostering meaningful user interactions (Latifi et al., 2021). Additionally, platforms should provide tailored services for users at different stages. For novice users, platforms could enhance engagement by offering appealing video content, more interactive opportunities, and well-defined learning paths and evaluation standards. For advanced users, courses should align with their motivations and aspirations, providing in-depth, challenging content and recognised certifications.

5.5.2 *Conduct tailored course design*

Building on the insights from clustering and weight analysis, online course design should prioritise personalised learning pathways to cater to the varied needs of learners. For instance, users who engage more in discussions but watch fewer videos may benefit from more interactive content, while those inclined toward independent learning could be better served by an expanded library of high-quality video resources (Nilashi et al., 2022). Such targeted designs are instrumental in meeting diverse user needs and elevating overall course quality. The study further reveals that online education platforms can tailor course content and instructional strategies based on learner background attributes, such as MedianHoursForCertification and PercentBachelor, to create a more customised and effective learning experience for different user groups. To ensure excellence in course delivery, high-quality online courses should not only provide rich and diverse content but also balance content depth with student engagement. Course designers should focus on crafting engaging (Zhang and Chen, 2022), accessible video content while exploring varied interaction formats, including group projects, online assessments, and live lectures, to boost student participation and learning efficacy.

5.5.3 Sustained strategies for enhancing course quality

High-quality online courses go beyond providing exceptional content and certifications, they also focus on course structure and the overall learning experience to ensure ongoing student engagement. Continuously refining courses to meet the diverse needs of user groups can significantly improve the quality of online education platforms and increase user satisfaction. Regular analysis of new user behaviour patterns and course quality ratings offers a solid foundation for iterative enhancements to course content and design. Data-driven course quality strategies, powered by advanced data mining techniques, enable platforms to detect subtle user behaviours and trends in course quality over time. This insight allows platforms to adapt course structures proactively. For instance, a sustained drop in quality ratings for a particular course category signals the need for a detailed investigation and targeted improvements. Leveraging data-informed decisions, online education platforms can enhance personalised learning experiences, strengthen user acquisition and retention, and refine their marketing and advertising strategies for better outcomes.

6 Conclusions

This study presents a novel perspective for advancing theoretical research in online education while delivering practical, data-driven solutions with significant theoretical and real-world implications.

On the one hand, this paper proposes a RF-FCE method within a clustering framework. The method innovatively combines fuzzy comprehensive evaluation with machine learning techniques and applies it to the assessment of online course quality. It enables objective evaluations despite the presence of uncertainty and subjective factors. The RF-FCE model does not rely on pre-existing labels but evaluates online course quality across multiple dimensions (such as course content, user information, and learning outcomes) through fuzzy evaluation and clustering analysis, without requiring large amounts of empirical data. This integrated data mining approach strengthens the reliability and practical value of the evaluation results, allowing online courses to better address the diverse needs of users.

On the other hand, this paper explores the multi-dimensional factors influencing online course quality and their interrelationships through the mining and analysis of online course data. It identifies key factors such as content depth, learner age distribution, interaction participation, and certification requirements as critical elements in course quality assessment. The study demonstrates that high-quality courses typically feature more comprehensive content, greater appeal, and higher student engagement, which is reflected in significant improvements in metrics such as TotalCourseHours, participants, PercentPlayedVideo, and PercentPostedInForum. Online courses from renowned universities stand out in terms of TotalCourseHours, reflecting learners' strong demand for rich learning resources. Additionally, course quality is closely tied to the age structure of learners, with mature learners more likely to choose high-quality courses and place more emphasis on certification outcomes. This finding suggests that course design should carefully consider the needs of learners across different age groups, particularly the learning motivations and certification needs of older learners.

This study introduces a novel RF-FCE method, which provides an interpretable and scalable framework for evaluating online course quality without reliance on pre-existing labels. Compared to recent machine learning-based educational assessment models, such as the deep learning-driven engagement predictor proposed by Savchenko et al (2022) and Wu et al. (2024), which achieves high accuracy but operates as a ‘black box’, RF-FCE maintains transparent decision pathways through fuzzy logic integration. This aligns with the growing demand for explainable AI in education, as emphasised by Chaush et al. (2023), who argue that interpretability is critical for fostering educator trust and facilitating pedagogical intervention.

Furthermore, while existing ensemble and clustering methods in learning analytics often prioritise structural partitioning over educational insight – such as the feature-agnostic clustering approach reviewed by Križanić (2020) – RF-FCE enables granular, behaviour-aware evaluation across multiple dimensions including content depth, interaction patterns, and learner demographics. This advances the theoretical conversation on data informed instructional design, moving beyond predictive performance toward actionable, context sensitive quality assessment.

In summary, RF-FCE contributes a methodologically transparent and pragmatically adaptable tool for online course evaluation, extending the scope of intelligent educational assessment with an emphasis on interpretability and real world applicability. The data mining technique proposed in this paper overcomes the reliance of traditional models on historical evaluation labels. Based on existing online education evaluation methods, a new evaluation approach with higher interpretability is introduced. This approach is particularly suitable for complex educational quality assessments and offers significant practical value and broad application potential. However, the RF-FCE model also has certain limitations. For example, while the model excels in terms of accuracy and interpretability, it is still limited by computational resource constraints when handling ultra-large datasets.

Acknowledgements

This work was supported by the Doctoral Launch Plan of Liaoning Provincial Department of Science and Technology (User behaviour Monitoring of Online Education Platform Based on Data Mining Technology, grant No. 2023-BSBA-306).

Declarations

The dataset used is a publicly available data resource that can be obtained from the Alibaba Cloud Tianchi repository. All data is completely anonymous and does not involve individual participants or ethical issues. In addition, ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

All authors declare that they have no conflicts of interest.

References

- Abdollahzadeh, B., Khodadadi, N., Barshandeh, S., Trojovský, P., Gharehchopogh, F.S., El-kenawy, E.S.M., ... and Mirjalili, S. (2024) 'Puma optimizer (PO): a novel metaheuristic optimization algorithm and its application in machine learning', *Cluster Computing*, Vol. 27, No. 4, pp.5235–5283.
- Alam, A. (2023) 'Improving learning outcomes through predictive analytics: enhancing teaching and learning with educational data mining', in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp.249–257.
- Aldowah, H., Al-Samarraie, H. and Fauzy, W.M. (2019) 'Educational data mining and learning analytics for 21st century higher education: a review and synthesis', *Telematics and Informatics*, Vol. 37, No. 1, pp.13–49.
- Alhothali, A., Albsisi, M., Assalahi, H. and Aldosemani, T. (2022) 'Predicting student outcomes in online courses using machine learning techniques: a review', *Sustainability*, Vol. 14, No. 10, p.6199.
- Ali, N.W. and Simmou, W. (2025) 'Framework for measuring the impact of a firm's artificial intelligence capability on creativity within the organizational and performance of the firm', *Journal of Intelligence Technology and Innovation*, Vol. 3, No. 1, pp.74–92.
- Cai, L., Lu, S. and Chen, B. (2022) 'Constructing technology commercialization capability: the critical role of user engagement and big data analytics capability', *Journal of Organizational and End User Computing (JOEUC)*, Vol. 34, No. 9, pp.1–21.
- El-Kenawy, E.S.M., Khodadadi, N., Mirjalili, S., Abdelhamid, A.A., Eid, M.M. and Ibrahim, A. (2024) 'Greylag goose optimization: nature-inspired optimization algorithm', *Expert Systems with Applications*, Vol. 238, No. 1, p.122147.
- Feng, G., Fan, M. and Chen, Y. (2022) 'Analysis and prediction of students' academic performance based on educational data mining', *IEEE Access*, Vol. 10, No. 1, pp.19558–19571.
- Kizilcec, R.F., Pérez-Sanagustín, M. and Maldonado, J.J. (2017) 'Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses', *Computers & Education*, Vol. 104, No. 1, pp.18–33.
- Križanić, S. (2020) 'Educational data mining using cluster analysis and decision tree technique: a case study', *International Journal of Engineering Business Management*, Vol. 12, No. 1, p.1847979020908675.
- Latifi, S., Noroozi, O. and Talaee, E. (2021) 'Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes', *British Journal of Educational Technology*, Vol. 52, No. 2, pp.768–784.
- Li, G., Wang, X., Bi, D. and Hou, J. (2022) 'Risk measurement of the financial credit industry driven by data: based on DAE-LSTM deep learning algorithm', *Journal of Global Information Management (JGIM)*, Vol. 30, No. 11, pp.1–20.
- Luo, Z., Shao, X. and Ma, X. (2024) 'Enhancing learners' performance in contest through knowledge mapping algorithm: the roles of artificial intelligence and blockchain in scoring and data integrity', *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp.1–21.
- Mai, T.T., Bezbradica, M. and Crane, M. (2022) 'Learning behaviours data in programming education: community analysis and outcome prediction with cleaned data', *Future Generation Computer Systems*, Vol. 127, No. 1, pp.1–12.
- Mohamed Nafuri, A.F., Sani, N.S., Zainudin, N.F.A., Rahman, A.H.A. and Aliff, M. (2022) 'Clustering analysis for classifying student academic performance in higher education', *Applied Sciences*, Vol. 12, No. 19, p.9467.
- Nalli, G., Amendola, D., Perali, A. and Mostarda, L. (2021) 'Comparative analysis of clustering algorithms and moodle plugin for creation of student heterogeneous groups in online university courses', *Applied Sciences*, Vol. 11, No. 13, p.5800.

- Nilashi, M., Abumalloh, R.A., Zibarzani, M., Samad, S., Zogaan, W.A., Ismail, M.Y., ... and Akib, N.A.M. (2022) 'What factors influence students satisfaction in massive open online courses? Findings from user-generated content using educational data mining', *Education and Information Technologies*, Vol. 27, No. 7, pp.9401–9435.
- Okewu, E., Adewole, P., Misra, S., Maskeliūnas, R. and Damasevičius, R. (2020) 'Artificial neural networks for educational data mining in higher education: a systematic literature review', *Computers and Education*, Vol. 149, No. 1, p.103789.
- Rakic, S., Tasic, N., Marjanovic, U., Softic, S., Lüftenegger, E. and Turcin, I. (2020) 'Student performance on an E-Learning platform: mixed method approach', *International Journal of Emerging Technologies in Learning*, Vol. 15, No. 2, pp.1–17.
- Rizvi, R.A. and Khalid, W. (2025) 'Unlocking innovation in manufacturing: the impact of data analytics maturity and knowledge-oriented leadership', *Journal of Management Science and Operations*, Vol. 3, No. 2, pp.17–43, DOI: <https://doi.org/10.30210/JMSO.202503.002>.
- Romero, C. and Ventura, S. (2020) 'Educational data mining and learning analytics: an updated survey', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 10, No. 3, p.e1355.
- Savchenko, A.V., Savchenko, L.V. and Makarov, I. (2022) 'Classifying emotions and engagement in online learning based on a single facial expression recognition neural network', *IEEE Transactions on Affective Computing*, Vol. 13, No. 4, pp.2132–2143.
- Shang, S. and Du, C. (2023) 'Online course quality evaluation from the perspective of knowledge management: analysis of online reviews', *Library Hi Tech*, Vol. 41, No. 6, pp.1725–1747.
- Shu, M. (2025) 'Utilizing transfer learning for deep learning based image classification', *Journal of Intelligence Technology and Innovation*, Vol. 3, No. 1, pp.58–73.
- Soffer, T., Kahan, T. and Livne, E. (2017) 'E-assessment of online academic courses via students' activities and perceptions', *Studies in Educational Evaluation*, Vol. 54, No. 1, pp.83–93.
- Towfek, S.K., Khodadadi, N., Abualigah, L. and Rizk, F.H. (2024) 'AI in higher education: insights from student surveys and predictive analytics using PSO-guided WOA and linear regression', *Journal of Artificial Intelligence in Engineering Practice*, Vol. 1, No. 1, pp.1–17.
- Wook, M., Razali, N.A.M., Ramli, S., Wahab, N.A., Hasbullah, N.A., Zainudin, N.M. and Talib, M.L. (2020) 'Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback)', *Education and Information Technologies*, Vol. 25, No. 4, pp.3215–3232.
- Wright, A.C., Carley, C., Alarakyia-Jivani, R. and Nizamuddin, S. (2023) 'Features of high-quality online courses in higher education: a scoping review', *Online Learning*, Vol. 27, No. 1, pp.46–70.
- Wu, C-H. (2024) 'Eco-economic predictions: applying QPSO-BiLSTM and attention mechanisms for accurate renewable energy forecasting', *Journal of Management Science and Operations*, Vol. 2, No. 4, pp.28–47, <https://doi.org/10.30210/JMSO.202402.011>.
- Wu, Z., She, Q. and Zhou, C. (2024) 'Intelligent customer service system optimization based on artificial intelligence', *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp.1–27.
- Zhang, H. and Li, M. (2020) 'Using data mining to assess the effectiveness of online courses in higher education', *International Journal of Educational Technology in Higher Education*, Vol. 17, No. 1, pp.1–15.
- Zhang, X. and Chen, Y. (2022) 'Research on teaching quality evaluation and group characteristics of blended courses based on data mining technology', in *2022 10th International Conference on Orange Technology (ICOT)*, pp.1–4, IEEE.
- Zhou, N., Zhang, Z. and Li, J. (2020) 'Analysis on course scores of learners of online teaching platforms based on data mining', *Ingénierie des Systèmes d'Inf.*, Vol. 25, No. 5, pp.609–617.
- Zimmerman, W., Altman, B., Simunich, B., Shattuck, K. and Burch, B. (2020) 'Evaluating online course quality: a study on implementation of course quality standards', *Online Learning*, Vol. 24, No. 4, pp.147–163.