Forum for Linguistic Studies | Volume 07 | Issue 02 | February 2025

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

K-Means Clustering: A Tool for English Language Teaching Innovations

Minjuan Chen [1]***, Tan Wee Hoe*** [2*]

[1] *Faculty of Social Sciences & Liberal Arts, UCSI University, Kuala Lumpur 56000, Malaysia*

[2] *International Institute of Science Diplomacy & Sustainability, UCSI University, Kuala Lumpur 56000, Malaysia*

ABSTRACT

How can we ensure teachers receive precise evaluations to excel in classrooms? This study addresses inaccuracies in traditional English teaching competence evaluations by introducing a big data-driven estimation algorithm that employs fuzzy K-means clustering and information fusion. First, we build a model that analyzes key indicators of teaching ability with certain constraints. These constraints help us focus on the most important factors. Then, we use a step-by-step quantitative method to evaluate the teaching competence in our data model. This allows us to extract valuable "fingerprints" of teaching ability. Think of it like finding unique patterns that help us understand teaching effectiveness better. Ultimately, by combining big data information fusion with the K-means clustering algorithm, we cluster and consolidate the indicator parameters of English teaching competence, devise tailored teaching resource allocation strategies, and conduct English teaching competence assessment. Experimental findings indicate that the utilization of this approach for evaluating English teaching competence demonstrates robust information fusion analysis capabilities, thereby enhancing the precision of competence assessment and optimizing the utilization efficiency of teaching resources.

Keywords: English Teaching; Teaching Ability; Information Fusion; Data Clustering

*CORRESPONDING AUTHOR:

Tan Wee Hoe, International Institute of Science Diplomacy & Sustainability, UCSI University, Kuala Lumpur 56000, Malaysia;
Email: tanwh@ucsiuniversity.edu.my

ARTICLE INFO

Received: 10 December 2024 | Revised: 29 January 2025 | Accepted: 31 January 2025 | Published Online: 25 February 2025
DOI: https://doi.org/10.30564/fls.v7i2.8379

CITATION

Chen, M., Hoe, T.W., 2025. K-Means Clustering: A Tool for English Language Teaching Innovations. Forum for Linguistic Studies. 7(2): 988–998.
DOI: https://doi.org/10.30564/fls.v7i2.8379

1. Introduction

The use of information processing technology and big data analysis for teaching evaluation and resource allocation plays a crucial role in enhancing the overall efficiency and precision of educational management [1]. This not only improves the quantitative management capabilities of teaching processes but also aids in effective planning and decision-making by providing accurate data insights [1, 2]. Consequently, addressing the challenges in English teaching ability assessment through big data analysis has become an important focus of research. With the increasing complexity and diversity of teaching environments, traditional assessment methods are often inadequate to capture the nuanced performance metrics of English teaching [3]. Firstly, traditional assessment methods predominantly rely on simplistic quantitative metrics, such as students' test scores or teacher performance ratings. Secondly, these methods often lack adaptability to dynamic changes. Teaching environments are continuously evolving, influenced by emerging instructional technologies, policy shifts, and changing societal demands, all of which impact teaching competence. Furthermore, a significant limitation of traditional assessment approaches lies in their inability to integrate multiple data sources effectively. The evaluation of teaching competence requires a comprehensive analysis of diverse information, including student feedback, classroom observations, and the utilization of teaching resources. To overcome these limitations, advanced quantitative techniques are required to evaluate and optimize teaching strategies effectively.

Specifically, this paper aims to explore solutions for assessing English teaching abilities using big data analysis frameworks. Imagine holding a large collection of puzzle pieces that represent different aspects of teaching. Traditional methods may struggle to assemble these pieces accurately. With big data analysis, it becomes possible to create a clearer picture of teaching abilities. One of the key approaches involves constructing parameter models and big data analysis models [4] that constrain and reflect the actual levels of English teaching. Imagine sorting apples and oranges to ensure similar items are compared, leading to accurate results. Sophisticated methods such as data fusion and clustering algorithms make this possible. Data fusion works like blending different fruit juices to create a complete flavor profile, combining multiple information sources to provide a comprehensive view of teaching dynamics. Clustering works like grouping similar fruits together, identifying patterns and trends in teaching performance by organizing similar data points. Additionally, developing objective functions and statistical analysis models is like creating a recipe to ensure consistency and accuracy in assessments. These models play an important role in standardizing and enhancing the accuracy of English teaching assessments These models not only enable precise evaluation of teaching abilities but also improve the reliability of prediction capabilities, which are crucial for future planning and interventions [3, 5].

Moreover, these analytical models contribute to a better understanding of the strengths and weaknesses in English teaching practices. Through parameter modeling and the application of big data techniques, teaching performance can be mapped and measured effectively. For example, data-driven insights can identify areas requiring improvement, such as curriculum design, pedagogical approaches, or teacher training programs. By leveraging these insights, it becomes possible to optimize teaching methods and allocate resources more strategically. Ultimately, the integration of parameter models, big data analysis, data fusion, and clustering methods enhances the overall framework for English teaching ability assessment. This comprehensive approach enables not only a better understanding of the existing teaching landscape but also the development of actionable strategies to improve it [6].

To further refine the accuracy of English teaching ability assessment, it is essential to establish robust information sampling models that effectively capture the key factors influencing teaching ability [7, 8]. These models serve as the foundation for conducting in-depth analyses of teaching effectiveness. By combining nonlinear information fusion techniques with time series analysis methods [9, 10], the statistical properties of English teaching ability can be systematically explored. The constraint indicator parameters of English teaching ability, which represent a set of nonlinear time series, provide critical insights into the factors influencing teaching performance. For instance, the temporal analysis of teaching practices can reveal patterns in instructional quality over time, while nonlinear fusion methods integrate these patterns to derive meaningful conclusions.

To represent the complex relationships between these parameters, a high-dimensional feature distribution space is constructed. This space allows for the visualization and modeling of parameter indicators associated with English teaching ability assessment. Key constraint indicators include teacher qualification levels, the quality and extent of investment in teaching facilities, and the alignment of teaching practices with relevant policy frameworks. Each of these indicators significantly impacts the overall effectiveness of teaching and must be evaluated in conjunction with others to derive a comprehensive assessment. To quantify these relationships, a differential equation is developed to model the flow of information within the key factors that influence teaching ability[11, 12]. This equation serves as a mathematical representation of the dynamic interactions between various factors influencing English teaching ability, offering a structured approach to analyzing and optimizing teaching performance (**Figure 1**).

$$\frac{dC\,(t)}{dt} = k \bullet (T(t) - C\,(t)) \tag{1}$$

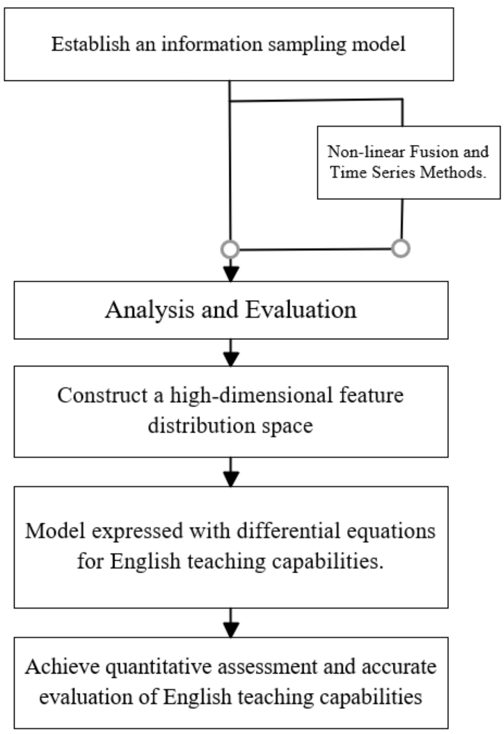

Figure 1. Information flow model of key factors influencing English teaching capabilities.

The equation: Let h(.) be the multivariate value function for English teaching ability assessment, and[13] be the error measurement function for evaluation. The solution vector for English teaching ability assessment is calculated in the high-dimensional feature distribution space using correlation fusion methods, Subset of characteristic training for obtaining teaching ability assessment S_i (i = 1,2,...,L), Meet the following conditions:

$$\begin{gathered}\sum = diag(\delta_1, \delta_2, ..., \delta_r), \delta_i = \sqrt{\lambda_i}, \forall i \neq j \\ \bigcup_{i=1}^{L} S_i = V - v_s \\ \text{Command} x_{m+1} = \mu x_m (1 - x_m)\end{gathered} \tag{2}$$

The conjugate solution of a statistical information model for English teaching ability assessment satisfies the initial value characteristic decomposition condition.

$$\begin{gathered}U = \{u(\,t\,)|\; u\,(t) \in X\; ||\; u\; || \leq dt \in I\} \\ \text{among them } (I_i)_{i \in N} = \{x_1 x_2 ... x_n\}\end{gathered} \tag{3}$$

Data Information Flow Model for English Teaching Ability Assessment is constructed based on the statistical feature distribution sequence x(m) of a set of multivariate variables for English teaching ability assessment, using previous statistical measurement values:

$$\begin{gathered}C_{1x}\,(T) = E\,\{x(\,m\,)\} = 0 \\ C_{2x}\,(T) = E\,\{x(\,m\,)x(\,m + T\,)\} = r\,(T) \\ C_{kx}(T_1 T_2 ... T_{n-1}) \equiv 0 n \geq 3\end{gathered} \tag{4}$$

When q = 2, The level of teaching ability assessment and distribution of teaching resources in English meet the (2+1)-dimensional continuous functional condition, that is, the English teaching ability assessment has a convergent solution, subject to constraints.

$$\Psi_x\,(\omega) = ln\varphi_x\,(\omega) = -\frac{1}{2}\omega^2\sigma^2 \tag{5}$$

The control objective function for predicting and estimating English teaching ability is constructed using quantitative recursive analysis method in the big data information model for English teaching ability assessment[14] based on the data flow model built for English teaching ability assessment.

$$\max_{x_{a,b,d,p}} \sum\nolimits_{a \in A} \sum\nolimits_{b \in B} \sum\nolimits_{d \in D} \sum\nolimits_{p \in P} x_{a,b,d,pVp} \tag{6}$$

$$\text{s.t.} \sum\nolimits_{a \in A} \sum\nolimits_{d \in D} \sum\nolimits_{p \in P} x_{abdp} R_p^{bw} \leq K_b^{bw}\,(S)\,,\; b \in B \tag{7}$$

Quantitative recursive evaluation of English teaching ability level is conducted using the grey model. The historical data representing the distribution of English teaching ability is assumed to be $\{x_i\}_{i=1}^{M}$, The probability density functional[15] for estimating the predictive ability of English teaching, given a certain initial value of perturbation features, is as follows:

$$u_c\left(t\right) = Nx_c\left(t\right) \tag{8}$$

In the high-dimensional feature distribution space, the continuous function for predicting the English teaching ability estimation statistical model is… $u : I \times IR^d$→IR, After N-1 iterations, where N $\geq$ 1, the grayscale sequence of English teaching ability assessment satisfies M(N) < Y. By using quantitative recursive analysis method, the N neighboring sample values of the output indicator distribution big data information flow for English teaching ability assessment are obtained:

$$P_{1J} = \sum\nolimits_{d_i \in NMM} Sim(x, d_i\,) y(\,d_i, C_j) \tag{9}$$

Using the big data information fusion method, we construct an inter-domain classification objective function [16] for the distribution of big data information flow in English teaching ability assessment. Specifically, the objective function is the big data clustering objective function [17]:

$$J_m(U, V) = \sum\nolimits_{N=1}^{m} \sum\nolimits_{i=1}^{c} \mu_{iN}^{z}(d_{in}) \tag{10}$$

Distribution sequence of the correlation of indices related to the assessment of English teaching ability in research $\{x_m\}_{m=1}^{M}$. Quantitative recursive feature extraction results for teaching ability evaluation are obtained through quantitative analysis combined with N-value optimization methods (**Figure 2**):

$$x_m = a_0 + \sum\nolimits_{i=1}^{Z_{AR}} a_i x_{m-j} + \sum\nolimits_{j=0}^{N_{NA}} b_j \eta_{m-j} \tag{11}$$

In the equation: Z_{AR} is the sampling amplitude for initial English teaching ability assessment; N_{VA} is a scalar time series; *bj* is the oscillation decay value for English teaching ability assessment.

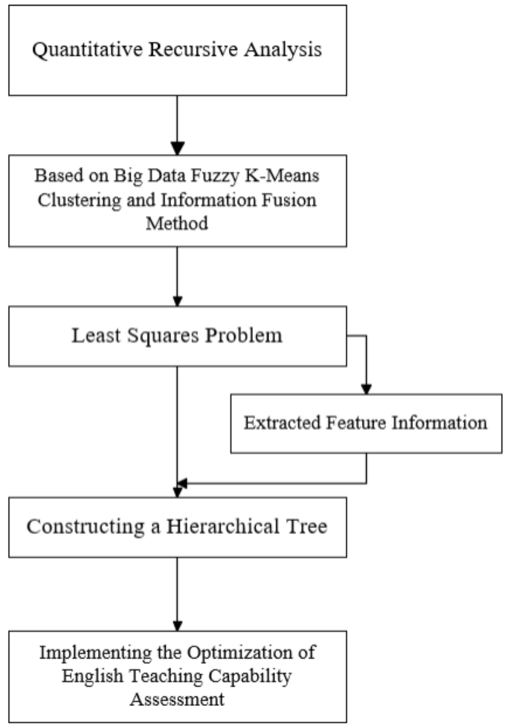

Figure 2. The Process of teaching capability assessment.

2. Optimization of English Teaching Ability Evaluation Model

In order to improve the quantitative evaluation ability of English teaching level, this study is based on the analysis of the English teaching ability evaluation research based on big data information model. It adopts the quantitative recursive analysis method. Imagine trying to measure the height of a tree while only seeing part of it. Quantitative recursive analysis acts like a ladder, providing a better view with each step and refining measurements. In order to further improve the accuracy of English teaching ability evaluation, this study proposes an English teaching ability estimation method based on big data fuzzy K-means clustering and information fusion. Clustering works like sorting a pile of mixed seeds into groups of similar types, while fuzzy K-means clustering allows some seeds to belong to multiple groups if they share similarities. This captures nuances in teaching abilities. This method transforms the problem of English teaching ability evaluation into the problem of solving the least squares estimation problem of the objective function of K-means clustering. The so-called least squares [18] problem is to find the consistency estimation value of the English teaching ability evaluation resource constrained vector to minimize the Euclidean norm of F-. The extracted value of the entropy characteristic of the English teaching ability constraint feature information is obtained as:

$$P_{loss} = 1 - \frac{1-p_0}{\rho} = \frac{p_0+\rho-1}{\rho} = \sum\nolimits_{m=1}^{M} P_{M,z} \quad (12)$$

The estimation equation for English teaching ability is transformed into finding the least square solution, given d_i as the perturbation feature vector for teaching ability assessment:

$$B(t) = x(t) + iy(t) = a(t)\, e^{i\theta(t)} + m(t) \quad (13)$$

In the equation, x(t) represents the real part of the time series for evaluating the distribution of big data, while y(t) represents the imaginary part of the constraint indicator sequence for evaluating English teaching ability.

The utilization rate of English teaching resources can be calculated based on the subset of the Nth category obtained through amplitude randomization using the alternative data method for assessing teaching abilities in English. This can be achieved by perturbing the empirical distribution data of teaching abilities in the Nth category [19]:

$$U_{util} = \gamma \overline{X} \quad (14)$$

Building a hierarchical tree, using big data analysis methods to establish the principal component features for assessing English teaching abilities, and utilizing fuzzy proximity filling to solve the similarity of teaching resource distribution:

$$Sim_1(d_i, d_{1j}) = \frac{\sum_{N=1}^{Z} W_{in} \times W_{1jn}}{\sqrt{\sum_{k=1}^{Z} W_{in}^2} \cdot \sqrt{\sum_{n=1}^{Z} W_{1jn}^2}} \quad (15)$$

3. Simulation Experiment Analysis

By using Matlab simulation analysis methods to test the big data analysis performance of English teaching ability assessment, and using statistical analysis methods to sample data for English teaching ability assessment, the decision threshold value for teaching ability assessment is obtained. The associated parameters for setting the distribution of English teaching resources are $\xi_{c_1}^{d_2} = 3/5, \xi_{c_2}^{d_2} = 2/5, \xi_{c_3}^{d_2} = 2/5$, $\mathrm{maxg}_{c_1}(d_2) = 6/5$, $\mathrm{maxg}_{c_2}(d_2) = 3/8$, $\mathrm{maxg}_{c_3}(d_2) = 1/10$, The sampling frequency $f_0 = 600Hz$, the initial step size for adaptive learning is ρ = 0.97, and the correlation coefficient for the distribution of teaching resources' features is B = 1.14. Based on the above parameter settings, the big data reconstruction focuses on the key factors influencing English teaching ability assessment. The time-domain waveform of the big data distribution is shown in the figure below (refer to **Figure 3**).

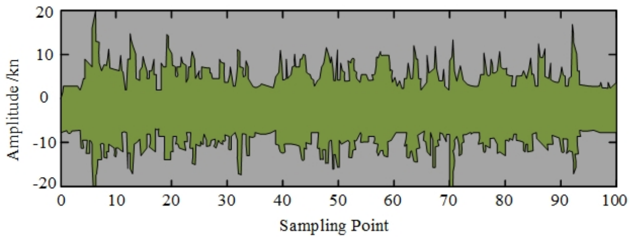

Figure 3. Temporal distribution of big data temporal waveform.

Figure 3 illustrates the temporal waveform of the big data distribution, which captures how teaching competence metrics evolve over time. Key observations include: Rising peaks, these indicate periods of enhanced teaching resource allocation, showcasing the effectiveness of the proposed clustering and fusion approach. Sudden drops, these may highlight moments of underutilization or disruptions in the resource distribution process, requiring further analysis. Utilizing the statistical outcomes of the indicator parameters for English teaching competence assessment illustrated in **Figure 3** as the focal point of investigation, data clustering and information fusion procedures are employed to facilitate the assessment of teaching proficiency. The evaluation accuracy and other pertinent indicators are presented in **Table 1** to showcase the test outcomes. Examination of these findings demonstrates that the methodology adopted in this study exhibits superior accuracy in assessing teaching proficiency and enhances the efficient utilization of teaching resources.

4. Discussion

This study introduced an innovative methodology for assessing English teaching competence by integrating big data fuzzy K-means clustering and information fusion techniques. The experimental findings demonstrate that this approach significantly enhances the accuracy of competence assessment while optimizing the allocation of teaching resources. Compared to traditional evaluation methods, the model not only achieved higher evaluation precision but also exhibited improved efficiency in resource utilization, as evi-

Table 1. Performance test comparison.

Evaluation Cycle	Methodology of This Article		Literature [20]		Literature [21]	
	Evaluation Accuracy/%	Utilization Rate/%	Evaluation Accuracy/%	Utilization Rate/%	Evaluation Accuracy/%	Utilization Rate/%
1	98.22	98.03	87.44	89.13	83.24	86.34
2	97.10	97.68	86.56	87.35	82.13	87.01
3	96.34	99.04	88.77	89.32	86.10	79.32
4	98.55	96.35	89.44	87.68	88.24	78.93

denced by the comparative results presented in **Table 1**. The methodology shows great promise for addressing the challenges of evaluating teaching competence in an increasingly data-driven educational environment. This section discusses the implications, potential applications, and limitations of the proposed approach, while offering suggestions for future research directions to build upon and enhance its effectiveness.

4.1. Implications of the Study

Artificial intelligence (AI) has significantly impacted education by enhancing personalized learning, improving accessibility, and supporting innovative teaching practices. AI-driven systems create adaptive learning environments tailored to individual students' needs, enabling more effective engagement and understanding [22]. In higher education, AI applications like ChatGPT and other tools support academic tasks such as data analysis, writing assistance, and personalized feedback, fostering greater efficiency and accessibility. Additionally, AI-powered solutions like virtual reality, big data analytics, and cloud-based platforms revolutionize traditional teaching by enabling collaborative, immersive, and data-informed learning experiences [23, 24].

The successful application of big data analysis and clustering algorithms to teaching competence evaluation carries several noteworthy implications for both educational research and practice. First, the study underscores the critical role of advanced data processing methods in enhancing educational research methodologies. The use of fuzzy K-means clustering effectively tackles the challenge of analyzing high-dimensional and complex educational datasets, which are often difficult to interpret using traditional methods. This approach allows for the identification of underlying patterns in the data that may otherwise remain hidden. Additionally, the integration of information fusion further strengthens the model's ability to synthesize multiple, diverse data sources, leading to a more comprehensive assessment of teaching capabilities. This capability is particularly valuable in a rapidly evolving educational landscape, where decision-makers require accurate, real-time insights to guide policy and practice. The model's ability to integrate and process diverse data sets is consistent with current trends in education technology, which increasingly emphasize the role of data-driven decision-making in improving teaching outcomes and resource management.

Second, the research contributes to the field of educational assessment by introducing a quantitative recursive analysis framework that facilitates dynamic, real-time adjustments in teaching competence evaluation. This recursive approach is crucial because it ensures that the evaluation process can adapt to the continuously changing nature of educational environments. As teaching conditions evolve—whether due to technological advancements, shifts in educational policy, or changes in societal needs—the model remains flexible and responsive. For example, if a new pedagogical approach is introduced or if there is a shift in policy that affects teaching standards, the model can accommodate these changes by adjusting its constraints and assessment criteria. This adaptability is particularly valuable for education administrators and policymakers, as it enables them to optimize resource allocation and make informed decisions in real time, improving overall educational outcomes [25].

Third, the study emphasizes the potential of combining machine learning techniques with traditional statistical methods to enhance the robustness of educational evaluation models. By transforming the assessment problem into a least-squares optimization task, the methodology ensures consistency and reliability in the results. This hybrid approach represents a significant advancement in the field, bridging the gap between theoretical research and practical, real-world applications. The ability to apply both machine learning algorithms and traditional statistical techniques to teaching competence evaluation not only strengthens the accuracy of assessments but also facilitates a more nuanced understanding of educational dynamics.

4.2. Practical Applications

The proposed methodology offers several promising practical applications in the context of English teaching and beyond. One of the key applications is in the formulation of data-driven resource allocation strategies. By clustering and consolidating various indicator parameters, the model enables educational administrators to identify areas in need of improvement and strategically allocate resources. For example, regions or institutions with lower teaching competence scores can be prioritized for additional teacher training programs, infrastructure upgrades, or targeted policy interventions. This data-driven approach ensures that resources are used efficiently and effectively, promoting equitable access to quality education across different regions and institutions.

Imagine a school district that has implemented the proposed model to assess the teaching competence of its English teachers. The model identifies that one particular school has a lower competence score compared to others in the district. Upon further analysis, the clustering algorithm reveals that the underperformance is primarily due to inadequate teacher training and outdated teaching materials. Using this information, the school district can take targeted actions, such as providing specialized professional development workshops for the teachers at that school and updating their teaching resources. Over time, the model can track the progress of these interventions and adjust the resource allocation as needed. This approach not only improves the teaching quality at the underperforming school but also ensures that resources are used effectively and efficiently. Additionally, while the methodology relies on big data and advanced computational techniques, it can still be adapted for use in rural schools or areas with limited data infrastructure. In rural schools with limited data infrastructure, the model could be implemented using a phased approach. Initially, the focus could be on collecting basic data manually and using simplified clustering techniques to identify key areas for improvement. For example, teachers could be trained to collect data on student performance and classroom interactions using simple tools like spreadsheets or mobile apps. This data could then be analyzed using a basic version of the clustering algorithm to identify patterns and allocate resources more effectively. Over time, as more data becomes available and infrastructure improves, the model could be expanded to include more sophisticated data sources and clustering techniques. For instance, as schools gain access to more advanced technology, they could start incorporating data from online learning platforms [26], student surveys, and classroom observation tools. This would allow for a more comprehensive assessment of teaching competence and the development of more targeted interventions [27].

Another promising application lies in the development of personalized teaching strategies. The fuzzy K-means clustering algorithm can identify distinct patterns in teaching competence, allowing for the segmentation of educators into groups with similar strengths and weaknesses. Once these groups are identified, tailored professional development programs can be designed to address specific needs within each group, fostering continuous improvement in teaching quality. Furthermore, this approach could be extended to student performance analysis, enabling the creation of personalized learning plans that cater to the unique needs, learning styles, and preferences of individual students. This personalized approach would not only improve teaching outcomes but also enhance student engagement and success.

The methodology also has significant potential applications in monitoring and evaluation processes within educational systems. By providing a quantitative and objective framework for assessing teaching competence, the model can serve as a benchmarking tool for evaluating the effectiveness of various educational reforms, teacher training programs, and resource allocation initiatives. Furthermore, the integration of information fusion allows for the inclusion of diverse data sources—such as student feedback, classroom observations, and test scores—thereby enriching the evaluation process [28]. This multidimensional approach enhances the validity of assessments, ensuring that teaching competence is measured from a variety of perspectives.

4.3. Limitations

Despite its many advantages, the proposed methodology has certain limitations that warrant further exploration. One of the major limitations is the dependency on algorithmic parameters, such as the initial cluster centers and the number of clusters. These parameters can have a significant impact on the outcomes of the clustering process, potentially leading to variability in assessment results. While the fuzzy K-means algorithm partially mitigates this issue by allowing for overlapping clusters, the reliance on subjective parameter selection remains a challenge. The sensitivity of the model to these parameters highlights the need for further work on optimizing the clustering process to ensure more consistent and reliable results.

Another limitation concerns the model's reliance on historical data for competence evaluation. While the recursive analysis framework enhances the model's predictive capabilities, it may struggle to account for sudden, disruptive changes in teaching environments or educational policies. For example, unprecedented events such as the COVID-19 pandemic or other global crises could render historical data less relevant, requiring the model to incorporate real-time data to maintain its accuracy and reliability. This issue underscores the importance of continuously updating the data used for evaluation and ensuring that the model is capable of adapting to unforeseen disruptions.

Additionally, the computational complexity of the methodology may present challenges for its implementation in resource-constrained educational settings. The integration of big data analysis and clustering algorithms requires significant computational resources, which may not be readily available in all institutions, especially those with limited technological infrastructure. This limitation highlights the need for further research into optimizing the algorithm's efficiency and exploring alternative approaches that balance computational demands with assessment accuracy.

4.4. Future Research Directions

To address the limitations discussed above, future research should focus on several key areas. First, efforts should be directed toward enhancing the robustness of the clustering algorithm by developing automated parameter selection methods. Techniques such as grid search, cross-validation, or metaheuristic optimization algorithms could be employed to identify the most optimal parameter configurations. This would reduce the reliance on subjective inputs, improving the consistency and reliability of the clustering process, and ultimately enhancing the quality of assessment results.

Second, the model's adaptability to dynamic educational environments could be further enhanced by incorporating real-time data streams. The integration of sensors, Internet of Things (IoT) devices, and real-time analytics platforms could enable continuous monitoring of teaching competence, providing decision-makers with timely feedback for more agile decision-making. This approach would ensure that the evaluation process remains relevant and responsive to changing educational conditions, allowing for quick adjustments in resource allocation and teaching strategies.

Third, future research could explore the application of advanced machine learning techniques, such as deep learning and natural language processing, to enhance feature extraction and pattern recognition capabilities. For example, sentiment analysis of student feedback or automated assessment of classroom interactions could provide valuable additional data that could further enrich the teaching competence evaluation process. Incorporating these techniques into the model would improve the precision and depth of the analysis, offering a more comprehensive view of teaching performance.

Finally, the methodology could be extended to other educational contexts beyond English teaching. Future studies could explore its application in assessing teaching competence in non-English language subjects or evaluating institutional performance. Comparative studies across different cultural and educational settings would provide valuable insights into the generalizability and scalability of the approach. Moreover, interdisciplinary collaborations with experts in education, computer science, and policy studies would contribute to refining the methodology and expanding its applications, further advancing the field of educational assessment.

5. Conclusions

This study aimed to address the challenges in evaluating English teaching competence by introducing an innovative methodology combining big data fuzzy K-means clustering and information fusion. The proposed approach demonstrated its ability to process high-dimensional data, cluster and integrate indicator parameters, and provide precise teaching competence assessments. By extracting meaningful patterns from the relevant feature information and leveraging recursive quantitative analysis, the methodology achieved high accuracy and resource utilization efficiency, as validated by simulation experiments.

The significance of this research lies not only in its ability to enhance assessment precision but also in its broader implications for educational resource management. The clustering and fusion techniques facilitated the development of targeted resource allocation strategies, ensuring that teaching resources are effectively utilized to address disparities in teaching quality. This advancement underscores the role of data-driven approaches in promoting equitable education and optimizing teaching outcomes. However, while the proposed methodology showcases robust analytical capabilities, it is essential to acknowledge its reliance on algorithm parameters and historical data. These dependencies may limit the model's adaptability in dynamic educational settings, especially when confronted with sudden shifts or incomplete data. Furthermore, the computational complexity of big data clustering techniques poses potential challenges for widespread adoption, particularly in resource-constrained environments. Addressing these challenges will require further refinement of the methodology, such as the incorporation of automated parameter tuning and real-time data integration mechanisms. The findings from this research open several avenues for future exploration. For instance, integrating advanced machine learning techniques like deep learning or graph neural networks could enhance the model's capacity to extract meaningful patterns from unstructured data sources, such as teacher feedback or classroom recordings. Additionally, expanding the scope of this methodology to other educational domains, such as STEM education or multilingual teaching environments, could further validate its versatility and scalability. Collaborative research across disciplines, involving education professionals and data scientists, could also play a significant role in refining the methodology and addressing its current limitations.

In conclusion, this study contributes a novel framework for English teaching competence assessment, combining innovative algorithmic approaches with practical insights for resource optimization. The results highlight the transformative potential of big data analysis to address complex educational challenges. By leveraging data-driven insights, we can move beyond traditional limitations and create more equitable and effective education systems. Imagine a world where every classroom, regardless of location or resources, has access to the tools and strategies needed to thrive. This research is a step towards that vision—a vision where data empowers educators, informs policymakers, and ultimately, enriches the learning experience for every student. By building upon these findings, future research can further bridge the gap between educational theory and practice, paving the way for adaptive, inclusive, and data-driven educational systems that unlock the full potential of both teachers and learners.

Author Contributions

Conceptualization, M.C.; Data curation, M.C.; Formal analysis, M.C.; Funding acquisition, M.C.; Investigation, M.C.; Methodology, M.C.; Visualization, M.C.; Writing—original draft, M.C.; Project administration, T.W.H.; Resources, T.W.H.; Software, T.W.H.; Supervision, T.W.H.; Validation, T.W.H.; Writing—review & editing, T.W.H. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Some or all data that support the findings of this study are available from the author upon reasonable request.

Conflicts of Interest

There are no conflicts of interest involved in this study.

References

[1] Jiang, L., 2022. Effective Utilization of Computer-Mediated Communication Technology in Network-based Foreign Language Teaching. Wireless Communications and Mobile Computing. 1–7. DOI:

https://doi.org/10.1155/2022/1048311

[2] Onan, A., 2023. Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion. Journal of King Saud University-Computer and Information Sciences. 35(7), 101610.

[3] Jiang, L., Lv, M., Cheng, M., et al., 2024. Design and implementation effects of SPOC-based blended teaching from the perspective of deep teaching: A case study of EFL students. Innovations in Education and Teaching International. 1–15. DOI: https://doi.org/10.1080/14703297.2024.2350519

[4] Onan, A., 2022. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. Journal of King Saud University-Computer and Information Sciences. 34(5), 2098–2117.

[5] Onan, A., 2019. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. IEEE Access. 7, 145614–145633.

[6] Wang, J., Xu, C., Zhang, J., et al., 2022. Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems. 62, 738–752.

[7] Lüdecke, D., Ben-Shachar, M.S., 2021. Performance: An R package for assessment, comparison and testing of statistical models. Journal of Open Source Software. 6(60).

[8] Shang, R., Ara, B., Zada, I., et al., 2021. Analysis of simple K-mean and parallel K-mean clustering for software products and organizational performance using education sector dataset. Scientific Programming. 2021, 1–20.

[9] Ouyang, X., Huo, J., Xia, L., et al., 2020. Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. IEEE Transactions on Medical Imaging. 39(8), 2595–2605.

[10] Onan, A., 2023. GTR-GA: Harnessing the power of graph-based neural networks and genetic algorithms for text augmentation. Expert Systems with Applications. 232, 120908.

[11] Ghaderpour, E., Pagiatakis, S.D., Hassan, Q.K., 2021. A survey on change detection and time series analysis with applications. Applied Sciences. 11(13), 6141.

[12] Zou, Y., Donner, R.V., Marwan, N., 2019. Complex network approaches to nonlinear time series analysis. Physics Reports. 787, 1–97.

[13] Onan, A., 2023. SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. Journal of King Saud University-Computer and Information Sciences. 35(7), 101611.

[14] Zhu, J., Wu, P., 2022. BIM/GIS data integration from the perspective of information flow. Automation in Construction. 136, 104166.

[15] Bravyi, S., Sheldon, S., Kandala, et al., 2021. Mitigating measurement errors in multiqubit experiments. Physical Review A. 103(4), 042605.

[16] Kar, A.K., Dwivedi, Y.K., 2020. Theory building with big data-driven research–Moving away from the "What" towards the "Why". International Journal of Information Management. 54, 102205.

[17] Chen, C. H., Song, F., Hwang, F. J., et al., 2020. A probability density function generator based on neural networks. Physica A: Statistical Mechanics and Its Applications. 541, 123344.

[18] Onan, A., Korukoğlu, S., Bulut, H., 2017. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. Information Processing & Management. 53(4), 814–833.

[19] Alguliyev, R.M., Aliguliyev, R.M., Sukhostat, L.V., 2020. Efficient algorithm for big data clustering on single machine. CAAI Transactions on Intelligence Technology. 5(1), 9–14.

[20] Pan, J., Liu, S., Shu, J., et al., 2022. Hierarchical recursive least squares estimation algorithm for second-order Volterra nonlinear systems. International Journal of Control, Automation and Systems. 20(12), 3940–3950.

[21] Feng, L., Hu, P., Chen, M.M., et al., 2023. Quantifying cumulative changes in water quality caused by small floodgates in Taihu Lake Basin—A case in Wuxi. Science of The Total Environment. 900, 165608.

[22] Ikotun, A.M., Ezugwu, A.E., Abualigah, L., et al., 2023. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences. 622, 178–210.

[23] Wen, J., Yang, J., Jiang, B., et al., 2020. Big data driven marine environment information forecasting: A time series prediction network. IEEE Transactions on Fuzzy Systems. 29(1), 4–18.

[24] Papadakis, S., Kiv, A.E., Kravtsov, H., et al., 2023. Revolutionizing education: using computer simulation and cloud-based smart technology to facilitate successful open learning. In Joint Proceedings of the 10th Illia O. Teplytskyi Workshop on Computer Simulation in Education, and Workshop on Cloud-based Smart Technologies for Open Education (CoSinEi and CSTOE 2022) co-located with ACNS Conference on Cloud and Immersive Technologies (No. 3358, pp. 1–18). CEUR Workshop Proceedings.

[25] Lavidas, K., Papadakis, S., Manesis, D., et al., 2022. The effects of social desirability on students' self-reports in two social contexts: Lectures vs. lectures and lab classes. Information. 13(10), 491.

[26] Jiang, L., Liang, F., Wu, D., 2024. Effects of technology-aided teaching mode on the development of CT skills of EFL students in Higher Vocational Colleges: A case study of English argumentative writing. Thinking Skills and Creativity. 53, 101594. DOI: https://doi.org/10.1016/j.tsc.2024.101594

[27] Lavidas, K., Voulgari, I., Papadakis, S., et al., 2024. De-

terminants of Humanities and Social Sciences Students' Intentions to Use Artificial Intelligence Applications for Academic Purposes. Information. 15(6), 314.
[28] Jiang, L., Qu, Y., 2024. Assessing core competencies of business English students in Chinese higher vocational colleges. Scientific Reports. 14(1), 29328. DOI: https://doi.org/10.1038/s41598-024-78898-y